APLICATIVO WEB PARA EL ANÁLISIS PREDICTIVO DE LAS BRECHAS DE GÉNERO EN LAS PRUEBAS SABER 11, PRO Y TYT

ANGIE PAOLA MANRIQUE RAVELO

ALISSON CATALINA CELEITA GUTIÉRREZ

CESAR YESID BARAHONA RODRÍGUEZ

UNIVERSIDAD DE CUNDINAMARCA

Facultad de Ingeniería

Programa de Ingeniería de Sistemas

Facatativá, 2020

APLICATIVO WEB PARA EL ANÁLISIS PREDICTIVO DE LAS BRECHAS DE GÉNERO EN LAS PRUEBAS SABER 11, PRO Y TYT

Angie Paola Manrique Ravelo

Alisson Catalina Celeita Gutiérrez

Director: Ing. Cesar Yesid Barahona Rodríguez, Especialista en Gestión de proyectos, Magister en Sistemas Computacionales

GRUPO DE INVESTIGACIÓN DE SISTEMAS Y TECNOLOGÍA DE FACATATIVA (GISTFA)

UNIVERSIDAD DE CUNDINAMARCA

Facultad de Ingeniería

Programa de Ingeniería de Sistemas

Facatativá, 2020

Notas de aceptación:

_____

_____

_____

_____

_____

_____

Firma del Jurado:

_____

Firma del Jurado:

_____

AGRADECIMIENTOS

A todas aquellas personas que contribuyeron en cada una de las etapas de aprendizaje durante esta carrera, docentes y compañeros que intervinieron en este proceso de crecimiento educativo.

Y especialmente, a mi familia quiénes estuvieron presentes en cada paso dado, a nuestro director de proyecto Cesar Barahona, por estar guiándonos de forma continua y a mi amiga Alisson Celeita, con quién realicé el proyecto, por ser una compañía constante y un apoyo incondicional dentro y fuera de las canchas.

Angie Paola Manrique Ravelo

AGRADECIMIENTOS

Este documento va dedicado a cada una de las personas que han estado conmigo en los diferentes pasos que he dado para llegar a este punto de mi carrera, todos aquellos quienes me ayudaron cuando sentía que no podía culminar y me enseñaron muchas de las bases para ser la persona y la profesional que soy hoy en día. Los agradecimientos especiales son para mi compañera de proyecto, mi director de proyecto de grado y mi familia. Para mi compañera y mejor amiga Angie Manrique gracias por todos los momentos que hemos pasado y soportado juntas, en especial aquellos chascarrillos que a pesar de ser un poco amargos tu lograbas darle ese toque de comedia que me alegraba los días. Para mi director de proyecto Cesar Barahona, gracias por ser ese guía y apoyo incondicional en la universidad, por ser del tipo de docente que realmente busca lo mejor para sus muchachos. Y para mi familia, gracias por apoyarme en esta montaña rusa desde el comienzo de la carrera, sin ustedes no hubiera logrado la mayoría de las cosas que he conseguido hasta el día de hoy.

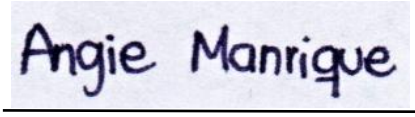
Alisson Catalina Celeita Gutiérrez

COMPROMISO DE AUTOR

Yo Angie Paola Manrique Ravelo, con cédula de identidad No. 1073178861 y con cód. 461216142 estudiante del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que:

El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma:

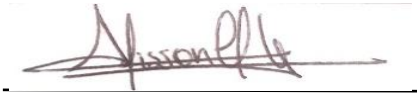


COMPROMISO DE AUTOR

Yo Alisson Catalina Celeita Gutiérrez, con cédula de identidad No. 1070973007 y con cód. 461216120 estudiante del programa de Ingeniería de Sistemas de la Universidad de Cundinamarca, declaro que:

El contenido del presente documento es un reflejo de mi trabajo personal y manifiesto que, ante cualquier notificación de plagio, copia o falta a la fuente original, soy responsable directo legal, económico y administrativo sin afectar al director del trabajo, a la Universidad y a cuantas instituciones hayan colaborado en dicho trabajo, asumiendo las consecuencias derivadas de tales prácticas.

Firma:



RESUMEN

El proyecto busca hacer uso de la información suministrada por el ICFES de sus pruebas estandarizadas a nivel nacional (Saber 11, Saber Pro y Saber TyT); y a partir de dichos datos, analizar la existencia de brechas de género en los resultados de las mismas, uno de los aspectos importantes durante el análisis es determinar la relevancia de esta problemática en las distintas áreas rurales y urbanas del país, evidenciando así el comportamiento presentado en los últimos años en la educación media y superior. Con la realización de este proyecto se espera tener como resultado los indicadores de brechas de género, mirar la incidencia entre ellos (si es que existe), encontrar una relación entre las pruebas de educación básica, media y superior, y posteriormente poder predecir futuros comportamientos en las pruebas. Haciendo que este proceso quede plasmado de forma automática en un aplicativo web. La información resultante de los análisis radica su importancia al dejar en evidencia las debilidades y fortalezas de los estudiantes en las áreas evaluadas con el fin de que la representación clara de los resultados sea usada por las diferentes personas e instituciones interesadas. Y, asimismo, brindar información actualizada de las pruebas para futuras investigaciones, siendo ésta una forma de evaluar el funcionamiento de los mecanismos que imparte el gobierno y las instituciones a la población estudiantil en el país.

Palabras claves: Minería de datos, Brechas de género, Educación, Análisis descriptivo, Análisis predictivo.

ABSTRACT

The project seeks to make use of the information provided by the ICFES of its nationally standardized tests (Saber 11, Saber Pro and Saber TyT); and from these data, analyze the existence of gender gaps in their results, one of the important aspects during the analysis is to determine the relevance of this problem in the different rural and urban areas of the country, thus demonstrating the behavior presented in recent years in middle and higher education. With the completion of this project, it is expected to have as a result the indicators of gender gaps, look at the incidence among them (if there is one), find a relationship between the tests of basic, middle and higher education, and then be able to predict future behaviors in the tests. Making this process automatically reflected in a web application. The information resulting from the analysis lies in its importance in exposing the weaknesses and strengths of the students in the areas evaluated so that the clear representation of the results is used by the different interested persons and institutions. And, also, to provide updated information of the tests for future investigations, this being a way of evaluating the functioning of the mechanisms that the government and the institutions impart to the student population in the country.

Keywords: Data mining, Gender gaps, Education, Descriptive analysis, Predictive analysis.

ÍNDICE GENERAL

RESUMEN ---------- 15
INTRODUCCIÓN ---------- 31
I. INFORME DE INVESTIGACIÓN ---------- 33
1.1. ESTADO DE ARTE ---------- 33
1.2. LÍNEA DE INVESTIGACIÓN ---------- 39
1.3. PLANTEAMIENTO DE PROBLEMA Y PREGUNTA DE INVESTIGACIÓN 40
1.4. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS ---------- 41
1.4.1. OBJETIVO GENERAL ---------- 41
1.4.2. OBJETIVOS ESPECÍFICOS ---------- 41
1.5. ALCANCE E IMPACTO DEL PROYECTO ---------- 42
1.6. METODOLOGÍA ---------- 43
1.7. MARCOS DE REFERENCIA ---------- 44
1.7.1. MARCO TEÓRICO ---------- 44
1.7.1.1. BIG DATA Y MINERIA DE DATOS ---------- 44
1.7.1.2. TÉCNICAS DE MINERIA DE DATOS ---------- 45
1.7.1.3. METODOLOGÍAS ---------- 46
1.7.1.4. SISTEMA ---------- 48
1.7.2. MARCO LEGAL ---------- 49
II. MODELO DESCRIPTIVO ---------- 51
2.1. EXTRACCIÓN DE LOS DATOS ---------- 51
2.2. PREPROCESAMIENTO DE LOS DATOS ---------- 56
2.3. IDENTIFICACIÓN Y ANÁLISIS DE LOS INDICADORES ---------- 58
2.4. INDICADORES DE BRECHA DE GÉNERO ---------- 69
III. DOCUMENTACION DEL SOFTWARE ---------- 73
3.1. PLAN DE PROYECTO ---------- 73
3.2. DETERMINACIÓN DE REQUERIMIENTOS ---------- 73
3.3. ESPECIFICACIÓN DEL DISEÑO ---------- 77
3.3.1. MODELO DE ENTIDAD RELACION (MER) ---------- 77
3.3.2. DIAGRAMAS DE CASOS DE USO ---------- 82
3.3.3. DIAGRAMAS DE SECUENCIAS ---------- 87
3.3.4. DIAGRAMAS DE ACTIVIDADES ---------- 97

3.3.5. DIAGRAMA DE CLASES ---------- 112
3.3.5.1. DIAGRAMA DE CLASES USUARIO ---------- 112
3.3.5.2. Diagrama de clases Archivos. ---------- 113
3.3.5.3. Diagrama de clases Análisis Descriptivo. ---------- 114
3.3.5.4. Diagrama de clases Análisis Predictivo. ---------- 115
3.3.5.5. Diagrama de clases componentes inicio. ---------- 116
3.4. DISEÑO DE LOS CASOS DE PRUEBA ---------- 117
3.5. ESTIMACIÓN DE RECURSOS ---------- 125
IV. RESULTADOS ANÁLISIS DESCRIPTIVO ---------- 126
4. 1. PRUEBAS SABER 11 ---------- 126
4.2. PRUEBAS SABER TYT ---------- 151
4.3. PRUEBAS SABER PRO ---------- 183
V. MODELO PREDICTIVO ---------- 226
VI. RESULTADOS MODELO PREDICTIVO ---------- 231
6.1. SABER 11 ---------- 231
6.2. SABER TYT ---------- 244
6.3. SABER PRO ---------- 254
CONCLUSIONES ---------- 267
BIBLIOGRAFÍA ---------- 269
7. ANEXOS ---------- 274
7.1. MANUAL DE USUARIO ---------- 274
INTRODUCCIÓN ---------- 276
1. Inicio de Sesión ---------- 277
2. Contáctenos ---------- 278
3. Página principal usuario ---------- 279
4. Análisis descriptivo de pruebas Saber 11 ---------- 281
5. Análisis descriptivo de pruebas Saber TyT ---------- 285
6. Análisis descriptivo de pruebas Saber Pro ---------- 289
7. Análisis predictivo de las Pruebas Saber Superior ---------- 293
8. Inicio de sesión Administrador ---------- 298
9. Insertar usuarios nuevos ---------- 299
10. Listar usuarios ---------- 300
11. Listado de pruebas ---------- 301

12. Cargar archivos pruebas ---------- 302
13. Otras funciones Administrador ---------- 305
7.2. MANUAL TÉCNICO ---------- 307
7.3. ARTÍCULO CACIED ---------- 323
7.4. ARTÍCULO EAN ---------- 329
7.5. ARTÍCULO CIETA ---------- 332
7.6. ARTÍCULO REVISTA PENSAMIENTO UDECINO ---------- 341
7.7. ARTÍCULO CICI ---------- 351
7.8. ARTÍCULO ENCUENTRO DE SEMILLEROS ---------- 355
7.9 FORMATOS DE SEGUIMIENTO ---------- 358

LISTA DE TABLAS

Tabla 1 Requerimientos funcionales---------- 73
Tabla 2 Descripción tabla usuario ---------- 78
Tabla 3 Descripción tabla analisis_sabertyt ---------- 78
Tabla 4 Descripción tabla analisis_saberpro ---------- 79
Tabla 5 Descripción tabla sb11_[anio] ---------- 79
Tabla 6 Descripción tabla sabertyt_[anio] ---------- 80
Tabla 7 Descripción tabla saberpro[anio]---------- 80
Tabla 8 Descripción tabla sb11_muestra---------- 81
Tabla 9 Descripción tabla sabertyt_muestra---------- 81
Tabla 10 Descripción tabla saberpro_muestra---------- 82
Tabla 11 Descripción tabla aplicacion---------- 82
Tabla 12 Descripción de roles propuestos ---------- 82
Tabla 13 Descripción diagrama de casos de uso administrador---------- 83
Tabla 14 Descripción diagrama de casos de uso aplicativo---------- 85
Tabla 15 Descripción diagrama de casos de uso usuario ---------- 86
Tabla 16 Descripción de los objetos de los diagramas de secuencia---------- 96
Tabla 17 Descripción actividad Iniciar sesión ---------- 97
Tabla 18 Descripción actividad Crear usuario---------- 98
Tabla 19 Descripción actividad Consultar usuario ---------- 99
Tabla 20 Descripción actividad Actualizar usuarios ---------- 100
Tabla 21 Descripción actividad Eliminar usuarios---------- 101
Tabla 22 Descripción actividad Consultar listado pruebas---------- 102
Tabla 23 *Descripción actividad Consultar cantidad de estudiantes* ---------- 103
Tabla 24 *Descripción actividad Carga de los datos* ---------- 104
Tabla 25 *Descripción actividad Cerrar sesión*---------- 105
Tabla 26 *Descripción actividad Calcular indicadores de brecha de género*------- 106
Tabla 27 *Descripción actividad Realizar análisis predictivo* ---------- 107
Tabla 28 *Descripción actividad Limpieza y tratamiento* ---------- 108
Tabla 29 *Descripción actividad Visualizar resultados* ---------- 109
Tabla 30 *Descripción actividad Consultar cantidad de mujeres*---------- 110
Tabla 31 *Descripción actividad Generar reporte* ---------- 111
Tabla 32 *Estimación de recursos proyecto*---------- 125

LISTA DE FIGURAS

Figura 1. Situación actual de las escuelas y desempeño en pruebas Piza --- 35
Figura 2. Aplicaciones de minería de datos en la educación --- 37
Figura 3. Fases de la metodología CRISP-DM --- 47
Figura 4. Script descarga --- 51
Figura 5. Archivo .bat --- 52
Figura 6. Archivo VBScript --- 52
Figura 7. Conexión FTP Java --- 53
Figura 8. Descarga archivos Java --- 54
Figura 9. Acceso al FTP con WinSCP --- 55
Figura 10. Sincronización de carpetas WinSCP --- 55
Figura 11. Comparación archivos WinSCP --- 56
Figura 12. Extración resultados de pruebas --- 57
Figura 13. Preprocesamiento en WEKA, Saber 11 2016-1 --- 58
Figura 14. Preprocesamiento en WEKA, Saber TyT 2016-1 --- 59
Figura 15. Resultados Apriori en WEKA, Saber TyT 2016-1 --- 60
Figura 16. Nuevas reglas Apriori en WEKA, Saber TyT 2016-1 --- 61
Figura 17. Estrato evaluado con K-Means --- 62
Figura 18. Género de colegio evaluado con K-Means --- 63
Figura 19. Nivel de Sisbén evaluado con K-Means --- 63
Figura 20. Ingreso familiar evaluado con K-Means --- 64
Figura 21. Jornada de colegio evaluada con K-Means --- 64
Figura 22. Estado civil evaluado con K-Means --- 65
Figura 23. Valor matricula evaluado con K-Means --- 65
Figura 24. Tipo de pago de matrícula evaluado con K-Means --- 66
Figura 25. Grupo de referencia evaluado con K-Means --- 66
Figura 26. Cabeza de familia evaluado con K-Means --- 67
Figura 27. Número de personas a cargo evaluado con K-Means --- 67
Figura 28. Lectura diaria evaluado con K-Means --- 68
Figura 29. Indicadores Saber 11 --- 69
Figura 30. Indicadores Saber TyT --- 69
Figura 31. Indicadores Saber Pro --- 70
Figura 32. Plan de proyecto realizado en Project --- 73
Figura 33. Modelo Entidad Relación --- 77
Figura 34. Diagrama de casos de uso del Administrador --- 83
Figura 35. Diagrama de casos de uso del Aplicativo --- 84
Figura 36. Diagrama de casos de uso del Usuario --- 85
Figura 37. Diagrama de secuencia Iniciar sesión --- 87
Figura 38. Diagrama de secuencia Crear usuarios --- 88
Figura 39. Diagrama de secuencia Consultar usuarios --- 88
Figura 40. Diagrama de secuencia Carga de los datos --- 90
Figura 41. Diagrama de secuencia Consultar listado pruebas --- 91
Figura 42. Diagrama de secuencia Consultar cantidad de estudiantes --- 91

Figura 43. Diagrama de secuencia Cerrar sesión ---------- 92
Figura 44. Diagrama de secuencia Calcular indicadores de brecha de género ---- 92
Figura 45. Diagrama de secuencia Realizar análisis predictivo ---------- 93
Figura 46. Diagrama de secuencia Limpieza y tratamiento de los datos ---------- 94
Figura 47. Diagrama de secuencia Generar reportes ---------- 95
Figura 48. Diagrama de secuencia Consultar cantidad de mujeres ---------- 95
Figura 49. Diagrama de secuencia Visualizar resultados ---------- 96
Figura 50. Diagrama de actividad Iniciar sesión ---------- 97
Figura 51. Diagrama de actividad Crear usuario ---------- 98
Figura 52. Diagrama de actividad Consultar usuario ---------- 99
Figura 53. Diagrama de actividad Actualizar usuarios ---------- 100
Figura 54. Eliminar usuarios ---------- 101
Figura 55. Diagrama de actividad Consultar listado pruebas ---------- 102
Figura 56. Diagrama de actividad Consultar cantidad de estudiantes ---------- 103
Figura 57.Diagrama de actividad Carga de los datos ---------- 104
Figura 58. Diagrama de actividad Cerrar sesión ---------- 105
Figura 59. Diagrama de actividad Calcular indicadores de brecha de género --- 106
Figura 60. Diagrama de actividad Realizar análisis predictivo ---------- 107
Figura 61. Diagrama de actividad Limpieza y tratamiento ---------- 108
Figura 62. Diagrama de actividad Visualizar resultados ---------- 109
Figura 63. Diagrama de actividad Consultar cantidad de mujeres ---------- 110
Figura 64. Diagrama de actividad Generar reporte ---------- 111
Figura 65. Diagrama de Clases Componente Usuario ---------- 112
Figura 66. Diagrama de Clases Componente Archivos ---------- 113
Figura 67. Diagrama de Clases Análisis Descriptivo ---------- 114
Figura 68. Diagrama de Clases Análisis Predictivo ---------- 115
Figura 69. Diagrama de Clases Componentes inicio ---------- 116
Figura 70. Resultados generales CALISOFT ---------- 117
Figura 71. Comparativa resultados promedio ---------- 118
Figura 72. Resultados evaluación Modelado ---------- 118
Figura 73. Resultados pruebas Diagrama de clases ---------- 119
Figura 74. Resultados pruebas Diagrama de casos de uso ---------- 120
Figura 75. Resultados pruebas Diagramas de secuencia ---------- 121
Figura 76. Resultados pruebas Diagramas de actividades ---------- 122
Figura 77. Resultados pruebas MER ---------- 123
Figura 78. Resultados pruebas Base de datos ---------- 123
Figura 79. Resultados SonarQube ---------- 124
Figura 80. Género vs Estrato de vivienda, Saber 11 ---------- 126
Figura 81. Matemáticas vs Estrato de vivienda, Saber 11 ---------- 127
Figura 82. Lectura Critica vs Estrato vivienda, Saber 11 ---------- 127
Figura 83. Ciencias Sociales vs Estrato de vivienda, Saber 11 ---------- 128
Figura 84. Ciencias Naturales vs Estrato de vivienda, Saber 11 ---------- 128
Figura 85. Inglés vs Estrato de vivienda, Saber 11 ---------- 129
Figura 86. Género vs Servicio TV, Saber 11 ---------- 130

Figura 87. Lectura crítica vs Género, Saber 11 ---------- 130
Figura 88. Matemáticas vs Servicio TV, Saber 11 ---------- 131
Figura 89. Ciencias naturales vs Servicio TV, Saber 11 ---------- 131
Figura 90. Ciencias sociales vs Servicio TV, Saber 11 ---------- 132
Figura 91. Inglés vs Servicio TV, Saber 11 ---------- 132
Figura 92. Género vs Tiene computador, Saber 11 ---------- 133
Figura 93. Lectura crítica vs Tiene computador, Saber 11 ---------- 133
Figura 94. Matemáticas vs Tiene computador, Saber 11 ---------- 134
Figura 95. Ciencias naturales vs Tiene computador, Saber 11 ---------- 134
Figura 96. Ciencias sociales vs Tiene computador, Saber 11 ---------- 135
Figura 97. Inglés vs Tiene computador, Saber 11 ---------- 135
Figura 98. Género vs Área ubicación colegio, Saber 11 ---------- 136
Figura 99. Lectura crítica vs Área ubicación colegio, Saber 11 ---------- 136
Figura 100. Matemáticas vs Área ubicación colegio, Saber 11 ---------- 137
Figura 101. Ciencias Naturales vs Área ubicación colegio, Saber 11 ---------- 137
Figura 102. Ciencias sociales vs Área ubicación colegio, Saber 11 ---------- 138
Figura 103. Inglés vs Área ubicación colegio, Saber 11---------- 138
Figura 104. Género vs Género del colegio, Saber 11---------- 139
Figura 105. Lectura crítica vs Género del colegio, Saber 11---------- 139
Figura 106. Matemáticas vs Género del colegio, Saber 11---------- 140
Figura 107. Ciencias naturales vs Género del colegio, Saber 11---------- 140
Figura 108. Ciencias sociales vs Género del colegio, Saber 11 ---------- 141
Figura 109. Inglés vs Género del colegio, Saber 11---------- 141
Figura 110. Género vs Jornada de colegio, Saber 11 ---------- 142
Figura 111. Lectura crítica vs Jornada de colegio, Saber 11---------- 142
Figura 112. Matemáticas vs Jornada de colegio, Saber 11 ---------- 143
Figura 113. Ciencias naturales vs Jornada de colegio, Saber 11---------- 143
Figura 114. Ciencias sociales vs Jornada de colegio, Saber 11 ---------- 144
Figura 115. Inglés vs Jornada de colegio, Saber 11 ---------- 144
Figura 116. Género vs Educación madre, Saber 11 ---------- 145
Figura 117. Lectura crítica vs Educación madre, Saber 11---------- 145
Figura 118. Matemáticas vs Educación madre, Saber 11 ---------- 146
Figura 119. Ciencias naturales vs Educación madre, Saber 11---------- 146
Figura 120. Ciencias sociales vs Educación madre, Saber 11 ---------- 147
Figura 121. Inglés vs Educación madre, Saber 11 ---------- 147
Figura 122. Género vs Educación padre, Saber 11 ---------- 148
Figura 123. Lectura crítica vs Educación padre, Saber 11 ---------- 149
Figura 124. Matemáticas vs Educación padre, Saber 11 ---------- 149
Figura 125. Ciencias naturales vs Educación padre, Saber 11 ---------- 150
Figura 126. Ciencias sociales vs Educación padre, Saber 11 ---------- 150
Figura 127. Inglés vs Educación padre, Saber 11 ---------- 151
Figura 128. Género vs Estrato de vivienda, Saber TyT ---------- 151
Figura 129. Lectura crítica vs Estrato de vivienda, Saber TyT---------- 152
Figura 130. Razonamiento cuantitativo vs Estrato de vivienda, Saber TyT------- 152

Figura 131. Comunicación escrita vs Estrato de vivienda, Saber TyT---------- 153
Figura 132. Competencias ciudadanas vs Estrato de vivienda, Saber TyT------- 153
Figura 133. Inglés vs Estrato de vivienda, Saber TyT ---------- 154
Figura 134. Género vs Pago matrícula a crédito, Saber TyT ---------- 154
Figura 135. Lectura crítica vs Pago matrícula a crédito, Saber TyT ---------- 155
Figura 136. Razonamiento cuantitativo vs Pago matrícula a crédito, Saber TyT 155
Figura 137. Comunicación escrita vs Pago matrícula a crédito, Saber TyT ------ 156
Figura 138. Competencias ciudadanas vs Pago matrícula a crédito, Saber TyT 156
Figura 139. Inglés vs Pago matrícula a crédito, Saber TyT ---------- 157
Figura 140. Género vs Tiene internet, Saber TyT---------- 157
Figura 141. Lectura crítica vs Tiene internet, Saber TyT---------- 158
Figura 142. Razonamiento cuantitativo vs Tiene internet, Saber TyT ---------- 158
Figura 143. Comunicación escrita vs Tiene internet, Saber TyT---------- 159
Figura 144. Competencias ciudadanas vs Tiene internet, Saber TyT ---------- 159
Figura 145. Inglés vs Tiene internet, Saber TyT---------- 160
Figura 146. Género vs Área de residencia, Saber TyT ---------- 160
Figura 147. Lectura crítica vs Área de residencia, Saber TyT ---------- 161
Figura 148. Razonamiento cuantitativo vs Área de residencia, Saber TyT ------- 161
Figura 149. Comunicación escrita vs Área de residencia, Saber TyT ---------- 162
Figura 150. Competencias ciudadanas vs Área de residencia, Saber TyT ------- 162
Figura 151. Inglés vs Área de residencia, Saber TyT ---------- 163
Figura 152. Género vs Estado civil, Saber TyT ---------- 163
Figura 153. Lectura crítica vs Estado civil, Saber TyT ---------- 164
Figura 154. Razonamiento cuantitativo vs Estado civil, Saber TyT ---------- 164
Figura 155. Comunicación escrita vs Estado civil, Saber TyT ---------- 165
Figura 156. Competencias ciudadanas vs Estado civil, Saber TyT ---------- 165
Figura 157. Inglés vs Estado civil, Saber TyT ---------- 166
Figura 158. Género vs Cabeza de familia, Saber TyT ---------- 166
Figura 159. Lectura crítica vs Cabeza de familia, Saber TyT ---------- 167
Figura 160. Razonamiento cuantitativo vs Cabeza de familia, Saber TyT -------- 167
Figura 161. Comunicación escrita vs Cabeza de familia, Saber TyT ---------- 168
Figura 162. Competencias ciudadanas vs Cabeza de familia, Saber TyT -------- 168
Figura 163. Inglés vs Cabeza de familia, Saber TyT ---------- 169
Figura 164. Género vs Número personas a cargo, Saber TyT ---------- 169
Figura 165. Lectura crítica vs Número personas a cargo, Saber TyT ---------- 170
Figura 166. Razonamiento cuantitativo vs Número personas a cargo, Saber TyT ---------- 170
Figura 167. Comunicación escrita vs Número personas a cargo, Saber TyT ---- 171
Figura 168. Competencias ciudadanas vs Número personas a cargo, Saber TyT ---------- 171
Figura 169. Inglés vs Número personas a cargo, Saber TyT ---------- 172
Figura 170. Género vs Educación madre, Saber TyT---------- 172
Figura 171. Lectura crítica vs Educación madre, Saber TyT---------- 173
Figura 172. Razonamiento cuantitativo vs Educación madre, Saber TyT --------- 173

Figura 173. Comunicación escrita vs Educación madre, Saber TyT ---------- 174
Figura 174. Competencias ciudadanas vs Educación madre, Saber TyT --------- 174
Figura 175. Inglés vs Educación madre, Saber TyT ---------- 175
Figura 176. Género vs Educación padre, Saber TyT ---------- 175
Figura 177. Lectura crítica vs Educación padre, Saber TyT ---------- 176
Figura 178. Razonamiento cuantitativo vs Educación padre, Saber TyT ---------- 176
Figura 179. Comunicación escrita vs Educación padre, Saber TyT ---------- 177
Figura 180. Competencias ciudadanas vs Educación padre, Saber TyT ---------- 177
Figura 181. Inglés vs Educación padre, Saber TyT ---------- 178
Figura 182. Género vs Dedicación lectura diaria, Saber TyT ---------- 178
Figura 183. Lectura crítica vs Dedicación lectura diaria, Saber TyT ---------- 178
Figura 184. Razonamiento cuantitativo vs Dedicación lectura diaria, Saber TyT 179
Figura 185. Comunicación escrita vs Dedicación lectura diaria, Saber TyT ------ 179
Figura 186. Competencias ciudadanas vs Dedicación lectura diaria, Saber TyT 180
Figura 187. Inglés vs Dedicación lectura diaria, Saber TyT ---------- 180
Figura 188. Género vs Área de la carrera, Saber TyT ---------- 181
Figura 189. Lectura crítica vs Área de la carrera, Saber TyT ---------- 181
Figura 190. Razonamiento cuantitativo vs Área de la carrera, Saber TyT -------- 182
Figura 191. Comunicación escrita vs Área de la carrera, Saber TyT ---------- 182
Figura 192. Competencias ciudadanas vs Área de la carrera, Saber TyT -------- 183
Figura 193. Inglés vs Área de la carrera, Saber TyT ---------- 183
Figura 194. Género vs Valor matrícula universidad, Saber Pro ---------- 184
Figura 195. Lectura crítica vs Valor matrícula universidad, Saber Pro ---------- 185
Figura 196. Razonamiento cuantitativo vs Valor matrícula universidad, Saber Pro ---------- 185
Figura 197. Comunicación escrita vs Valor matrícula universidad, Saber Pro --- 186
Figura 198. Competencias ciudadanas vs Valor matrícula universidad, Saber Pro ---------- 186
Figura 199. Inglés vs Valor matrícula universidad, Saber Pro ---------- 187
Figura 200. Género vs Pago matrícula a crédito, Saber Pro ---------- 187
Figura 201. Lectura crítica vs Pago matrícula a crédito, Saber Pro ---------- 188
Figura 202. Razonamiento cuantitativo vs Pago matrícula a crédito, Saber Pro 188
Figura 203. Comunicación escrita vs Pago matrícula a crédito, Saber Pro ------- 189
Figura 204. Competencias ciudadanas vs Pago matrícula a crédito, Saber Pro 189
Figura 205. Inglés vs Pago matrícula a crédito, Saber Pro ---------- 190
Figura 206. Género vs Estrato de vivienda, Saber Pro ---------- 190
Figura 207. Lectura crítica vs Estrato de vivienda, Saber Pro ---------- 191
Figura 208. Razonamiento cuantitativo vs Estrato de vivienda, Saber Pro ------- 191
Figura 209. Comunicación escrita vs Estrato de vivienda, Saber Pro ---------- 192
Figura 210. Competencias ciudadanas vs Estrato de vivienda, Saber Pro ------- 192
Figura 211. Inglés vs Estrato de vivienda, Saber Pro ---------- 193
Figura 212. Género vs Tiene internet, Saber Pro ---------- 194
Figura 213. Lectura crítica vs Tiene internet, Saber Pro ---------- 194
Figura 214. Razonamiento cuantitativo vs Tiene internet, Saber Pro ---------- 195

Figura 215. Comunicación escrita vs Tiene internet, Saber Pro ---------- 195
Figura 216. Competencias ciudadanas vs Tiene internet, Saber Pro---------- 196
Figura 217. Inglés vs Tiene internet, Saber Pro ---------- 196
Figura 218. Género vs Área de residencia, Saber Pro ---------- 197
Figura 219. Lectura crítica vs Área de residencia, Saber Pro---------- 197
Figura 220. Razonamiento cuantitativo vs Área de residencia, Saber Pro-------- 198
Figura 221. Comunicación escrita vs Área de residencia, Saber Pro---------- 198
Figura 222. Competencias ciudadanas vs Área de residencia, Saber Pro-------- 199
Figura 223. Inglés vs Área de residencia, Saber Pro ---------- 199
Figura 224. Género vs Estado civil, Saber Pro ---------- 200
Figura 225. Lectura crítica vs Estado civil, Saber Pro ---------- 200
Figura 226. Razonamiento cuantitativo vs Estado civil, Saber Pro---------- 201
Figura 227. Comunicación escrita vs Estado civil, Saber Pro ---------- 201
Figura 228. Competencias ciudadanas vs Estado civil, Saber Pro---------- 202
Figura 229. Inglés vs Estado civil, Saber Pro ---------- 202
Figura 230. Género vs Cabeza de familia, Saber Pro ---------- 203
Figura 231. Lectura crítica vs Cabeza de familia, Saber Pro---------- 203
Figura 232. Razonamiento cuantitativo vs Cabeza de familia, Saber Pro--------- 204
Figura 233. Comunicación escrita vs Cabeza de familia, Saber Pro---------- 204
Figura 234. Competencias ciudadanas vs Cabeza de familia, Saber Pro--------- 205
Figura 235. Inglés vs Cabeza de familia, Saber Pro ---------- 205
Figura 236. Género vs Número personas a cargo, Saber Pro---------- 206
Figura 237, Lectura crítica vs Número personas a cargo, Saber Pro---------- 206
Figura 238. Razonamiento cuantitativo vs Número personas a cargo, Saber Pro ---------- 207
Figura 239. Comunicación escrita vs Número personas a cargo, Saber Pro----- 207
Figura 240. Competencias ciudadanas vs Número personas a cargo, Saber Pro ---------- 208
Figura 241. Inglés vs Número personas a cargo, Saber Pro---------- 208
Figura 242. Género vs Educación madre, Saber Pro ---------- 209
Figura 243. Lectura crítica vs Educación madre, Saber Pro ---------- 210
Figura 244. Razonamiento cuantitativo vs Educación madre, Saber Pro---------- 210
Figura 245. Comunicación escrita vs Educación madre, Saber Pro ---------- 211
Figura 246. Competencias ciudadanas vs Educación madre, Saber Pro---------- 211
Figura 247. Inglés vs Educación madre, Saber Pro ---------- 212
Figura 248. Género vs Educación padre, Saber Pro ---------- 213
Figura 249. Lectura crítica vs Educación padre, Saber Pro ---------- 213
Figura 250. Razonamiento cuantitativo vs Educación padre, Saber Pro ---------- 214
Figura 251. Comunicación escrita vs Educación padre, Saber Pro ---------- 214
Figura 252. Competencias ciudadanas vs Educación padre, Saber Pro ---------- 215
Figura 253. Inglés vs Educación padre, Saber Pro ---------- 215
Figura 254. Género vs Dedicación lectura diaria, Saber Pro---------- 216
Figura 255. Lectura crítica vs Dedicación lectura diaria, Saber Pro---------- 217
Figura 256. Razonamiento cuantitativo vs Dedicación lectura diaria, Saber Pro 217

Figura 257. Comunicación escrita vs Dedicación lectura diaria, Saber Pro ------- 218
Figura 258. Competencias ciudadanas vs Dedicación lectura diaria, Saber Pro 218
Figura 259. Inglés vs Dedicación lectura diaria, Saber Pro ---------- 219
Figura 260. Género vs Área de la carrera, Saber Pro ---------- 219
Figura 261. Lectura crítica vs Área de la carrera, Saber Pro ---------- 220
Figura 262. Razonamiento cuantitativo vs Área de la carrera, Saber Pro --------- 220
Figura 263. Comunicación escrita vs Área de la carrera, Saber Pro ---------- 221
Figura 264. Competencias ciudadanas vs Área de la carrera, Saber Pro --------- 221
Figura 265. Inglés vs Área de la carrera, Saber Pro ---------- 222
Figura 266. Género vs Método programa, Saber Pro ---------- 222
Figura 267. Lectura crítica vs Método programa, Saber Pro ---------- 223
Figura 268. Razonamiento cuantitativo vs Método programa, Saber Pro --------- 223
Figura 269. Comunicación escrita vs Método programa, Saber Pro ---------- 224
Figura 270. Competencias ciudadanas vs Método programa, Saber Pro --------- 224
Figura 271. Inglés vs Método programa, Saber Pro ---------- 225
Figura 272. Matriz de confusión Random Forest ---------- 226
Figura 273. Random Tree WEKA ---------- 226
Figura 274. Reglas Random Tree ---------- 227
Figura 275. Matriz de confusión Random Tree ---------- 227
Figura 276. Árbol J48 WEKA ---------- 228
Figura 277. Matriz de confusión J48 ---------- 228
Figura 278. Árbol filtrado J48 ---------- 229
Figura 279. Reglas J48 Java ---------- 230
Figura 280. Probabilidad de las reglas generadas ---------- 230
Figura 281. Lectura crítica vs Educación madre, Predicción Saber 11 ---------- 231
Figura 282.Matemáticas vs Educación madre, Predicción Saber 11 ---------- 232
Figura 283. Ciencias naturales vs Educación madre, Predicción Saber 11 ------- 232
Figura 284. Ciencias sociales vs Educación madre, Predicción Saber 11 -------- 233
Figura 285. Inglés vs Educación madre, Predicción Saber 11 ---------- 233
Figura 286. Lectura crítica vs Educación padre, Predicción Saber 11 ---------- 234
Figura 287. Matemáticas vs Educación padre, Predicción Saber 11 ---------- 235
Figura 288. Ciencias naturales vs Educación padre, Predicción Saber 11 ------- 235
Figura 289. Ciencias sociales vs Educación padre, Predicción Saber 11 --------- 236
Figura 290. Inglés vs Educación padre, Predicción Saber 11 ---------- 236
Figura 291. Lectura crítica vs Estrato de vivienda, Predicción Saber 11 ---------- 237
Figura 292. Matemáticas vs Estrato de vivienda, Predicción Saber 11 ---------- 237
Figura 293. Ciencias naturales vs Estrato de vivienda, Predicción Saber 11 ---- 238
Figura 294. Ciencias sociales vs Estrato de vivienda, Predicción Saber 11 ------ 238
Figura 295. Inglés vs Estrato de vivienda, Predicción Saber 11 ---------- 239
Figura 296. Lectura crítica vs Tiene computador, Predicción Saber 11 ---------- 239
Figura 297. Matemáticas vs Tiene computador, Predicción Saber 11 ---------- 240
Figura 298. Ciencias naturales vs Tiene computador, Predicción Saber 11 ------ 240
Figura 299. Ciencias sociales vs Tiene computador, Predicción Saber 11 ------- 241
Figura 300. Inglés vs Tiene computador, Predicción Saber 11 ---------- 241

Figura 301. Lectura crítica vs Jornada de colegio, Predicción Saber 11---------- 242
Figura 302. Matemáticas vs Tiene computador, Predicción Saber 11---------- 242
Figura 303. Ciencias naturales vs Tiene computador, Predicción Saber 11------ 243
Figura 304. Ciencias sociales vs Tiene computador, Predicción Saber 11 ------- 243
Figura 305. Inglés vs Tiene computador, Predicción Saber 11---------- 244
Figura 306. Lectura crítica vs Educación madre, Predicción Saber TyT---------- 245
Figura 307. Razonamiento cuantitativo vs Educación madre, Predicción Saber TyT
---------- 246
Figura 308. Comunicación escrita vs Educación madre, Predicción Saber TyT- 246
Figura 309. Competencias ciudadanas vs Educación madre, Predicción Saber TyT
---------- 247
Figura 310. Inglés vs Educación madre, Predicción Saber TyT---------- 248
Figura 311. Lectura crítica vs Educación padre, Predicción Saber TyT---------- 248
Figura 312. Razonamiento cuantitativo vs Educación padre, Predicción Saber TyT
---------- 249
Figura 313. Comunicación escrita vs Educación padre, Predicción Saber TyT-- 250
Figura 314. Competencias ciudadanas vs Educación padre, Predicción Saber TyT
---------- 250
Figura 315. Inglés vs Educación padre, Predicción Saber TyT ---------- 251
Figura 316. Lectura crítica vs Estrato vivienda, Predicción Saber TyT---------- 252
Figura 317. Razonamiento cuantitativo vs Estrato vivienda, Predicción Saber TyT
---------- 252
Figura 318. Comunicación escrita vs Estrato vivienda, Predicción Saber TyT--- 253
Figura 319. Competencias ciudadanas vs Estrato vivienda, Predicción Saber TyT
---------- 253
Figura 320. Inglés vs Estrato vivienda, Predicción Saber TyT---------- 254
Figura 321. Lectura crítica vs Educación madre, Predicción Saber Pro ---------- 255
Figura 322. Razonamiento cuantitativo vs Educación madre, Predicción Saber Pro
---------- 255
Figura 323. Comunicación escrita vs Educación madre, Predicción Saber Pro - 256
Figura 324. Competencias ciudadanas vs Educación madre, Predicción Saber Pro
---------- 257
Figura 325. Inglés vs Educación madre, Predicción Saber Pro ---------- 257
Figura 326. Lectura crítica vs Educación padre, Predicción Saber Pro ---------- 258
Figura 327. Razonamiento cuantitativo vs Educación padre, Predicción Saber Pro
---------- 259
Figura 328. Comunicación escrita vs Educación padre, Predicción Saber Pro -- 259
Figura 329. Competencias ciudadanas vs Educación padre, Predicción Saber Pro
---------- 260
Figura 330. Inglés vs Educación padre, Predicción Saber Pro ---------- 261
Figura 331. Lectura crítica vs Estrato de vivienda, Predicción Saber Pro --------- 261
Figura 332. Razonamiento cuantitativo vs Estrato de vivienda, Predicción Saber Pro
---------- 262
Figura 333. Comunicación escrita vs Estrato de vivienda, Predicción Saber Pro 262

Figura 334. Competencias ciudadanas vs Estrato de vivienda, Predicción Saber Pro ---------- 263
Figura 335. Inglés vs Estrato de vivienda, Predicción Saber Pro ---------- 263
Figura 336. Lectura crítica vs Valor matrícula, Predicción Saber Pro ---------- 264
Figura 337. Razonamiento cuantitativo vs Valor matrícula, Predicción Saber Pro ---------- 265
Figura 338. Comunicación escrita vs Valor matrícula, Predicción Saber Pro ---- 265
Figura 339. Competencias ciudadanas vs Valor matrícula, Predicción Saber Pro ---------- 266
Figura 340. Inglés vs Valor matrícula, Predicción Saber Pro ---------- 267
Figura 341. Formato de seguimiento 1---------- 358
Figura 342. Formato de seguimiento 2---------- 359
Figura 343. Reuniones virtuales 1---------- 360
Figura 344. Reuniones virtuales 2---------- 360
Figura 345. Reuniones virtuales 3---------- 361

LISTA DE ANEXOS

ANEXO 1 Manual de usuario----------259

ANEXO 2 Manual técnico----------292

ANEXO 3 Artículo CACIED----------315

ANEXO 4 Artículo EAN----------321

ANEXO 5 Artículo CIETA----------324

ANEXO 6 Artículo REVISTA PENSAMIENTO UDECINO----------334

ANEXO 7 Artículo CICI----------344

ANEXO 8 Artículo ENCUENTRO DE SEMILLEROS----------347

ANEXO 9 Formatos de seguimiento----------358

INTRODUCCIÓN

En Colombia, a través del tiempo se han presentado diversas brechas de género afectando múltiples campos tantos económicos como socioculturales, y siendo por lo tanto una problemática difícil de acabar. Entrando en contexto, según el Banco de Desarrollo de América Latina, en el mercado laboral de la región las mujeres tienen una probabilidad más alta de estar ocupadas de manera informal, o de tiempo parcial, con menor remuneración, y en consecuencia siendo subrepresentadas en puestos gerenciales y ejecutivos, con respecto a los hombres. Desafortunadamente esta situación persiste a pesar de que en los últimos 50 años se han hecho esfuerzos por reducirlas. En un intento por poner una cifra a los costos para lograr una mayor igualdad de género en América Latina, Mckinsey estima que cerrar las brechas de participación laboral entre hombres y mujeres equivaldría a una ganancia de 26% en el PIB mundial (Marchionni et al., 2018). Panorama que resulta desalentador y le da una dimensión a la problemática.

En este proyecto se busca hacer énfasis en el ámbito educativo, más específicamente en las áreas de la ciencia, tecnología, ingeniería y matemáticas, también conocidas por sus siglas en inglés como STEM. Este enfoque se justifica, por múltiples cifras o estadísticas que muestran el desequilibrio que se ha venido presentando; por ejemplo, en América Latina y el Caribe (ALC) para el año 2015 el índice de mujeres graduadas en estos núcleos estuvo comprendido entre 20% y 50%; mientras que en otros campos de estudio como lo son las ciencias naturales y afines, el porcentaje es mucho mayor encontrándose en un rango del 50% y 80% (López-Bassols et al., 2018).

Por consiguiente, a través de la minería de datos, se encontrarán indicadores de brecha de género en las pruebas Saber realizadas en el grado 11, la educación superior, técnica y tecnóloga; con el fin de evaluar su comportamiento y mirar que factores contribuyen en su crecimiento. A partir de este análisis, se procederá a encontrar una relación con los núcleos temáticos en común para predecir cuál será su tendencia en las próximas pruebas.

La solución tecnológica dada para esta problemática es un software que automatice todo el procedimiento a realizarse desde la publicación de los datos de las pruebas por parte del ICFES, hasta la visualización de los resultados encontrados en las mismas. En otras palabras, consiste en un aplicativo web que de manera automática realice los análisis, y que cada vez que haya información disponible de cada una de las pruebas, estos datos sean usados para su posterior evaluación. No sin antes hacer una debida etapa de preprocesamiento, que es la encargada de la limpieza de información, su integración y transformación (García et al., 2016).

Se ha demostrado en diversos estudios internacionales que el comportamiento equitativo en los diferentes campos de acción humana contribuye a que el país avance económica y socioculturalmente, generando a su vez mejores resultados tanto en la educación como en la calidad de vida de la sociedad. De allí deriva la importancia en la realización de este proyecto. La ONU en relación con los ODS, sostiene que es fundamental “hacer una descripción analítica sobre las brechas de género en Colombia” (ONU Mujeres et al., 2016).

En este sentido, los indicadores que se presentan y analizan en el presente documento, contribuyen a entender que está pasando en la educación colombiana, cómo abordar las diferencias halladas y resultando determinante para la reducción progresiva de la desigualdad en el contexto nacional.

Como se mencionó anteriormente, la presente investigación hace uso de la minería de datos, en específico algunos de sus algoritmos y técnicas para el análisis descriptivo de las pruebas. En el proceso de descubrimiento del comportamiento de los indicadores son consideradas variables o factores que sean determinantes en las brechas de género y son clasificadas en tres grandes ramas, las cuales son de tipo: económicas, socio académicas y familiares.

Posteriormente, en el estudio se pretende relacionar los indicadores de brecha género generados con el fin de poder predecir su comportamiento y de esta manera mirar que áreas deben ser fortalecidas para así poder cerrar poco a poco las diferencias entre hombres y mujeres en términos académicos, haciendo que la educación mejore su calidad.

I. INFORME DE INVESTIGACIÓN

1.1. ESTADO DE ARTE

El Instituto Colombiano para el Fomento de la Educación Superior, conocido por sus siglas ICFES es la entidad encargada de realizar las pruebas estandarizadas en Colombia. Dichas pruebas son exámenes realizados por el estado con el fin de evaluar el nivel educativo de los estudiantes y a su vez identificar la calidad y posicionamiento que tienen cada una de las instituciones educativas a nivel nacional.

Las pruebas Saber 11 son realizadas en el último nivel de la Educación Media, con el objetivo de comprobar las competencias desarrolladas por los estudiantes en su proceso escolar y verificar que tienen los conocimientos necesarios para la Educación Superior. Saber Pro son los exámenes que se presentan en el último año del nivel educativo Superior en el país, tiene contenidos genéricos para toda carrera y específicos de acuerdo con su área de formación. Mientras que, Saber TyT son mecanismos de evaluación diseñados para comprobar el desarrollo de las competencias específicas para estudiantes que cursan técnicos y tecnólogos.

A través del tiempo, investigadores han asegurado que hay relación directa entre la baja proporción de mujeres en las áreas de las ciencias e ingenierías con la segregación y desigualdad en la sociedad (Loury, 2016). Por su parte, Schroter & Skyt (2013) afirman que entre más habilidad se demuestre en estas áreas más posibilidad hay de que sean estudiadas por las mismas y el acceso a la universidad en estas facultades aumente.

Según un informe del World Economic Forum a finales de 2018, las brechas entre hombres y mujeres se han cerrado un 66% en los últimos 30 años. Se evidencia que mientras más desarrollado sea el país, la equidad entre géneros es mayor; sin embargo, las brechas no son solo en los países pobres. En el STEM (Ciencias, tecnología, ingeniería y matemáticas) las diferencias de acceso y trabajo son todavía grandes. Un estudio que presentó LinkedIn también a finales de 2018 muestra una brecha de género evidente entre los profesionales de AI (Inteligencia Artificial), donde las mujeres representan solo el 22% de la fuerza laboral (Hurtado, 2019).

Haciendo enfoque en el área de las ciencias exactas, los investigadores de la Universidad de Wisconsin-Madison analizaron en 86 diferentes países los resultados de pruebas de habilidad matemática encontrando que, en las culturas donde existía mayor igualdad de oportunidades y trato entre ambos sexos, la diferencia era mínima en los resultados. En el caso de los Estados Unidos se ha venido disminuyendo la brecha. Si en 1970 sólo una de cada 13 personas capaces de obtener 700 puntos en la prueba SAT (Scholarship Aptitude Test) antes de los 13 años era mujer, hoy la cifra es de 1 a 3 y se sigue cerrando camino a la igualdad. Un caso diferente es el de Marruecos, país donde las pruebas demostraron una igualdad entre los resultados obtenidos por hombres y mujeres. En este país musulmán, niños y niñas asisten a escuelas diferentes y, por lo tanto, no existe presión ni competencia entre ambos sexos en la sala de clases. En términos generales el estudio indica que, la igualdad de oportunidades contribuye con la equidad de género en los resultados de pruebas de matemáticas (Universia, 2013).

No solo se evidencian brechas de género, en el caso de la educación rural y urbana también hay diferencias sustanciales no sólo en los resultados de pruebas sino también en el acceso a las mismas. Según cifras dadas por el Ministerio de Educación obtenidas gracias al Índice Sintético de Calidad Educativa (ISCE), en Colombia la brecha entre la educación rural y urbana se hace evidente desde la educación temprana; en el 2017 los años promedio de educación para la población mayor de 15 años en zonas rurales fue de 6, mientras que en las zonas urbanas fue

de casi 10. También se ven afectados los resultados en las pruebas estandarizadas, en las instituciones educativas cerca del 50% tiene un desempeño inferior o bajo. Además, solo entre el 2 y 3 por ciento de la población rural han realizado estudios universitarios, el 10 por ciento son bachilleres y el alfabetismo solo alcanza el 12 o 13 por ciento en mayores de 15 años (Educación, 2016).

De hecho, “en el censo agropecuario que realizó el DANE en 2014 encontró que 20 por ciento de la población entre los 5 y 16 años no asistió a ninguna institución como jardín, escuela, colegio o universidad” (Educación, 2019). Esta situación, como es de esperarse, profundiza las brechas sociales y de desigualdad. Algunos índices preocupantes tienen que ver con el acceso a los servicios básicos tales como la electricidad, las líneas telefónicas, el internet y el agua potable, confirmando que, lejos de cerrarse, estas brechas se mantienen y se complejizan (Hernández Bonilla, 2018).

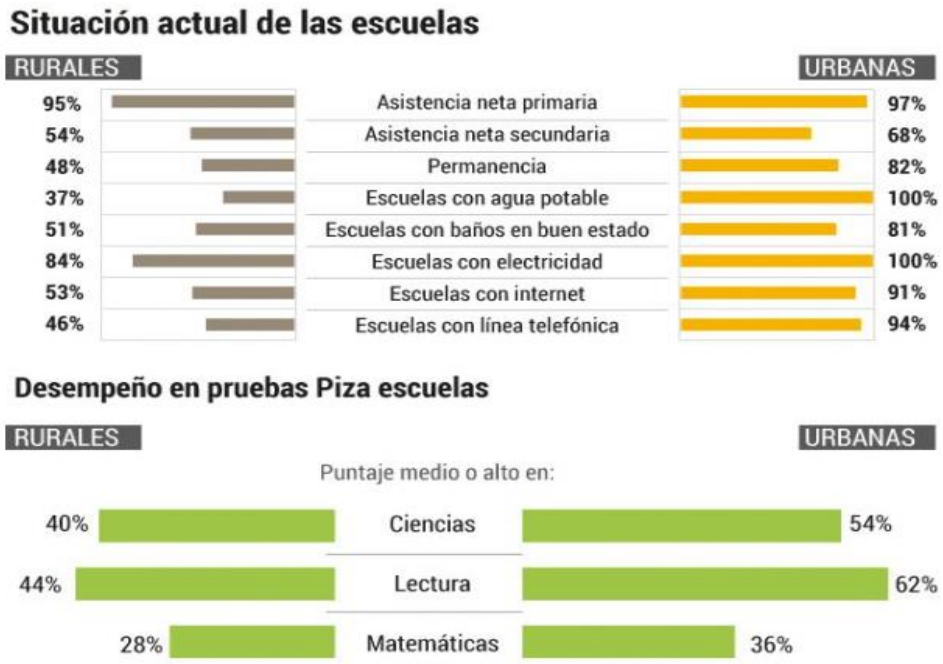


Figura 1. Situación actual de las escuelas y desempeño en pruebas Piza
Autores: (Hernández Bonilla, 2018)

Así como se puede ver en la Figura 1. la situación académica ha sido principalmente estudiada y se ha mostrado mayor interés en la educación básica, dejando en segundo lugar la educación media y la educación superior. Desde el punto de vista teórico es importante tener en cuenta que hay tres tipos de variables que son factores determinantes o predictores, las cuales son las psicológicas, las sociológicas y las didácticas (Zavala Guirado et al., 2018).

Son muchos los estudios que organizaciones e investigadores han realizado a nivel mundial sobre la desigualdad, pero a nivel de software no han sido desarrollados gran cantidad de aplicativos que contribuyan en la detección de dichas brechas y ayuden en su predicción. Hasta el momento existen softwares de predicción de brechas de género en el campo corporativo que se enfocan específicamente en las diferencias salariales como, por ejemplo, CuroGPG y CEO de GapSquare. Así mismo, también se puede encontrar múltiples aplicativos que hacen uso de la IA para la detección y predicción en varias áreas, tales como la meteorología, la música, las matemáticas, las finanzas y la medicina.

Sin embargo, lo más usado para predicción es la minería de datos que según Pérez & Santín se define como “un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos” (Pérez López & Santín González, 2008). Para lograr su propósito hace uso de técnicas las cuales de forma automática analizan la información almacenada en las bases de datos organizándola y descubriendo patrones, perfiles y tendencias. Las técnicas más conocidas son: árboles de decisión, redes neuronales, regresión lineal y reglas de asociación.

En primer lugar, los árboles de decisión son un mecanismo que se encarga de agrupar los datos partiendo de una condición inicial, sus decisiones tienen como respuesta un “sí” o “no” (Gutiérrez O. & Molina, 2016). Las redes neuronales por su parte son conjuntos de elementos interconectados entre sí que simulan los procesos neuronales (Cataldi et al., 2007), por otra parte, la regresión lineal se caracteriza por el uso de una variable aleatoria que es función lineal de otra variable X formando una ecuación, siendo X la variable predictora y la variable respuesta Y. Por último,

las reglas de asociación son las que encuentran posibles relaciones o correlaciones entre sucesos aparentemente independientes (Molina López & García Herrero, 2006).

Al centrarse en el ámbito educativo el uso de minería de datos se encuentra en auge, específicamente se ha venido enfocando en la pronosticación de calificaciones y rendimiento académico mayoritariamente. Algunos campos de estudio aplicando la minería de datos en la educación son los propuestos por Cristóbal Romero y Sebastián Ventura, mostrados en la Figura 2.

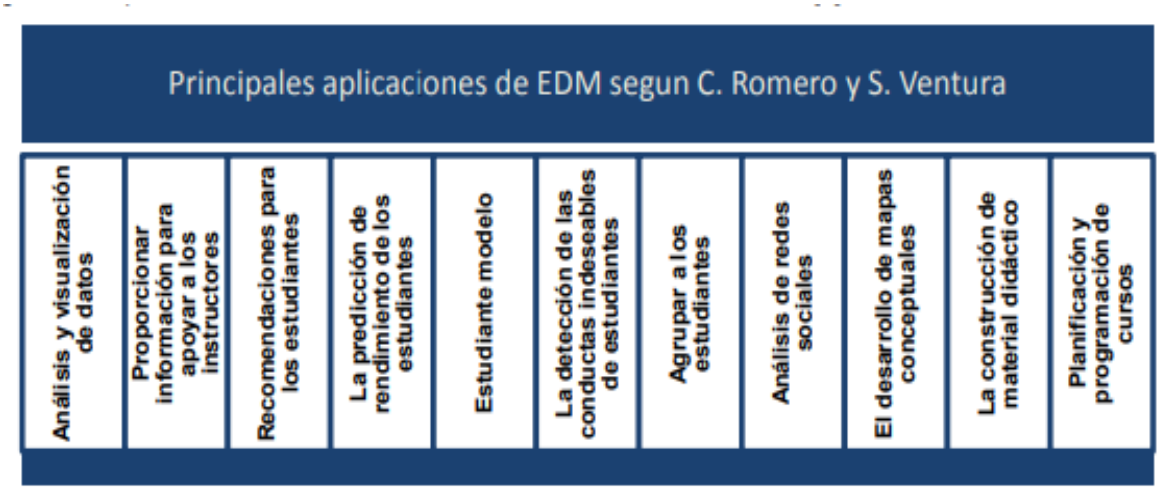


Figura 2. Aplicaciones de minería de datos en la educación
Autores: (Romero & Ventura, 2016)

Existen desarrollos en el mundo que la han aplicado como es el caso de Chile, donde se implementó un Data Warehouse y una arquitectura de Red Neuronal Artificial para el análisis y la predicción de rendimiento académico de los estudiantes de Ingeniería Civil en Computación e Informática de la Universidad de Atacama con la posibilidad de predecir el número de asignaturas aprobadas por los alumnos, a partir del uso de su propia información histórica (Zambrano et al., 2011).

En España, en la Universidad Complutense de Madrid se hizo uso de Big Data y algoritmos de aprendizaje automático para determinar las notas de los estudiantes, allí se logró hacer una comparación entre grupos, cursos y alumnos de forma

gráfica. También identificaron los estudiantes que iban a bajar su rendimiento académico para poder así enfocarse en ellos. En este proyecto no fue posible definir con exactitud las calificaciones de los alumnos, puesto que existen variables externas que no pueden ser evaluadas y que de una u otra forma alteran el resultado, pero si tiene una alta tasa de acierto (León, 2018).

A nivel nacional en el año 2013, el ICFES realizó un análisis a partir de las pruebas Saber 5, 9 y 11 con el fin de estimar los factores que generaban diferencias de género en los estudiantes colombianos en las áreas de las matemáticas y el lenguaje (enfocadas a las pruebas presentadas en el año 2009) aplicando modelos matemáticos para calcular el porcentaje de diferenciación (Caro Acero & Casas M., 2013). En la Universidad Nacional en el año 2015, realizaron un trabajo donde aplicando técnicas de análisis de agrupamiento de la minería de datos construyeron un modelo descriptivo para entender la estructura de los datos de la prueba Saber 11 (año 2012), enfocándose en el departamento del Cesar (Villafañe Blanco, 2015). En términos generales en los diversos estudios realizados en el país se han enfocado en el nivel educativo de los padres y el contexto socioeconómico de los estudiantes evaluados, entre otros factores.

En la Universidad de Cundinamarca extensión Facatativá para el año 2016, se realizó un proyecto de grado donde haciendo uso de técnicas de minería de datos se analizan y evalúan los patrones de conocimiento sobre caracterización de perfiles de aspirantes al programa de Ingeniería de Sistemas con el fin de determinar el porcentaje de deserción presentado en el programa (Chaves Sánchez et al., 2016). Asimismo, este proyecto contó en el año 2017 con una segunda fase que fue la construcción del aplicativo web encargado de la implementación de los algoritmos desarrollados en la fase 1 (Gómez Pinzón & Valenzuela Sabogal, 2017). También en la universidad, sede Fusagasugá, en el año 2017 como proyecto ya culminado se hizo uso de Machine Learning en un sistema recomendador web para la toma de decisiones en la adquisición de equipos de cómputo.

Se propone para nuestro proyecto el desarrollo de un aplicativo web de análisis predictivo para los resultados de las pruebas estandarizadas aplicadas en los

estudiantes colombianos específicamente Saber 11, Saber Pro y Saber TyT teniendo como principal variable en consideración, el género; puntualmente se quiere realizar una comparación de los resultados obtenidos por las mujeres con respecto a los de los hombres en los 4 módulos básicos que tienen en común las pruebas anteriormente mencionadas, que son razonamiento cuantitativo, competencias ciudadanas, inglés y lectura crítica para de esta manera, poder predecir el comportamiento futuro de los estudiantes en estos módulos e identificar las brechas de género y las diferencias existentes entre departamentos del país.

A nivel de software se tendrá una interfaz de administrador donde este podrá adjuntar los archivos .txt que contienen el compilado de resultados de las pruebas brindados el ICFES como datos abiertos. Por medio de gráficos y en reportes, el usuario podrá visualizar los resultados en cada una de las pruebas y por cada uno de los núcleos.

1.2. LÍNEA DE INVESTIGACIÓN

Se trata de un aplicativo web que se basa en el análisis de indicadores que afectan las brechas de género en las pruebas Saber 11, Saber Pro y Saber TyT, mediante el uso de algoritmos de minería de datos para realizar un análisis descriptivo, al igual que un análisis predictivo relacionando los núcleos temáticos de dichas pruebas. Por ello, este proyecto se desarrollará bajo la línea de investigación **Software, sistemas emergentes y nuevas tecnologías**.

1.3. PLANTEAMIENTO DE PROBLEMA Y PREGUNTA DE INVESTIGACIÓN

Desde hace tiempo en el contexto mundial, la mujer ha querido disminuir las desigualdades y estereotipos presentes en gran parte de los ámbitos socioculturales con respecto a los hombres, sin embargo, aún hay elementos en dónde se evidencian dichas diferencias. Ahora bien, si nos enfocamos en el sector educativo y de la tecnología, en la actualidad las mujeres no tienen la mejor percepción por parte de la sociedad en estos campos, así como tampoco en las matemáticas y ciencias exactas. Según un informe realizado para el día internacional de la mujer en el año 2019 por la firma IPSOS, siguen existiendo múltiples indicadores que preocupan y contribuyen con las brechas de género, tales como:

- Chile, Colombia y Argentina, son los países en Latinoamérica donde más ventaja competitiva tienen los hombres tanto en campos educativos, laborales y sociales.
- Colombia es el país en Latinoamérica que registra menos porcentaje de consideración sobre la idea de que las mujeres tienen más ventajas sobre los hombres.
- La mayoría de los encuestados creen que la discriminación hacia la mujer se acabará en 20 años gracias a la tecnología, la educación y la participación e inclusión en los diferentes campos laborales.
- En el año 2018, en el World Economic Forum se señaló que las diferencias se presentan más en la política, ciencias, educación y salud en Colombia, asimismo en los últimos años Colombia ha decaído hasta la posición 40 en el ranking global de brechas de género (Ipsos, 2019).

Teniendo en cuenta lo anterior, en el proyecto se quiere abordar esta problemática centrándose en las personas que presentan las pruebas estandarizadas nacionales desde el año 2015, debido a que en este año se presentó un cambio en el formato

de evaluación, y haciendo énfasis en los siguientes núcleos: lectura crítica, inglés, competencias ciudadanas y razonamiento cuantitativo. Dándole importancia tanto a los estudiantes que realizan sus carreras de pregrado como a los que estudian técnicos y tecnólogos, se realizará el estudio de la información para 3 de las 6 posibles pruebas existentes, las cuales son, Saber 11, Saber Pro y Saber TyT; todo esto con el objeto de determinar qué pasa con la situación académica de la población juvenil del país.

¿Cómo determinar mediante una herramienta informática indicadores de brecha de género en las pruebas saber de la educación media, superior, técnica y tecnológica?

1.4. OBJETIVO GENERAL Y OBJETIVOS ESPECÍFICOS

1.4.1. OBJETIVO GENERAL

Desarrollar un aplicativo web para el análisis predictivo usando técnicas de minería de datos, para la construcción de indicadores de brechas de género en el país, de acuerdo, a los datos dados por el ICFES de las pruebas saber 11, saber Pro y saber TyT.

1.4.2. OBJETIVOS ESPECÍFICOS

- Refinar los requerimientos necesarios para el funcionamiento del aplicativo de tal manera que se adapten para el análisis de las pruebas.

- Implementar modelo descriptivo de los índices de brechas de género, por medio de técnicas de minerías de datos.

- Diseñar el aplicativo web para dar a conocer los resultados del análisis basado en las pruebas del estado.

- Elaborar la herramienta informática según su diseño.

- Construir escenarios de prueba para analizar el modelo predictivo por medio de algoritmos basados en técnicas de minería de datos.

- Realizar pruebas del sistema a nivel de software sobre un modelo de calidad.

1.5. ALCANCE E IMPACTO DEL PROYECTO

En el ámbito mundial la ONU en su plan de desarrollo sostenible tiene como uno de sus objetivos principales promover la inclusión y el acceso para todas las personas a una educación de calidad, y así mismo eliminar las diferencias de género en este campo.

La realización de un proyecto que se enfoque en el análisis de las brechas de género existentes en el ámbito educativo ayuda a saber que tanto han funcionado las estrategias planteadas por el gobierno y las instituciones educativas, y como es la realidad de este sector.

De esta forma este proyecto, el cual se enfoca en las pruebas ICFES Saber 11, Saber Pro y Saber TyT y la predicción de los puntajes obtenidos clasificados por departamentos, permitirá reducir cifras preocupantes para el país como que tan solo el 25,9% de los estudiantes de ingenierías y arquitectura son mujeres en Colombia, y de esta manera contribuir en los siguientes objetivos de desarrollo sostenible:

4. Educación de calidad
5. Igualdad de género
9. Industria, innovación e infraestructura

En Colombia, la Ley 823 de 2003 es la encargada de proteger la igualdad de oportunidades para las mujeres en todos los sectores, eliminando todos los obstáculos que no permiten su cumplimiento. Este proyecto podrá brindarle información a futuras iniciativas que tengan como fin buscar soluciones a las problemáticas que se presentan en la educación colombiana y aquellas que afecten el empoderamiento de las mujeres.

1.6. METODOLOGÍA

Este proyecto se desarrollará bajo los parámetros de la metodología cuantitativa, en palabras de Hernández Sampieri “La investigación cuantitativa nos ofrece la posibilidad de generalizar los resultados más ampliamente, nos otorga control sobre los fenómenos, así como un punto de vista de conteo y las magnitudes de estos.” (Hernández Sampieri et al., 2006). Se elige esta metodología porque permite contabilizar los datos por medio de análisis matemático arrojando resultados con una precisión definida, además puede ser generalizable a partir de la muestra de una población determinada. Esto nos permitirá estudiar con mayor claridad la asociación o la relación entre las variables cuantificadas (Fernández & Díaz, 2002).

Para el desarrollo de un modelo descriptivo que muestre si existe relación entre ciertos factores y la brecha de género se hará uso de técnicas de minería de datos, específicamente técnicas de asociación, las cuales “establecen las posibles relaciones entre diferentes acciones aparentemente

independientes entre sí, reconociendo como pueden ocurrir los eventos o acciones a partir de la aparición de otros” (Reyes Nava et al., 2017). La metodología CRISP-DM, existe para orientar el proceso de minería de datos, facilita la realización de proyectos optimizando su planificación, reduciendo su complejidad y permitiendo un mejor seguimiento de estos. Tiene como ventaja la centralización de los objetivos realizando un análisis del problema para transformarlo en un problema técnico de minería de datos, para este proyecto ha sido escogida por su capacidad de definir

una estructura y orientar su ejecución (Cobos et al., 2010). Durante este proceso de construcción del modelo se evaluará cual algoritmo tiene mejor porcentaje de acierto, con el fin de obtener resultados lo más precisos posibles.

En la construcción del aplicativo web se implementará la metodología de desarrollo ágil SCRUM, ya que es un modelo general de gestión de entornos de producción basados en rutinas, donde el factor más importante es el conocimiento aplicado de las personas que hacen parte del proyecto, aún más primordial que el mismo proceso tecnológico (Palacio & Ruata, 2011). Por otra parte, su importancia radica también en las iteraciones que permiten el desarrollo incremental del proyecto desde su concepción hasta que se dé por terminado.

1.7. MARCOS DE REFERENCIA

1.7.1. MARCO TEÓRICO

1.7.1.1. BIG DATA Y MINERIA DE DATOS

Con la llegada de la llamada revolución digital y la generación masiva de información, ya hace algunos años, nace la **Big Data**, que es el término utilizado para hablar de grandes cantidades de datos, el cual se ha convertido en auge y día a día se encuentra en exploración contribuyendo a su crecimiento. Vale la pena aclarar que al hablar de Big Data no sólo se hace referencia al gran tamaño, sino que también abarca el volumen y variedad de datos, así como la velocidad de procesamiento de estos. Para la fase de almacenamiento se han venido generando nuevas estrategias de análisis, debido a que los sistemas convencionales no tienen la suficiente capacidad de gestión (Hernández-Leal et al., 2015). De lo anterior, se puede destacar la minería de datos como una de esas formas creadas para afrontar las tareas que conlleva procesar datos masivos.

Al hablar de minería de datos nos referimos a una tecnología que se creó con el fin de ayudar a la comprensión de aquellas bases de datos complejas por medio del

reconocimiento y extracción de patrones repetitivos de un considerable volumen de datos.

La minería de datos (Data Mining - DM) se conoce como una tecnología creada con el fin de ayudar a la comprensión de aquellas bases de datos complejas por medio del descubrimiento de nuevas relaciones, patrones y tendencias, haciendo uso de la estadística junto con diversas herramientas informáticas y diversos procedimientos que ayudan a la automatización. Dichas técnicas usadas ayudan a describir y comprender la información de mejor forma, haciendo posible la predicción de hechos futuros (Pérez López & Santín González, 2008).

1.7.1.2. TÉCNICAS DE MINERIA DE DATOS

Existen dos tipos de análisis que agrupan las técnicas anteriormente mencionadas según el resultado que se quiera obtener, y son descritos a continuación:

El análisis descriptivo se centra en la organización y clasificación de la información, busca irregularidades en los datos brindando información de las relaciones y características que surgen entre ellos; esto con el fin de extraer todo el conocimiento posible y de interés presentándolo de una forma comprensible. Las técnicas usadas en este tipo de análisis se centran en un aprendizaje no supervisado, por ello también es llamado modelo no supervisado (Carmona et al., 2015).

Una de las técnicas más usadas al momento de hablar de modelo descriptivo son *las reglas de asociación*. Con las reglas de asociación se realiza el análisis extrayendo los datos según las coincidencias que existan entre ellos, sus reglas siguen la estructura “si X entonces Y” ($X \Rightarrow Y$) (De Moya Amaris, ME, 2003). Cada uno de los atributos que se encuentra involucrado en las relaciones encontradas es llamado ítem y el conjunto de estos se llama itemsets. Existen varios algoritmos cuya función principal es la de identificar itemsets frecuentes y dividen esta técnica de minería de datos, entre los 3 más usados se encuentran: Algoritmo Apriori, FP-Growth y Eclat (Rodrigo, 2018).

El algoritmo Apriori primero identifica todos los itemsets que ocurren más frecuentemente dándole cierto límite y cuando ya ha finalizado convierte estos itemsets en reglas de asociación. FP-Growth por su parte emplea para el almacenamiento una estructura del tipo árbol, permitiendo comprimir la información hasta 200 veces, después de esto divide la información en varias bases de datos según el patrón frecuente, luego analiza cada una por separado y concatena los resultados obtenidos. Para finalizar, el algoritmo Eclat analiza todas las transacciones en formato vertical, donde cada línea contiene un solo ítem y las transacciones que lo contienen (Rodrigo, 2018).

Por otra parte, las técnicas aplicables al análisis predictivo se clasifican en dos ramas, que son técnicas de regresión y técnicas de aprendizaje computacional. Tienen como objetivo descubrir futuros comportamientos de atributos en específico. Algunos de los métodos principales son: Arboles de decisión, regresiones y series temporales, redes neuronales, método bayesianos y máquinas de soporte vectorial. El funcionamiento de los *árboles de decisión/clasificación* es analítico nodo por nodo, de forma que va buscando el mayor grado de pureza entre los grupos resultados; empezando por la raíz, que "contiene la totalidad de la información hasta los nodos terminales, donde se tiene la proporción de individuos ubicados en cada una de las categorías que toma la variable de interés" (Suárez Soto, 2014).

1.7.1.3. METODOLOGÍAS

Existen diferentes metodologías de DM, las más conocidas y usadas en este campo son KDD, CRISP-DM, SEMMA y Catalyst. En términos generales se puede decir que todas contemplan en primer lugar la selección y preparación de datos; para después pasar a la fase de hallazgo y evaluación de patrones. Pero al mirar en detalle su comportamiento: KDD, CRISP-DM y Catalyst analizan y comprenden el problema antes de comenzar el proceso de minería, mientras que SEMMA no realiza esta actividad en su modelo. Por otra parte, la fase de evaluación de los patrones hallados al ser la principal etapa para un proceso descriptivo de los datos

se encuentra presente en todas las metodologías. Pero la interpretación de estos patrones en SEMMA se realiza sobre la ejecución del modelo y no cuenta con una implementación de resultados, mientras que en las otras metodologías la evaluación se realiza en función del problema que se desea solucionar o examinar. En CRISP-DM, se propone además una planificación para el control futuro y un análisis de cierre del proyecto, conocido como “análisis postmortem”, este consiste en encontrar información objetiva acerca de la trayectoria de un proyecto (Moine et al., 2011). Este proceso tiene seis fases las cuales se encuentran representadas en el siguiente gráfico:



Figura 3. Fases de la metodología CRISP-DM
Autores: (Timaran & Jiménez, 2014)

En el análisis del problema se busca comprender cual es la situación que requiere una solución por medio de la minería de datos, recolectando de esta manera la información necesaria para interpretar los resultados. Ya teniendo definida la problemática, se pasa a la fase de análisis de los datos donde se hace la recolección inicial de estos, los cuales establecen las primeras hipótesis según las relaciones obtenidas.

En el periodo de preparación de los datos se realiza toda la transformación necesaria para que estos puedan ser manipulados; al finalizar, se da paso al modelado donde se seleccionan las técnicas más adecuadas según el proyecto que se esté tratando, aquí es donde hace uso de las técnicas y los algoritmos de minería de datos. En la fase de evaluación se interpretan los patrones que se lograron obtener, buscando con ello el consolidar todo el conocimiento que se pueda adquirir y si los resultados son los esperados, se procede a su explotación o implementación.

1.7.1.4. SISTEMA

Para el desarrollo del aplicativo web se va a hacer uso del lenguaje de programación Java. Este lenguaje orientado a objetos busca la creación de proyectos que se puedan ejecutar sin importar el sistema operativo que tenga el dispositivo. Se ha convertido en uno de los lenguajes más populares debido a las grandes posibilidades de utilización (Oracle Corporation, n.d.).

Como gestor de base de datos se utilizará PostgreSQL el cual se considera como el motor más actualizado. Para su administración se hace uso de PgAdmin que permite realizar consultas SQL y generar la base de datos por medio de interfaz gráfica (Platzi, 2015).

Por último, en relación con la aplicación de los algoritmos de DM, el software elegido ha sido WEKA (Waikato Environment for Knowledge Analysis), el cual cuenta con una interfaz gráfica de usuario de fácil uso o una API en Java que permite a través de sus múltiples herramientas integradas la realización de tareas para el aprendizaje automático, además que es de código abierto y es usado en campos tanto de enseñanza como industrializados (University of Waikato, n.d.).

1.7.2. MARCO LEGAL

Este proyecto está delimitado dentro de los aspectos legales relacionados con el desarrollo del software y la minería de datos. Así como también el uso de los datos suministrados por el ICFES, más exactamente la privacidad y protección de la información, sin dejar a un lado los derechos de autor de la obra. A continuación, se describen las normas y leyes colombianas que están involucradas en la presente investigación:

En primer lugar, el artículo 15 de la constitución política de Colombia de 1991 establece los principios fundamentales para la recolección, tratamiento y circulación de datos personales; así como los procedimientos que se encargan de respetar la libertad y demás garantías consagradas en la constitución (Corte Constitucional, 1991).

La ley 23 de 1982 dicta las disposiciones sobre protección de derechos de autor para obras literarias, científicas y artísticas sin importar su modo de expresión (LEY NÚMERO 23 DE 1982, 1982).

La ley Estatutaria 1266 de 2008, tipifica “las disposiciones generales del hábeas data y regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países además de otras disposiciones referidas al tema” (Ley 1266 de 2008, 2008).

El decreto número 1377 de 2013 reglamenta la ley 1581 de 2012, que contiene las “disposiciones generales para la protección de datos personales”, es una ley que promueve la integridad de la información para que no queden expuestos y susceptibles al ser tratados ya sea por entidades públicas como privadas. (Decreto Número 1377 de 2013, 2013).

La ley 1273 de 2009, estipula los delitos relacionados con la protección de la información y de los datos en sistemas que utilizan las tecnologías de la información y las comunicaciones, acondicionando así el Código Penal nacional (Ley 1273 de 2009, 2009).

En este caso, al hacer uso de los datos suministrados por el Instituto Colombiano para la Evaluación de la Educación (ICFES) se tiene en cuenta que dichos datos están protegidos por derechos de propiedad intelectual y derechos de autor a favor del ICFES. En consecuencia, el instituto prohíbe el uso o publicación total o parcial de este material con fines de lucro, y por ende solo autoriza a quienes hagan su uso para fines académicos e investigativos (Caro Acero & Casas M., 2013).

II. MODELO DESCRIPTIVO

2.1. EXTRACCIÓN DE LOS DATOS

Como primera opción para la automatización de la extracción de información se hizo uso de tareas programadas con la ejecución de archivos .bat, como se puede apreciar en la imagen este es un script que hace la autenticación en el FTP, busca la carpeta que desea copiar y el lugar de destino. Estas líneas de comando se encuentran en un archivo .txt, el cual es llamado por el archivo .bat. que a su vez es ejecutado por la tarea programada en búsqueda del servidor FTP.

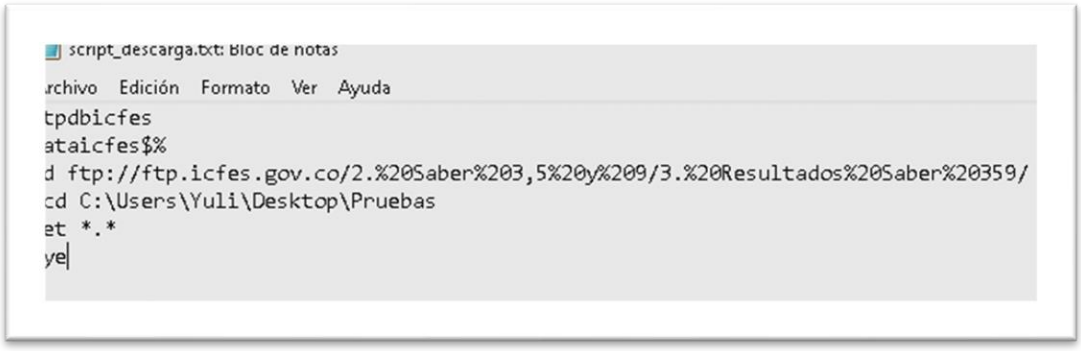


Figura 4. Script descarga
Autores: (Manrique, Celeita, Barahona, 2020)

En la actualidad este procedimiento tiene muchos contras, ya que el acceso al servidor por medio de comandos CMD puede ser bloqueado, ya sea por el firewall o bloqueo de puertos por los proveedores de internet. Por esta razón no pudo ser implementada esta forma de automatización.

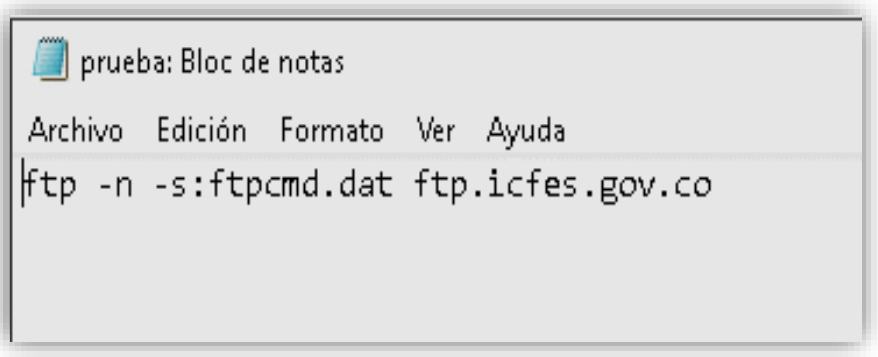


Figura 5. Archivo .bat
Autores: (Manrique, Celeita, Barahona, 2020)

Como segunda opción, se creó un VBScript, el cual intentaba hacer el mismo proceso de autenticación con el servidor en búsqueda de la obtención de su información, pero siendo éste otro tipo de archivo. Sin embargo, al igual que el anterior es afectado por el bloqueo de puertos.

other: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
'Option Explicit
'const progname="FTP upload script by Richard Finegold"
'const url = "ftp.icfes.gov.co"
'const rdir = "C:\Users\Yuli\Desktop\Pruebas"
'const user = "ftpdbicfes"
'const pass = "dataicfes$%"

'This is an example of ftp'ing without calling the external "FTP" command
'It uses InetCtrls.Inet.1 instead
'Included is a "hint" for simple downloading

'Sources:
'http://msdn.microsoft.com/library/partbook/ipwvb5/loggingontoftpserver.htm
'http://msdn.microsoft.com/library/partbook/egvb6/addinginternettransfercontrol.htm
'http://cwashington.netreach.net/ - search on "ftp" - inspiration only!

'Insist on arguments
dim objArgs
Set objArgs = Wscript.Arguments
If 0=objArgs.Count Then
   MsgBox "No files selected for operation!", vbOkOnly + vbCritical, progname
   WScript.Quit
End If
```

Figura 6. Archivo VBScript
Autores: (Manrique, Celeita, Barahona, 2020)

Después se creó una aplicación de escritorio en Java, con el fin de hacer la conexión y descarga de cada uno de los archivos existentes en el servidor.

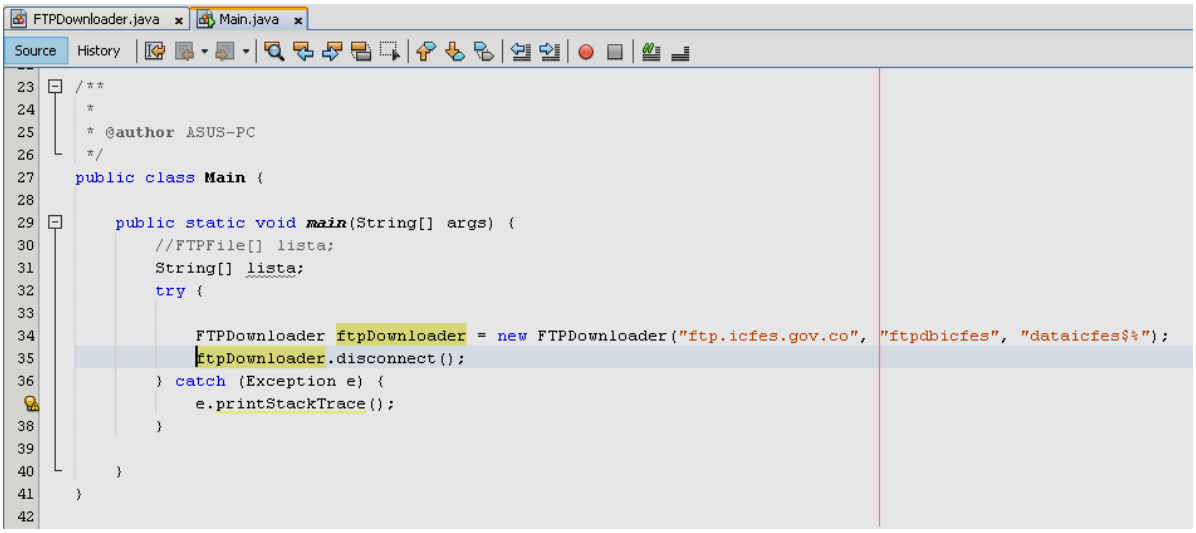


Figura 7. Conexión FTP Java
Autores: (Manrique, Celeita, Barahona, 2020)

Para lo cual fue llevado a cabo un procedimiento donde primero accedía al servidor y comenzaba a traer los archivos tipo “InputStream”; vale la pena aclarar que para poder ser guardado debe existir de forma local un archivo del mismo formato que permita ser sobrescrito. Este método restringe la navegación entre carpetas, lo que dificulta la descarga de todas las pruebas, puesto que éstas se encuentran almacenadas dentro de carpetas. Lo sucedido con este lenguaje también ocurre en otros tales como PHP.

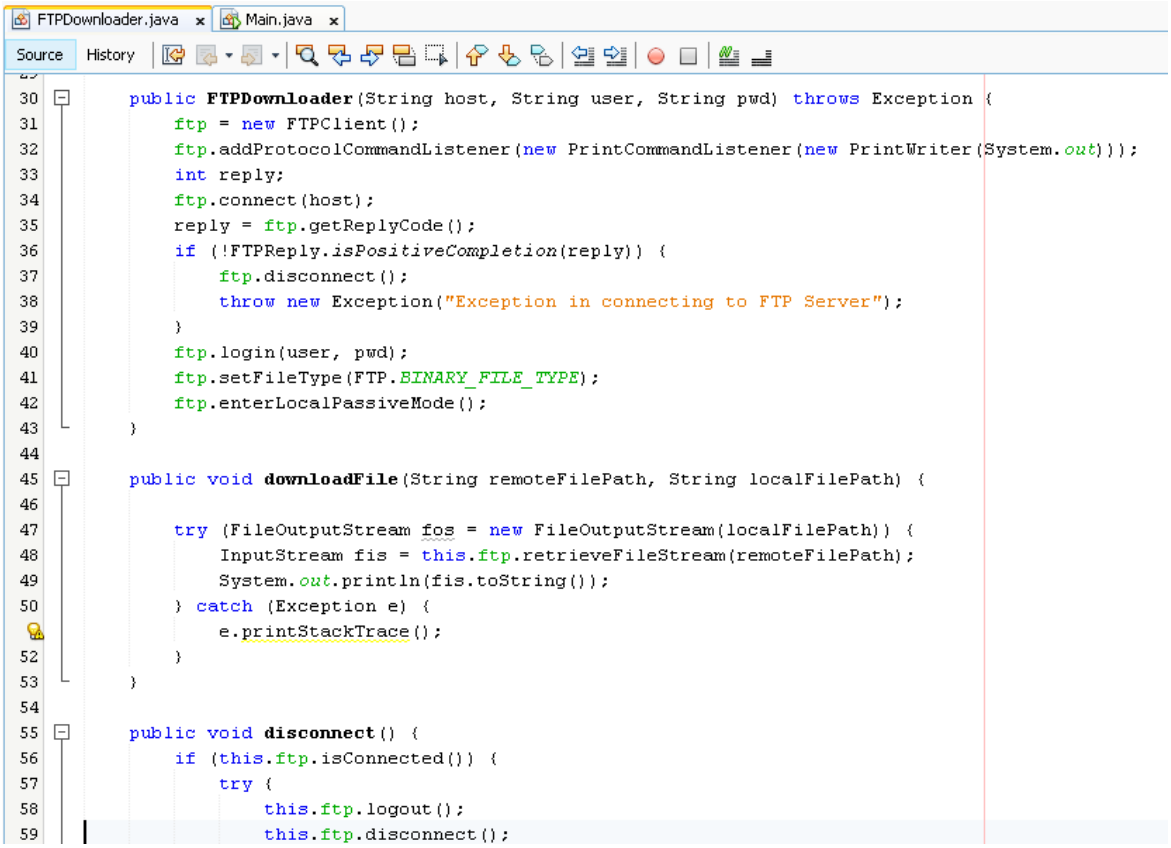


Figura 8. Descarga archivos Java
Autores: (Manrique, Celeita, Barahona, 2020)

Al no funcionar ninguna de las opciones mostradas anteriormente se procedió a hacer uso de un software libre para facilitarle el procedimiento de descarga de archivos al administrador.

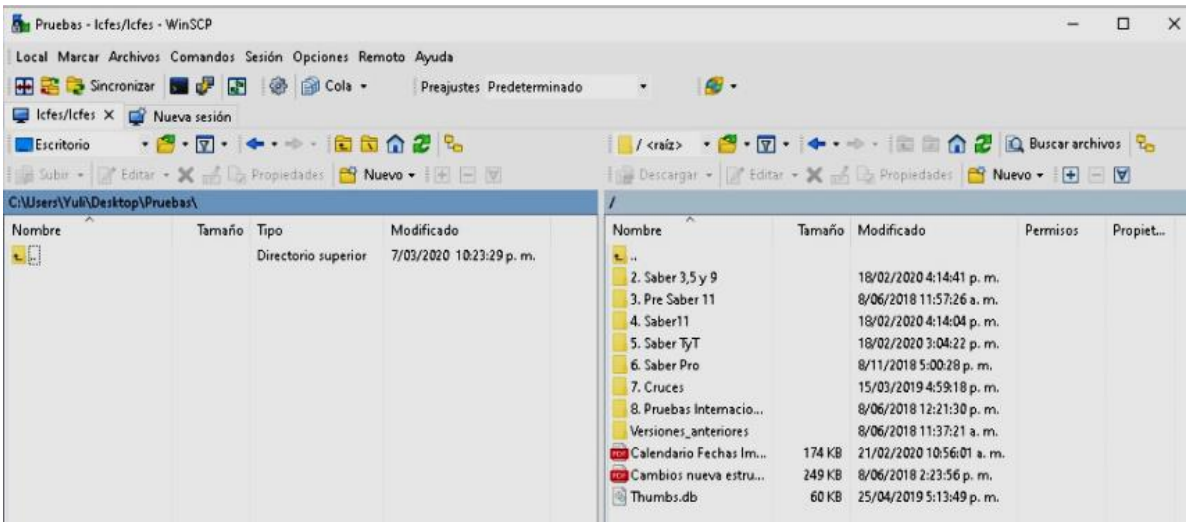


Figura 9. Acceso al FTP con WinSCP
Autores: (Manrique, Celeita, Barahona, 2020)

Se escogió WinSCP, porque este programa permite la conexión con servidores FTP. Mediante este software el administrador puede navegar entre carpetas y arrastrar los archivos que desee a la ubicación de preferencia. Para simplificar aún más el proceso, este aplicativo tiene la opción de sincronizar carpetas entre el servidor y el almacenamiento local, dando la posibilidad de elegir que documentos quiere archivar.

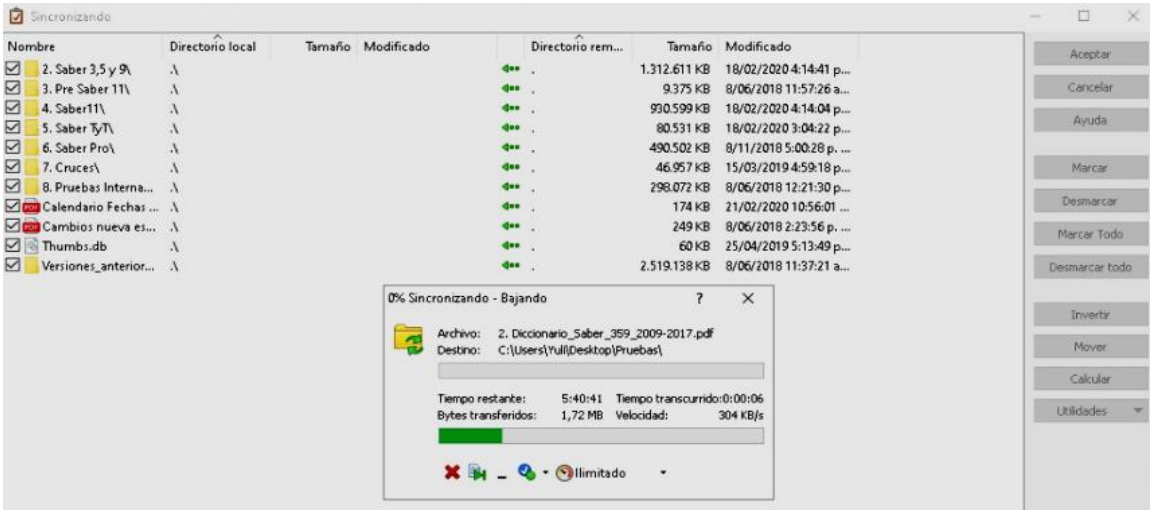


Figura 10. Sincronización de carpetas WinSCP
Autores: (Manrique, Celeita, Barahona, 2020)

Posteriormente, realiza una comparación de los archivos ya existentes y los faltantes, con el fin de conocer que información debe ser extraída como se puede observar en la siguiente imagen. Culminando esta primera etapa de ETL dando paso a la fase de transformación y limpieza de los datos.

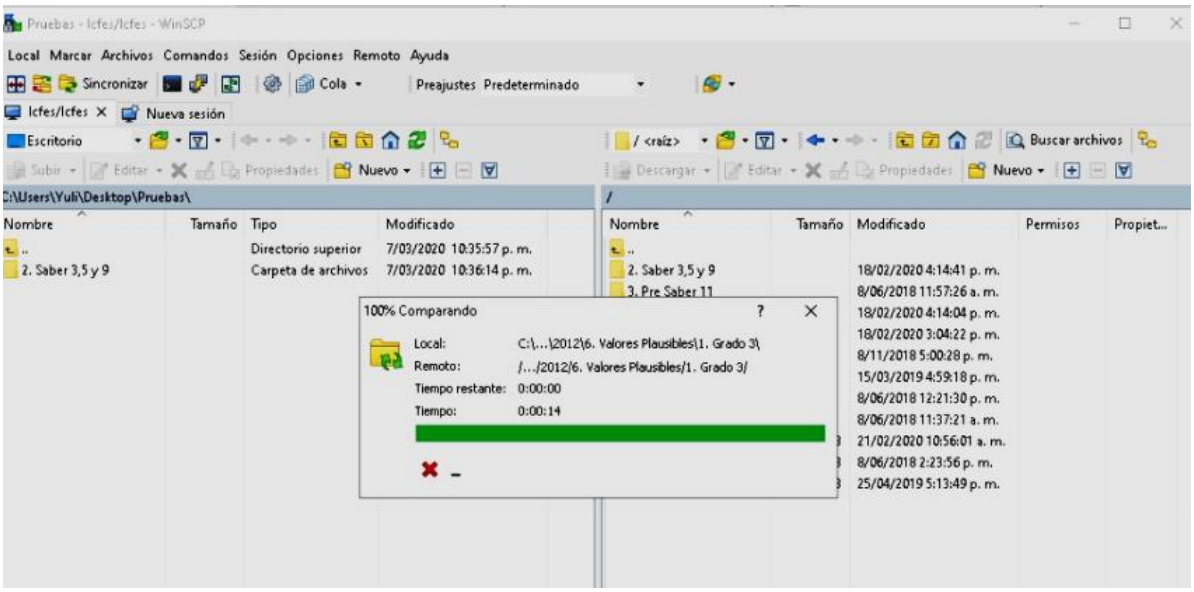


Figura 11. Comparación archivos WinSCP
Autores: (Manrique, Celeita, Barahona, 2020)

2.2. PREPROCESAMIENTO DE LOS DATOS

Esta fase es de suma importancia antes de empezar con la etapa de evaluación de algoritmos de minería de datos porque prepara la información para que el análisis a realizarse sea óptimo. El proceso de ETL (Extract, Transform and Load) consiste básicamente en la transformación de los datos para convertirlos en información útil en DM, ayudando así en la toma de decisiones y facilitando la generación de conocimiento (Franco, 2019). Para las pruebas Saber que se están manejando se tiene gran cantidad de información, haciendo que requiera un tratamiento más detallado que las pruebas realizadas en la educación primaria y secundaria. El procedimiento por realizarse es el siguiente:

- Extracción: Del repositorio del ICFES se descargan archivos en formato .rar de cada una de las pruebas y de cada uno de los períodos de presentación que se van a evaluar, posteriormente se procede a descomprimirlos dejando toda la información alojada en una misma carpeta del sistema.

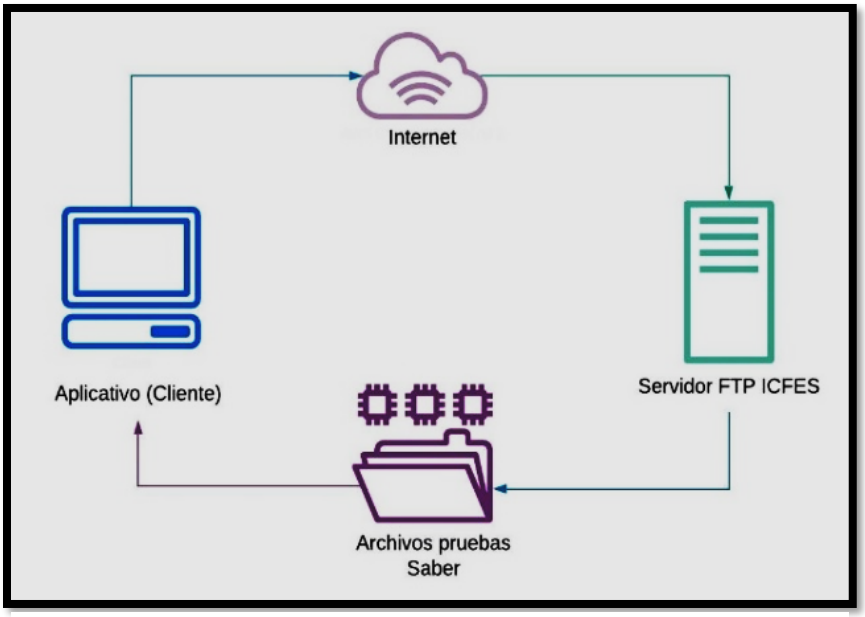


Figura 12. Extración resultados de pruebas
Autores: (Manrique, Celeita, Barahona, 2020)

- Transformación: En el software el administrador tiene un menú donde debe seleccionar el archivo que desea ser cargado a base de datos. Mientras que, de forma interna, el aplicativo se encarga de realizar un cambio de nombre al documento y guardarlo después de cambiar todos aquellos caracteres especiales encontrados en su interior.

- Limpieza: Como parte final de este proceso es necesario eliminar los registros donde hay múltiples campos sin contestar, así como las columnas que son innecesarias para el estudio, por ejemplo, todas aquellas donde el valor es el mismo para todos y/o cuando son códigos de instituciones o departamentos que no brindan un valor relevante de lo realizado en la prueba o del contexto del estudiante.

2.3. IDENTIFICACIÓN Y ANÁLISIS DE LOS INDICADORES

Haciendo uso del software de aprendizaje automático WEKA, en primer lugar, como parte del preprocesamiento de datos, se realizó un filtro conocido como "NumericToNominal" para todos los atributos de tipo numérico, con esto se logra que los datos sean discretizados y que en este caso el algoritmo Apriori que va a ser probado pueda iniciar su análisis. Este filtrado se hizo en 2 de las pruebas Saber, específicamente en la prueba Saber 11, realizada en el primer periodo del año 2016 y en la prueba Saber TyT, del mismo período de presentación.

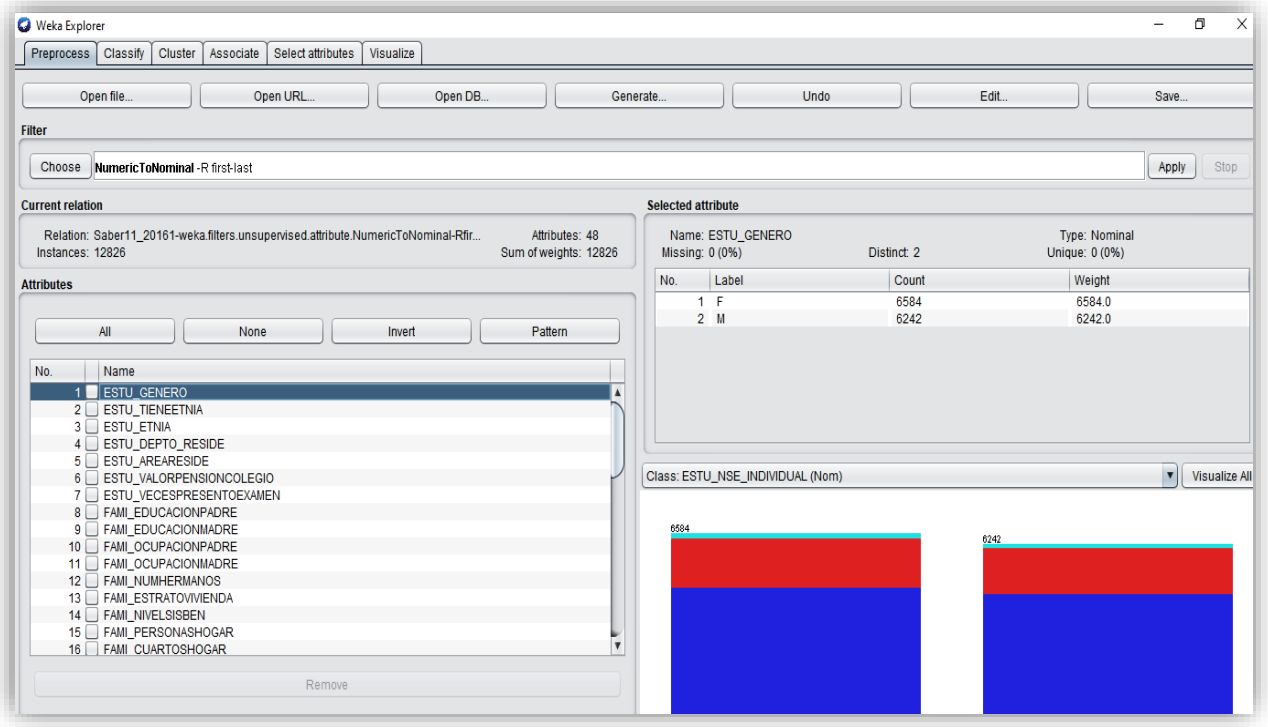


Figura 13. Preprocesamiento en WEKA, Saber 11 2016-1
Autores: (Manrique, Celeita, Barahona, 2020)

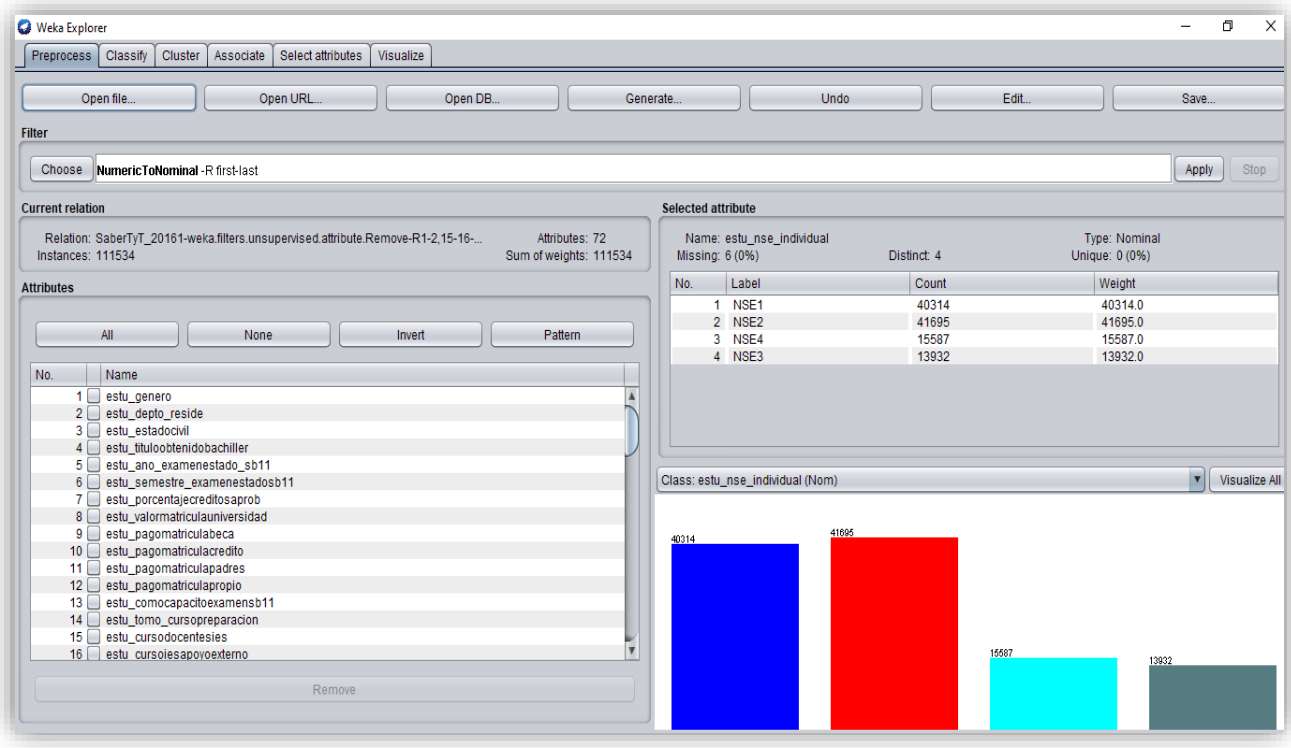


Figura 14. Preprocesamiento en WEKA, Saber TyT 2016-1
Autores: (Manrique, Celeita, Barahona, 2020)

Al tener los atributos discretizados se ejecutó el algoritmo Apriori con sus valores preestablecidos por defecto, con el fin de encontrar asociaciones entre los diferentes atributos existentes y mirar cuáles son las más representativas para el hallazgo de brechas de género; como resultado de esto 10 reglas fueron entregadas como información de salida, como se muestra a continuación:

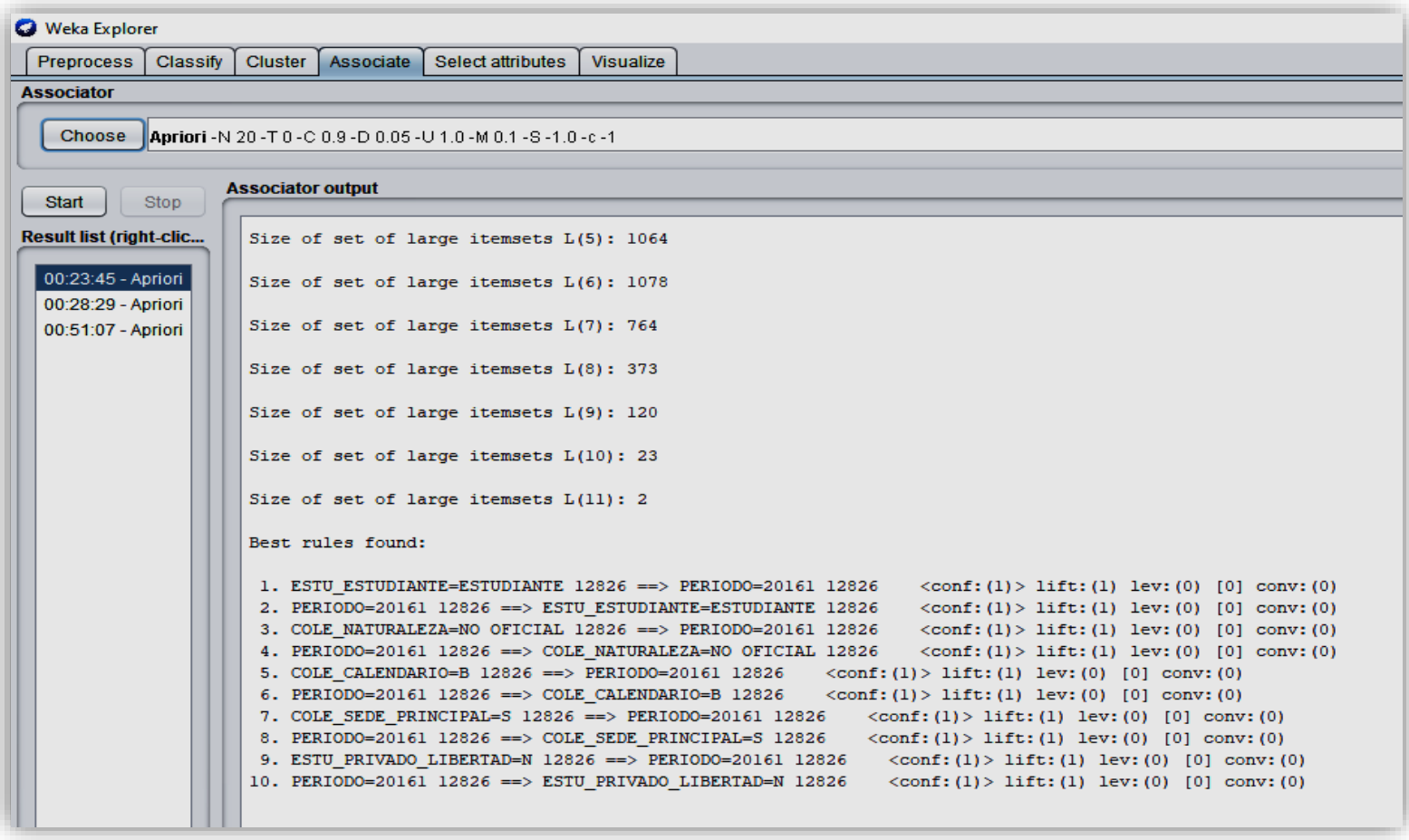


Figura 15. Resultados Apriori en WEKA, Saber TyT 2016-1
Autores: (Manrique, Celeita, Barahona, 2020)

Así mismo al ver que el algoritmo podía encontrar mayor cantidad de reglas, se procedió a ajustar su cantidad a 20, encontrando que las reglas de asociación generadas por WEKA relacionan atributos que no le dan valor agregado al estudio, es decir, no se están encontrando factores que influyan en el comportamiento de las brechas de género.

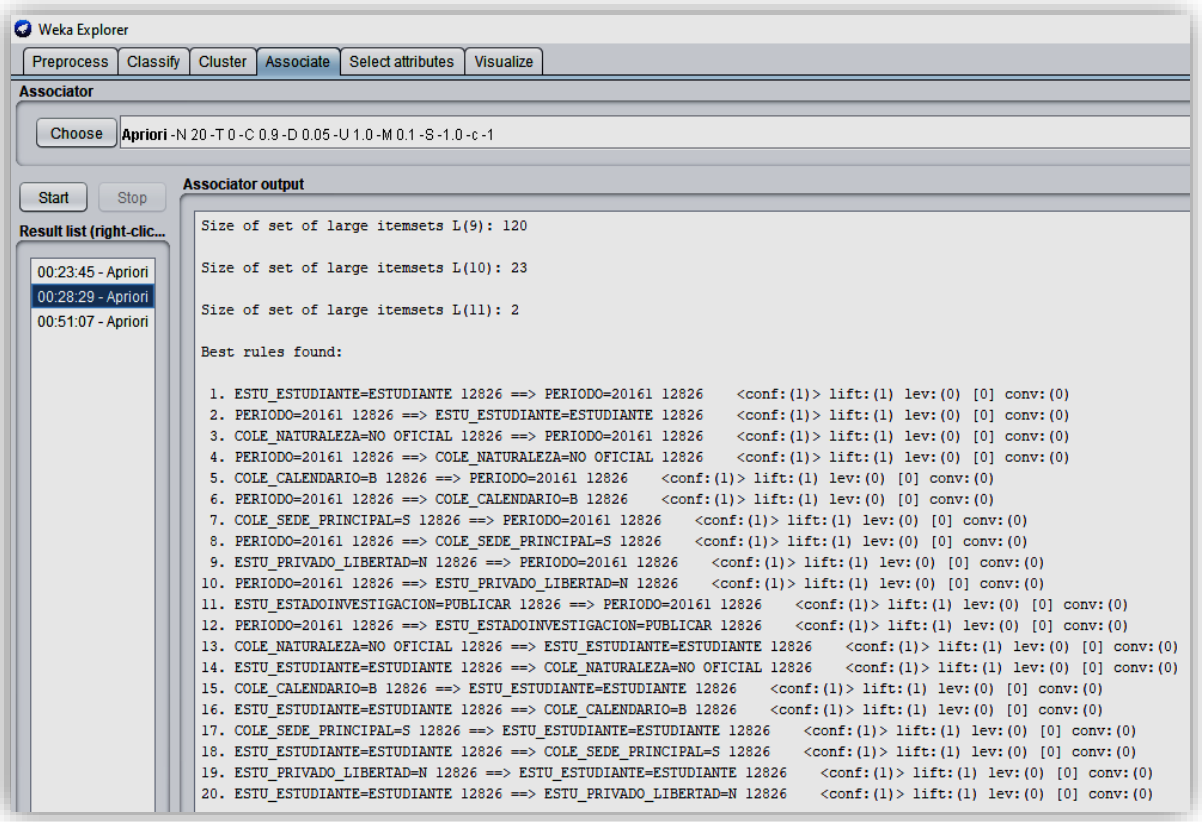


Figura 16. Nuevas reglas Apriori en WEKA, Saber TyT 2016-1
Autores: (Manrique, Celeita, Barahona, 2020)

Cabe aclarar que este mismo proceso de evaluación del algoritmo fue hecho en ambas pruebas por igual, y en diferentes oportunidades; con el objetivo de mirar como variaba el comportamiento de éste en diferentes escenarios. ¿Qué se pudo deducir de la realización de estos análisis con el algoritmo Apriori? Los resultados o en este caso las reglas encontradas no arrojan información relevante del tema que se está tratando, sino que muestran las asociaciones más significativas halladas en el conjunto de datos, pero desafortunadamente son triviales estas relaciones. Por lo tanto, se puede decir que el algoritmo Apriori al tener un comportamiento no supervisado, lo que hace es encontrar las relaciones con mayor cantidad de repeticiones, pero no aporta información relevante al proyecto y por esta razón no será usado para encontrar los indicadores de brecha de género en las distintas pruebas.

En consecuencia, se empezaron a realizar pruebas con el algoritmo de clasificación K-means o K-medias, debido a que este algoritmo se especializa en el descubrimiento de patrones y relaciones entre los datos que son desconocidas permitiéndonos clasificar los datos según las relaciones encontradas (Molina López & García Herrero, 2006). De nuevo, haciendo uso de WEKA, se evaluaron los resultados obtenidos con cada uno de los atributos, en este caso los clústeres.

A partir de las agrupaciones o clústeres más representativos se determinaron los diferentes indicadores para cada una de las pruebas.

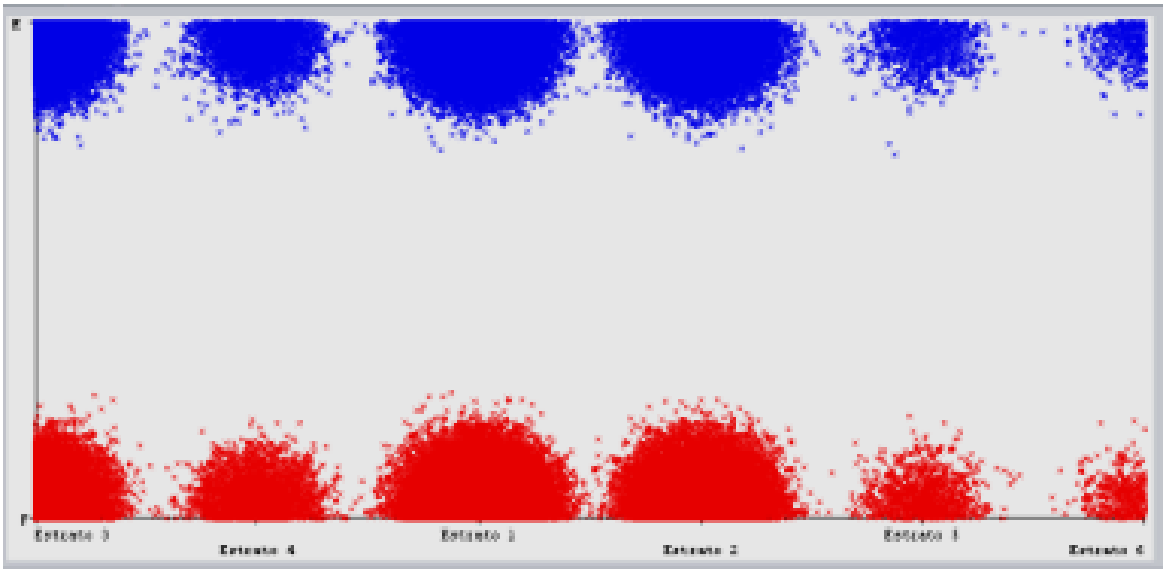


Figura 17. Estrato evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Evaluando el atributo de estrato se encontraron diferentes agrupaciones que permiten evidenciar brechas entre ambos géneros.

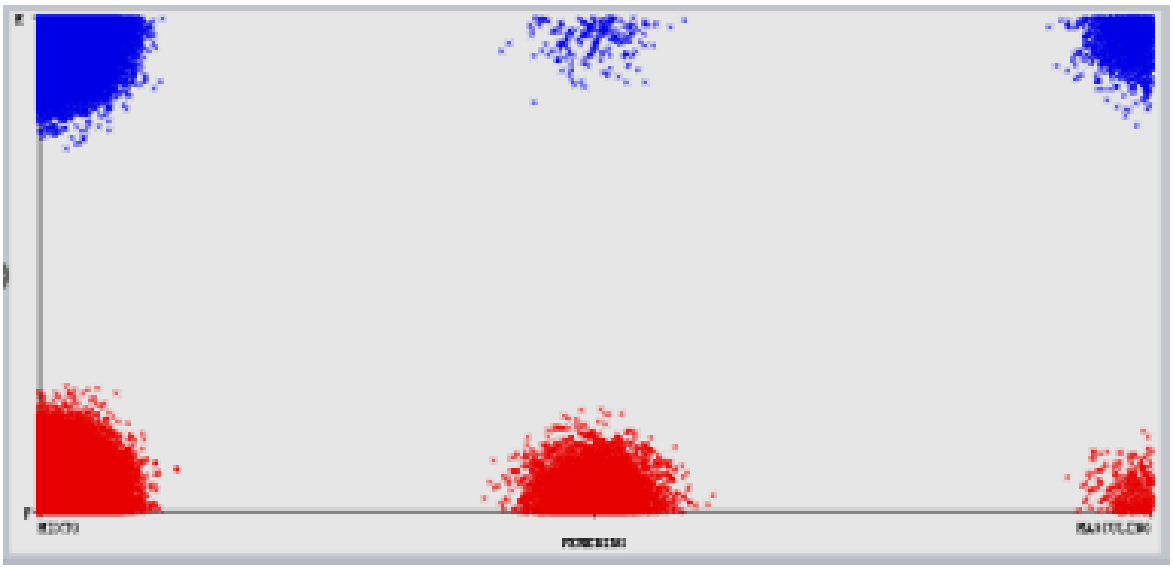


Figura 18. Género de colegio evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Por su parte, **Género de colegio** es un atributo que al ser evaluado tiene gran relevancia en el estudio que se está realizando, se puede decir que es uno de los indicadores que tiene más importancia en la prueba Saber 11.

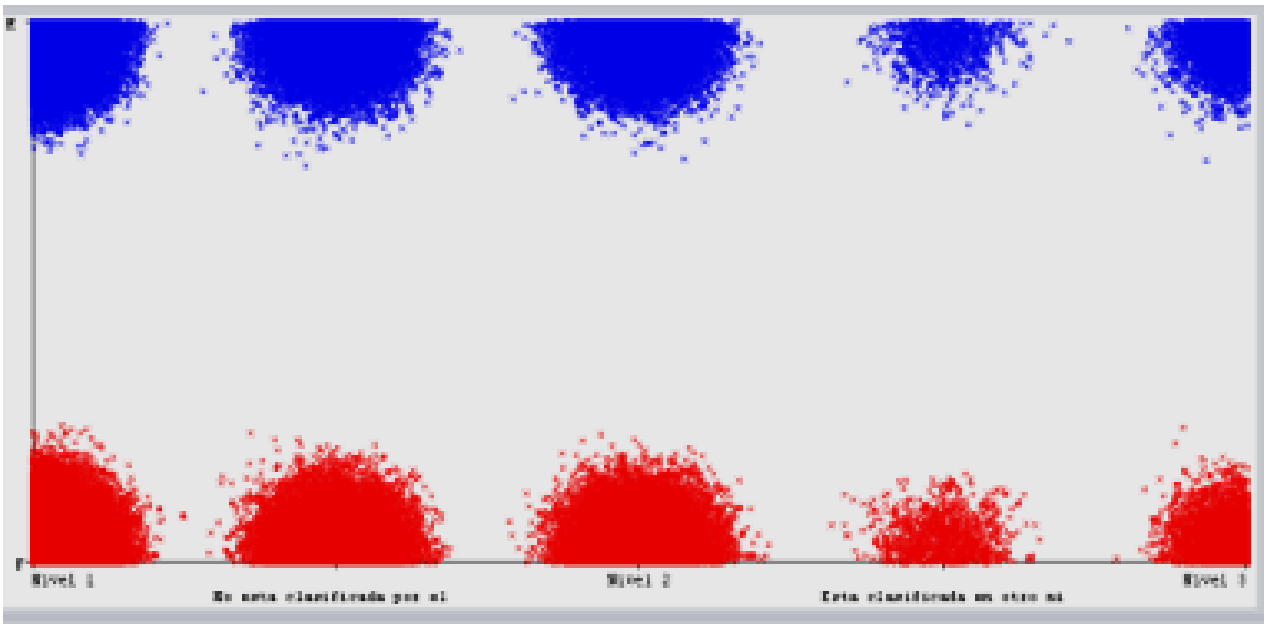


Figura 19. Nivel de Sisbén evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Este indicador, **Nivel de Sisbén,** permite ubicar desde un punto de vista económico a los estudiantes y saber que afectación tiene en los puntajes de las pruebas. De igual manera sucede con el atributo **Ingreso familiar**.

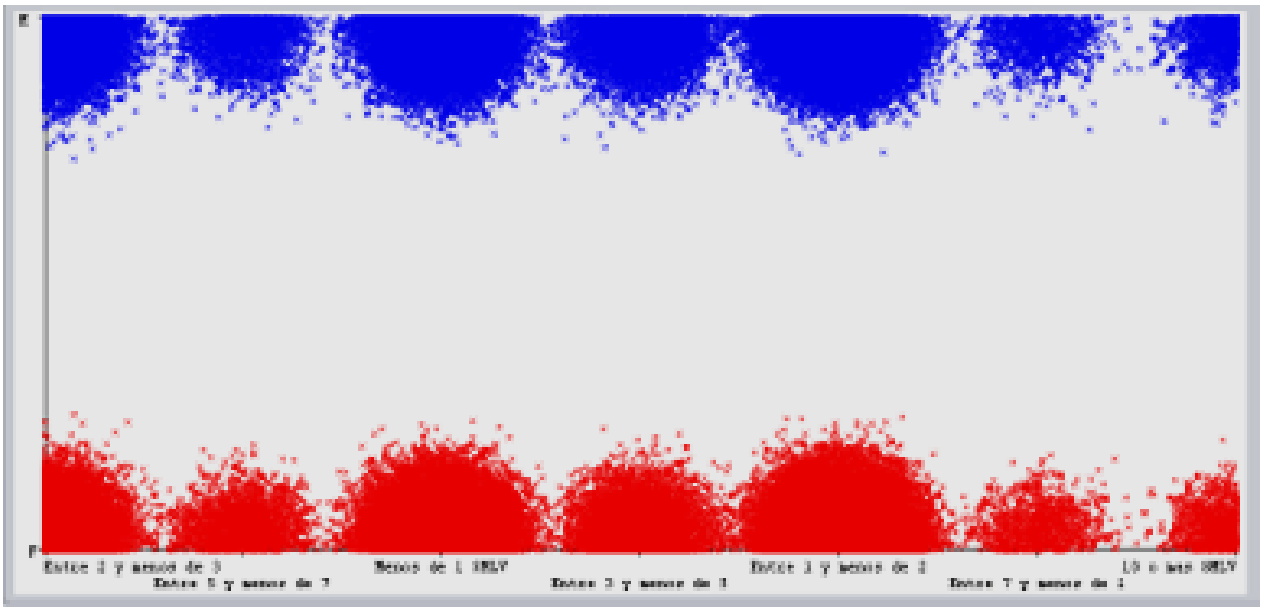


Figura 20. Ingreso familiar evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

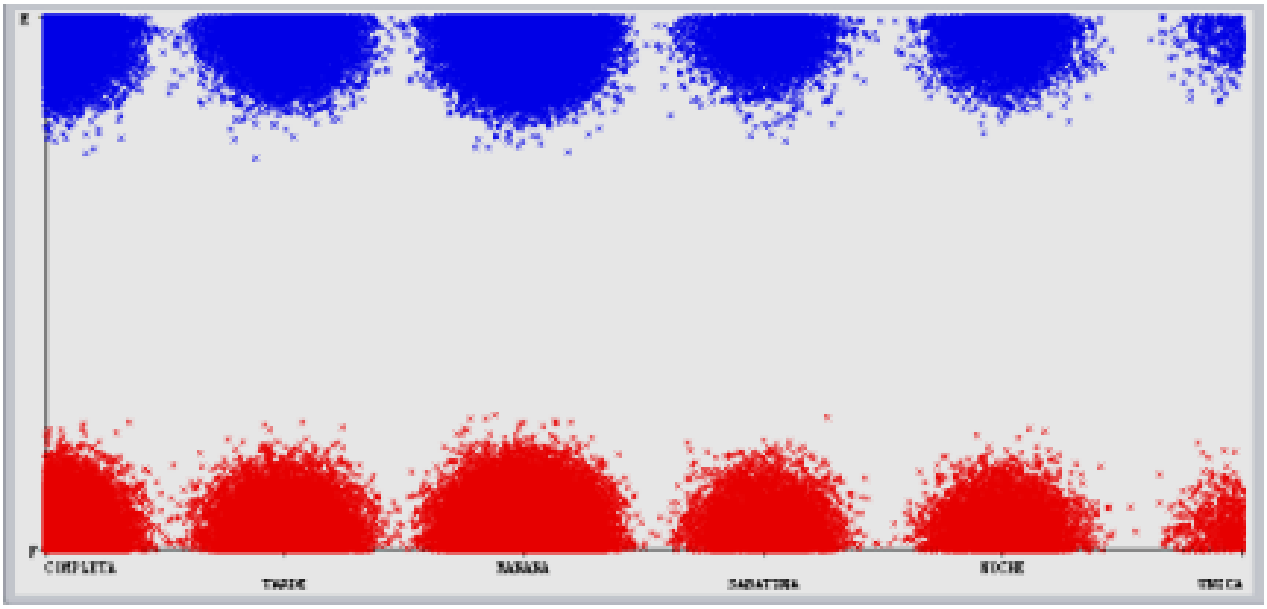


Figura 21. Jornada de colegio evaluada con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

El atributo **Jornada de colegio** permite realizar un comparativo de los resultados académicos que son obtenidos por los estudiantes en los diferentes horarios dispuestos por los colegios para poder mirar su afectación en la brecha de género.

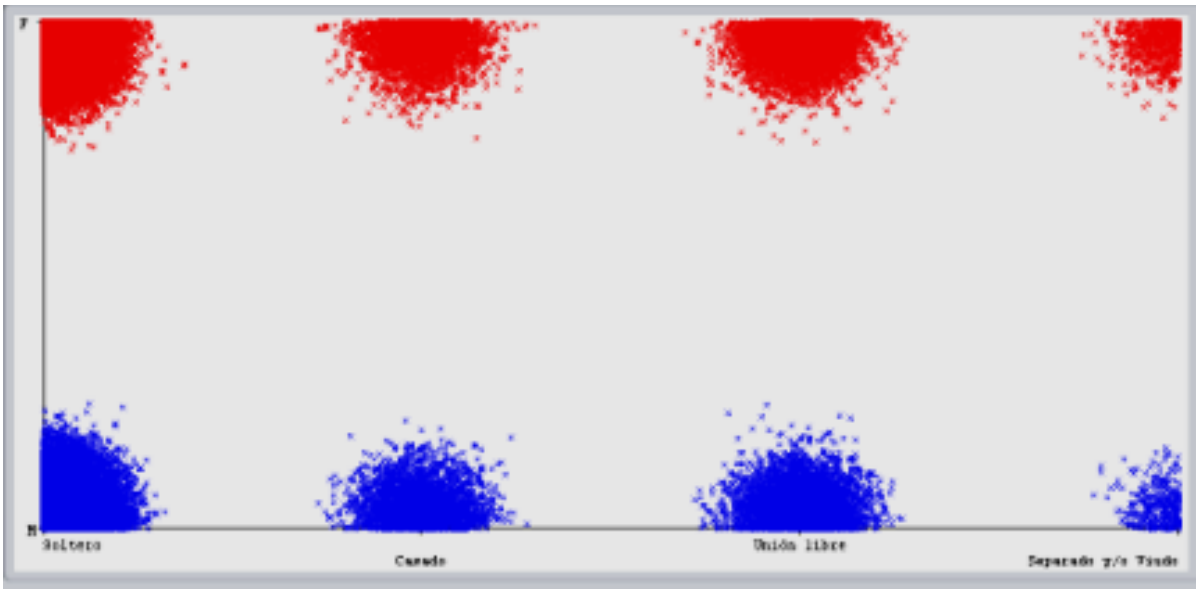


Figura 22. Estado civil evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Este atributo, **Estado civil**, permite ver afectación social y familiar para aquellos estudiantes que presentaron las pruebas Saber Pro y Saber TyT.

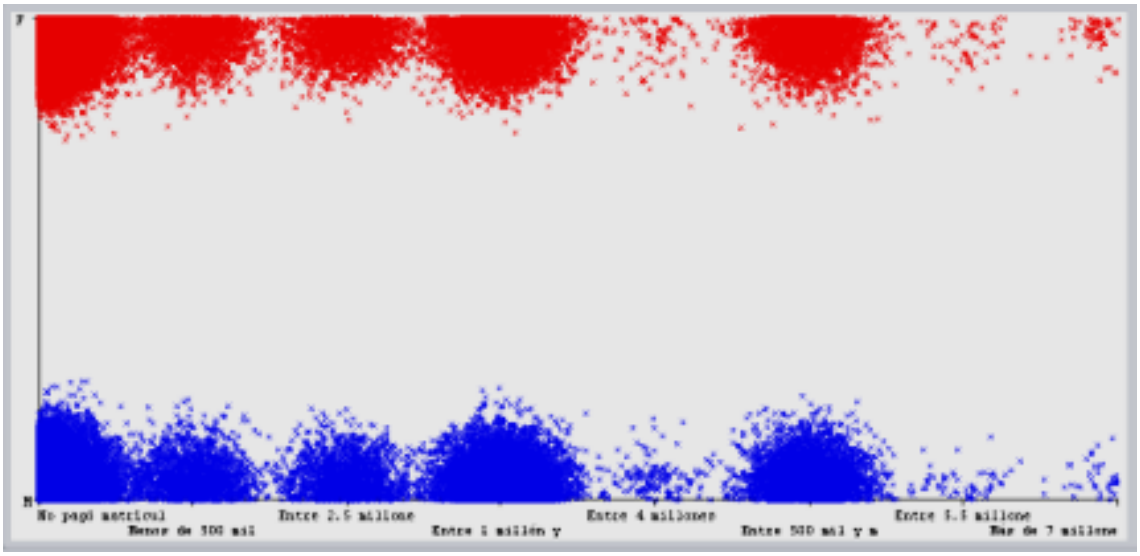


Figura 23. Valor matricula evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

El atributo **valor matricula** permite observar la situación económica de los estudiantes y así mismo saber cuál es el promedio en las matrículas de los diferentes colegios del país.

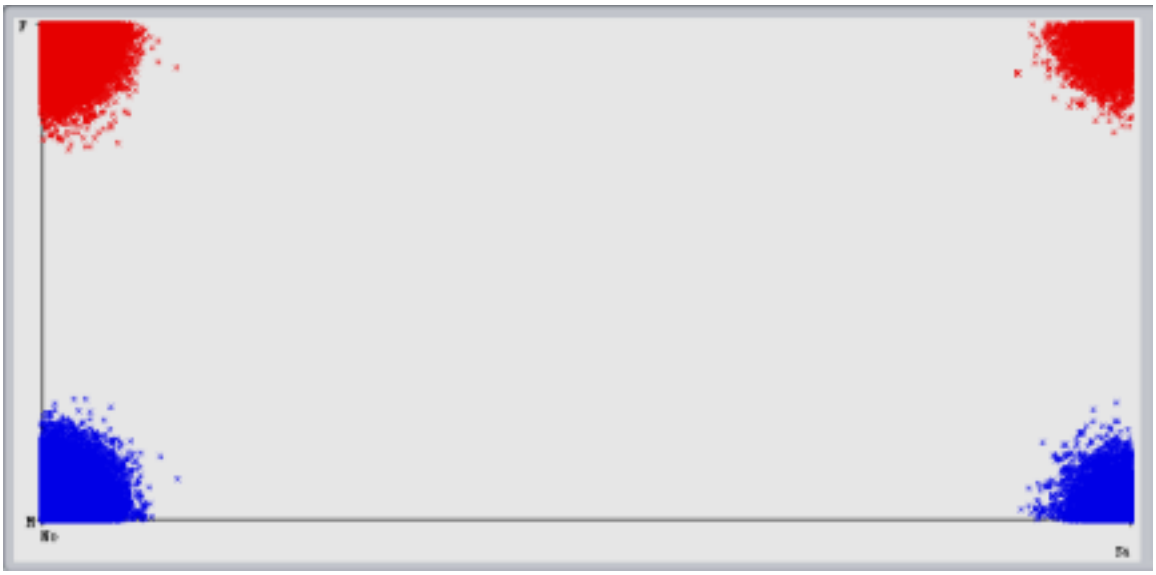


Figura 24. Tipo de pago de matrícula evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Este atributo **Tipo de pago de matrícula** nos da un indicio de cómo es la capacidad de endeudamiento de los estudiantes.

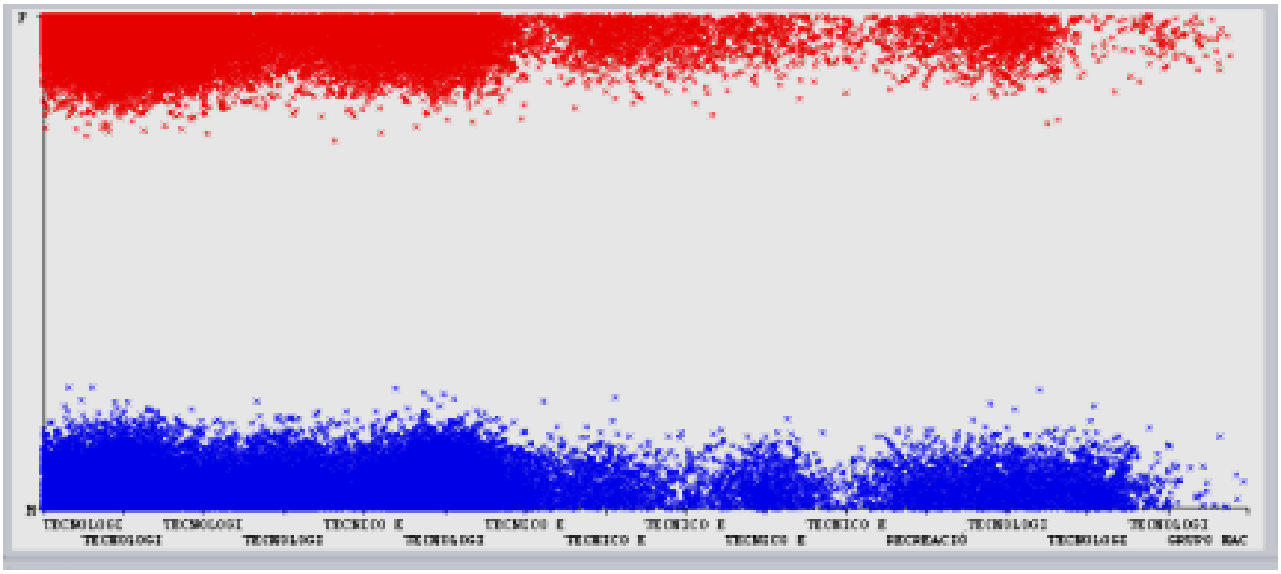


Figura 25. Grupo de referencia evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

El atributo **grupo de referencia** nos permite darnos una idea de las preferencias a la hora de escoger carrera y que enfoque tienen las mujeres y los hombres en cada caso.

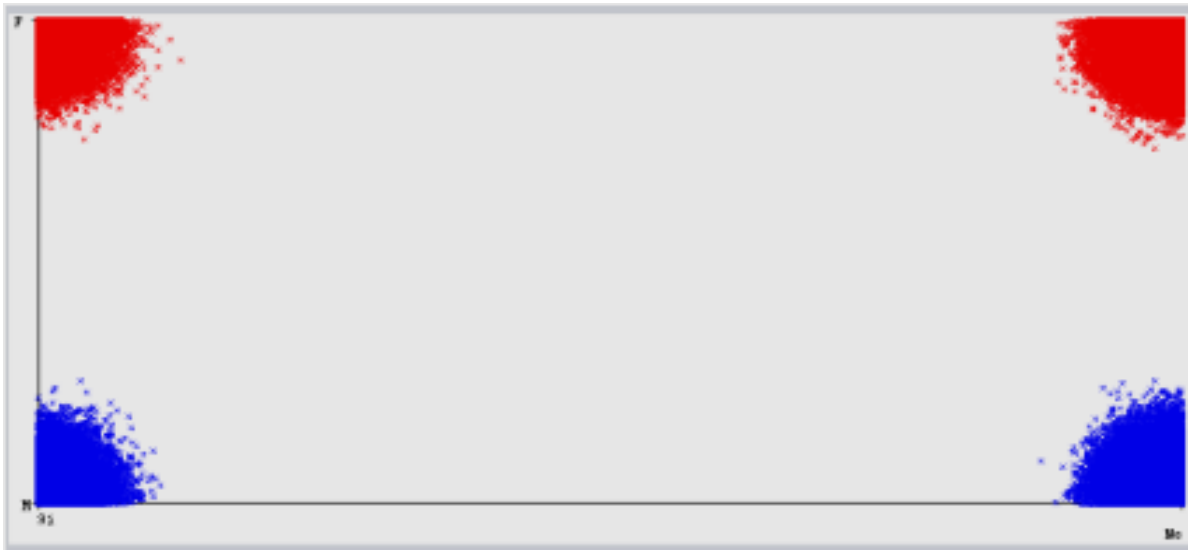


Figura 26. Cabeza de familia evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Cabeza de familia es un atributo que nos permite determinar si las obligaciones familiares de los estudiantes tiene afectación en la desigualdad de género.

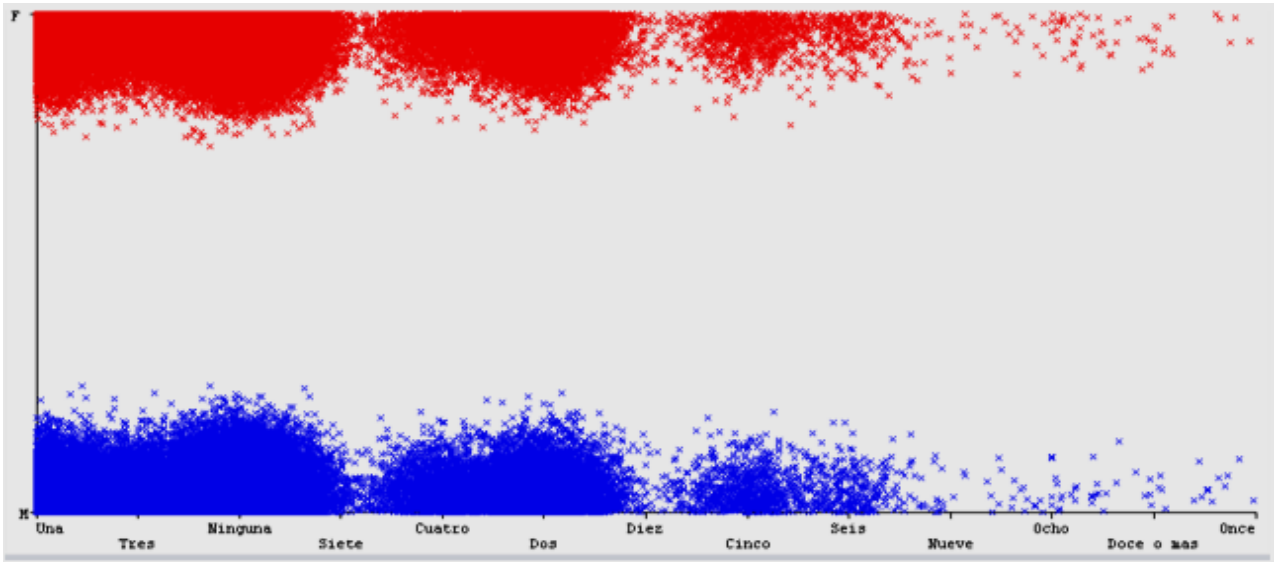


Figura 27. Número de personas a cargo evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

Una situación similar se observa con el atributo **Número de personas a cargo.** Mientras que, al analizar la **Lectura diaria** se puede observar como la dedicación a la lectura de los estudiantes afecta los resultados de las pruebas.

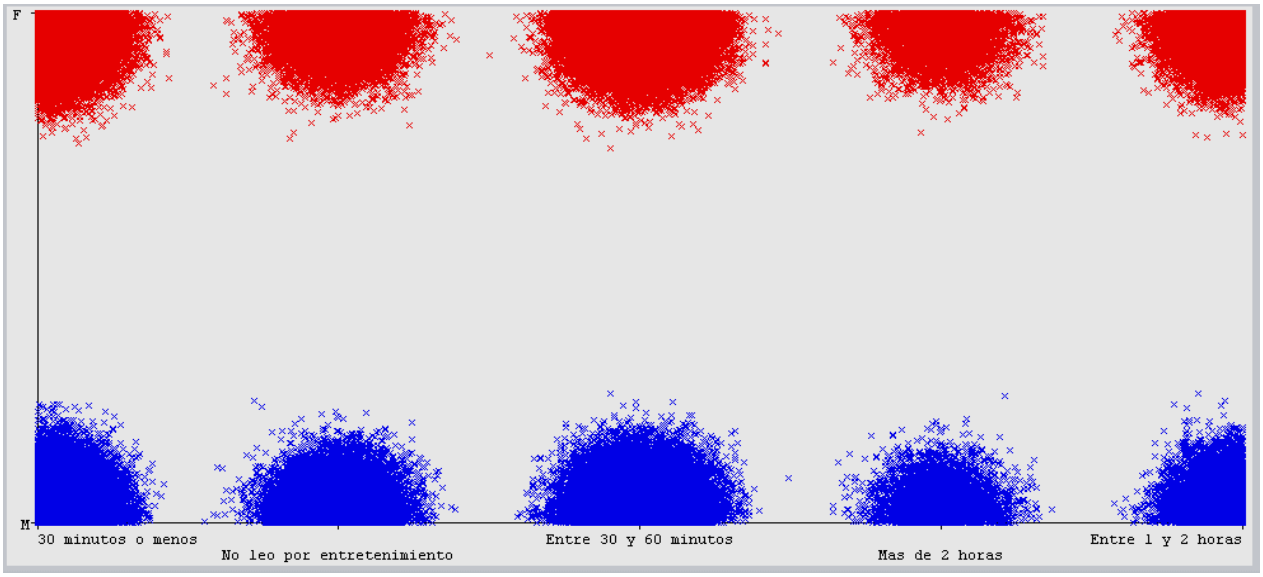


Figura 28. Lectura diaria evaluado con K-Means
Autores: (Manrique, Celeita, Barahona, 2020)

2.4. INDICADORES DE BRECHA DE GÉNERO

De acuerdo a los resultados observados en el software WEKA los indicadores que finalmente se determinaron en cada una de las pruebas son los siguientes:

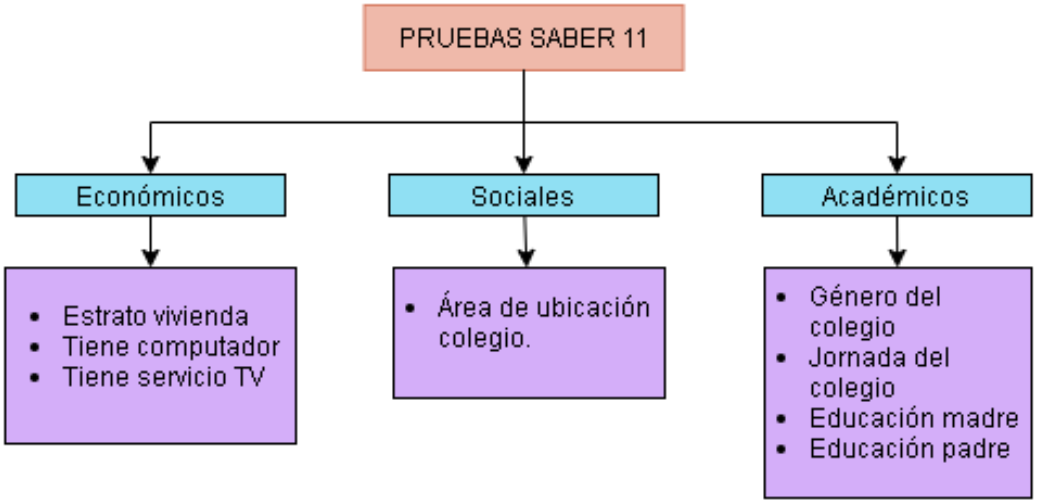


Figura 29. Indicadores Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

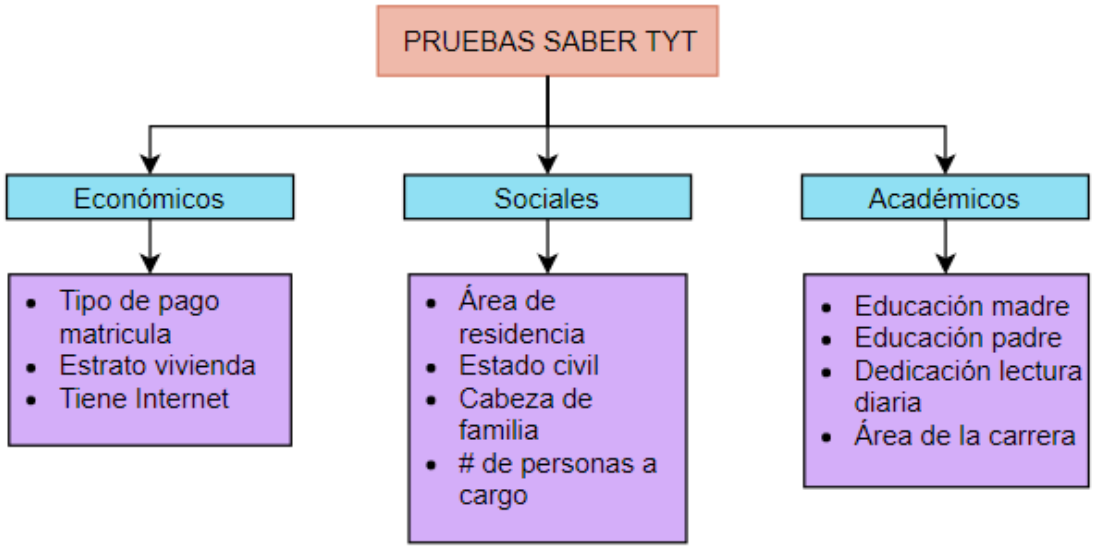


Figura 30. Indicadores Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

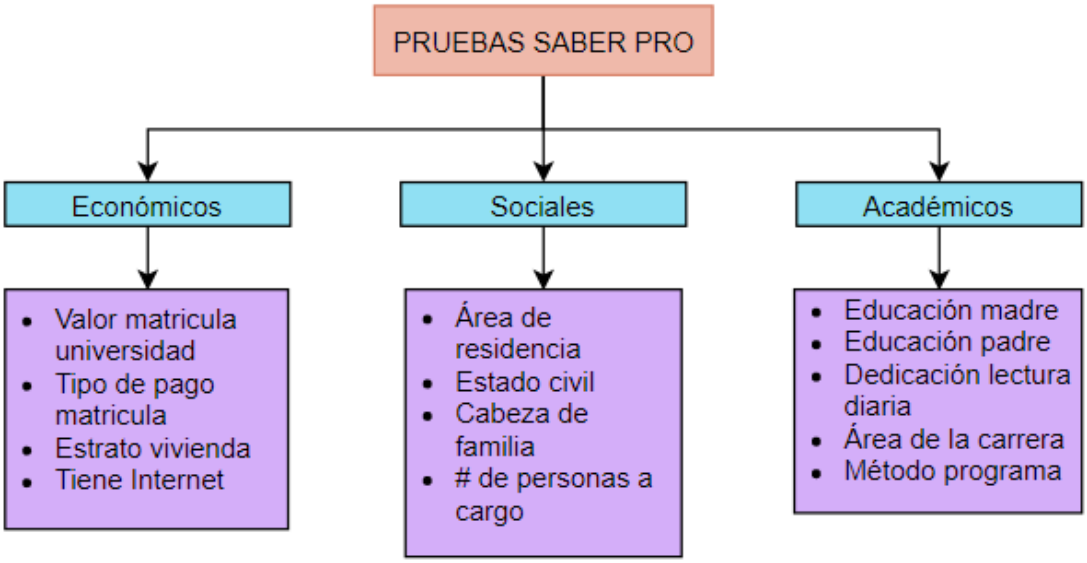


Figura 31. Indicadores Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

- Valor matricula universidad: Este indicador representa el rango de valores pagados por los estudiantes en la matricula del programa que están cursando en la universidad cada semestre.
- Tipo de pago matricula: En esta prueba se tienen varias opciones de respuesta de tipos de pago matricula entre los cuales se encuentra si es pagado con ayuda de padres, si es por medio de crédito, si fue pago propio o si el estudiante se encuentra becado. Entre estos métodos, el pago por crédito fue aquel indicador que arrojo una afectación significativa. Los atributos con los que cuenta el indicador son sí y no.
- Estrato de vivienda: Indica la clasificación socioeconómica jerarquizada del país, sus categorías comprenden desde el estrato 0 hasta el estrato 6.
- Tiene computador: Este indicador nos señala qué estudiantes cuentan con computador en sus hogares y quiénes no.

- Tiene internet y servicio TV: Este indicador nos señala qué estudiantes cuentan con el servicio de internet o el servicio de televisión en sus hogares y quiénes no.
- Estado civil: Representa la condición al momento de la realización de la prueba de los estudiantes determinada por sus relaciones de pareja, se dividen entre 4 opciones: si está casado, soltero, unión libre y separado y/o viudo.
- Cabeza de familia: Determina si la persona es quién tiene a cargo, o no, toda la responsabilidad económica del hogar.
- Número de personas a cargo: Este indicador está directamente relacionado con el de cabeza de familia, y pretende hacer la misma medición, pero desde un punto de vista cuantitativo. Estando en diferentes rangos comprendidos entre ninguna y11 personas.
- Área de ubicación de colegio: Se refiere a la zona donde queda ubicado el colegio donde el estudiante pertenecía al momento de presentación de la prueba. Sus atributos se dividen en Rural y Urbano.
- Género y jornada del colegio: Estos indicadores permiten identificar las condiciones de la institución educativa de los estudiantes que presentan la prueba Saber 11, más específicamente conocer cuál es el género del colegio y los horarios de asistencia a clase.
- Educación madre y educación padre: Estos dos indicadores nos muestran qué niveles educativos cuentan los padres de los estudiantes, estando clasificados en los siguientes atributos: ninguno, no sabe, no aplica, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica y tecnológica incompleta, técnica y tecnológica completa, educación profesional incompleta, educación profesional completa y postgrado.

- Dedicación lectura diaria: Muestra el tiempo diario que le dedican los estudiantes al ejercicio de la lectura, siendo clasificada por rangos de tiempo de la siguiente manera: 30 minutos o menos, no leo por entretenimiento, entre 30 y 60 minutos, más de 2 horas, y entre 1 y 2 horas.
- Área de residencia: Se refiere a la zona donde el estudiante vivía al momento de presentación de la prueba. Sus atributos se dividen en Rural y Urbano.
- Área de la carrera: Este indicador en las pruebas se conoce como grupo de referencia, y especifica en que área se encuentra clasificada la carrera que se está cursando. Existen múltiples categorías, pero para este estudio se escogieron 4: las cuales son Administración y afines, Contaduría y afines, psicología e ingeniería, en el caso de Saber Pro. Y para Saber TyT son: tecnológico en administración y turismo, tecnológico en ingeniería, industria y minas, técnico en administración y turismo, y técnico en ingeniería, industria y minas.
- Método del programa: Hace referencia al modo de acceso al programa educativo, los 3 tipos de clasificación son presencial, distancia y virtual.

III. DOCUMENTACION DEL SOFTWARE

3.1. PLAN DE PROYECTO

Modo de	Nombre de tarea	Duración	Comienzo	Fin
	Determinar indicadores a partir de metodos estadistic	5 días	mar 25/02/20	lun 2/03/20
	Base de datos	**9 días**	**mar 3/03/20**	**jue 12/03/20**
	Conexión servidor FTP y PostgreSQL	4 días	mar 3/03/20	sáb 7/03/20
	Crear tablas	1 día	lun 9/03/20	lun 9/03/20
	Carga de los resultados a las tablas	1 día	mar 10/03/20	mar 10/03/20
	Limpieza y transformación de los datos	2 días	mié 11/03/20	jue 12/03/20
	Aplicación de algoritmos de mineria de datos	13 días	vie 13/03/20	mar 31/03/20
	Crear modelo para la construcción de indicadores	13 días	mié 1/04/20	vie 17/04/20
	Desarrollo de software	**32 días**	**jue 23/04/20**	**jue 4/06/20**
	Crear inicio de sesión	1 día	jue 23/04/20	jue 23/04/20
	Administrador	**6 días**	**vie 24/04/20**	**vie 1/05/20**
	CRUD de usuarios	**4 días**	**vie 24/04/20**	**mié 29/04/20**
	Crear usuarios	1 día	vie 24/04/20	vie 24/04/20
	Consultar usuarios	1 día	lun 27/04/20	lun 27/04/20
	Modificar usuarios	1 día	mar 28/04/20	mar 28/04/20
	Eliminar usuarios	1 día	mié 29/04/20	mié 29/04/20
	Modificar información de acceso a FTP	1 día	vie 1/05/20	vie 1/05/20
	Realizar modulo de analisis descriptivo	**8 días**	**lun 4/05/20**	**mié 13/05/20**
	Calculo de indicadores de brecha de genero	6 días	lun 4/05/20	lun 11/05/20
	Implementación de modelo	2 días	mar 12/05/20	mié 13/05/20
	Realizar modulo de analisis predicitivo	**13 días**	**jue 14/05/20**	**vie 29/05/20**
	Desarrollo de modelo predictivo	11 días	jue 14/05/20	mié 27/05/20
	Implementación modelo	2 días	jue 28/05/20	vie 29/05/20
	Usuario	**4 días**	**lun 1/06/20**	**jue 4/06/20**
	Visualizar resultados	2 días	lun 1/06/20	mar 2/06/20
	Generar reportes	2 días	mié 3/06/20	jue 4/06/20

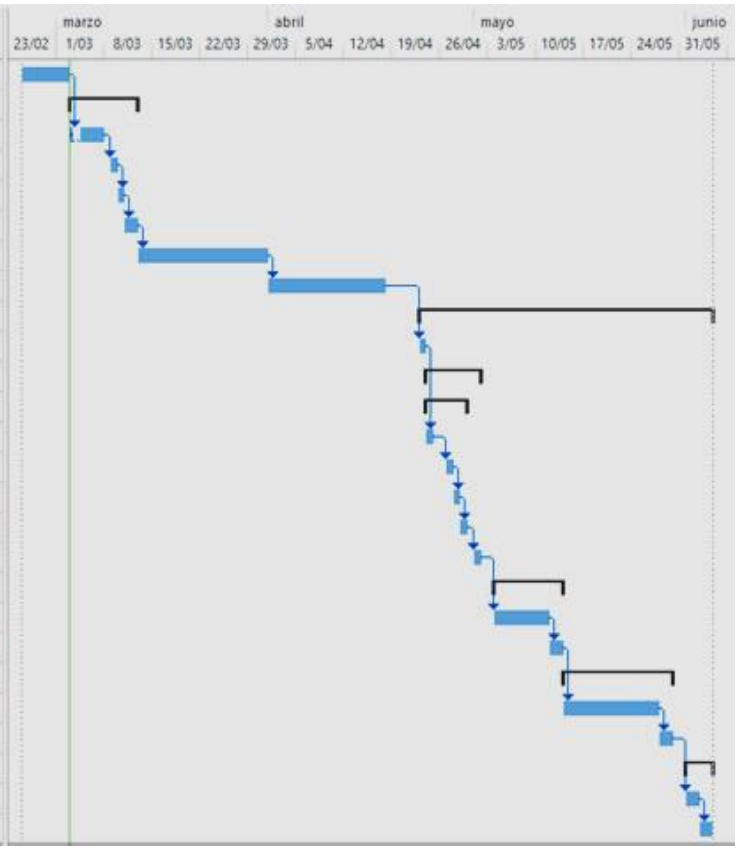


Figura 32. Plan de proyecto realizado en Project
Autores: (Manrique, Celeita, Barahona, 2020)

3.2. DETERMINACIÓN DE REQUERIMIENTOS

Tabla 1
Requerimientos funcionales

Identificación del Requerimiento	RF01
Nombre del requerimiento	Inicio de sesión
Descripción del requerimiento	El usuario podrá acceder a la plataforma a través de un Login, se le pedirá usuario y contraseña para el ingreso.
Identificación del Requerimiento	RF02
Nombre del requerimiento	Contáctenos

Descripción del requerimiento	El cliente podrá diligenciar un formulario con el fin de solicitar su registro en el aplicativo, además de enviar cualquier solicitud de comunicación con los desarrolladores del software.
Identificación del Requerimiento	RF03
Nombre del requerimiento	Creación usuarios
Descripción del requerimiento	El administrador podrá diligenciar un formulario con la información pertinente para realizar el registro de nuevos usuarios al aplicativo web. La información se le enviará al correo (usuario y contraseña).
Identificación del Requerimiento	RF04
Nombre del requerimiento	Consulta usuarios
Descripción del requerimiento	El administrador podrá visualizar el listado de usuarios existentes en el sistema.
Identificación del Requerimiento	RF05
Nombre del requerimiento	Modificación usuarios
Descripción del requerimiento	El administrador podrá actualizar y/o modificar la información de los usuarios, para ello deberá consultar y seleccionar el usuario que desea modificar.
Identificación del Requerimiento	RF06
Nombre del requerimiento	Eliminación usuarios
Descripción del requerimiento	El administrador podrá eliminar los usuarios, consultando previamente y seleccionando el usuario que desea eliminar. Antes de finalizar la operación saldrá una alerta de confirmación.
Identificación del Requerimiento	RF07
Nombre del requerimiento	Actualización datos de descarga
Descripción del requerimiento	El administrador se encargará de actualizar los datos de acceso (URL, usuario y contraseña) del repositorio donde están alojados los datos de las pruebas por parte del ICFES.
Identificación del Requerimiento	RF08
Nombre del requerimiento	Extracción de los datos

Descripción del requerimiento	El aplicativo se encargará de la extracción de los datos disponibles de las pruebas. De forma automática deberá descargar del repositorio del ICFES los archivos necesarios para su posterior análisis, estos vienen en formato .zip y al descomprimirlos queda un archivo .txt que contiene todos los datos de las pruebas según su nivel.
Identificación del Requerimiento	RF09
Nombre del requerimiento	Limpieza de los datos
Descripción del requerimiento	El aplicativo tendrá la función de limpiar los datos que no representan información relevante para el estudio y transformar los caracteres especiales de forma que sean legibles para la base de datos.
Identificación del Requerimiento	RF10
Nombre del requerimiento	Carga de los datos
Descripción del requerimiento	En esta actividad el aplicativo cargará los datos transformados a la base de datos para su posterior análisis.
Identificación del Requerimiento	RF11
Nombre del requerimiento	Cálculo de los indicadores de brecha de género
Descripción del requerimiento	El aplicativo se encargará de calcular los indicadores de brecha de género implementando los algoritmos de minería de datos pertenecientes a la técnica de reglas de agrupación, teniendo en cuenta las variables previamente elegidas para su evaluación.
Identificación del Requerimiento	RF12
Nombre del requerimiento	Análisis predictivo
Descripción del requerimiento	El aplicativo se encargará de calcular y determinar si existe relación entre los núcleos temáticos en común, desde las pruebas de la educación básica hasta las del nivel superior, para así poder predecir su comportamiento. Esto se realizará por medio de algoritmos de

	árboles de decisión, los cuales nos permitirán conocer la conexión, si es que existe, entre pruebas de cada una de las áreas evaluadas con respecto a la brecha de género.
Identificación del Requerimiento	RF13
Nombre del requerimiento	Visualización de los resultados
Descripción del requerimiento	El usuario podrá visualizar los resultados de forma gráfica y estadística filtrados por el año que desee obtener información y/o por los indicadores hallados durante la investigación y análisis.
Identificación del Requerimiento	RF14
Nombre del requerimiento	Reportes
Descripción del requerimiento	El aplicativo le permitirá al usuario visualizar y descargar los reportes de resultados de los análisis previamente realizados, incluyendo gráficas y estadísticas. El formato de descarga será PDF.

Atributos del Sistema

El aplicativo web dependerá del uso de la herramienta WEKA, para la implementación de los algoritmos de minería de datos.

El desarrollo del aplicativo se realizará en Java bajo los parámetros del MVC (modelo vista controlador), para la parte del diseño web se hará mediante lenguaje HTML.

El aplicativo hará uso de PostgreSQL el cual es un gestor de base de datos de buen rendimiento que soporta gran variedad de sistemas operativos y previene la corrupción de los datos.

3.3. ESPECIFICACIÓN DEL DISEÑO

3.3.1. MODELO DE ENTIDAD RELACION (MER)

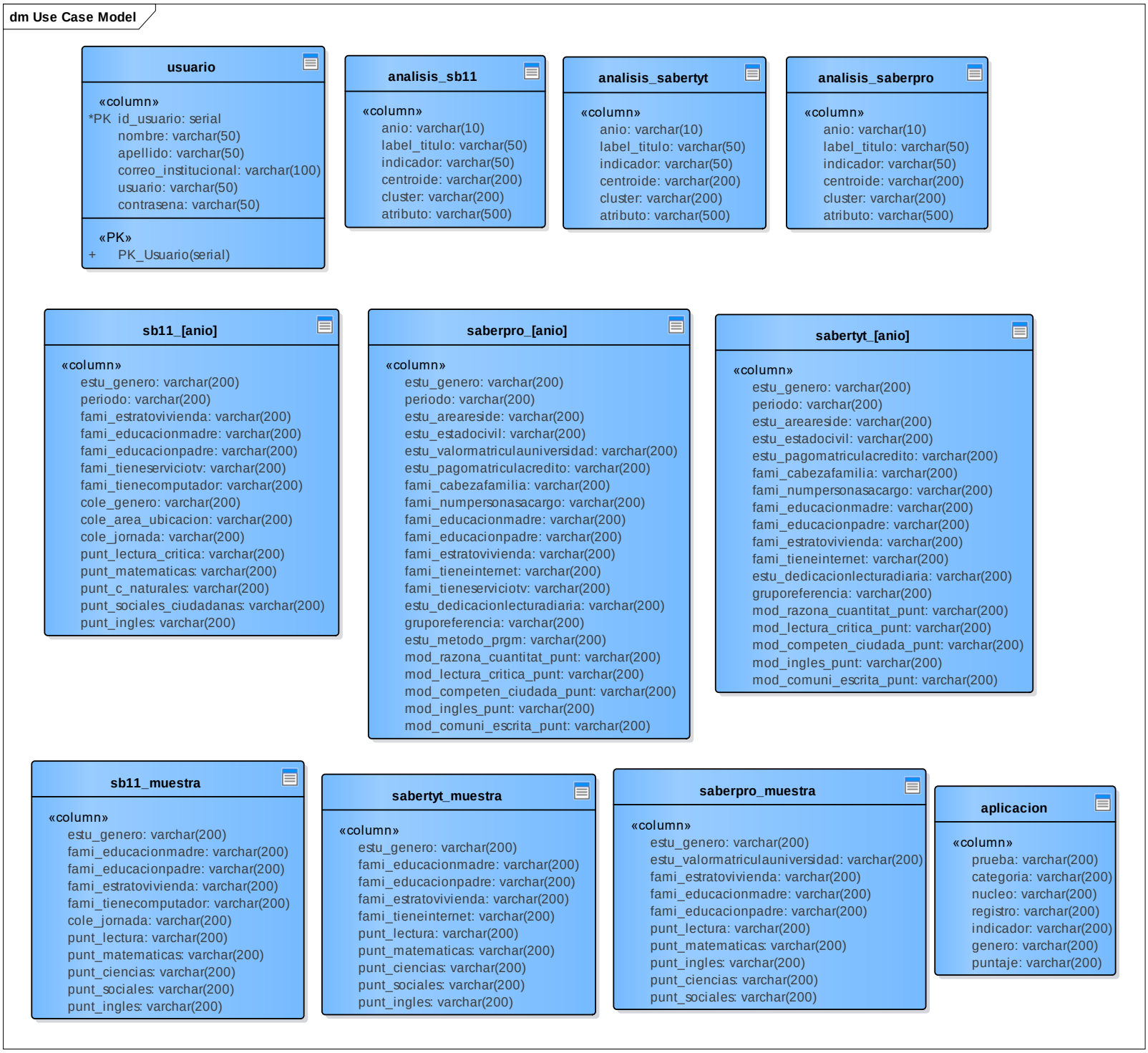


Figura 33. Modelo Entidad Relación

Autores: (Manrique, Celeita, Barahona, 2020)

Descripción tablas MER

Tabla 2
Descripción tabla usuario

Entidad	usuario
Llave primaria	**id_usuario:** Este atributo corresponde a un número serial, único e irrepetible.
Atributos	id_usuario, nombre, apellido, correo_institucional, usuario, contrasena
Descripción	Esta tabla contiene la información de los usuarios del aplicativo.

Tabla 3
Descripción tabla analisis_sb11

Entidad	analisis_sb11
Atributos	anio, label_titulo, indicador, centroide, cluster, atributo
Descripción	Esta tabla guarda los resultados generados por el análisis descriptivo en los diferentes años para la prueba Saber 11.

Tabla 3
Descripción tabla analisis_sabertyt

Entidad	analisis_sabertyt
Atributos	anio, label_titulo, indicador, centroide, cluster, atributo
Descripción	Esta tabla guarda los resultados generados por el análisis descriptivo en los diferentes años para la prueba Saber TyT.

Tabla 4
Descripción tabla analisis_saberpro

Entidad	analisis_saberpro
Atributos	anio, label_titulo, indicador, centroide, cluster, atributo
Descripción	Esta tabla guarda los resultados generados por el análisis descriptivo en los diferentes años para la prueba Saber Pro.

Tabla 5
Descripción tabla sb11_[anio]

Entidad	sb11_[anio]
Atributos	estu_genero, periodo, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, fami_tieneserviciotv, fami_tienecomputador, cole_genero, cole_area_ubicacion, cole_jornada, punt_lectura_critica, punt_matematicas, punt_c_naturales, punt_sociales_ciudadanas, punt_ingles
Descripción	Esta tabla contiene la información de los indicadores pertenecientes a las pruebas Saber 11 de los diferentes años.

Tabla 6
Descripción tabla sabertyt_[anio]

Entidad	sabertyt_[anio]
Atributos	estu_genero, periodo, estu_areareside, estu_estadocivil, estu_pagomatriculacredito, fami_cabezafamilia, fami_numpersonasacargo, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, fami_tieneinternet, estu_dedicacionlecturadiaria, gruporeferencia, mod_razona_cuantitat_punt, mod_lectura_critica_punt, mod_competen_ciudada_punt, mod_ingles_punt, mod_comuni_escrita_punt
Descripción	Esta tabla contiene la información de los indicadores pertenecientes a las pruebas Saber TyT de los diferentes años.

Tabla 7
Descripción tabla saberpro[anio]

Entidad	saberpro_[anio]
Atributos	estu_genero, periodo, estu_areareside, estu_estadocivil, estu_valormatriculauniversidad, estu_pagomatriculacredito, fami_cabezafamilia, fami_numpersonasacargo, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, fami_tieneinternet, fami_tieneserviciotv, estu_dedicacionlecturadiaria, gruporeferencia, estu_metodo_pgrm,

	mod_razona_cuantitat_punt, mod_lectura_critica_punt, mod_competen_ciudada_punt, mod_ingles_punt, mod_comuni_escrita_punt
Descripción	Esta tabla contiene la información de los indicadores pertenecientes a las pruebas Saber Pro de los diferentes años.

Tabla 8
Descripción tabla sb11_muestra

Entidad	sb11_muestra
Atributos	estu_genero, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, fami_tienecomputador, cole_jornada, punt_lectura, punt_matematicas, punt_ciencias, punt_sociales, punt_ingles
Descripción	Esta tabla guarda los resultados generados por el análisis predictivo para la prueba Saber 11.

Tabla 9
Descripción tabla sabertyt_muestra

Entidad	sabertyt_muestra
Atributos	estu_genero, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, fami_tieneinternet, punt_lectura, punt_matematicas, punt_ciencias, punt_sociales, punt_ingles
Descripción	Esta tabla guarda los resultados generados por el análisis predictivo para la prueba Saber TyT.

Tabla 10
Descripción tabla saberpro_muestra

Entidad	saberpro_muestra
Atributos	estu_genero, estu_valormatriculauniversidad, fami_estratovivienda, fami_educacionmadre, fami_educacionpadre, punt_lectura, punt_matematicas, punt_ciencias, punt_sociales, punt_ingles
Descripción	Esta tabla guarda los resultados generados por el análisis predictivo para la prueba Saber Pro.

Tabla 11
Descripción tabla aplicacion

Entidad	aplicacion
Atributos	prueba, categoria, nucleo, registro, indicador, genero, puntaje
Descripción	Esta tabla guarda los resultados generados por el análisis predictivo para las diferentes pruebas Saber.

3.3.2. DIAGRAMAS DE CASOS DE USO

Roles propuestos

Tabla 12
Descripción de roles propuestos

Administrador	Es la persona encargada de administrar el CRUD de usuarios del aplicativo web, también podrá ingresar la información para cargar las pruebas en la base de datos.
Aplicativo	El aplicativo se encarga de todos los cálculos necesarios para el análisis descriptivo, al igual que el análisis predictivo. Todo esto será por medio de algoritmos de minería de datos
Usuario	El usuario es la persona encargada de las consultas de los análisis al igual que las visualizaciones.

Diagramas

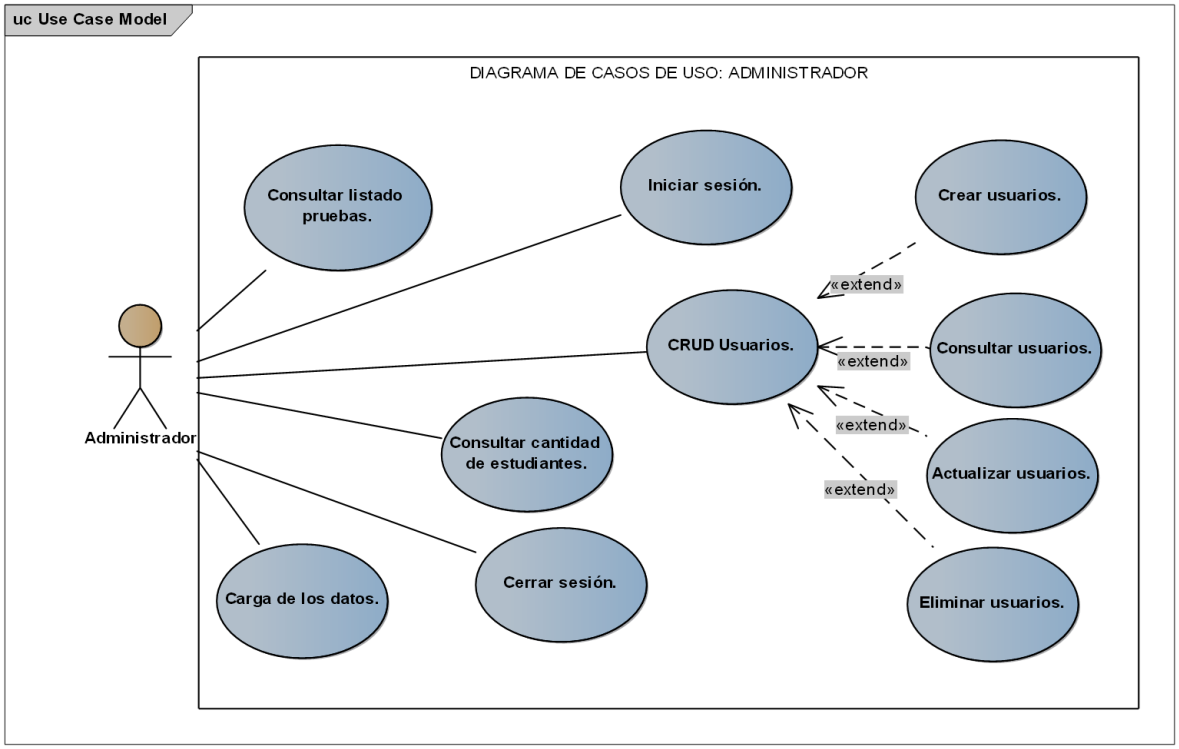


Figura 34. Diagrama de casos de uso del Administrador
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 13
Descripción diagrama de casos de uso administrador

DESCRIPCIÓN DIAGRAMA DE CASOS DE USO	
NOMBRE	**Iniciar sesión**
ACTOR(ES)	Administrador – Usuario
DESCRIPCIÓN	El administrador y usuario deben ingresar los datos de ingreso al aplicativo, los cuales son: usuario y contraseña. Si son correctos se redireccionará según su rol.
NOMBRE	**CRUD Usuarios.**
ACTOR(ES)	Administrador
DESCRIPCIÓN	El administrador se encarga de la gestión de los usuarios; puede crear, consultar, actualizar o eliminar los datos.

NOMBRE	**Consultar listado de pruebas**
ACTOR(ES)	Administrador
DESCRIPCIÓN	El administrador puede visualizar el listado de pruebas disponibles en el aplicativo.
NOMBRE	**Carga de los datos**
ACTOR(ES)	Administrador
DESCRIPCIÓN	Se encarga de la carga de los datos de las pruebas, para su posterior limpieza y tratamiento.
NOMBRE	**Consultar cantidad de estudiantes**
ACTOR(ES)	Administrador
DESCRIPCIÓN	El administrador puede visualizar la cantidad de estudiantes que presentaron las diferentes pruebas Saber.
NOMBRE	**Cerrar sesión.**
ACTOR(ES)	Administrador – Usuario
DESCRIPCIÓN	El administrador y usuario pueden finalizar su sesión.

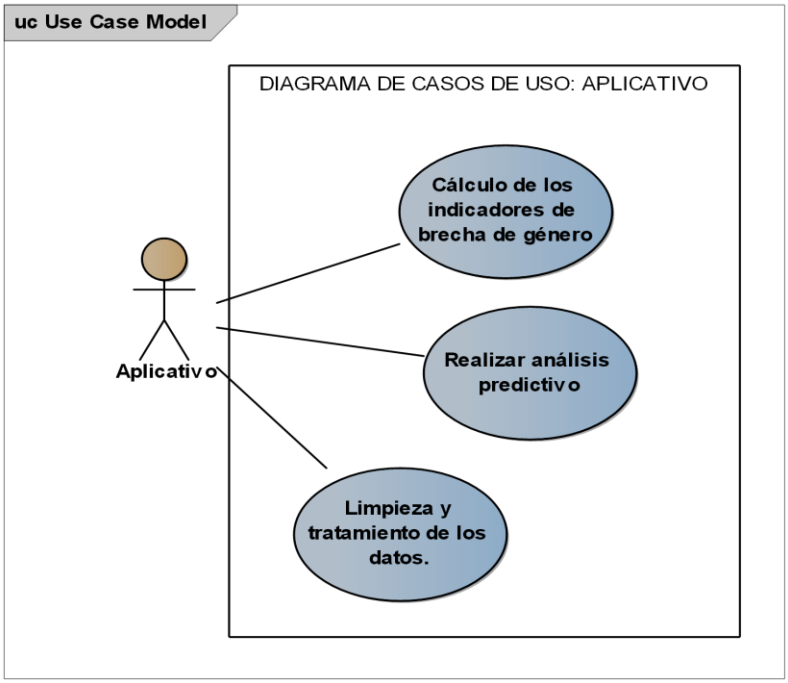


Figura 35. Diagrama de casos de uso del Aplicativo
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 14
Descripción diagrama de casos de uso aplicativo

DESCRIPCIÓN DIAGRAMA DE CASO DE USO	
NOMBRE	**Cálculo de los indicadores de brecha de género**
ACTOR(ES)	Aplicativo
DESCRIPCIÓN	Hacer uso de los algoritmos de minería de datos para encontrar el comportamiento de los indicadores de brecha de género en las pruebas.
NOMBRE	**Realizar análisis predictivo**
ACTOR(ES)	Aplicativo
DESCRIPCIÓN	Hacer uso del algoritmo de minería de datos para lograr predecir la relación entre núcleos temáticos de las pruebas que sean comunes en las diferentes pruebas.
NOMBRE	**Limpieza y tratamiento de los datos**
ACTOR(ES)	Aplicativo
DESCRIPCIÓN	El aplicativo realiza la limpieza de los datos previamente cargados por el administrador y procede a subirlos en la Base de Datos.

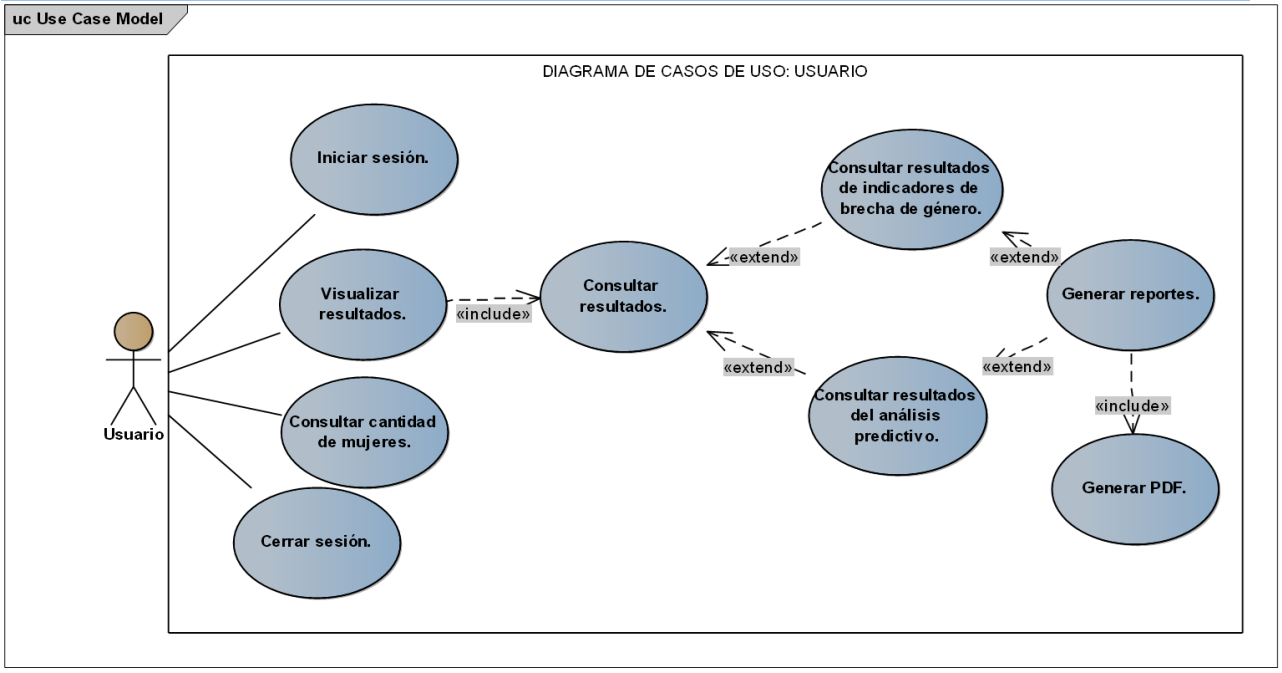


Figura 36. Diagrama de casos de uso del Usuario
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 15
Descripción diagrama de casos de uso usuario

DESCRIPCIÓN DIAGRAMA DE CASO DE USO	
NOMBRE	**Visualizar resultados**
ACTOR(ES)	Usuario
DESCRIPCIÓN	El usuario puede visualizar los resultados obtenidos tanto en el análisis descriptivo como predictivo, y puede filtrar por año o por indicadores en cada una de las pruebas. En el caso del análisis predictivo se filtra por núcleo temático.
NOMBRE	**Generar reportes**
ACTOR(ES)	Usuario
DESCRIPCIÓN	El usuario puede descargar en formato PDF los resultados obtenidos en el análisis de las pruebas, mostrándose gráficas de barras y tortas en estos reportes.
NOMBRE	**Consultar cantidad de mujeres**
ACTOR(ES)	Usuario
DESCRIPCIÓN	El usuario puede consultar la cantidad de mujeres que presentaron las pruebas según el año.

3.3.3. DIAGRAMAS DE SECUENCIAS

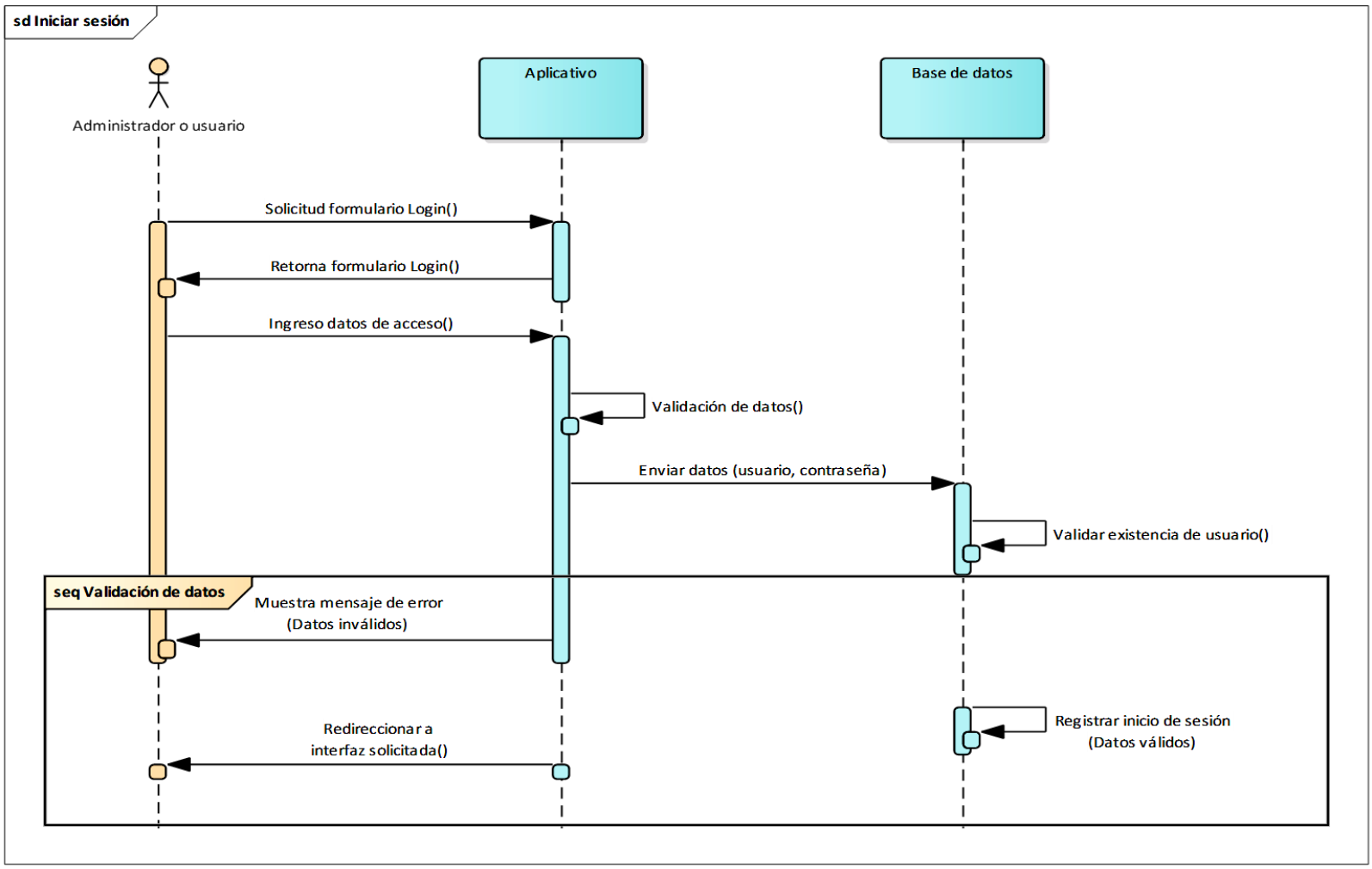


Figura 37. Diagrama de secuencia Iniciar sesión
Autores: (Manrique, Celeita, Barahona, 2020)

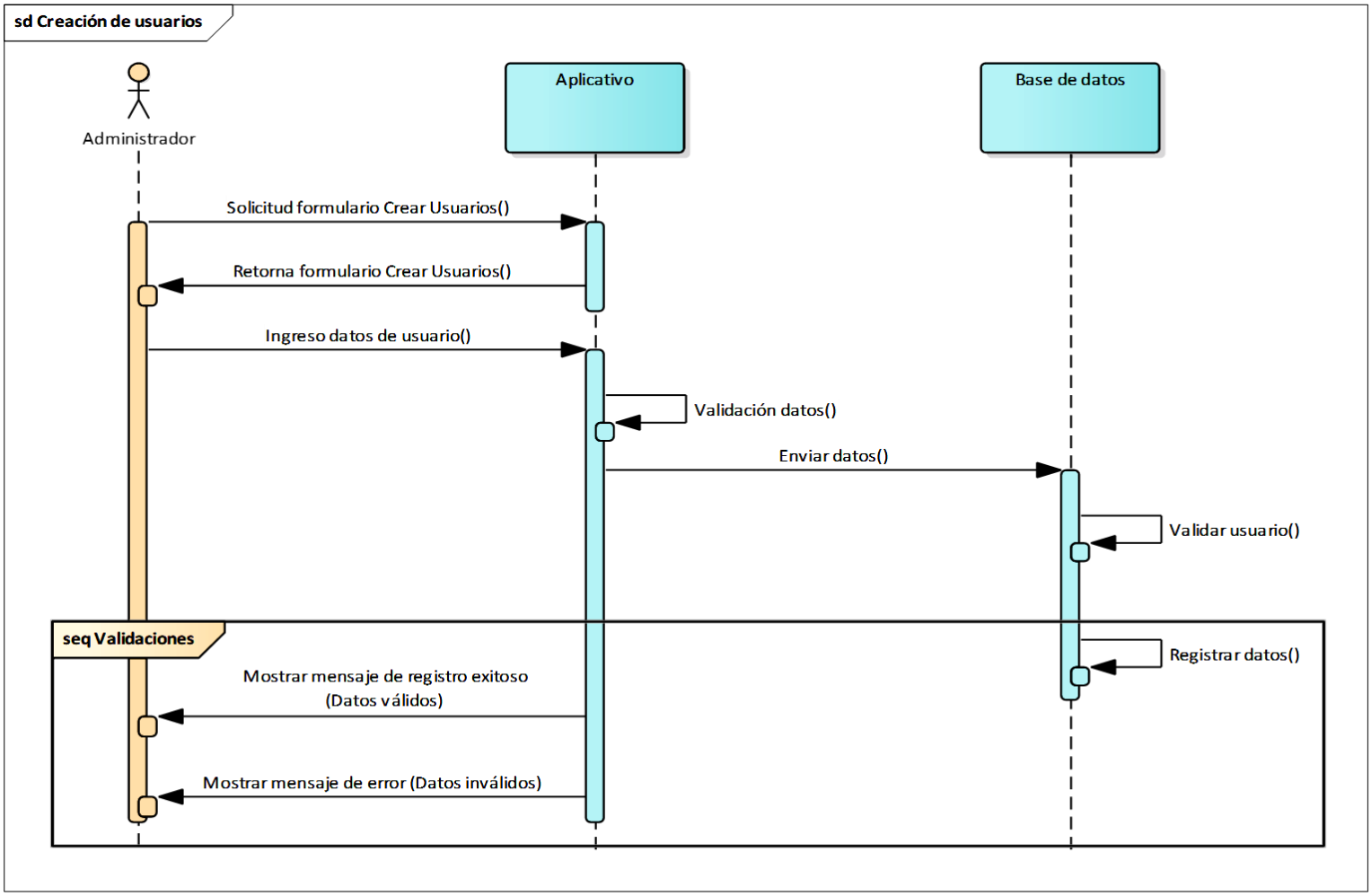


Figura 38. Diagrama de secuencia Crear usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

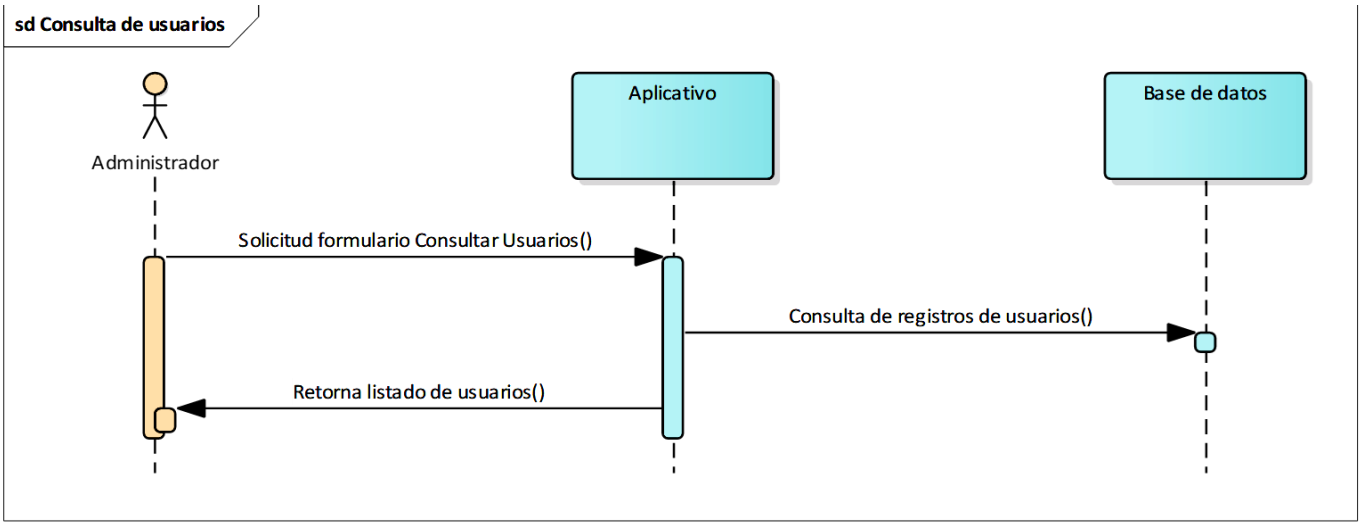


Figura 39. Diagrama de secuencia Consultar usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

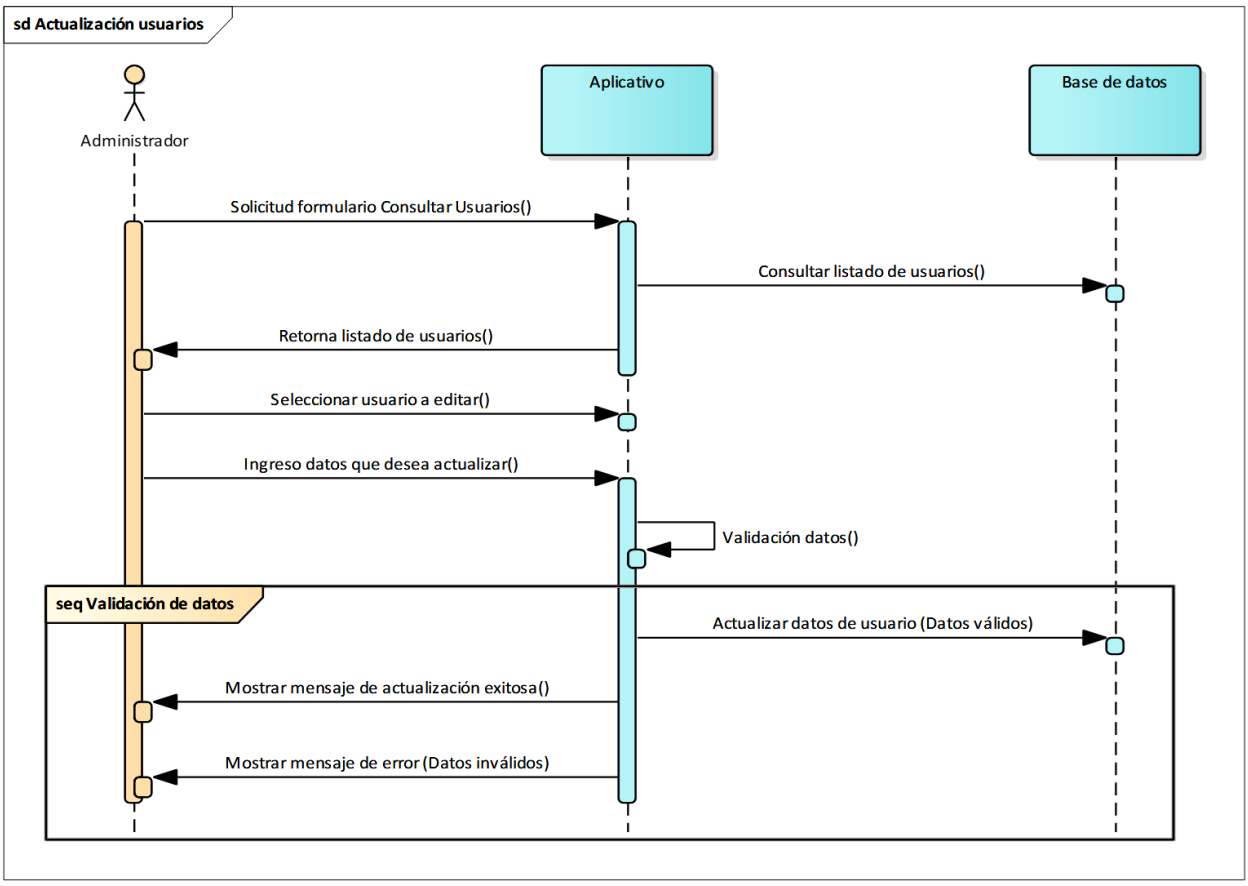


Figura 40. Diagrama de secuencia Actualizar usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

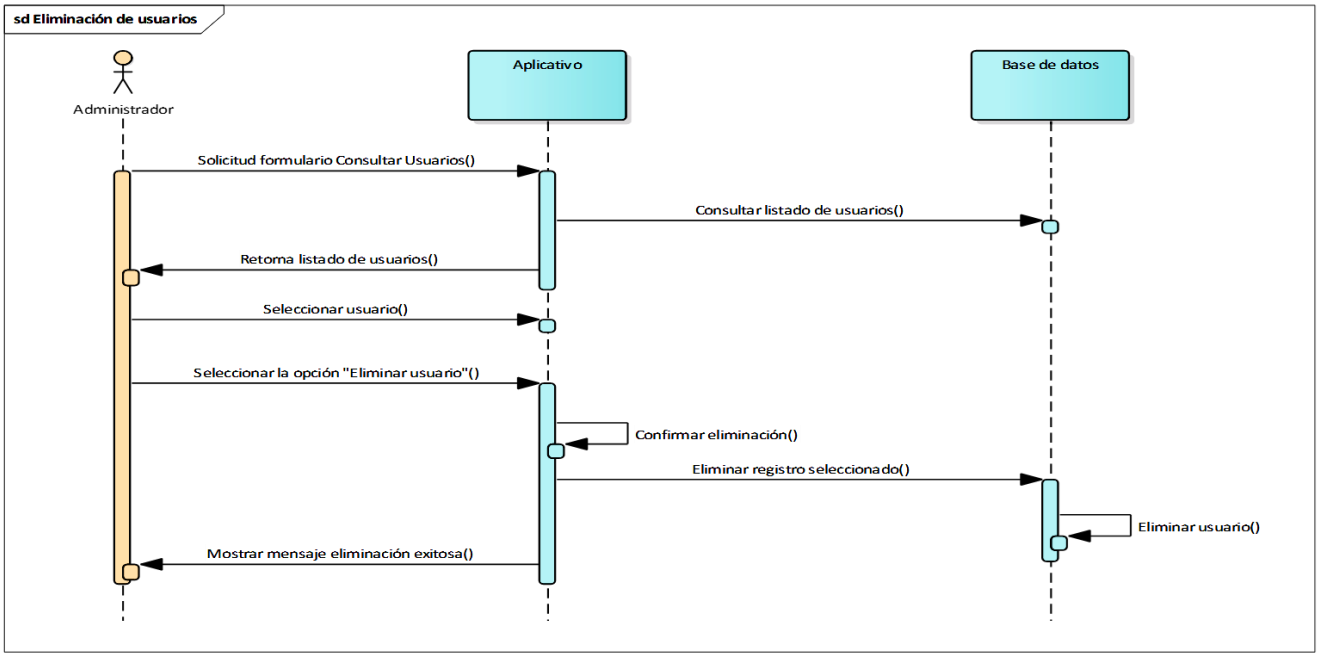


Figura 41. Diagrama de secuencia Eliminar usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

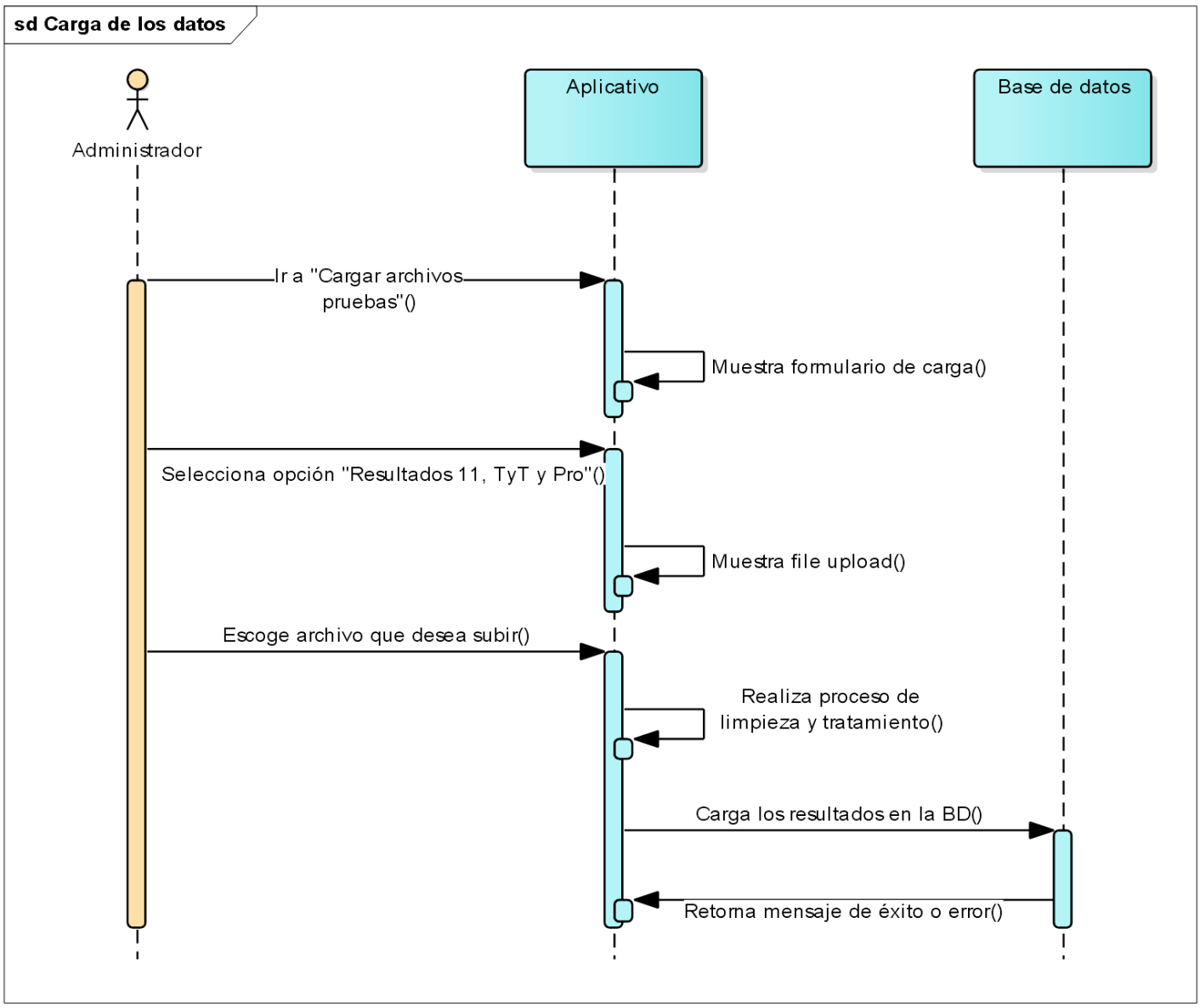


Figura 40. Diagrama de secuencia Carga de los datos
Autores: (Manrique, Celeita, Barahona, 2020)

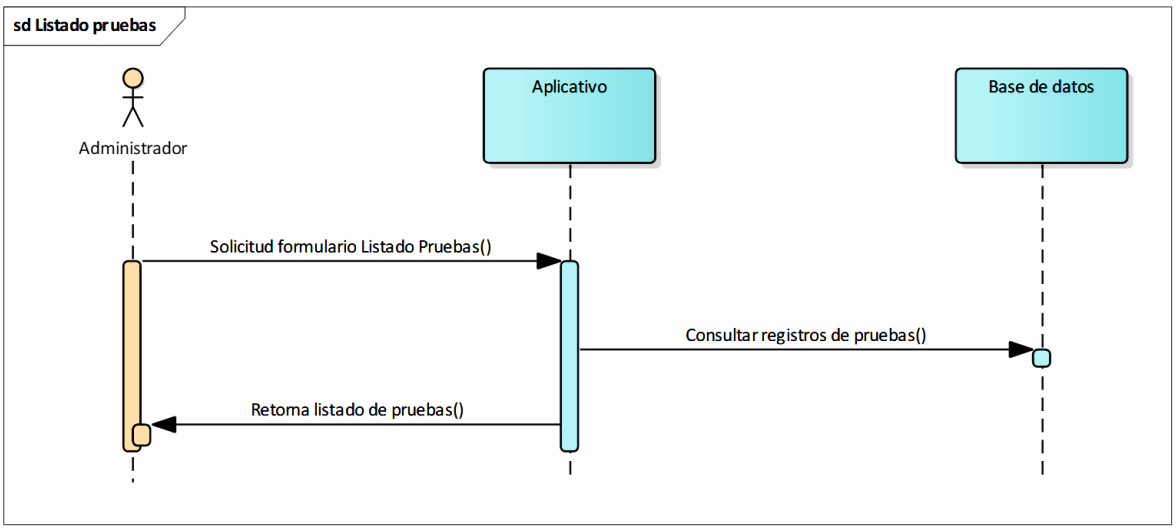


Figura 41. Diagrama de secuencia Consultar listado pruebas
Autores: (Manrique, Celeita, Barahona, 2020)

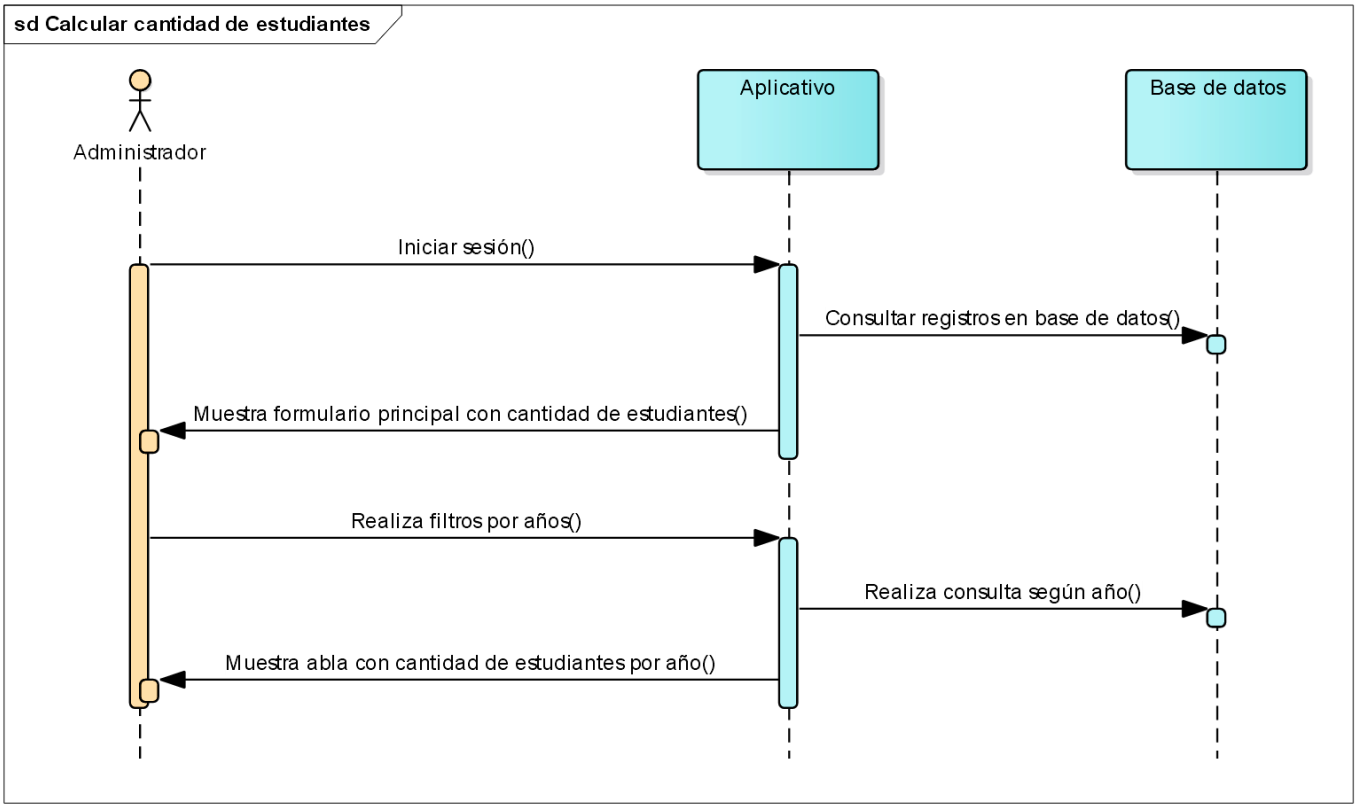


Figura 42. Diagrama de secuencia Consultar cantidad de estudiantes
Autores: (Manrique, Celeita, Barahona, 2020)

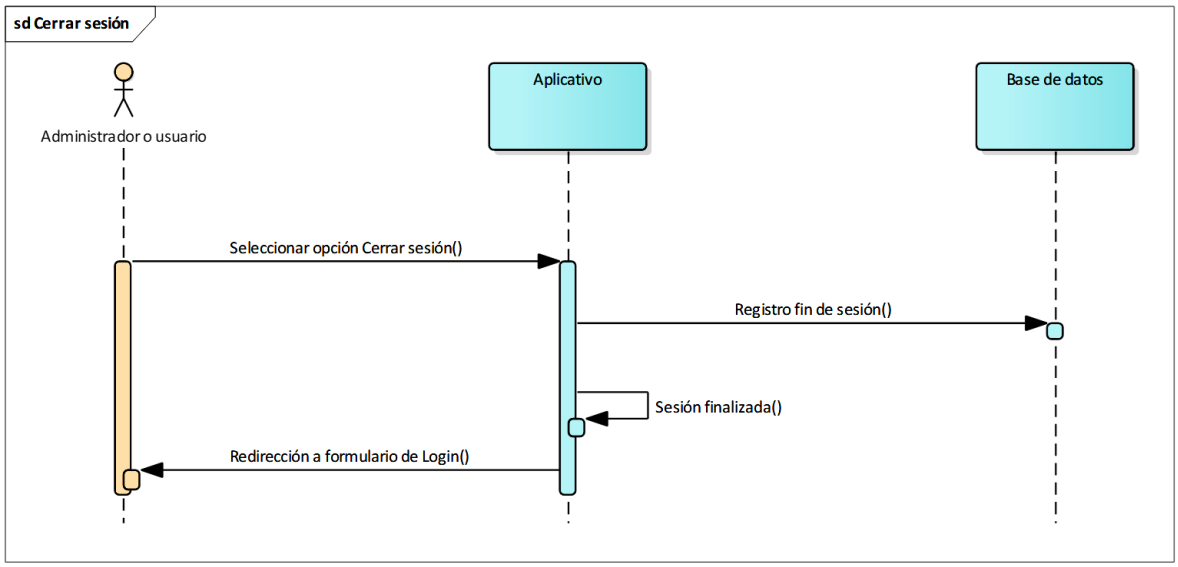


Figura 43. Diagrama de secuencia Cerrar sesión
Autores: (Manrique, Celeita, Barahona, 2020)

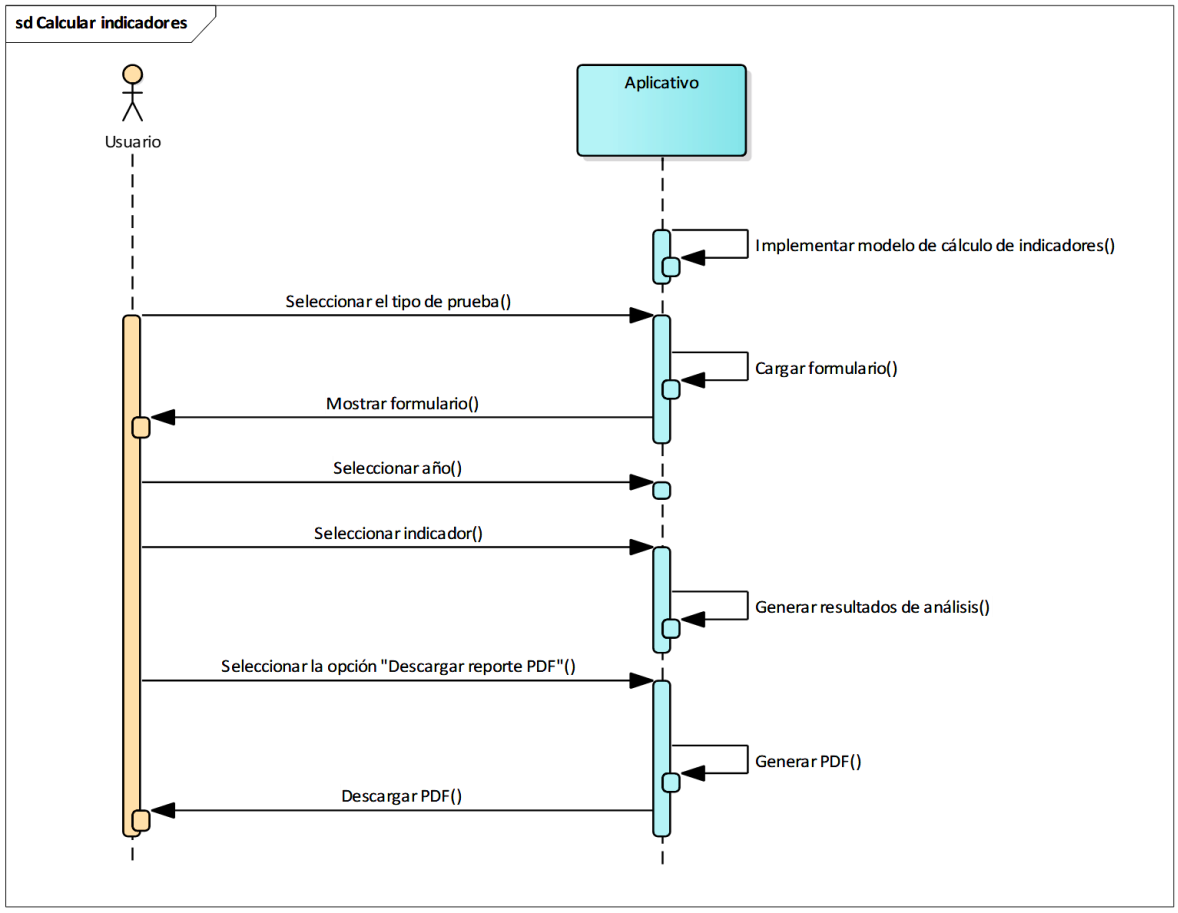


Figura 44. Diagrama de secuencia Calcular indicadores de brecha de género
Autores: (Manrique, Celeita, Barahona, 2020)

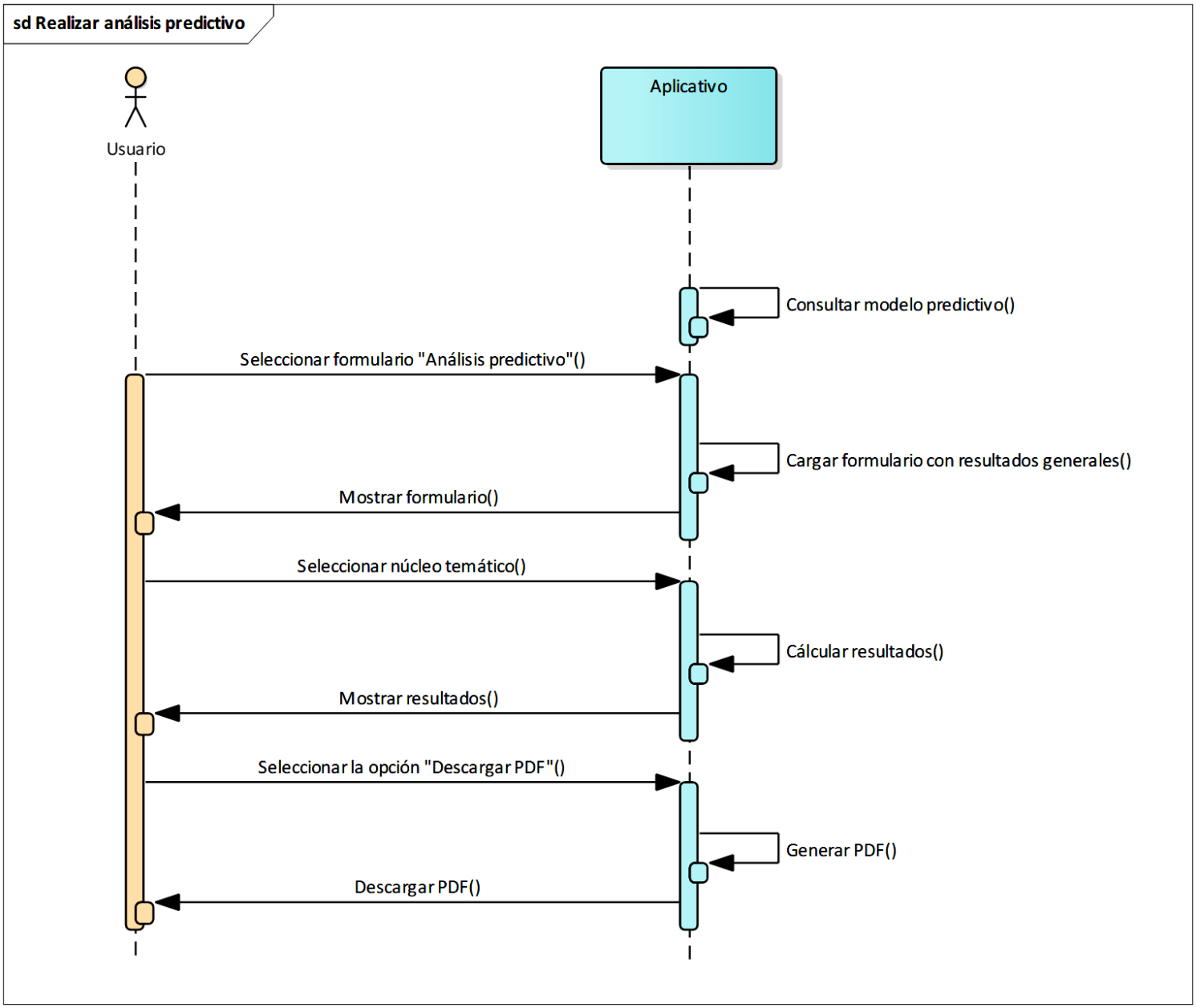


Figura 45. Diagrama de secuencia Realizar análisis predictivo
Autores: (Manrique, Celeita, Barahona, 2020)

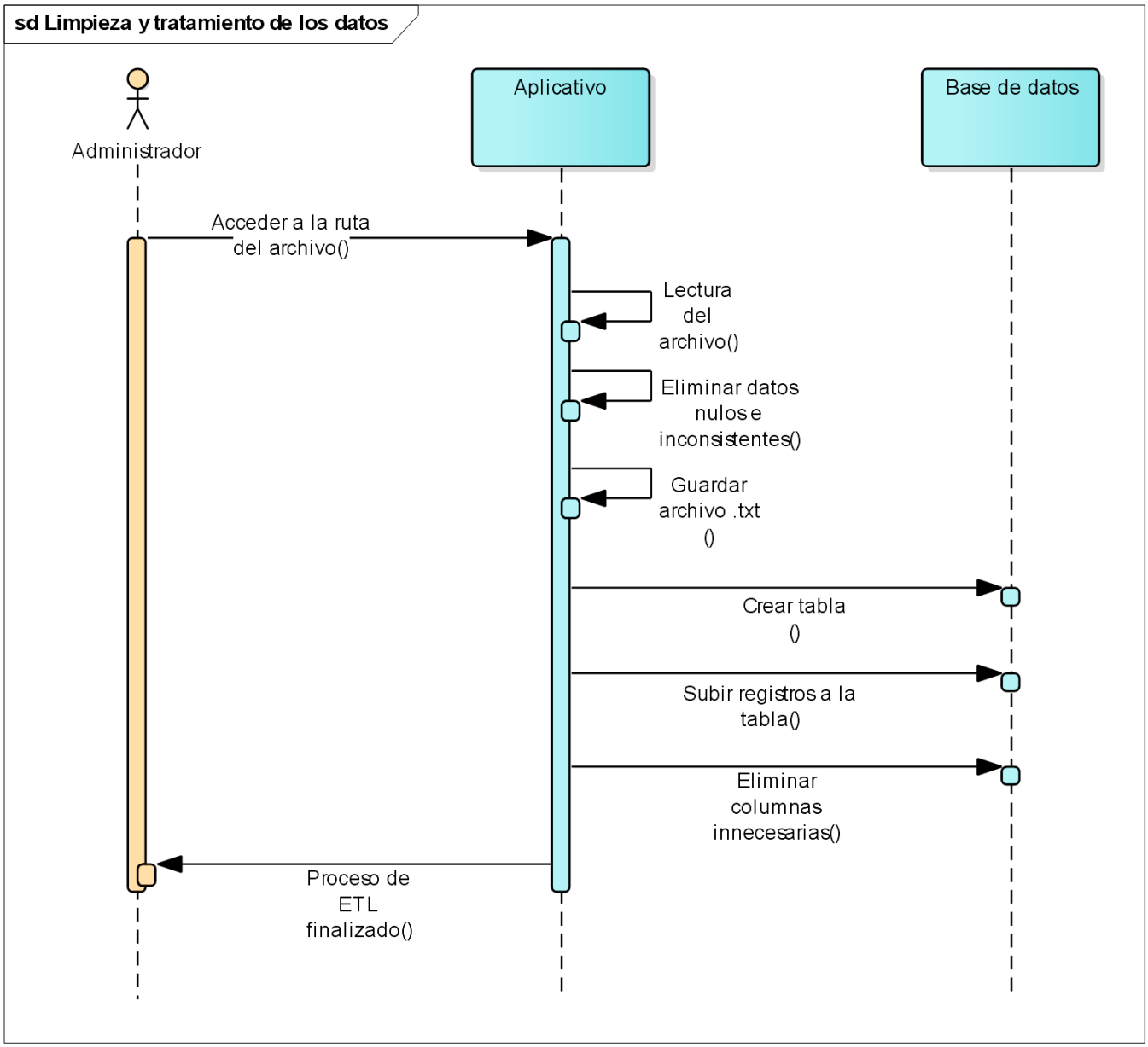


Figura 46. Diagrama de secuencia Limpieza y tratamiento de los datos
Autores: (Manrique, Celeita, Barahona, 2020)

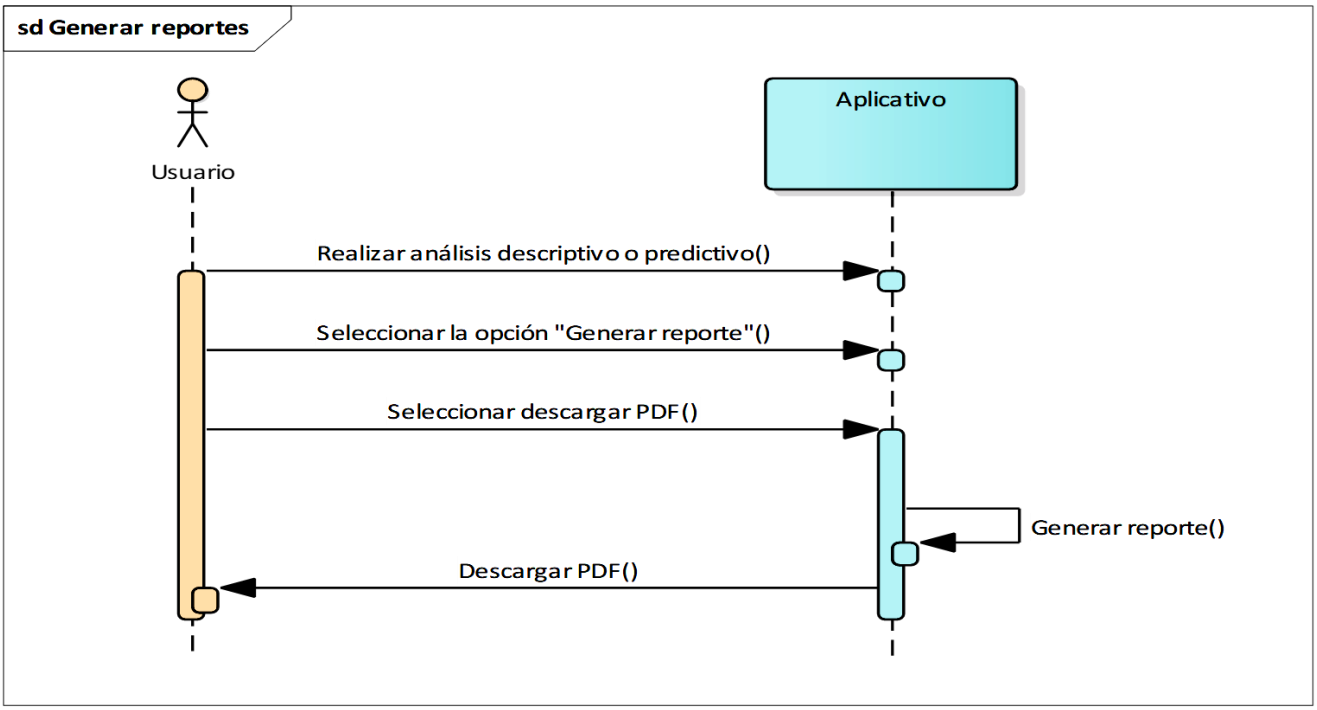


Figura 47. Diagrama de secuencia Generar reportes
Autores: (Manrique, Celeita, Barahona, 2020)

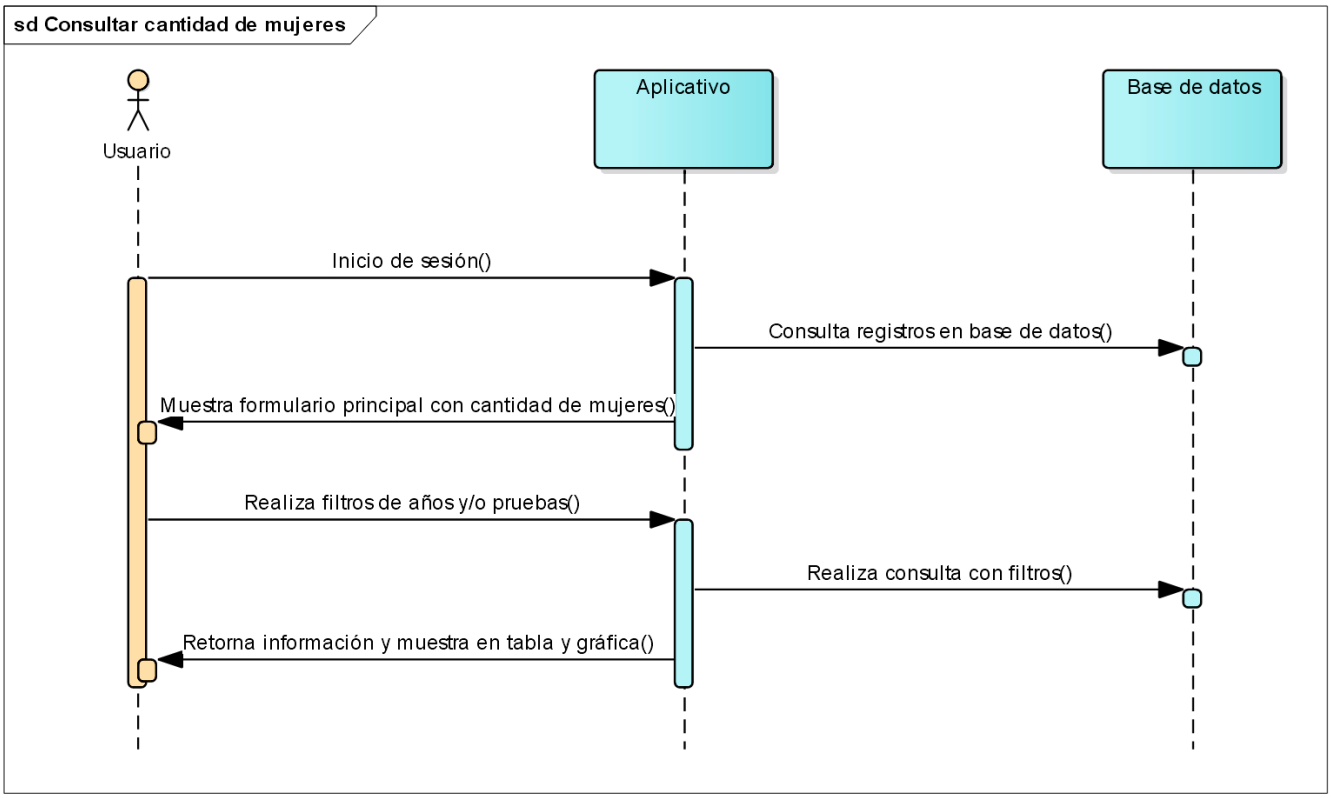


Figura 48. Diagrama de secuencia Consultar cantidad de mujeres
Autores: (Manrique, Celeita, Barahona, 2020)

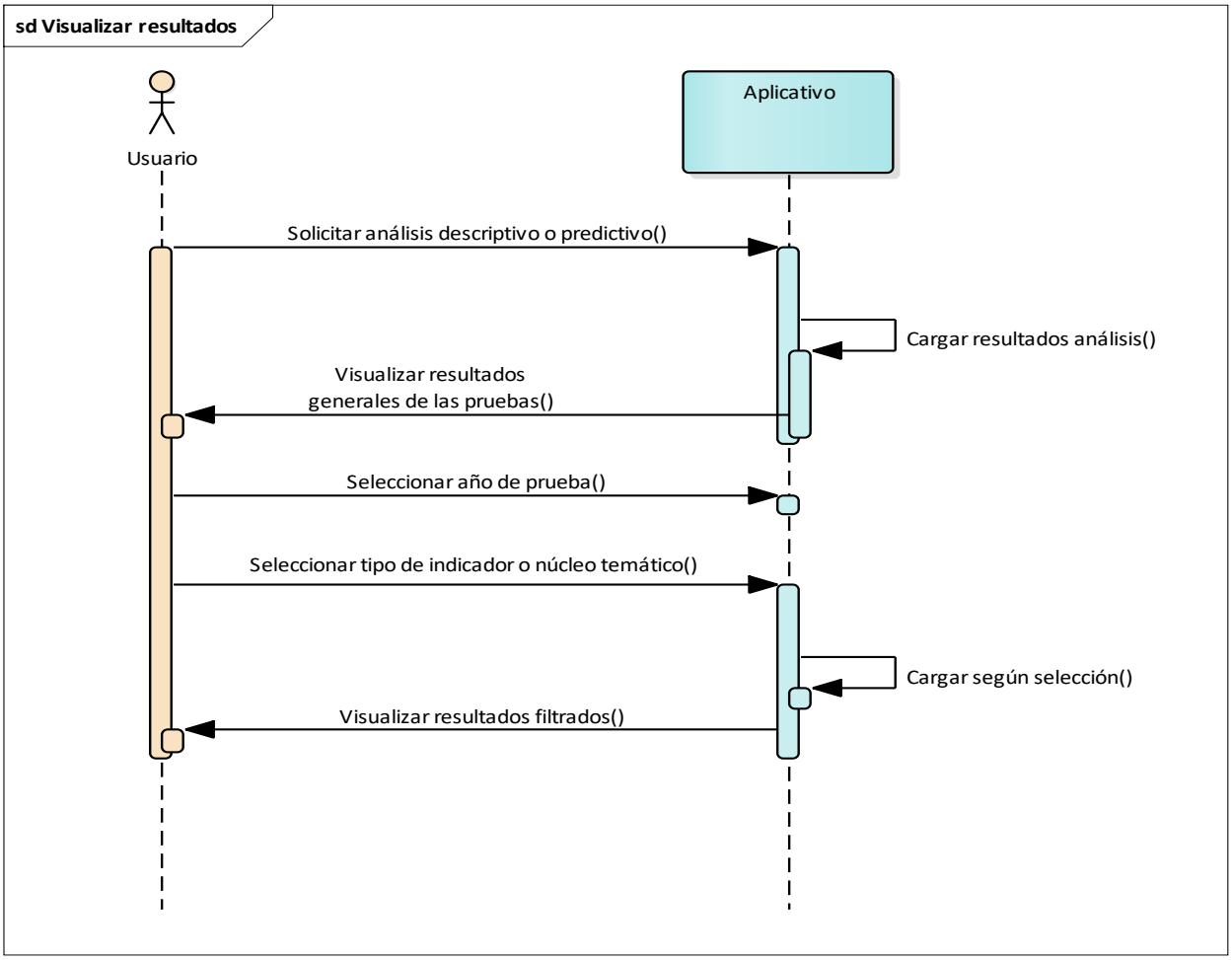


Figura 49. Diagrama de secuencia Visualizar resultados
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 16
Descripción de los objetos de los diagramas de secuencia

DESCRIPCIÓN OBJETOS DIAGRAMAS DE SECUENCIA	
OBJETO	Administrador
DESCRIPCIÓN	Cumple la función de gestionar el aplicativo y mantener la información actualizada de las pruebas. Así mismo se encarga de la gestión de usuarios.
OBJETO	Usuario
DESCRIPCIÓN	Este objeto es importante porque realiza múltiples peticiones de consulta en los diferentes diagramas tanto para visualizar los resultados de los análisis como para descargar reportes de esta misma información.

OBJETO	Aplicativo
DESCRIPCIÓN	Este objeto en términos generales tiene la función de calcular los indicadores de brecha de género y realizar el análisis predictivo, así como tareas básicas de mostrar formularios, entre otras.

3.3.4. DIAGRAMAS DE ACTIVIDADES

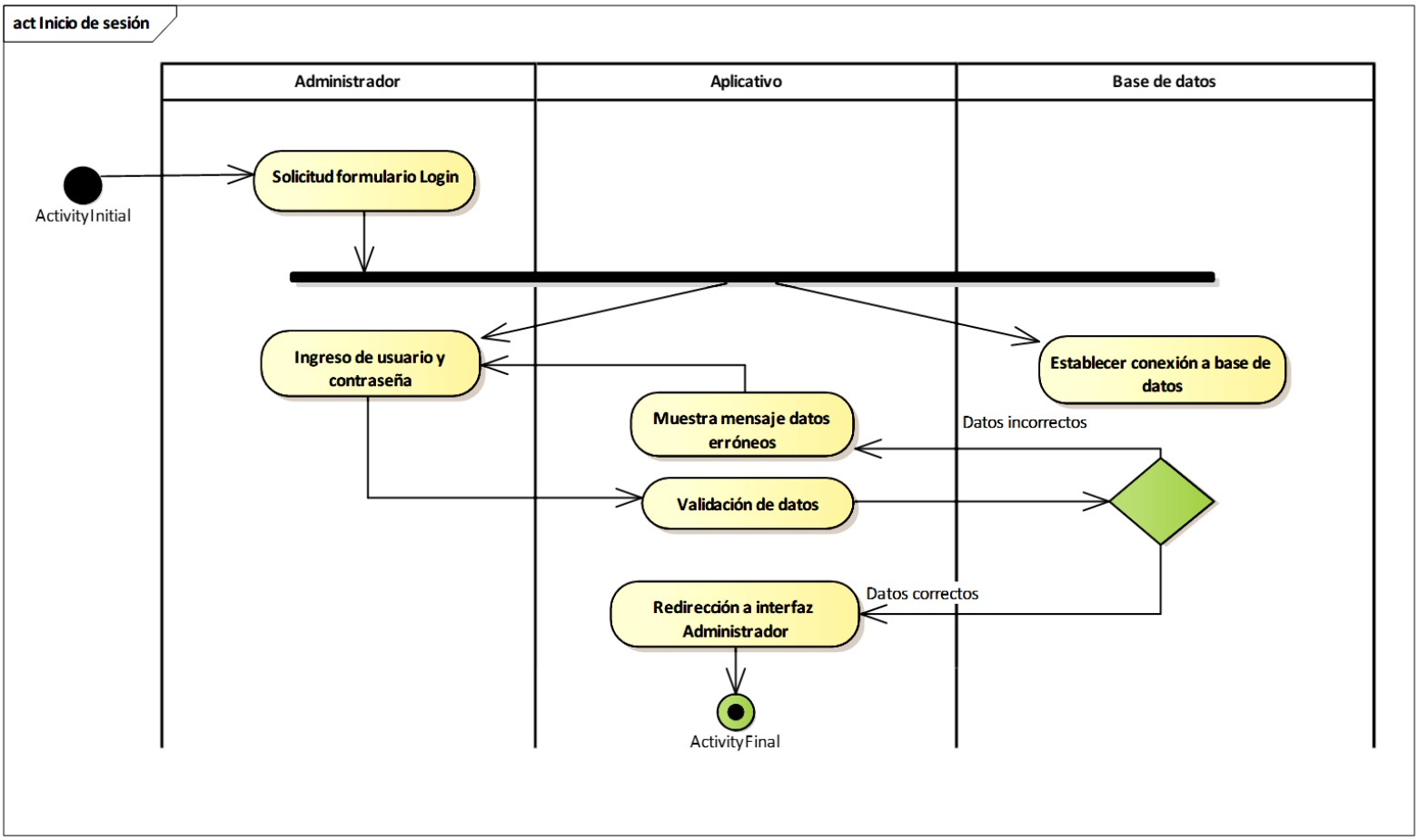


Figura 50. Diagrama de actividad Iniciar sesión
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 17
Descripción actividad Iniciar sesión

NOMBRE ACTIVIDAD	Iniciar sesión
DESCRIPCIÓN	Esta actividad hace referencia al ingreso del administrador o el usuario al aplicativo, para ello deben ingresar su usuario y contraseña para su posterior validación.

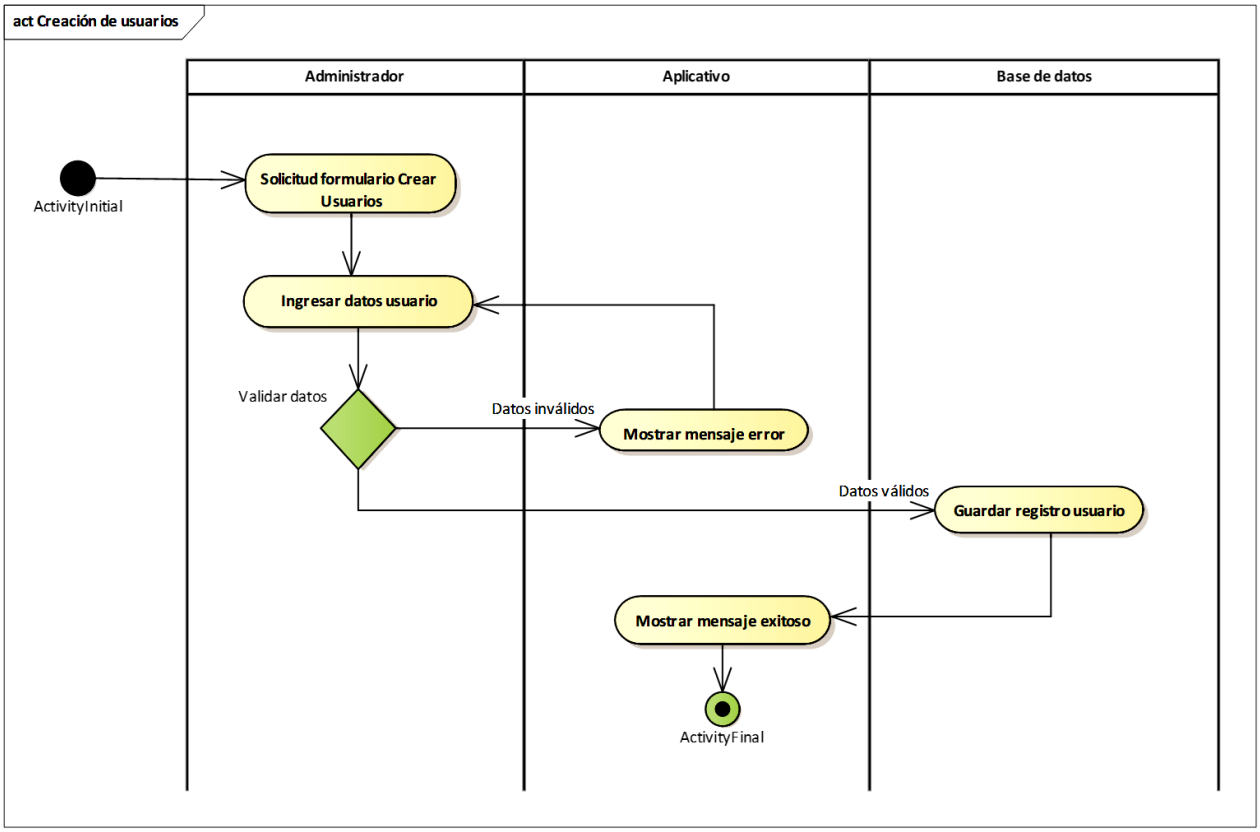


Figura 51. Diagrama de actividad Crear usuario
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 18
Descripción actividad Crear usuario

NOMBRE ACTIVIDAD	Crear usuario
DESCRIPCIÓN	En esta actividad se hace el registro de los usuarios del aplicativo, guardando sus respectivos datos.

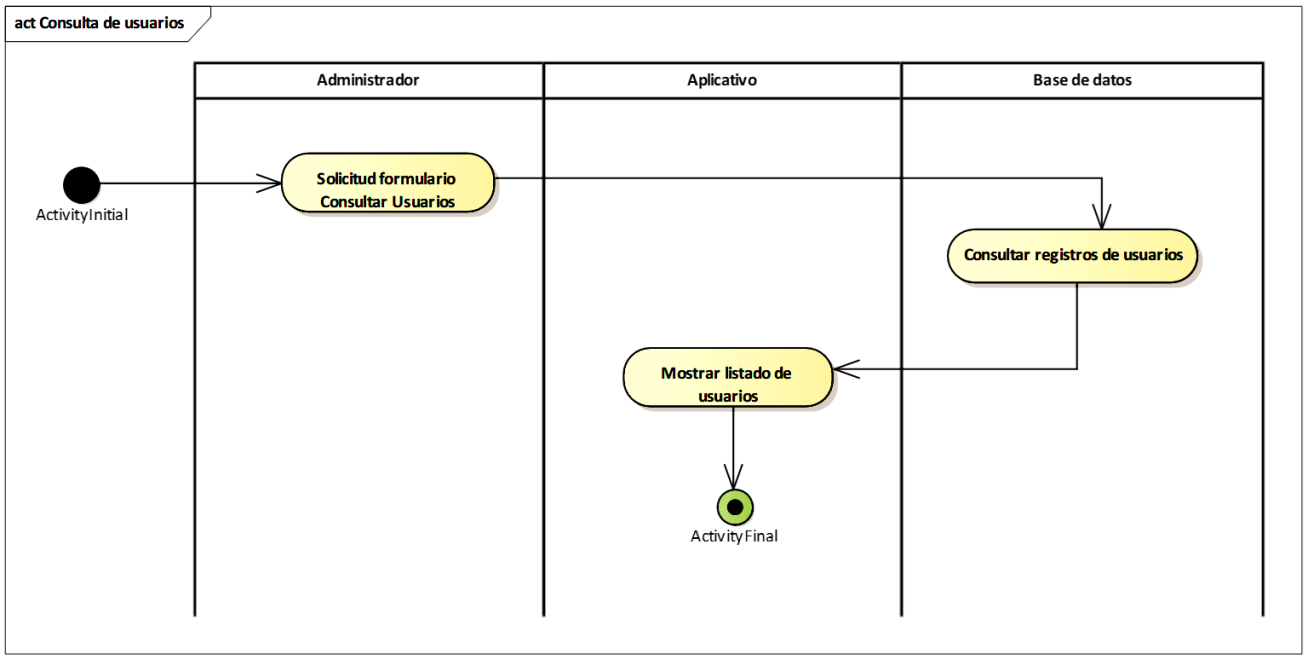


Figura 52. Diagrama de actividad Consultar usuario
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 19
Descripción actividad Consultar usuario

NOMBRE ACTIVIDAD	Consultar usuario
DESCRIPCIÓN	Se realiza una petición de consulta a la base de datos de forma que en el aplicativo se muestre el listado de usuarios existentes.

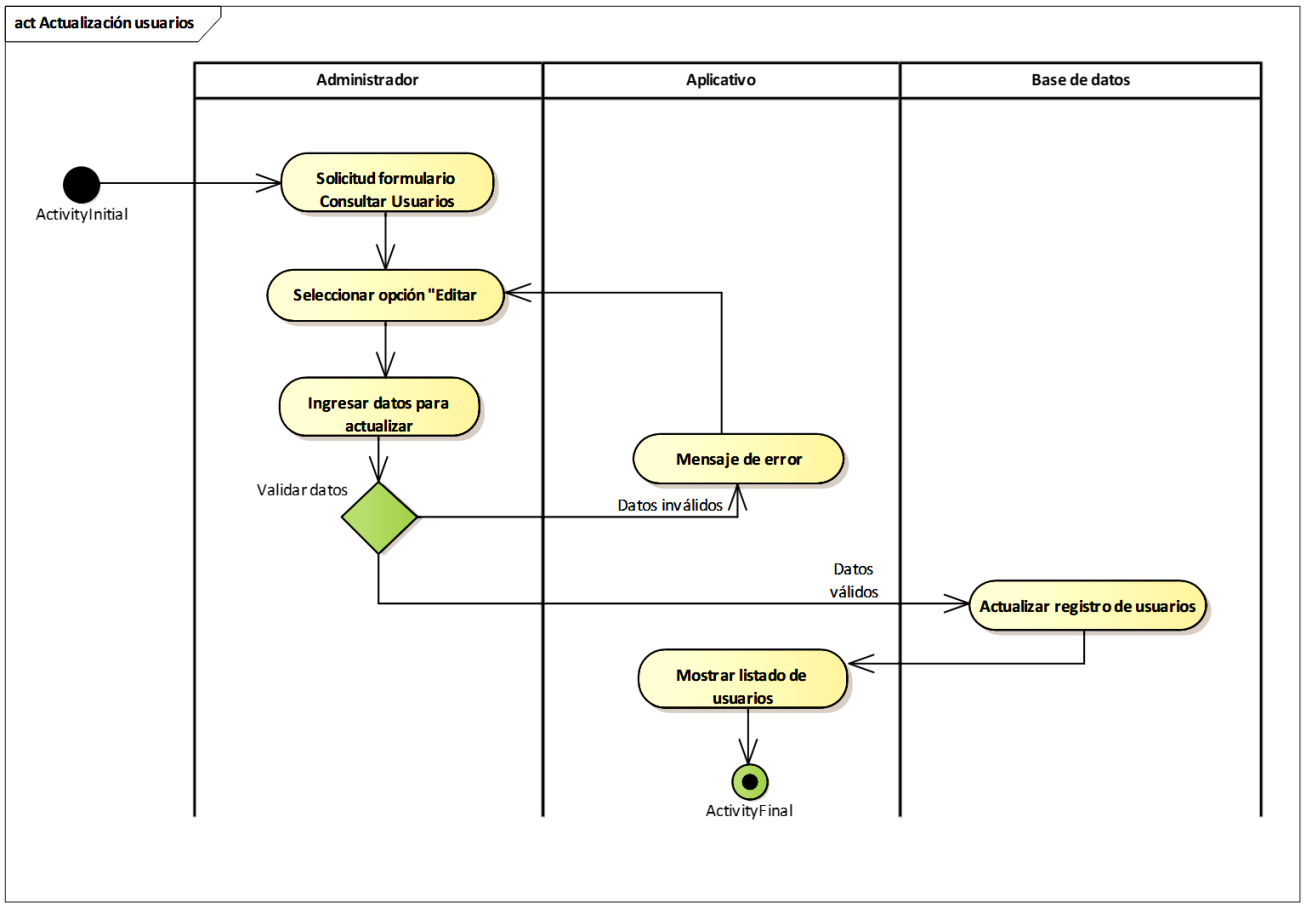


Figura 53. Diagrama de actividad Actualizar usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 20
Descripción actividad Actualizar usuarios

NOMBRE ACTIVIDAD	Actualizar usuarios
DESCRIPCIÓN	Esta actividad permite la edición de los datos actuales de los usuarios y si han sido ingresados de forma correcta actualiza los registros en la base de datos.

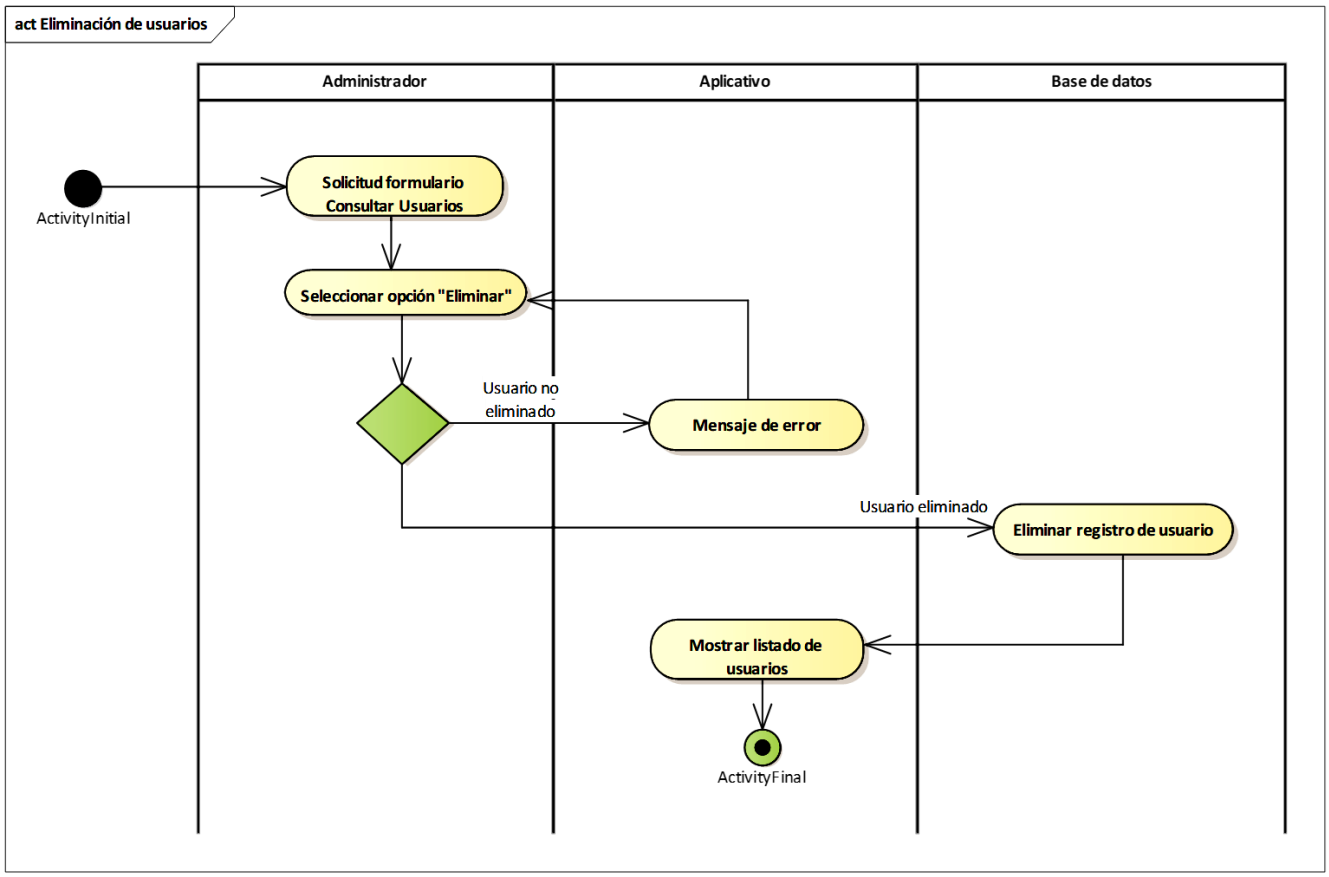


Figura 54. Eliminar usuarios
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 21
Descripción actividad Eliminar usuarios

NOMBRE ACTIVIDAD	Eliminar usuarios
DESCRIPCIÓN	En esta actividad se le permite al administrador poder eliminar los usuarios que desee, no sin antes pidiéndole una confirmación de la acción. Una vez aceptada se procede a borrar los registros de la base de datos.

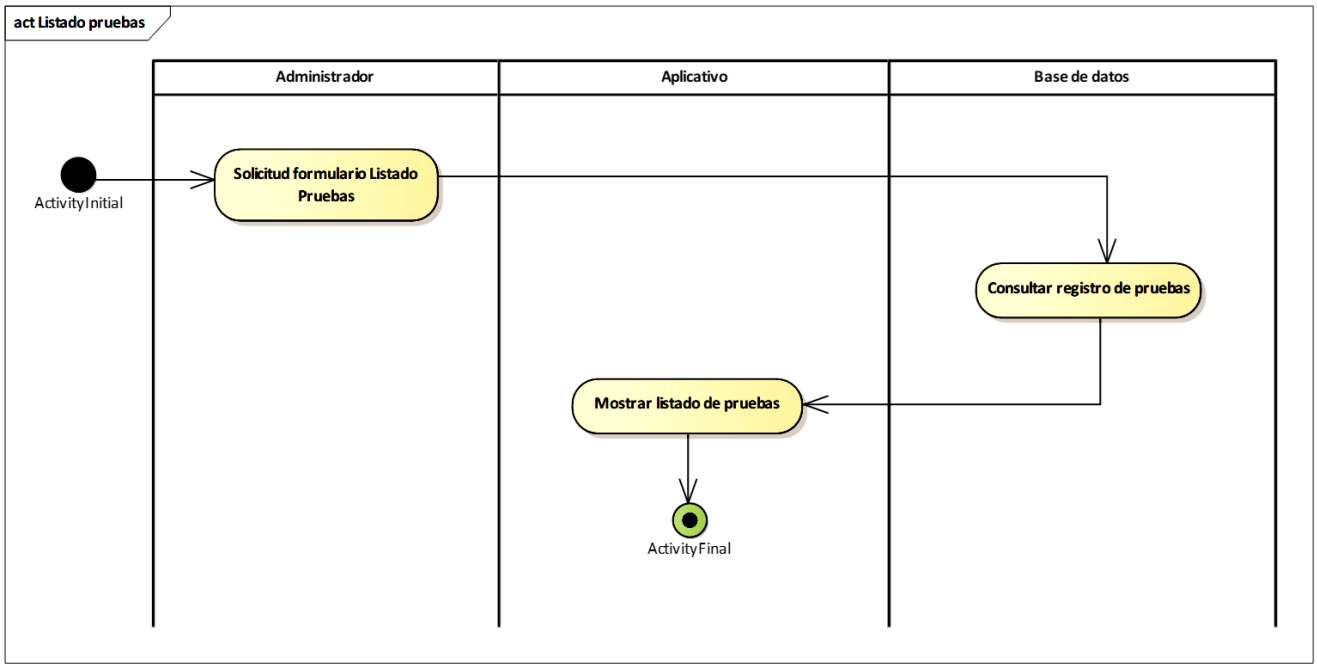


Figura 55. Diagrama de actividad Consultar listado pruebas
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 22
Descripción actividad Consultar listado pruebas

NOMBRE ACTIVIDAD	Consultar listado pruebas
DESCRIPCIÓN	En esta actividad se hace una petición de visualización por parte del administrador para traer el listado de pruebas disponibles en el aplicativo.

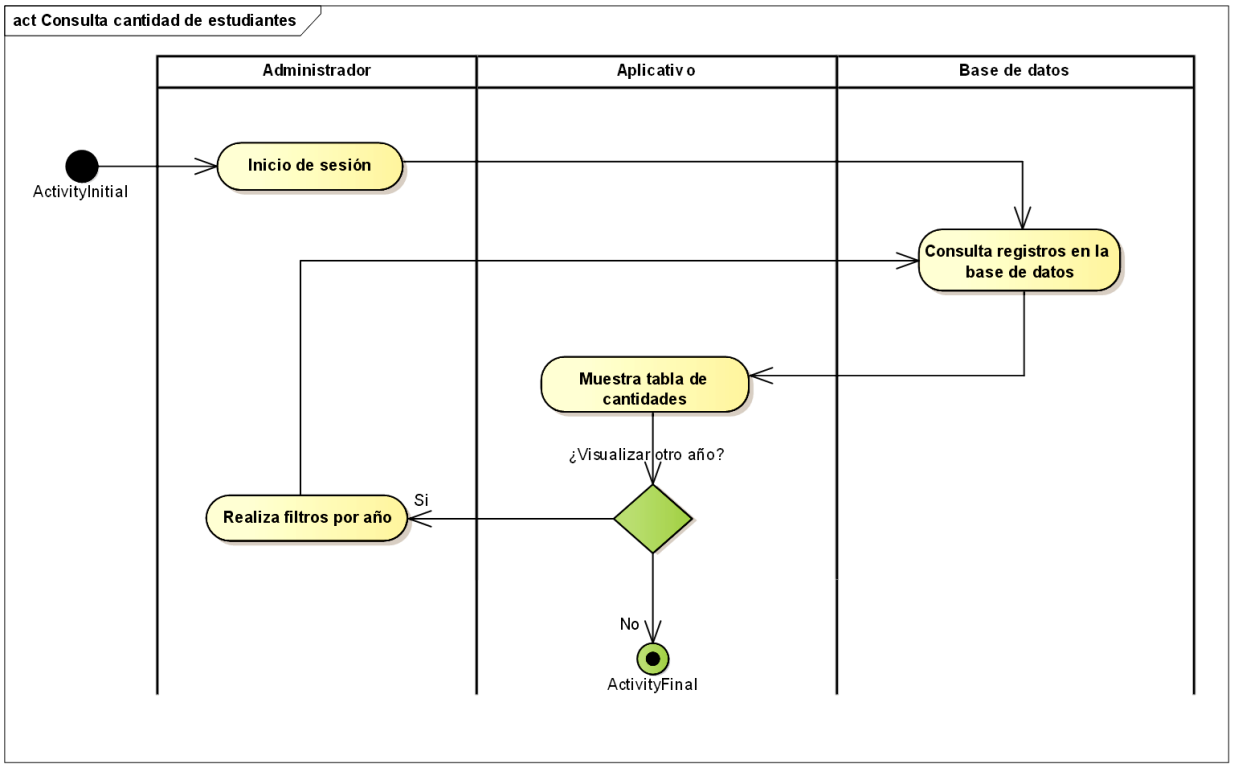


Figura 56. Diagrama de actividad Consultar cantidad de estudiantes
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 23
Descripción actividad Consultar cantidad de estudiantes

NOMBRE ACTIVIDAD	Consultar cantidad de estudiantes
DESCRIPCIÓN	En esta actividad se hace una petición de visualización por parte del administrador para mostrar la cantidad de estudiantes que realizaron las pruebas cada año.

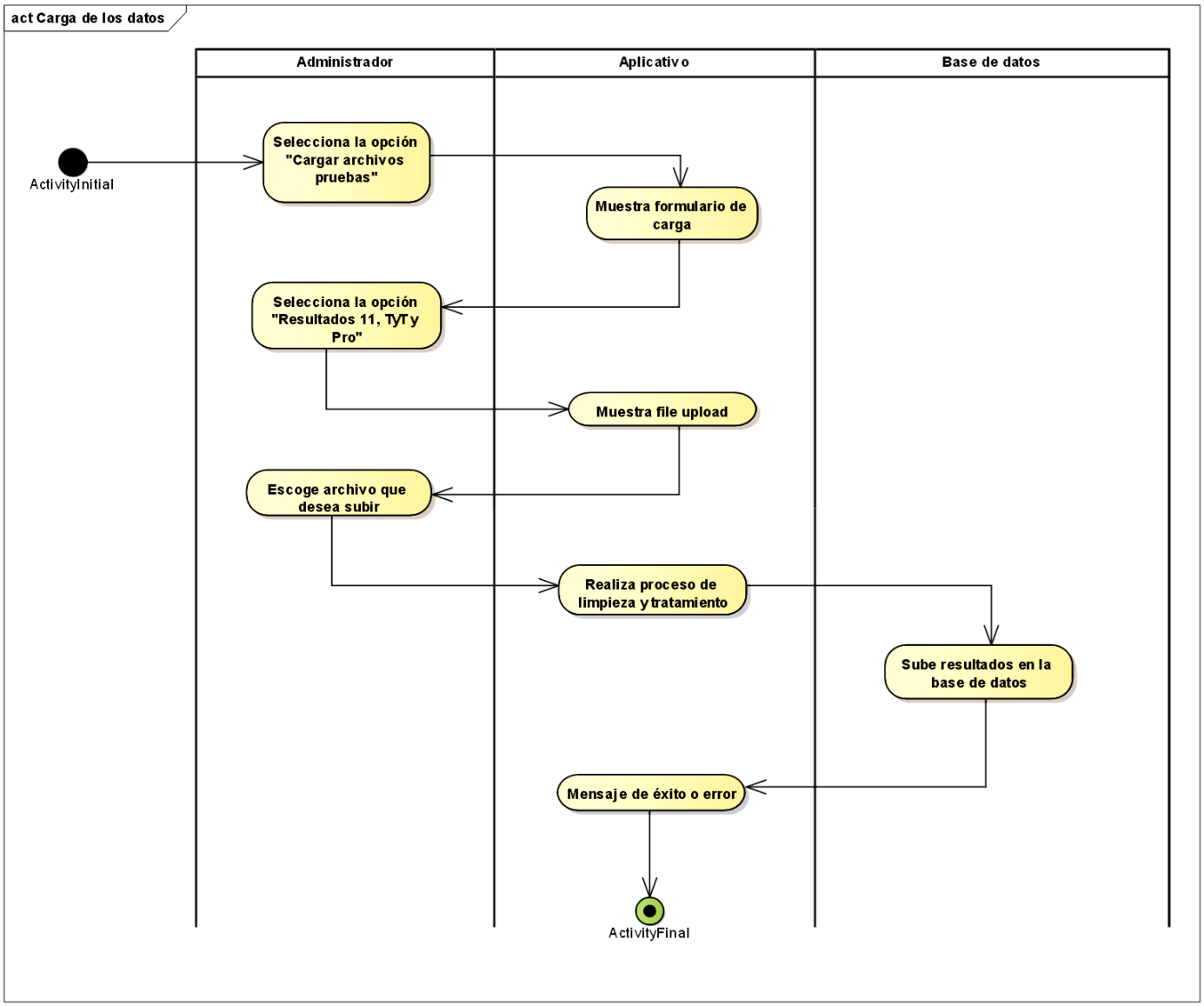


Figura 57.Diagrama de actividad Carga de los datos
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 24
Descripción actividad Carga de los datos

NOMBRE ACTIVIDAD	Carga de los datos
DESCRIPCIÓN	En esta actividad el administrador escoge el archivo que desea subir, en consecuencia, el aplicativo hace su respectiva limpieza y tratamiento para poder cargar los resultados a la base de datos.

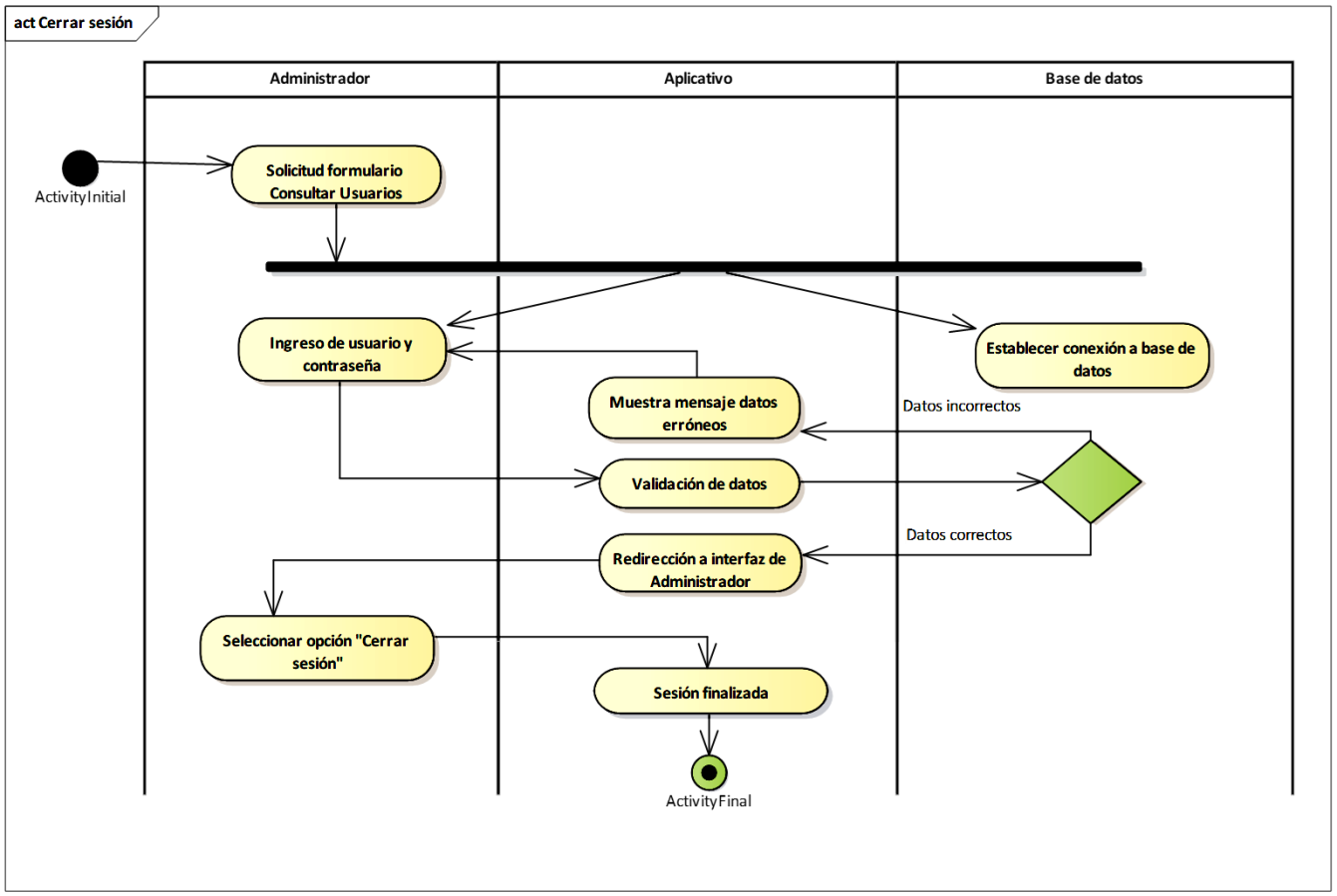


Figura 58. Diagrama de actividad Cerrar sesión
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 25
Descripción actividad Cerrar sesión

NOMBRE ACTIVIDAD	Cerrar sesión
DESCRIPCIÓN	En esta actividad se redirecciona al administrador o usuario al formulario de Login y se finaliza su sesión.

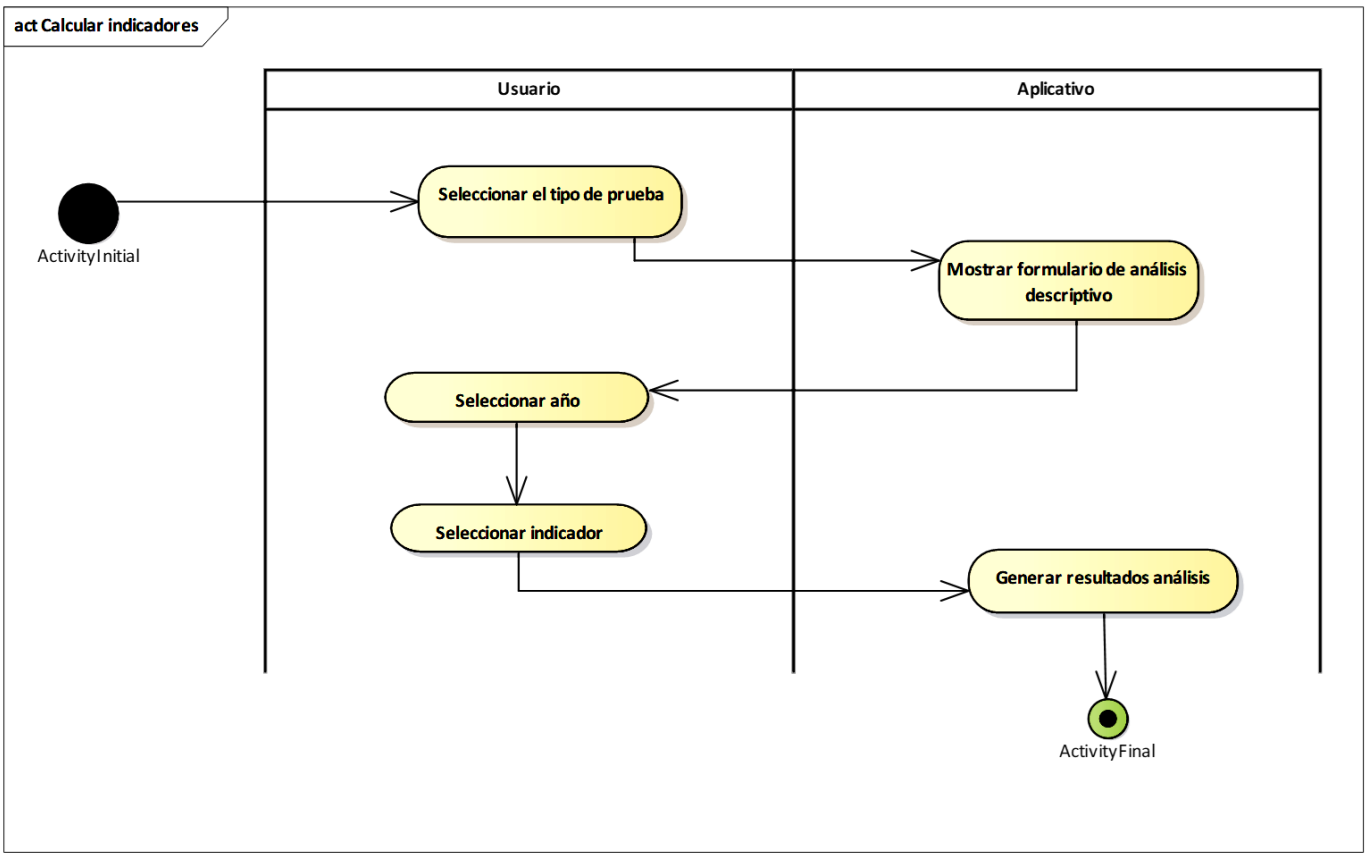


Figura 59. Diagrama de actividad Calcular indicadores de brecha de género
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 26
Descripción actividad Calcular indicadores de brecha de género

NOMBRE ACTIVIDAD	Calcular indicadores
DESCRIPCIÓN	En esta actividad el aplicativo se encarga de hacer el cálculo de los indicadores de forma general o según los filtros de año y/o tipo de indicador.

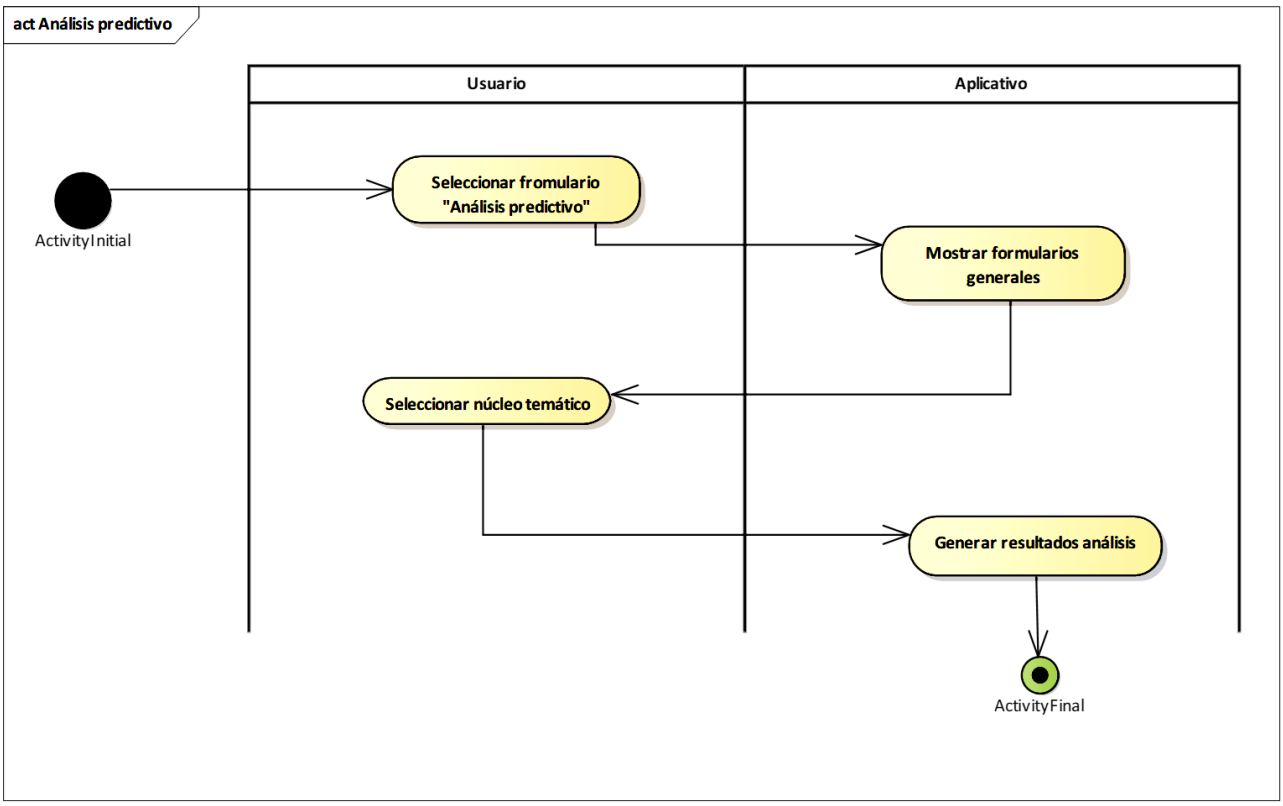


Figura 60. Diagrama de actividad Realizar análisis predictivo
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 27
Descripción actividad Realizar análisis predictivo

NOMBRE ACTIVIDAD	Realizar análisis predictivo
DESCRIPCIÓN	En esta actividad el aplicativo se encarga de realizar y mostrar los resultados del análisis predictivo en las pruebas de forma general o según los núcleos temáticos.

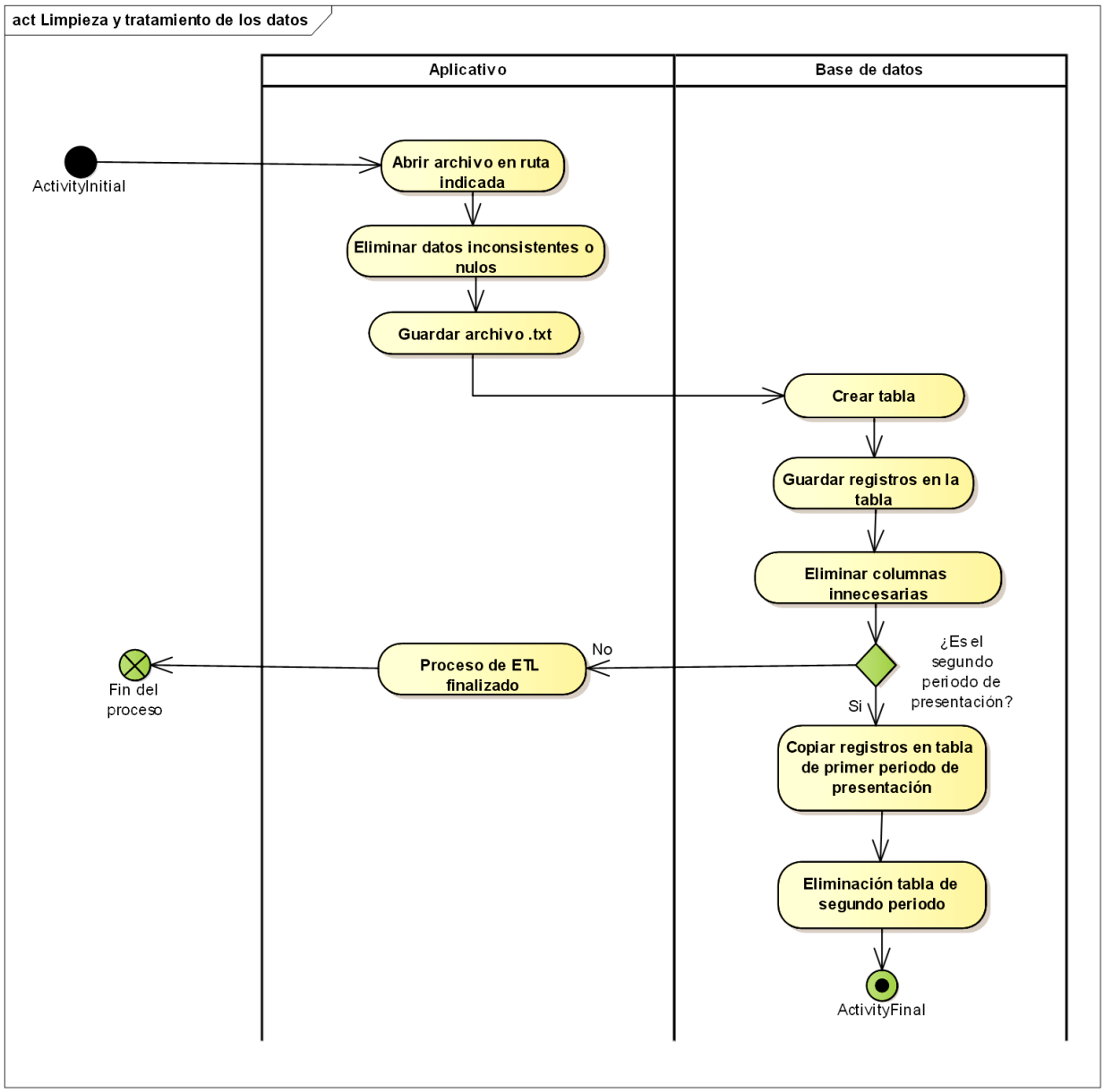


Figura 61. Diagrama de actividad Limpieza y tratamiento
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 28
Descripción actividad Limpieza y tratamiento

NOMBRE ACTIVIDAD	Limpieza y tratamiento
DESCRIPCIÓN	En esta actividad el aplicativo se encarga de limpiar el archivo escogido por el administrador para su carga, de forma que al tratarlo solo contenga la información útil para el análisis.

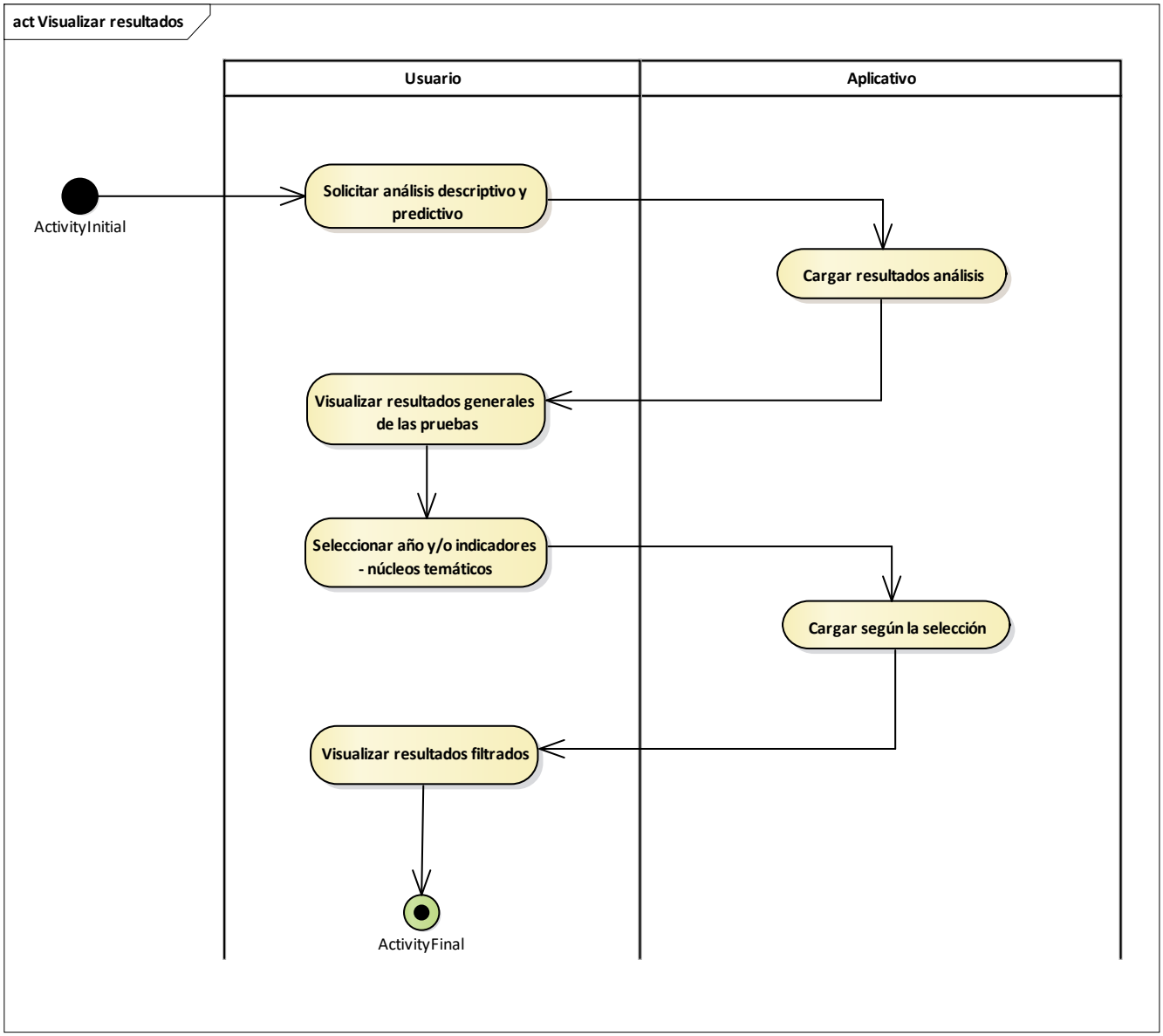


Figura 62. Diagrama de actividad Visualizar resultados
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 29
Descripción actividad Visualizar resultados

NOMBRE ACTIVIDAD	Visualizar resultados
DESCRIPCIÓN	En esta actividad se visualizan los resultados obtenidos de los análisis tanto de forma general como según su selección.

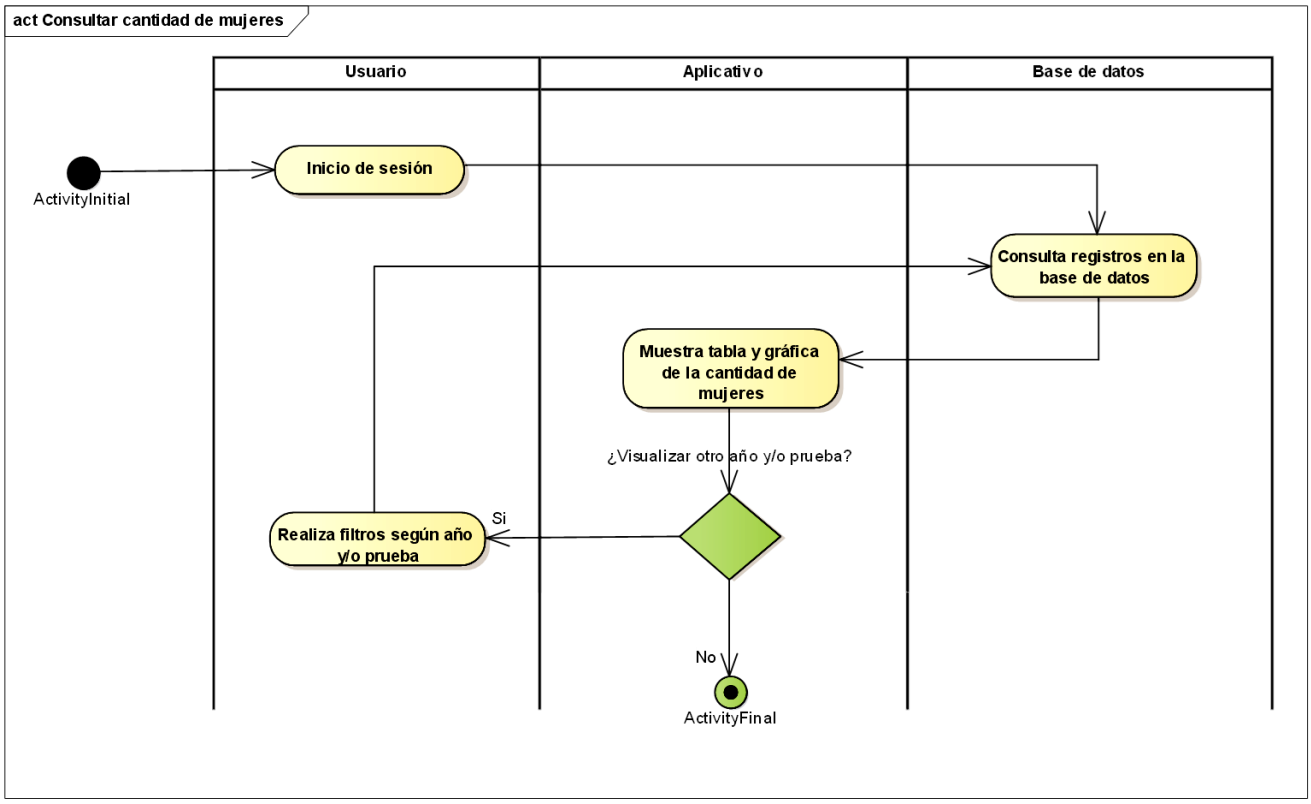


Figura 63. Diagrama de actividad Consultar cantidad de mujeres
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 30
Descripción actividad Consultar cantidad de mujeres

NOMBRE ACTIVIDAD	Consultar cantidad de mujeres
DESCRIPCIÓN	En esta actividad se hace una petición de visualización por parte del usuario para mostrar la cantidad de mujeres que realizaron las diferentes pruebas cada año.

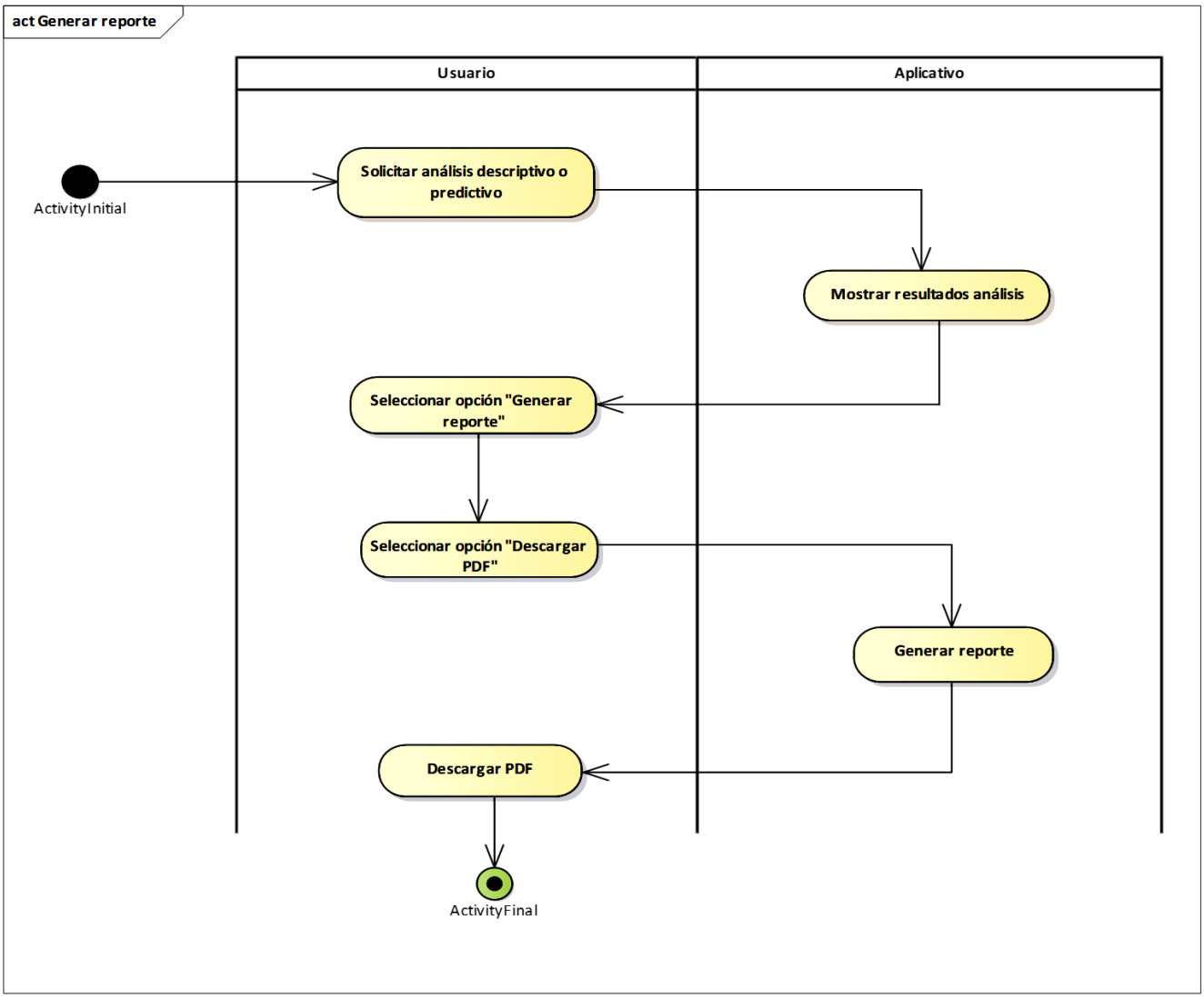


Figura 64. Diagrama de actividad Generar reporte
Autores: (Manrique, Celeita, Barahona, 2020)

Tabla 31
Descripción actividad Generar reporte

NOMBRE ACTIVIDAD	Generar reporte
DESCRIPCIÓN	En esta actividad el usuario puede generar reportes de los resultados de las pruebas y descargar un PDF de los mismos.

3.3.5. DIAGRAMA DE CLASES

3.3.5.1. DIAGRAMA DE CLASES USUARIO

Este diagrama representa la funcionalidad del CRUD de usuarios, además del contacto del usuario con el administrador del aplicativo.

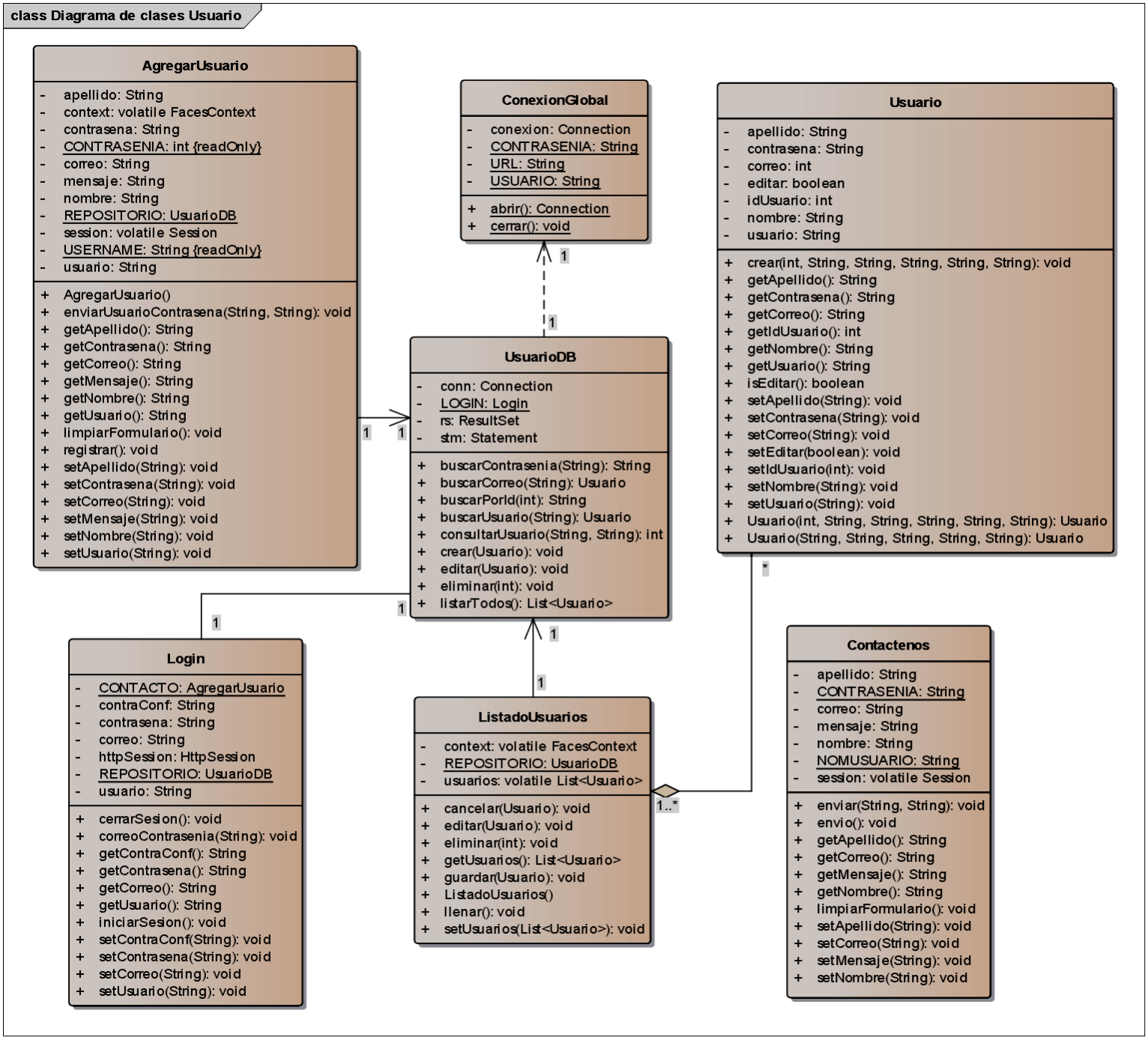


Figura 65. Diagrama de Clases Componente Usuario

Autores: (Manrique, Celeita, Barahona, 2020)

3.3.5.2. Diagrama de clases Archivos.

En este diagrama se administra el proceso de carga y transformación de los archivos .txt, los cuales contienen la información de las pruebas Saber.

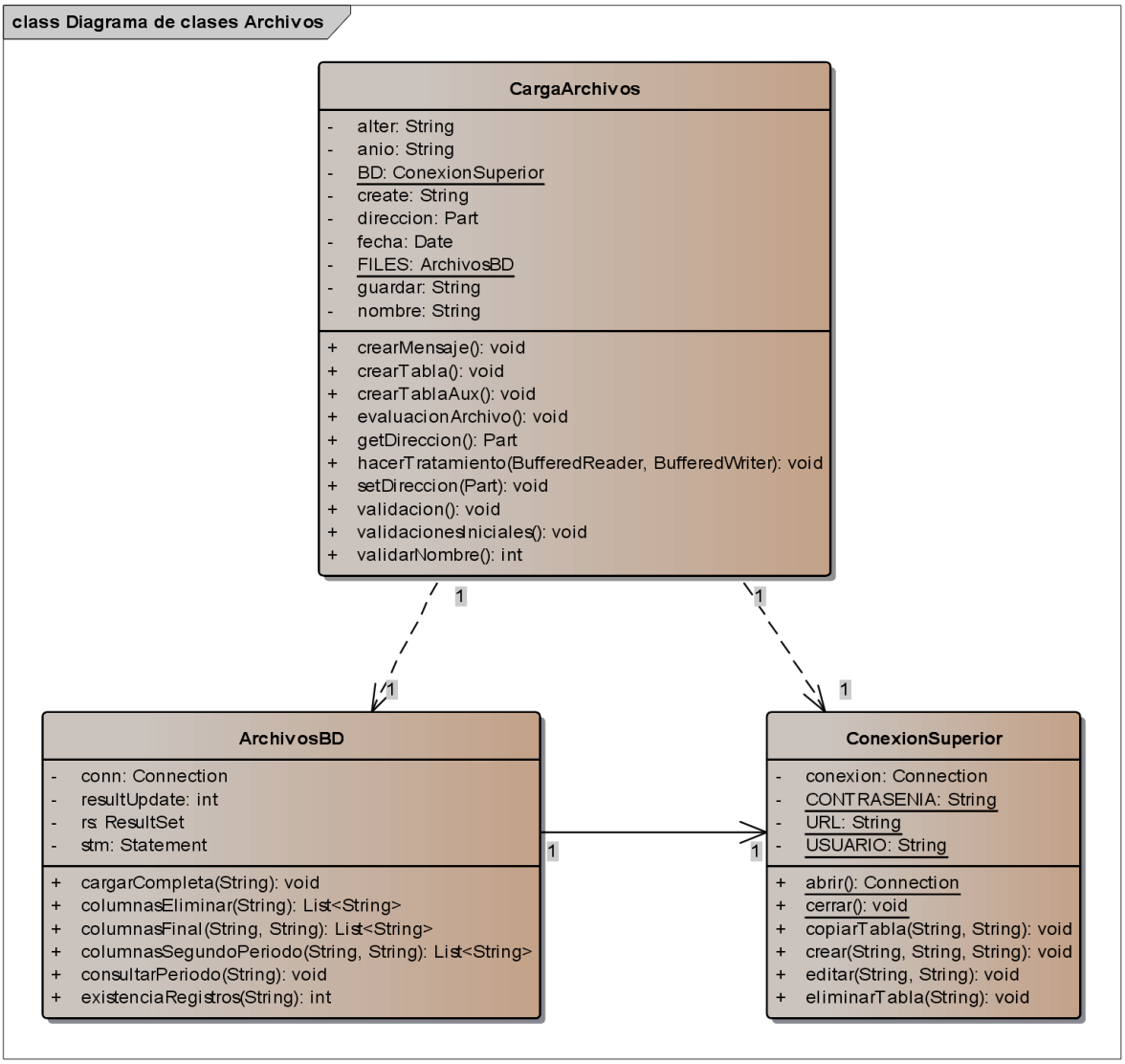


Figura 66. Diagrama de Clases Componente Archivos
Autores: (Manrique, Celeita, Barahona, 2020)

3.3.5.3. Diagrama de clases Análisis Descriptivo.

El siguiente diagrama representa el proceso de análisis descriptivo con su respectiva visualización por medio de graficas generadas a partir de los indicadores.

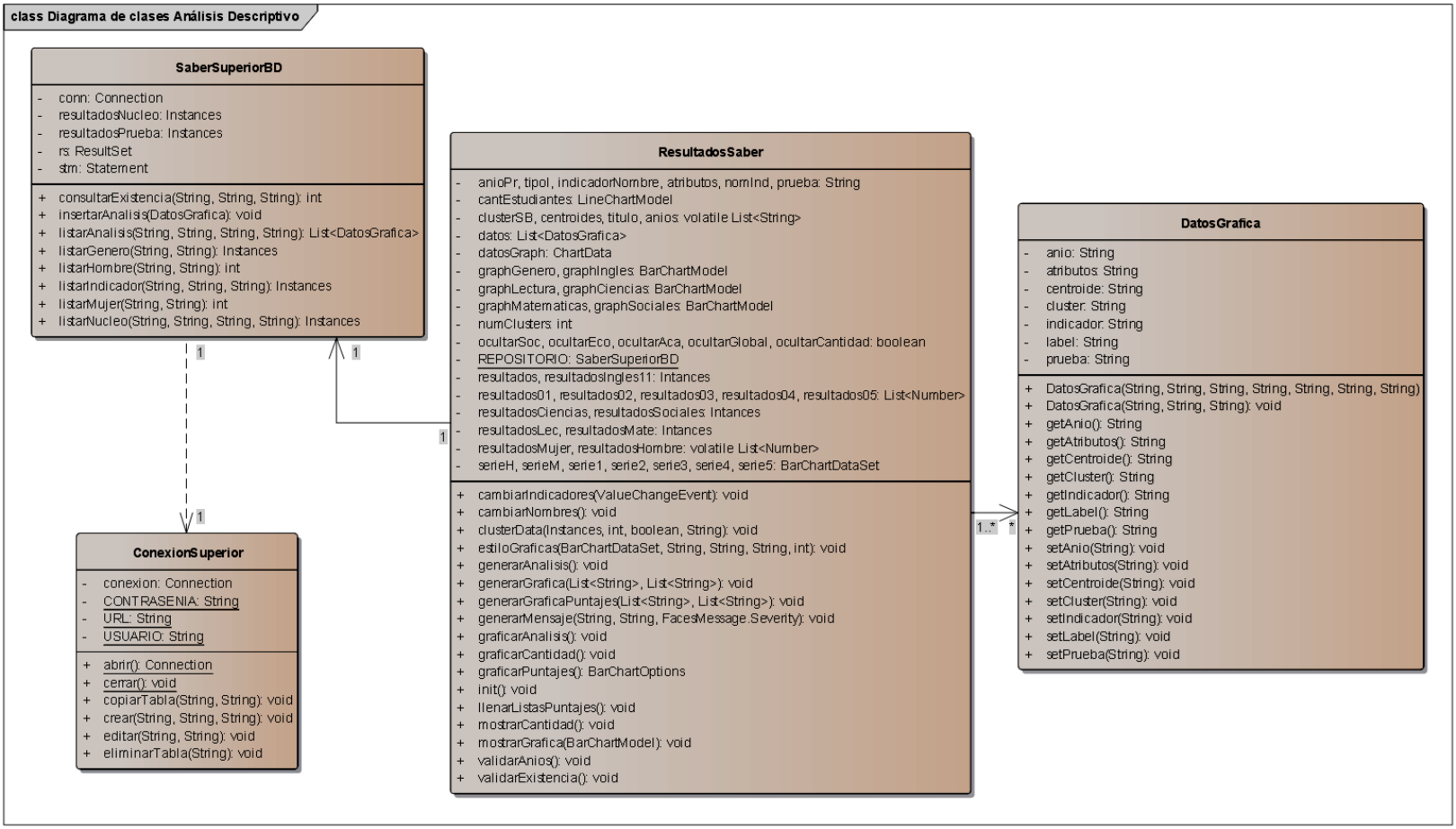


Figura 67. Diagrama de Clases Análisis Descriptivo

Autores: (Manrique, Celeita, Barahona, 2020)

3.3.5.4. Diagrama de clases Análisis Predictivo.

Este diagrama muestra las clases que administran cada una de las fases del análisis predictivo, así como su aplicación.

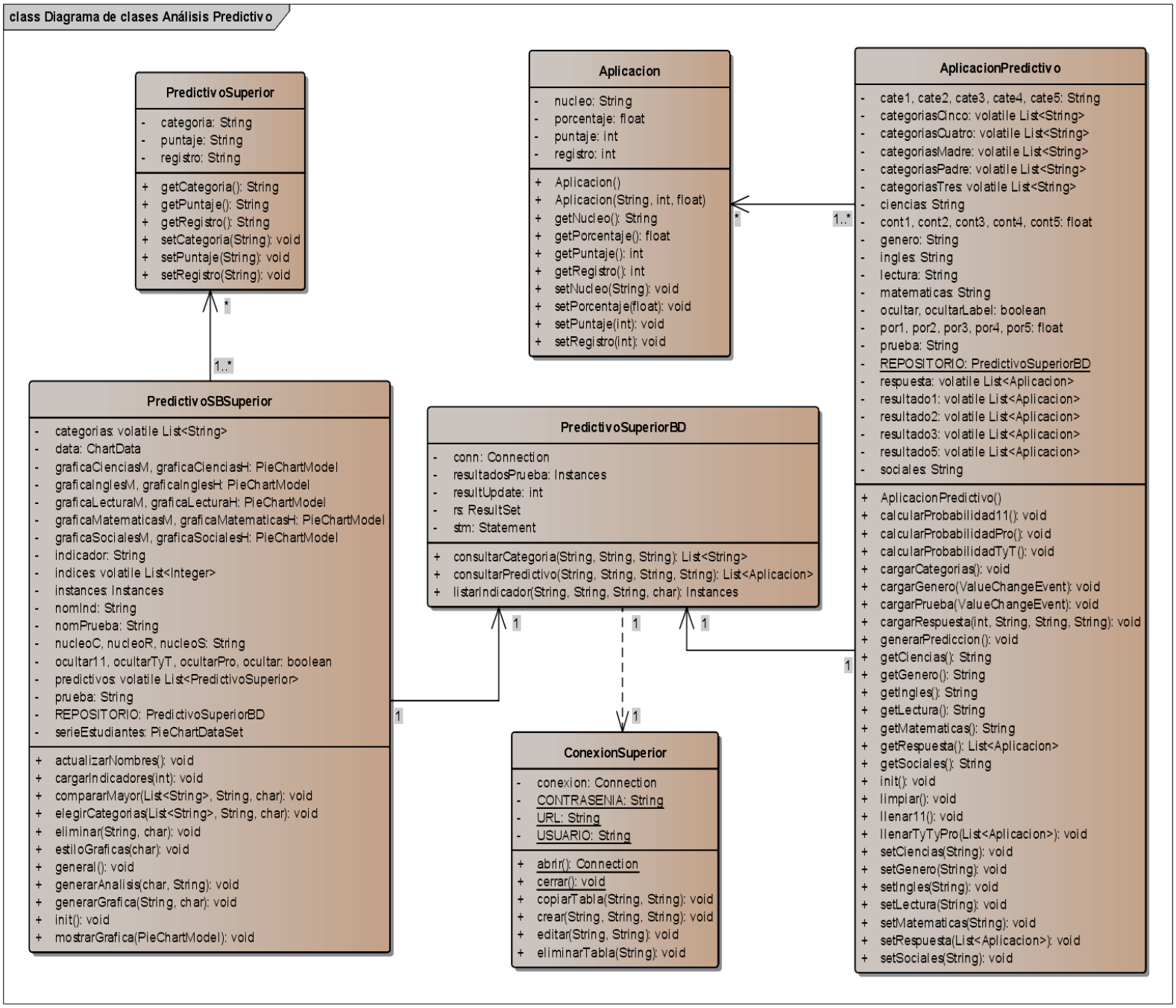


Figura 68. Diagrama de Clases Análisis Predictivo

Autores: (Manrique, Celeita, Barahona, 2020)

3.3.5.5. Diagrama de clases componentes inicio.

El siguiente diagrama ilustra las clases que manejan las paginas principales de cada uno de los módulos del aplicativo.

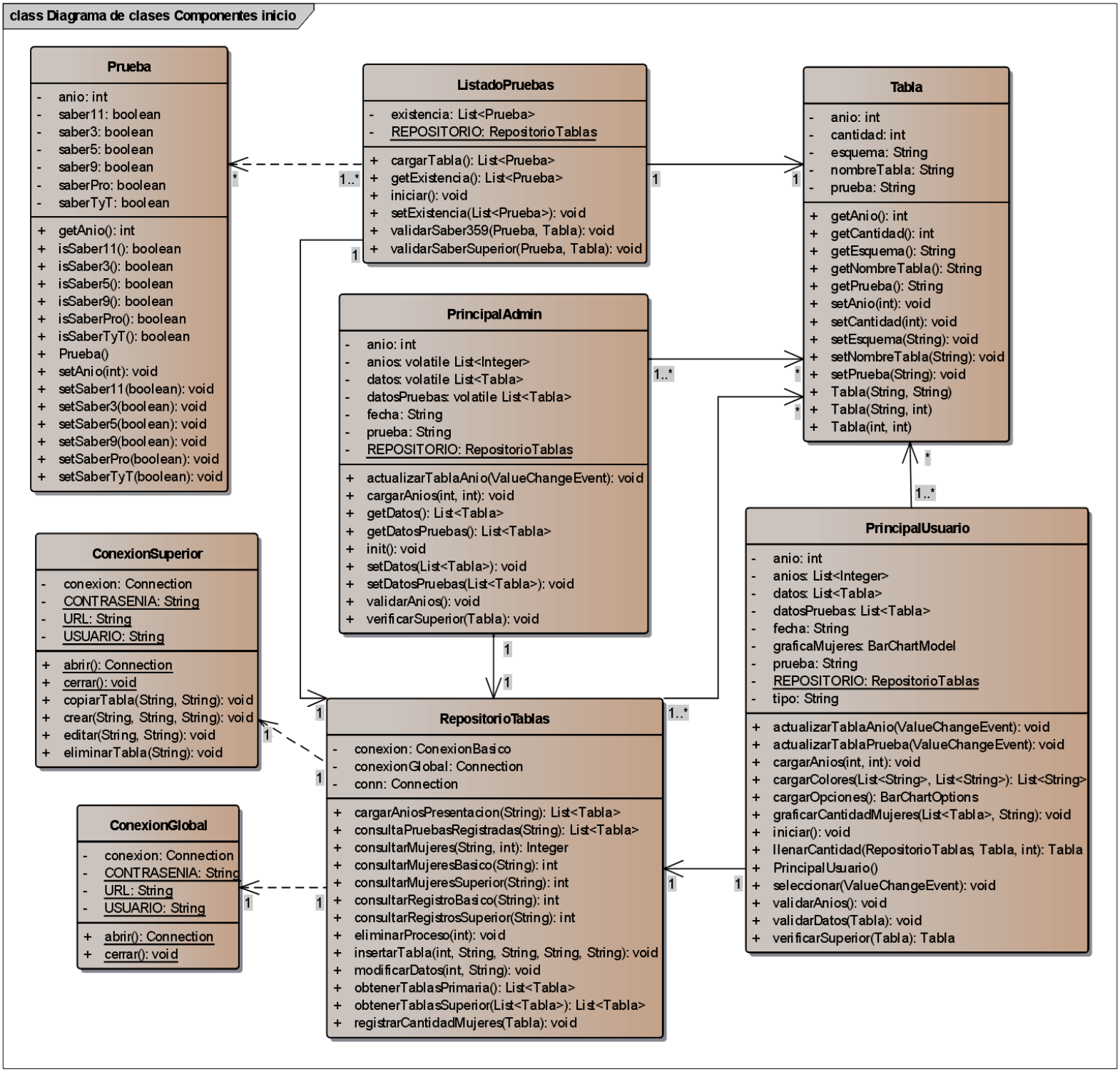


Figura 69. Diagrama de Clases Componentes inicio

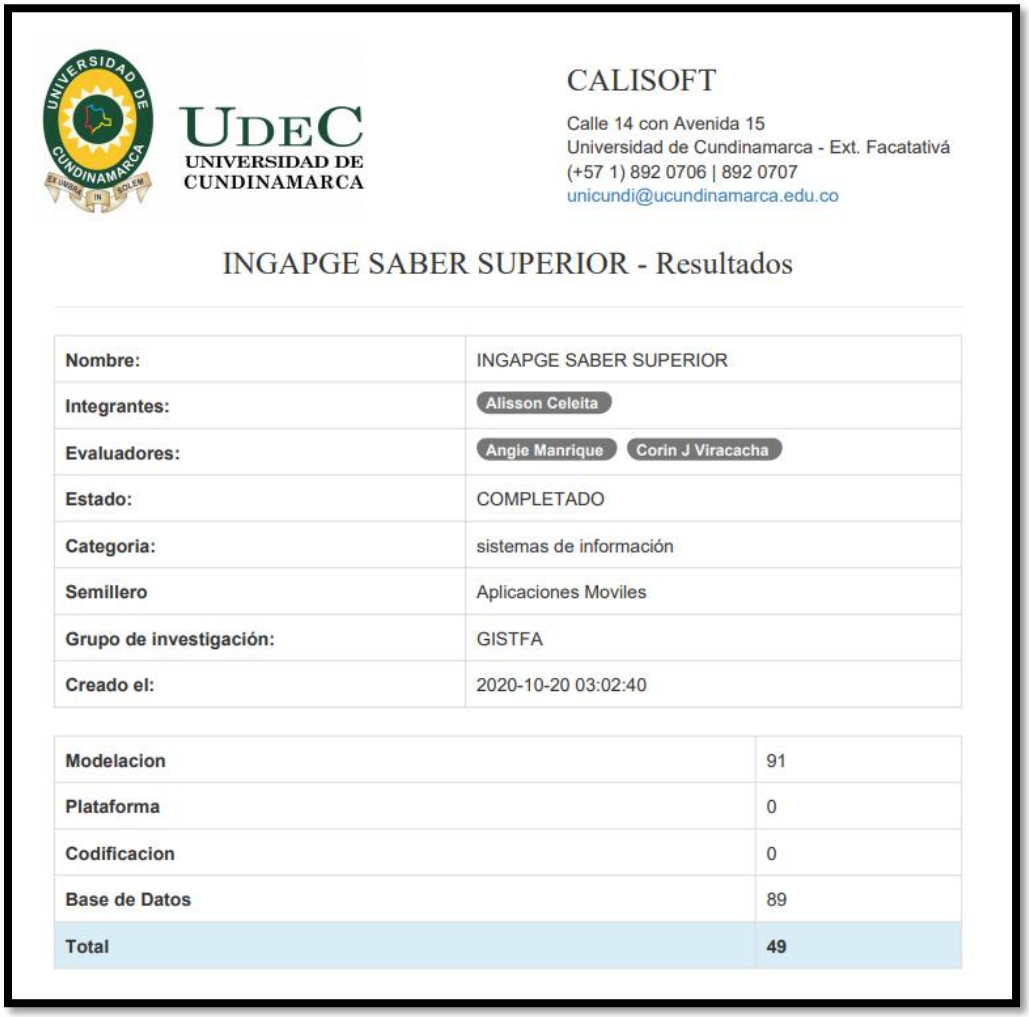
Autores: (Manrique, Celeita, Barahona, 2020)

3.4. DISEÑO DE LOS CASOS DE PRUEBA

En primer lugar, para la evaluación del modelado y estructura de la base de datos se realizó la valoración con el software de calidad desarrollado por la Universidad de Cundinamarca, conocido como **CALISOFT**. Mientras que, para el análisis del código del software se hizo uso de **SonarQube**, que es una plataforma de código libre para evaluar código fuente. A continuación, se presentan los resultados obtenidos:

- **Casos de prueba CALISOFT**

Con la plataforma CALISOFT la evaluación realizada corresponde a dos áreas: Modelación y Base de datos; puesto que, el lenguaje de desarrollo usado para este proyecto es JAVA, el cual no es calificable por este software. Es por esta razón que se ve en los ítems Plataforma y Codificación un puntaje de 0 y se ve afectado el puntaje total.



UDEC
UNIVERSIDAD DE CUNDINAMARCA

CALISOFT

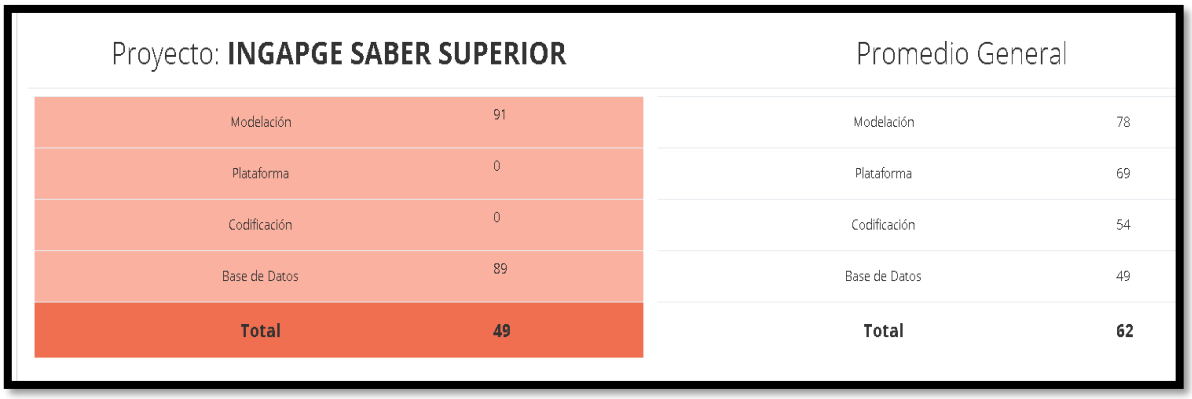
Calle 14 con Avenida 15
Universidad de Cundinamarca - Ext. Facatativá
(+57 1) 892 0706 | 892 0707
unicundi@ucundinamarca.edu.co

INGAPGE SABER SUPERIOR - Resultados

Nombre:	INGAPGE SABER SUPERIOR
Integrantes:	Alisson Celeita
Evaluadores:	Angie Manrique, Corin J Viracacha
Estado:	COMPLETADO
Categoria:	sistemas de información
Semillero	Aplicaciones Moviles
Grupo de investigación:	GISTFA
Creado el:	2020-10-20 03:02:40

Modelacion	91
Plataforma	0
Codificacion	0
Base de Datos	89
Total	**49**

Figura 70. Resultados generales CALISOFT
Autores: (Manrique, Celeita, Barahona, 2020)



Proyecto: **INGAPGE SABER SUPERIOR**

Modelación	91
Plataforma	0
Codificación	0
Base de Datos	89
Total	**49**

Promedio General

Modelación	78
Plataforma	69
Codificación	54
Base de Datos	49
Total	**62**

Figura 71. Comparativa resultados promedio
Autores: (Manrique, Celeita, Barahona, 2020)

Para el Modelado, se evaluó de forma separada cada uno de los diagramas realizados, obteniendo los siguientes resultados porcentuales.



EVALUACION DE MODELADO

Total: 91 %

DIAGRAMA_CLASES.pdf (Diagrama de clases)	80%
DIAGRAMAS_CASOS_USO.pdf (Diagrama de casos de uso)	100%
DIAGRAMAS_SECUENCIA.pdf (Diagrama de secuencias)	91%
DIAGRAMAS_ACTIVIDADES.pdf (Diagrama de actividades)	86%
MODELO_ENTIDAD_RELACION.pdf (Modelo Entidad Relación)	100%

Figura 72. Resultados evaluación Modelado
Autores: (Manrique, Celeita, Barahona, 2020)

Para el diagrama de clases se obtuvo un puntaje de 80%, el 20% faltante es debido a que no hay evidencia de generalizaciones y trazado en el diagrama.



Figura 73. Resultados pruebas Diagrama de clases
Autores: (Manrique, Celeita, Barahona, 2020)

En el diagrama de casos de uso se encontraron todas las caracteristicas necesarias para obtener una puntuación del 100%.



Figura 74. Resultados pruebas Diagrama de casos de uso
Autores: (Manrique, Celeita, Barahona, 2020)

Para el caso del diagrama de secuencia no se visualizaron mensajes con restricciones de tiempo, por ello, la calificacion es del 91%.

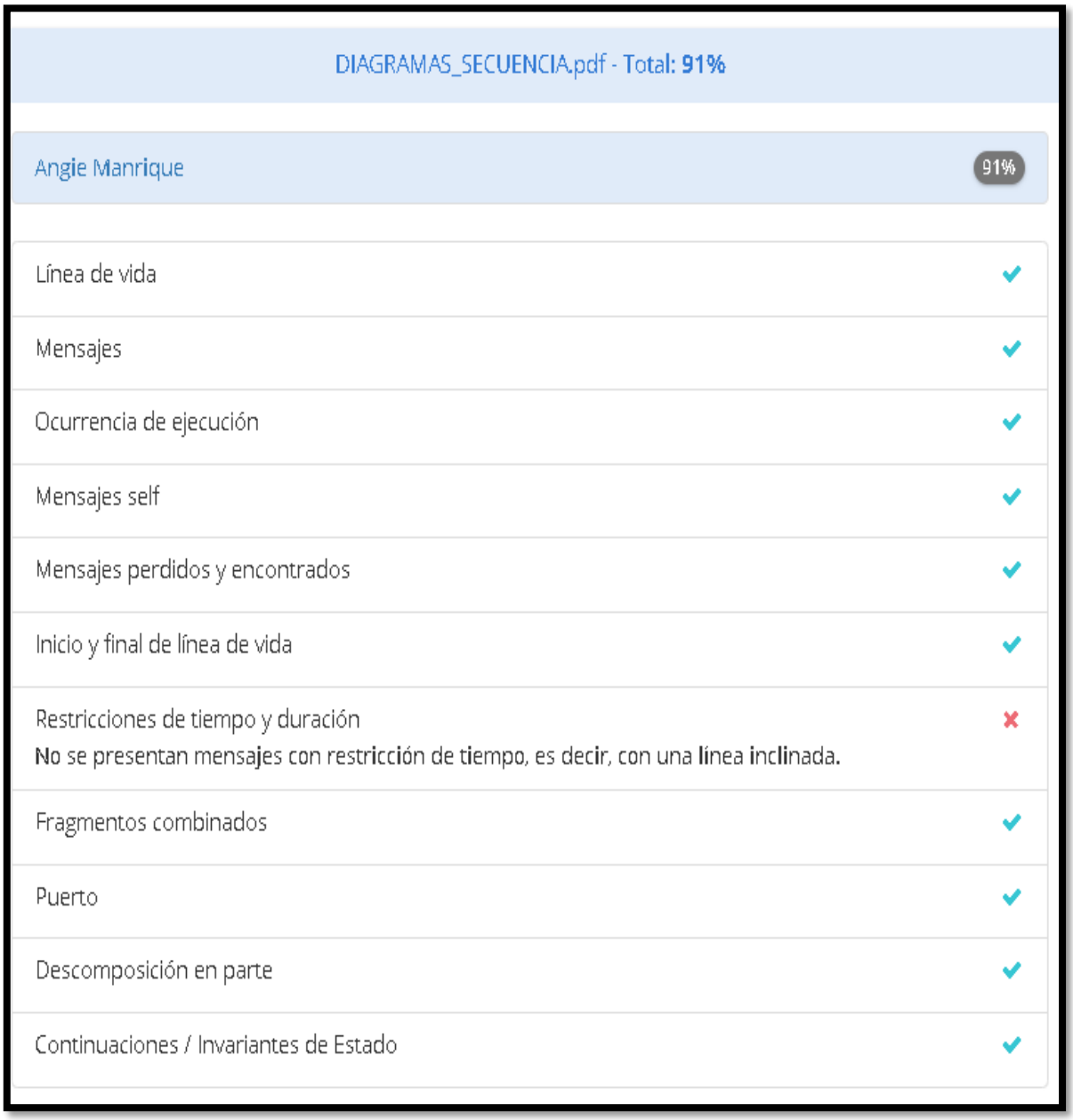


Figura 75. Resultados pruebas Diagramas de secuencia
Autores: (Manrique, Celeita, Barahona, 2020)

Los diagramas de actividades obtuvieron una calificacion de 86%, esto se debe a la ausencia del nodo final de flujo y la region de actividad interrumplible.

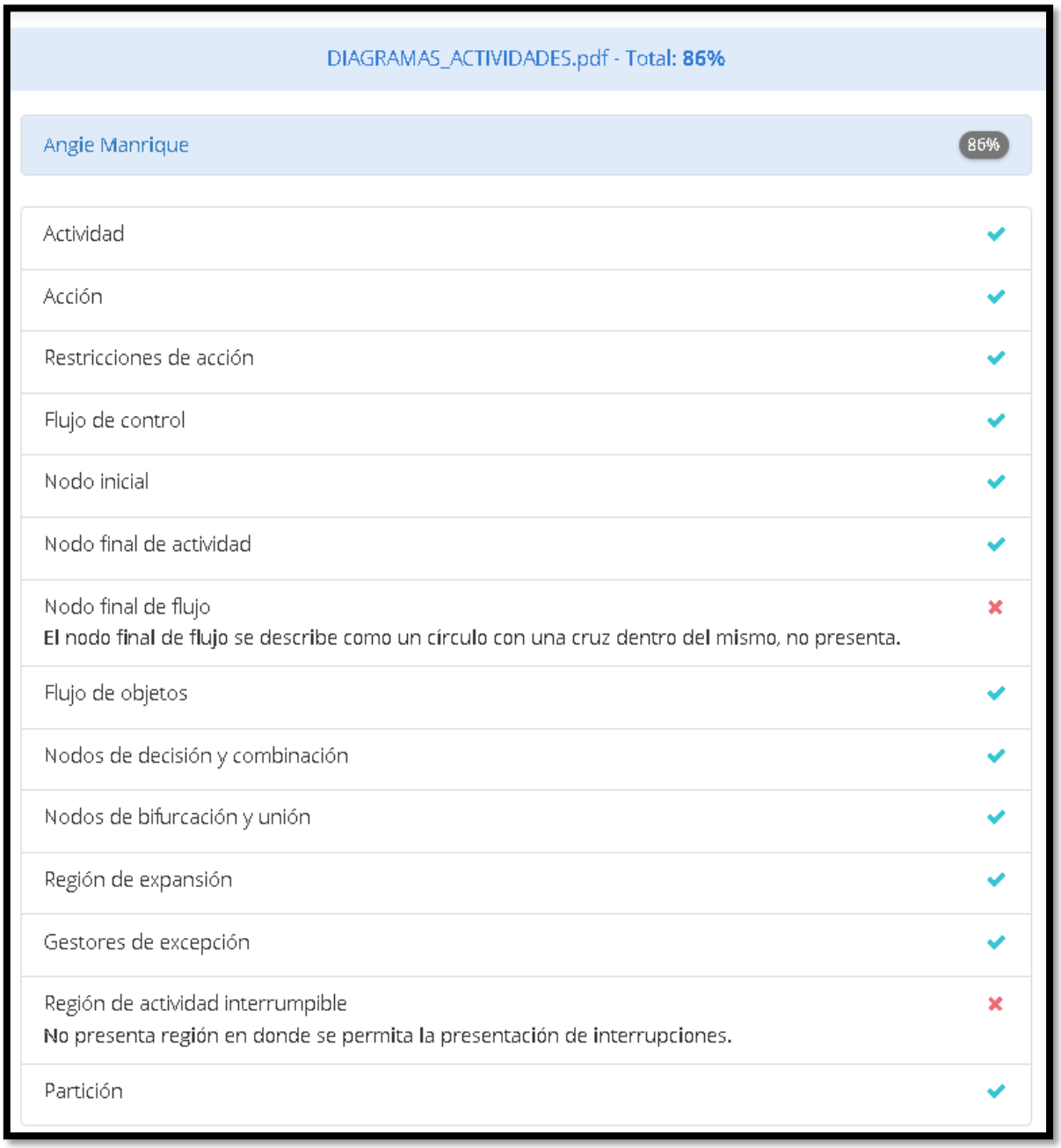


Figura 76. Resultados pruebas Diagramas de actividades
Autores: (Manrique, Celeita, Barahona, 2020)

En el caso del diagrama MER (Modelo Entidad Relación) se hallaron todas las caracteristicas necesarias para que su calificacion sea del 100%

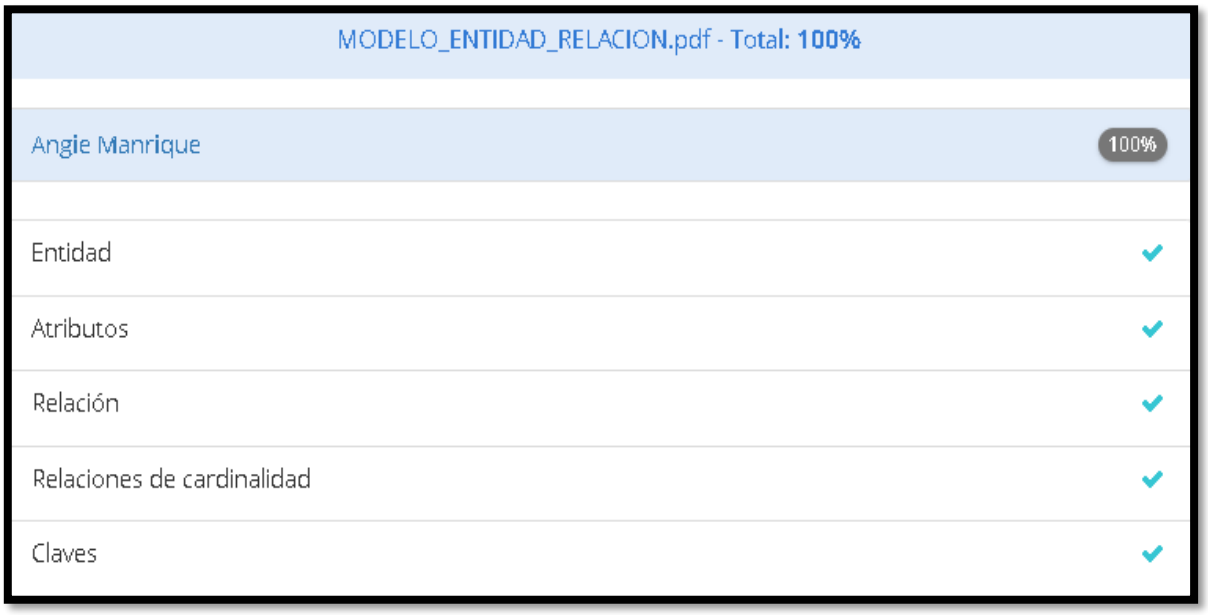


Figura 77. Resultados pruebas MER
Autores: (Manrique, Celeita, Barahona, 2020)

En la valoración realizada a la base de datos del proyecto, se obtuvo una puntuacion de 4.47, siendo 1 la calificación mas baja y 5 la más alta.

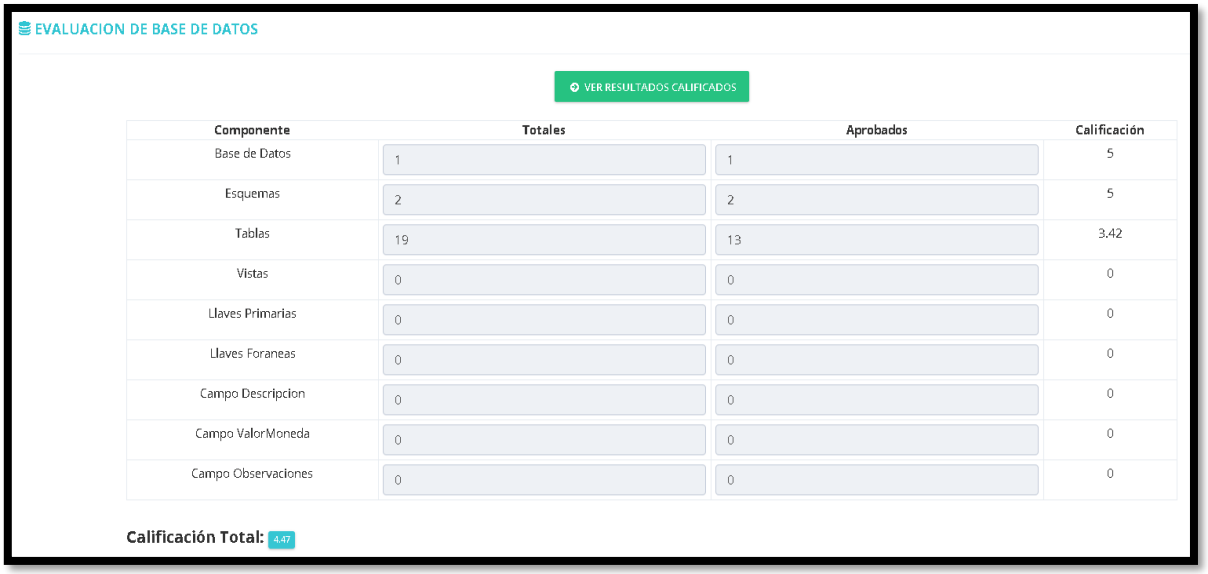


Figura 78. Resultados pruebas Base de datos
Autores: (Manrique, Celeita, Barahona, 2020)

Para evaluar el código del desarrollo del aplicativo, el cual fue desarrollador en el lenguaje Java, se hizo uso del software SonarQube con el fin de encontrar errores de sintaxis, código duplicado y errores en la lógica.

Obteniendo los siguientes resultados:

- Mantenibilidad: Esta métrica se clasifica en 5 opciones, siendo A la mejor calificación y E la más baja. Este proyecto, fue clasificado con una A, dicho nivel nos indica que el software presenta errores que pueden ser resueltos en corto tiempo y no superan el 5% del aplicativo.

- Complejidad: Esta métrica nos indica que tan comprensible es el seguimiento al flujo del control de las diferentes funciones existentes en el aplicativo. Para este caso se obtuvo como puntaje 1,339 en complejidad, señalando que su dificultad es baja.

- Duplicidad: Este parámetro señala la cantidad de líneas, bloques y/o archivos que se encuentran repetidos en el aplicativo. En este proyecto se obtuvo una calificación de 15.7% donde se visualiza que la duplicidad es baja.

- Deuda técnica: Esta característica está relacionada con el tiempo necesario para remediar problemas dentro del código. En términos porcentuales, este proyecto tiene una deuda técnica de 3.4.

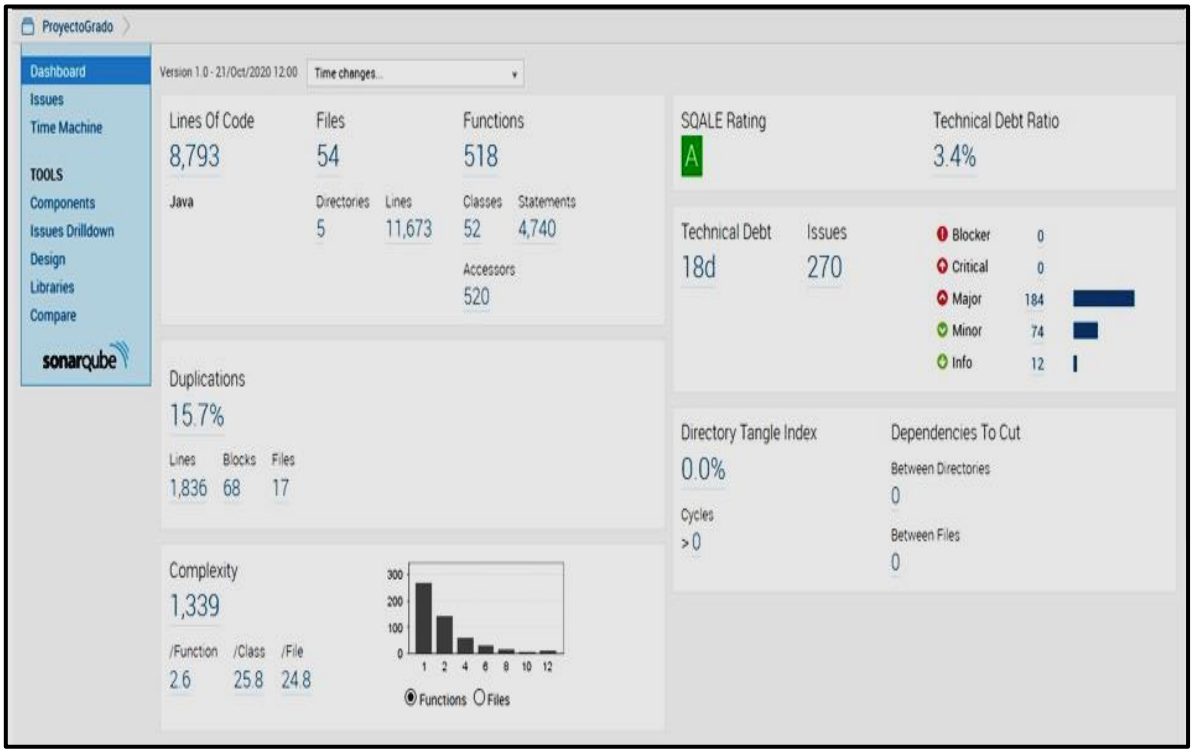


Figura 79. Resultados SonarQube
Autores: (Manrique, Celeita, Barahona, 2020)

3.5. ESTIMACIÓN DE RECURSOS

Tabla 32
Estimación de recursos proyecto

Rubros	Solicitado en efectivo a UDEC	Contrapartida en especie		Total
		UDEC	Otras Entidades	
PERSONAL				$0
EQUIPOS				$0
MATERIALES E INSUMOS				$0
SERVICIOS TECNOLOGICOS				$0
VIAJES				$0
OTROS				$0
TOTALES				$0

IV. RESULTADOS ANÁLISIS DESCRIPTIVO

4. 1. PRUEBAS SABER 11

En las pruebas Saber 11 según los indicadores de brecha de género hallados se realizó la aplicación del algoritmo K-Means, encontrando los resultados mostrados a continuación en las gráficas de barras.

a) Estrato de vivienda

Para el indicador de Estrato de vivienda, en relación con el género de los estudiantes que presentaron la prueba se puede observar que las mujeres son mayoría en cada uno de los niveles, presentando mayor diferencia respecto a los hombres en el estrato 1 y 2. Esto nos indica que la población estudiantil femenina (de la educación media del país) vive en condiciones socioeconómicas bajas.

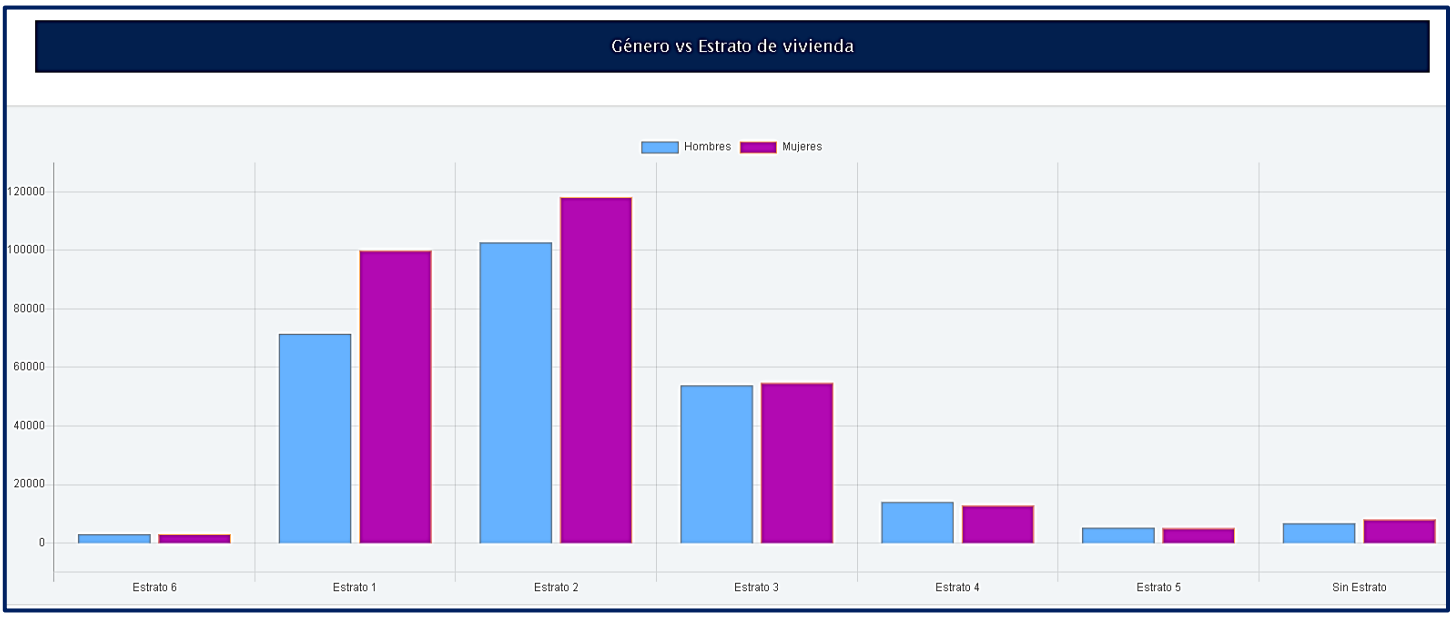


Figura 80. Género vs Estrato de vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Haciendo referencia a los puntajes obtenidos en cada uno de los núcleos temáticos de la prueba Saber 11; en primer lugar, se puede decir que este indicador muestra diferencias significativas en sus categorías, los estudiantes pertenecientes a los estratos 4, 5 y 6 del país son quiénes mejores resultados obtienen en Lectura Crítica, mientras que, su opuesto son los estratos más bajos, es decir los de nivel 1, 2 y 3 obtienen resultados más bajos. Vale la pena destacar que los estudiantes clasificados dentro de la categoria Sin estrato, son quienes obtienen peores resultados

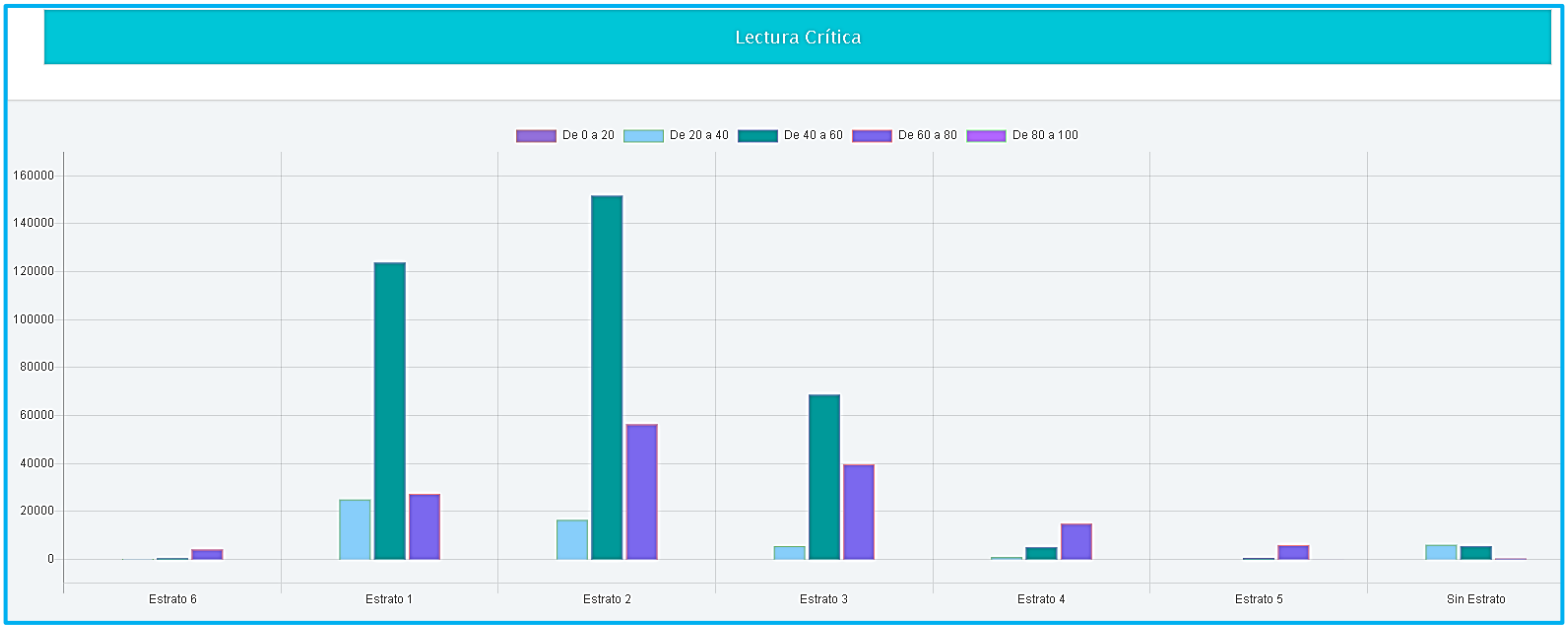


Figura 82. Lectura Critica vs Estrato vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de matemáticas, los resultados de los estudiantes que se encuentran en los estratos 1, 2, 3 y 6 son preocupantes; puesto que, la gran mayoría de ellos se encuentran en un rango de 40 a 60 sobre 100 puntos posibles. Los demás estratos, excluyendo la categoria Sin estrato, presentan puntajes en promedio entre 60 y 80, siendo estos resultados más destacables.

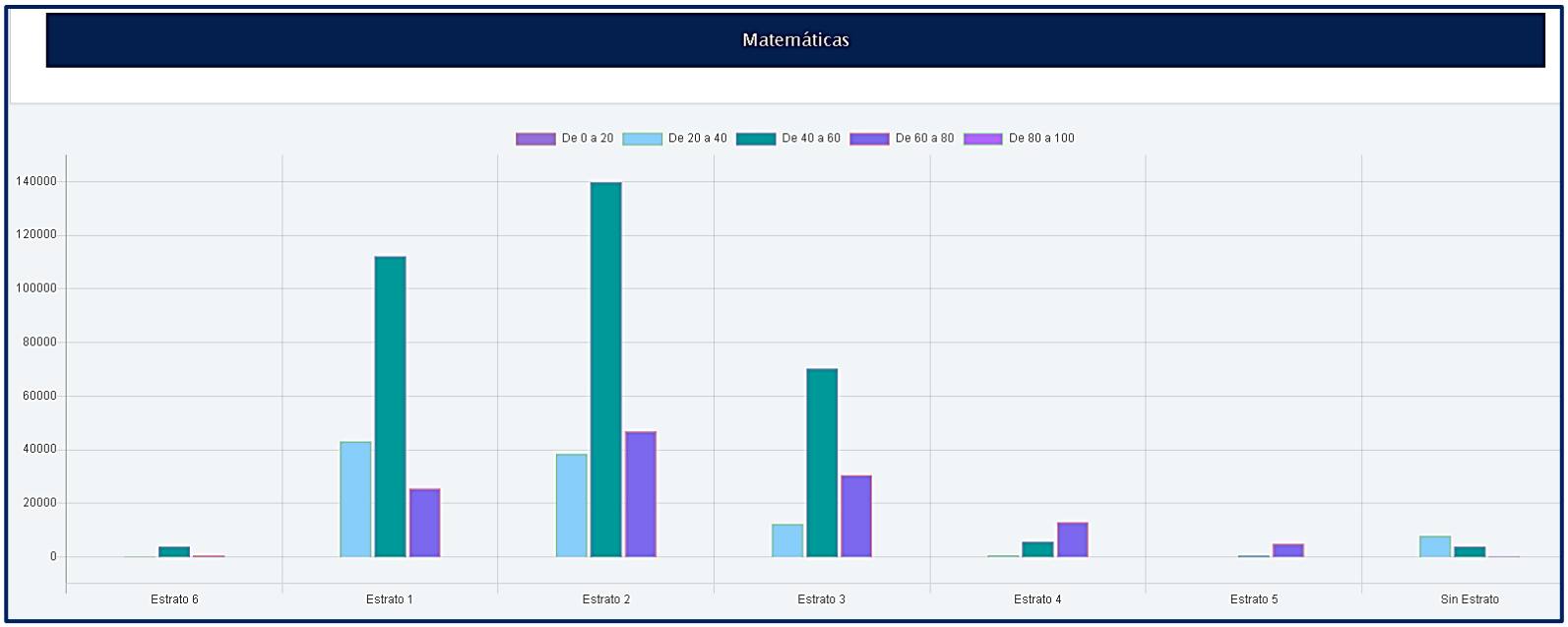


Figura 81. Matemáticas vs Estrato de vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias naturales, el comportamiento de los puntajes se divide en 2 grandes grupos, por un lado, los estratos 1, 2, 3 y 4 quiénes obtienen en promedio resultados dentro del rango de 40 a 60 y, por otro lado, los estratos 5 y 6 obtienen puntajes ubicados entre 60 y 80. Aquí se evidencia afectaciones para los estratos más bajos.

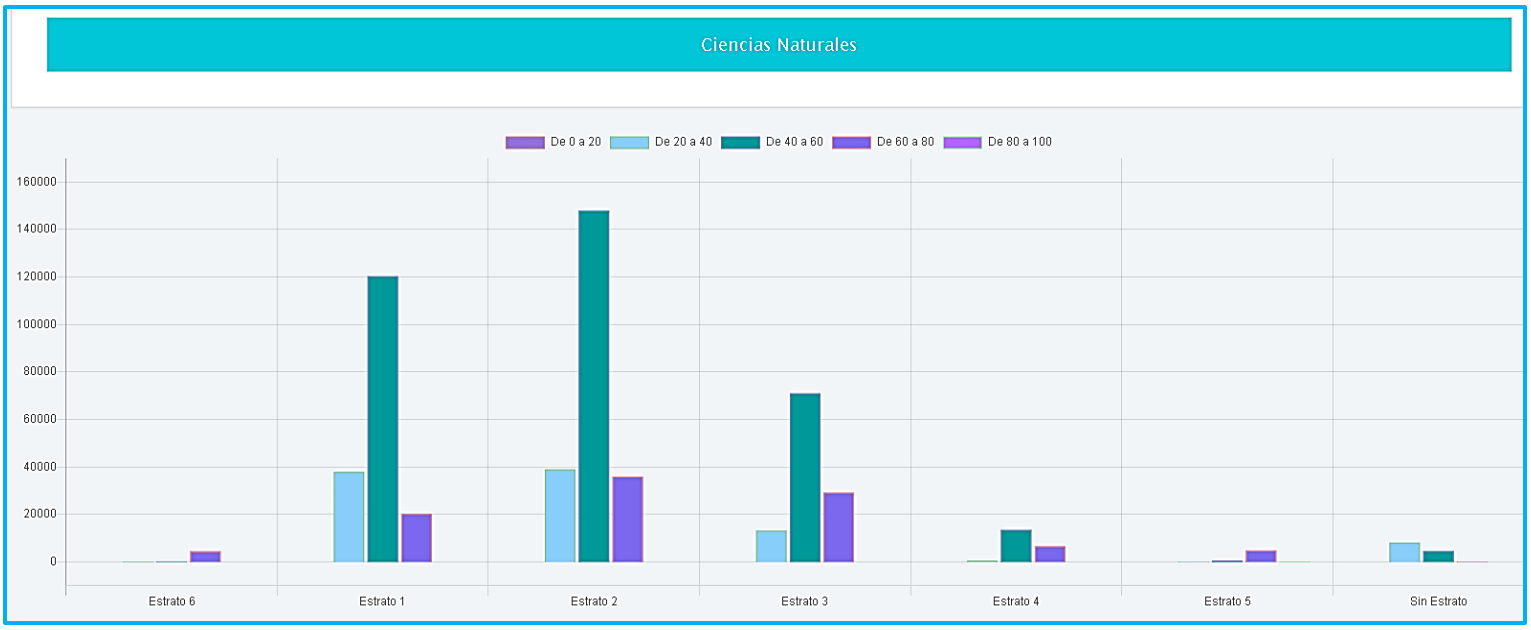


Figura 84. Ciencias Naturales vs Estrato de vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

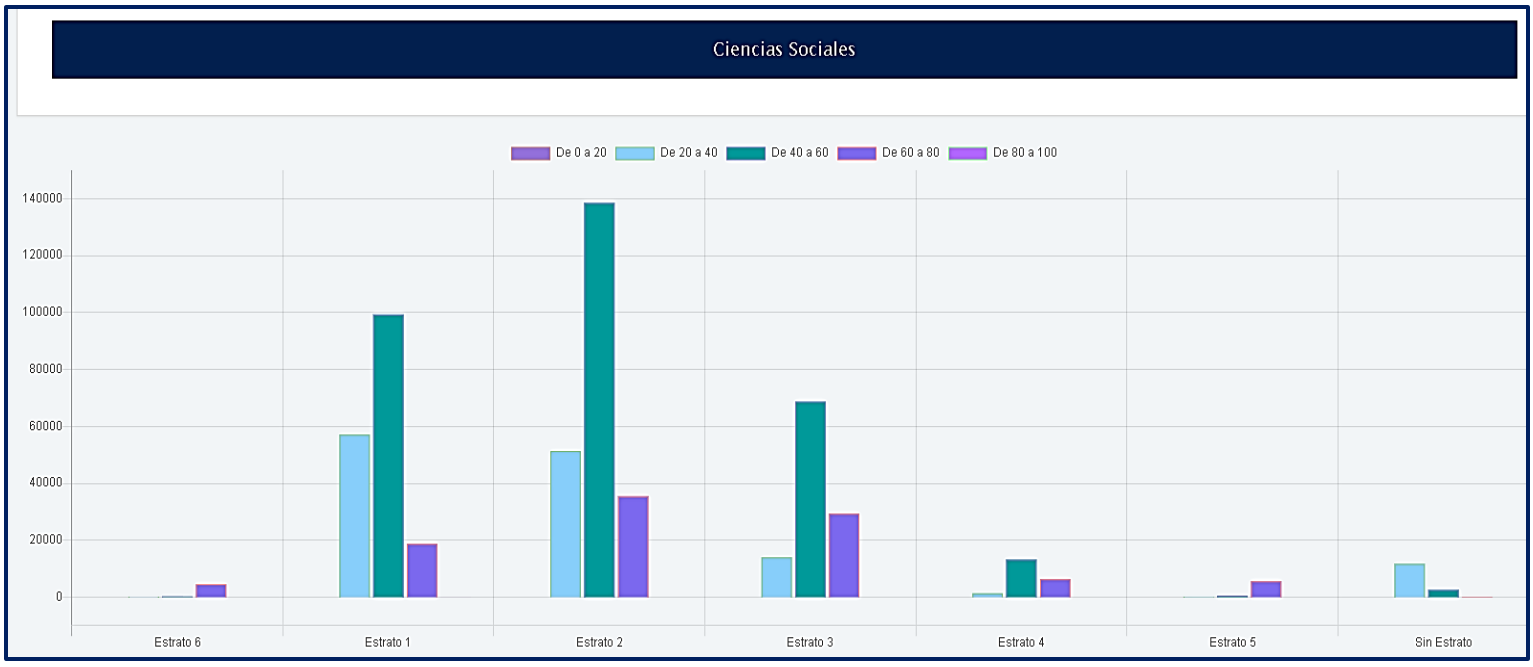


Figura 83. Ciencias Sociales vs Estrato de vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En la gráfica de Ciencias sociales se visualiza un comportamiento semejante al presentado en los resultados de Ciencias naturales, la diferencia se evidencia en el rango entre 20 y 40 donde hay mayor presencia de estudiantes. Los resultados en el núcleo de inglés, son similares a los presentados en el núcleo de Lectura crítica.

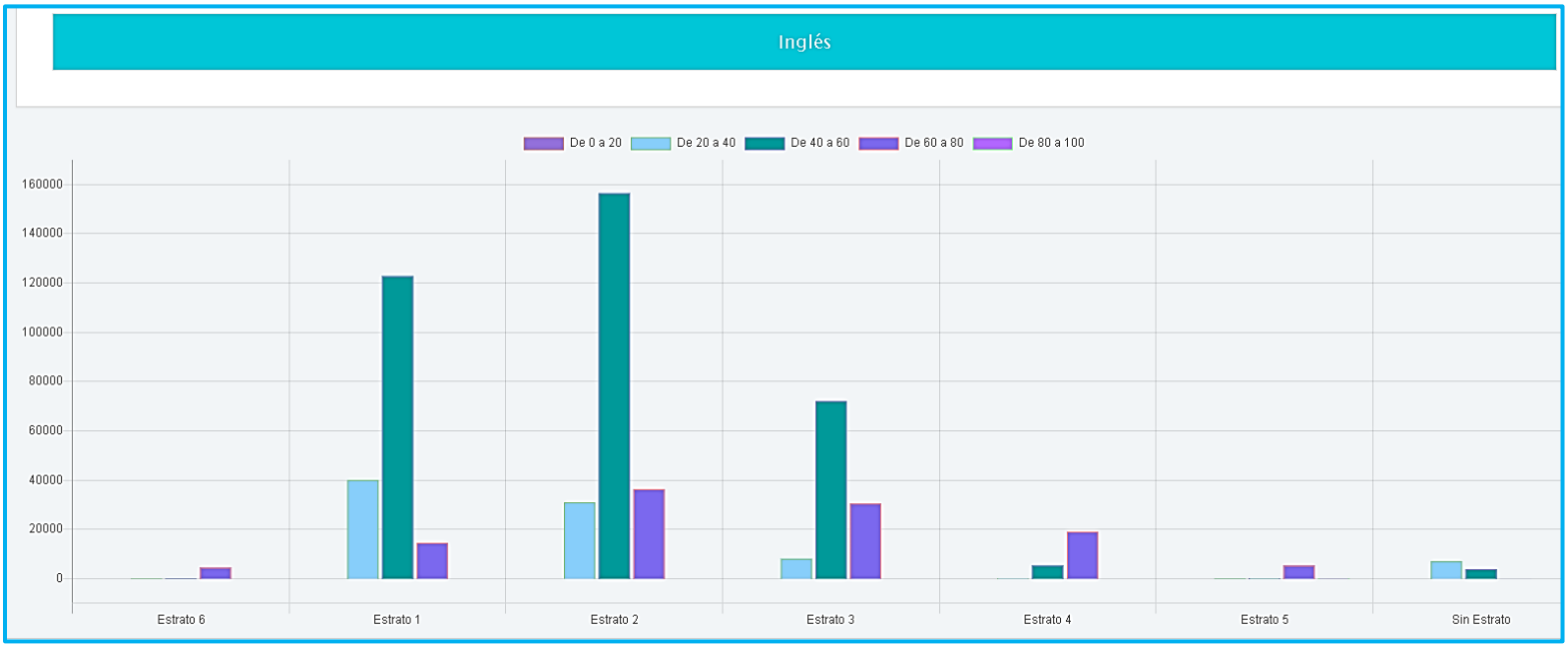


Figura 85. Inglés vs Estrato de vivienda, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En resumen, queda por resaltar para este indicador (Estrato de vivienda) que la gran mayoría de estudiantes en el país pertenece a los estratos 1 y 2, y son estos dos niveles los que peores resultados presentan en cada uno de los núcleos de la prueba Saber 11. Asimismo, las mujeres son la mayoría en cada una de las categorías, excepto en los estratos 4 donde se presenta una cantidad bastante similar entre ambos géneros y en el estrato 5, ligeramente sobrepasan en cantidad el género masculino. Las valoraciones más preocupantes en todos los núcleos se presentan en la categoria Sin estrato.

b) Servicio de TV

En este indicador, al relacionarlo con el género de los estudiantes se puede visualizar que, en las dos opciones de respuesta, las mujeres superan por una cantidad no muy significativa al género masculino. Además, se observa que la mayoría de evaluados si cuenta con el servicio de televisión en sus hogares.

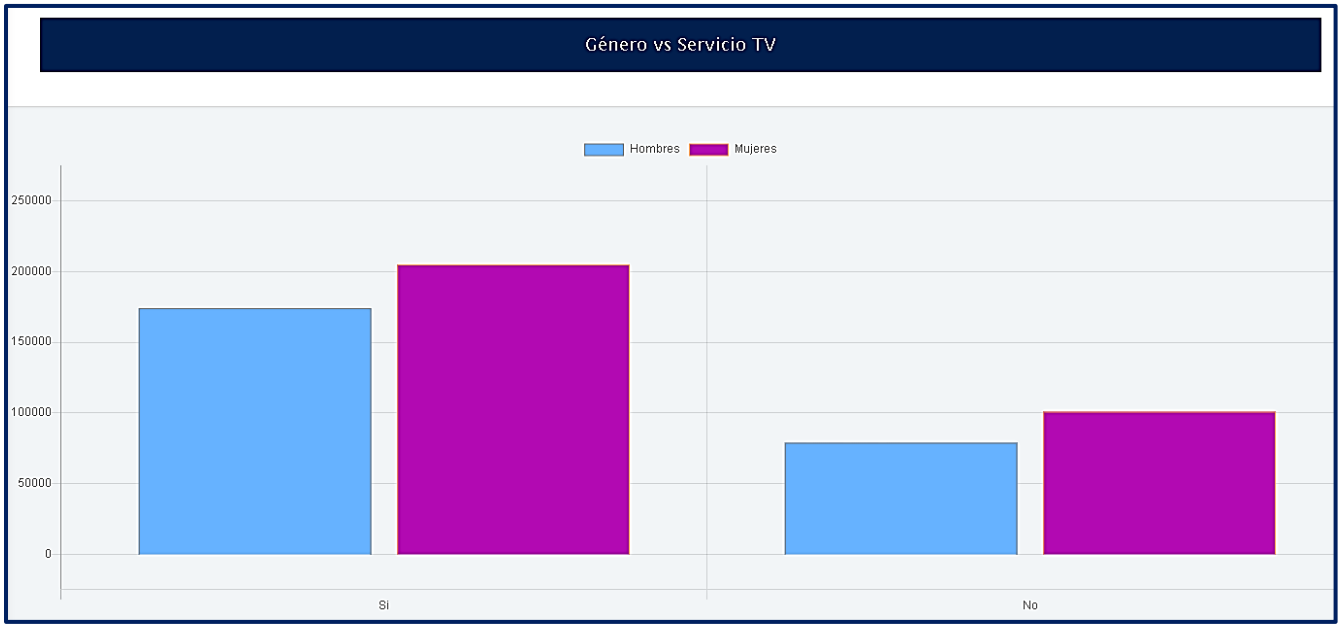


Figura 86. Género vs Servicio TV, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En las gráficas de puntajes de los diferentes núcleos temáticos, los resultados presentan un comportamiento similar, el cual es descrito por lo siguiente: los estudiantes que sí cuentan con el servicio de televisión en su mayoría obtienen puntajes en promedio de 40 a 60, seguidos por los ubicados dentro del rango de 60 a 80 y una presencia minoritaria entre 20 y 40.



Figura 87. Lectura crítica vs Género, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Por otra parte, los estudiantes que no cuentan con servicio de televisión también tienen en su mayoría puntajes dentro del rango de 40 a 60, pero a diferencia del otro grupo estos tienen resultados más bajos, puesto que hay más presencia en el rango de 20 a 40 que el de 60 a 80.

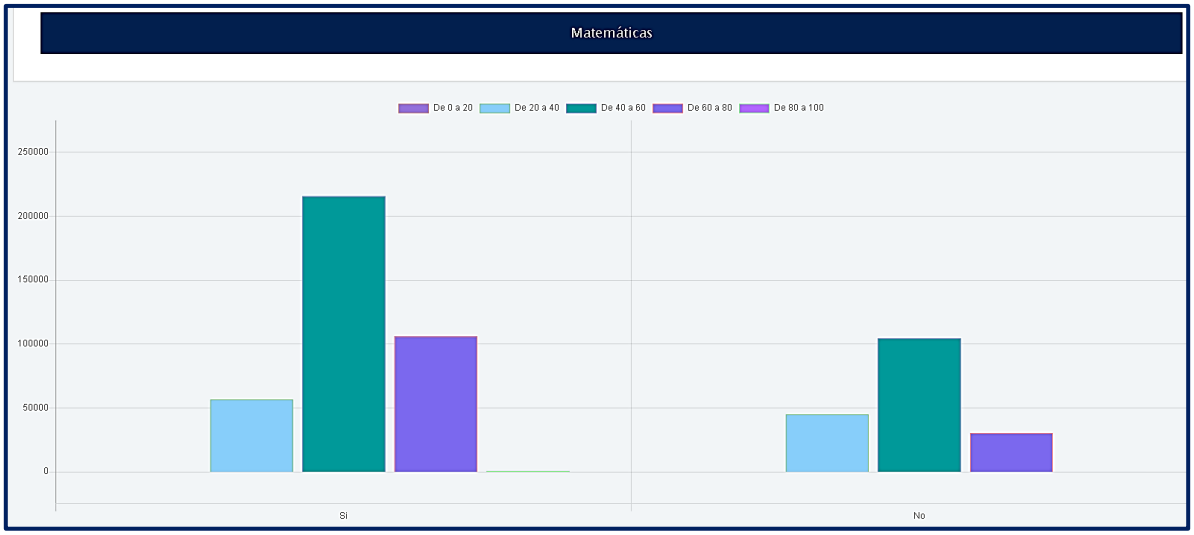


Figura 88. Matemáticas vs Servicio TV, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

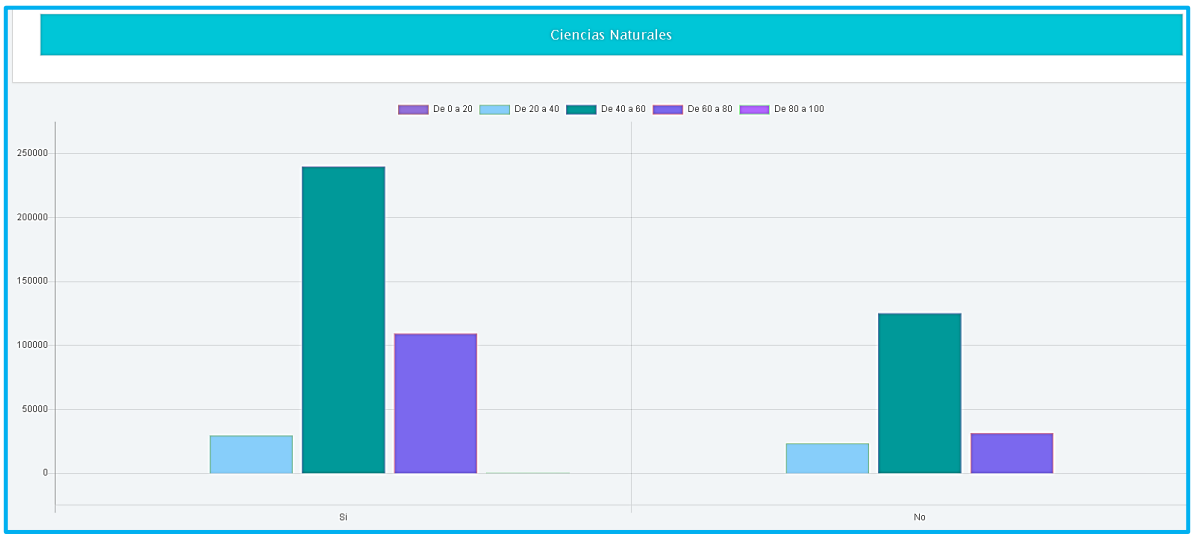


Figura 89. Ciencias naturales vs Servicio TV, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

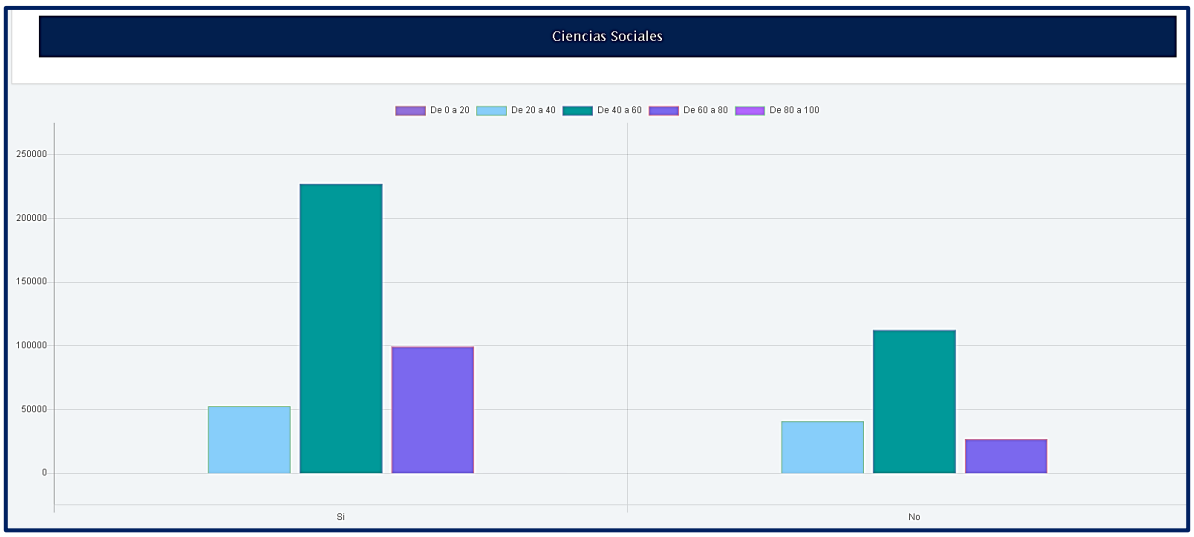


Figura 90. Ciencias sociales vs Servicio TV, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

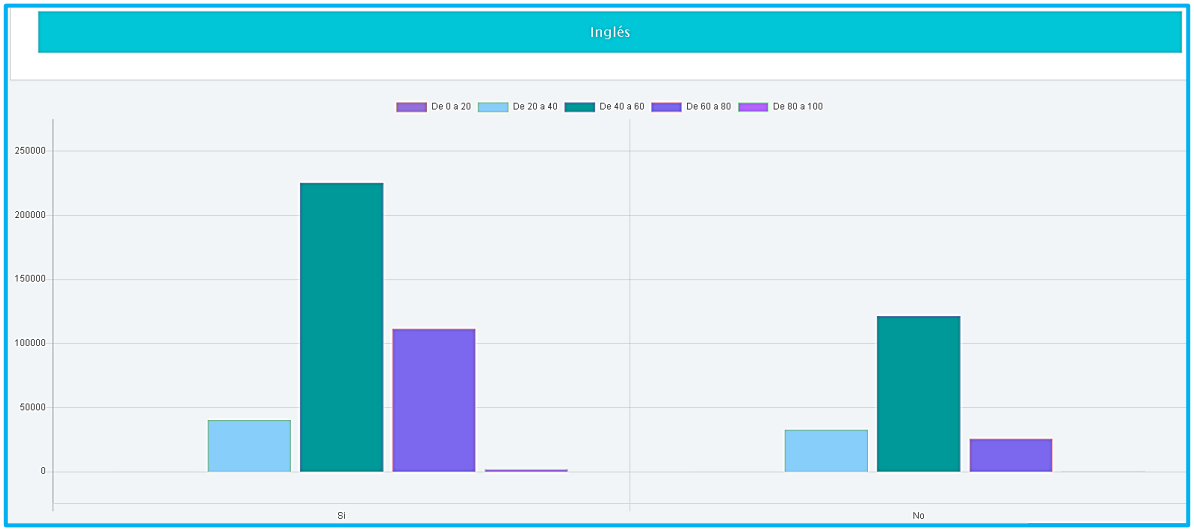


Figura 91. Inglés vs Servicio TV, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

c) Tiene computador

Los resultados con este indicador al relacionarlo con el género de los estudiantes muestran que, en las dos opciones de respuesta, las mujeres superan casi con la misma proporción al género masculino. Además, se observa que la mayoría de evaluados cuenta con computador en sus hogares.

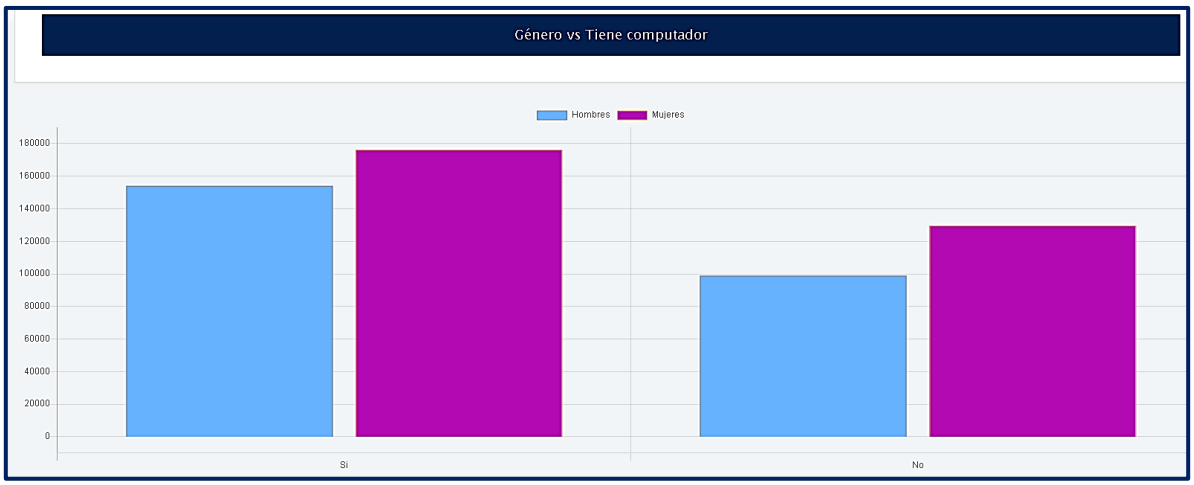


Figura 92. Género vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En las gráficas de puntajes de los diferentes núcleos temáticos, los resultados presentan un comportamiento similar, el cual es descrito por lo siguiente: los estudiantes que sí cuentan con computador en sus casas en su mayoría obtienen puntajes en promedio de 40 a 60, seguidos por los ubicados dentro del rango de 60 a 80 y una presencia minoritaria entre 20 y 40.

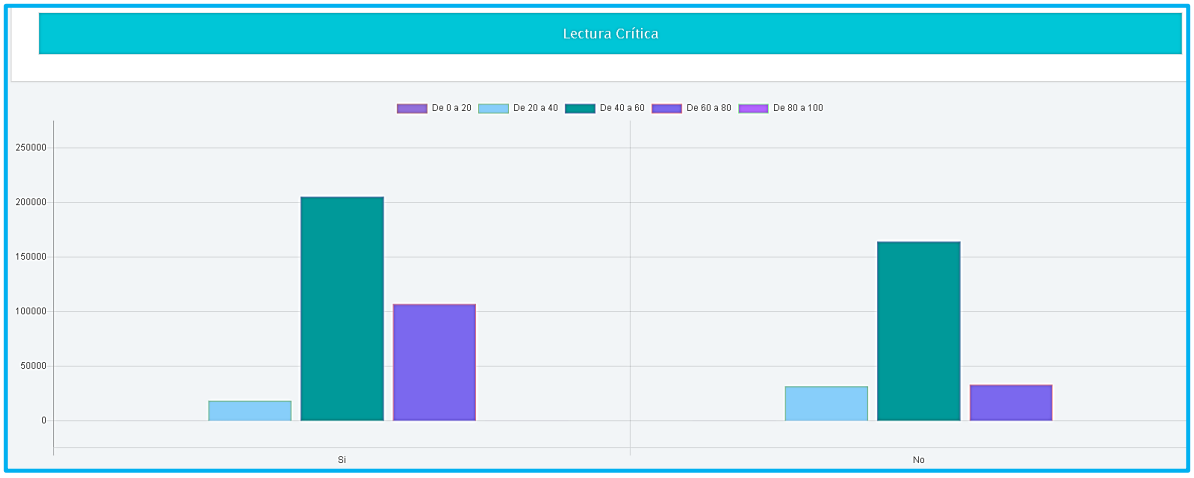


Figura 93. Lectura crítica vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Por otra parte, los estudiantes que no cuentan con computador también tienen en su mayoría puntajes dentro del rango de 40 a 60, pero a diferencia del otro grupo estos tienen resultados más bajos, puesto que hay más presencia en el rango de 20 a 40 que el de 60 a 80.

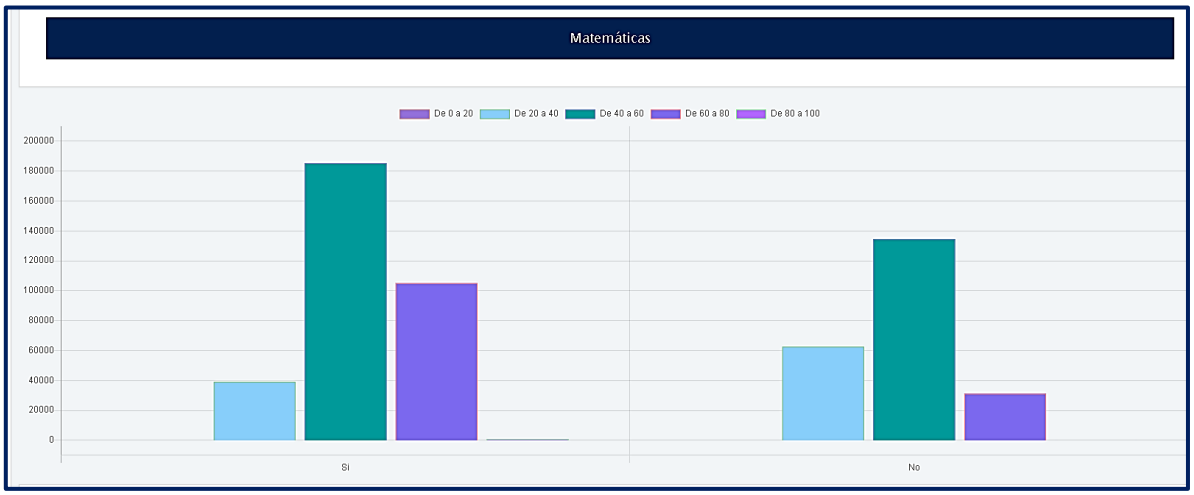


Figura 94. Matemáticas vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

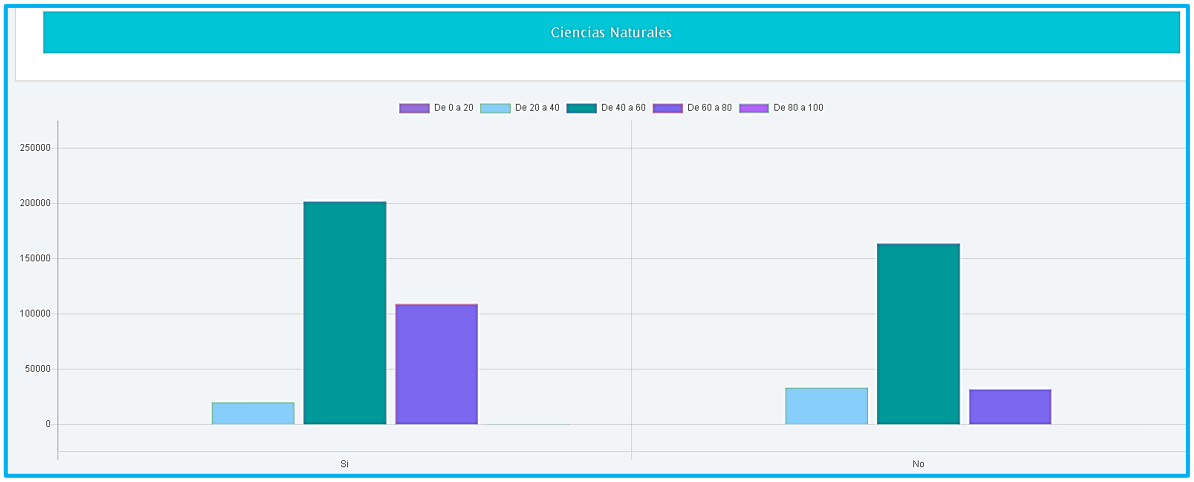


Figura 95. Ciencias naturales vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

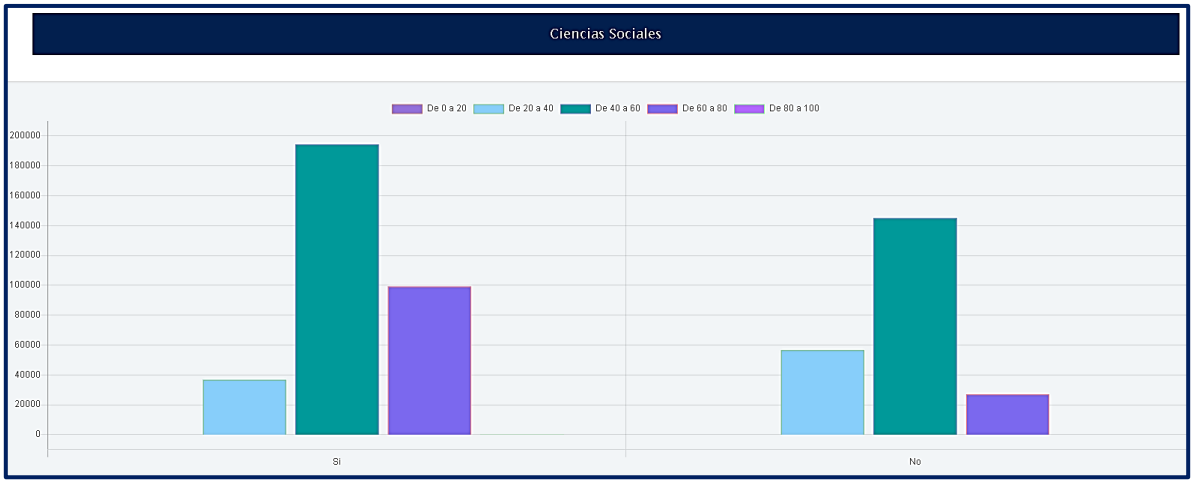


Figura 96. Ciencias sociales vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

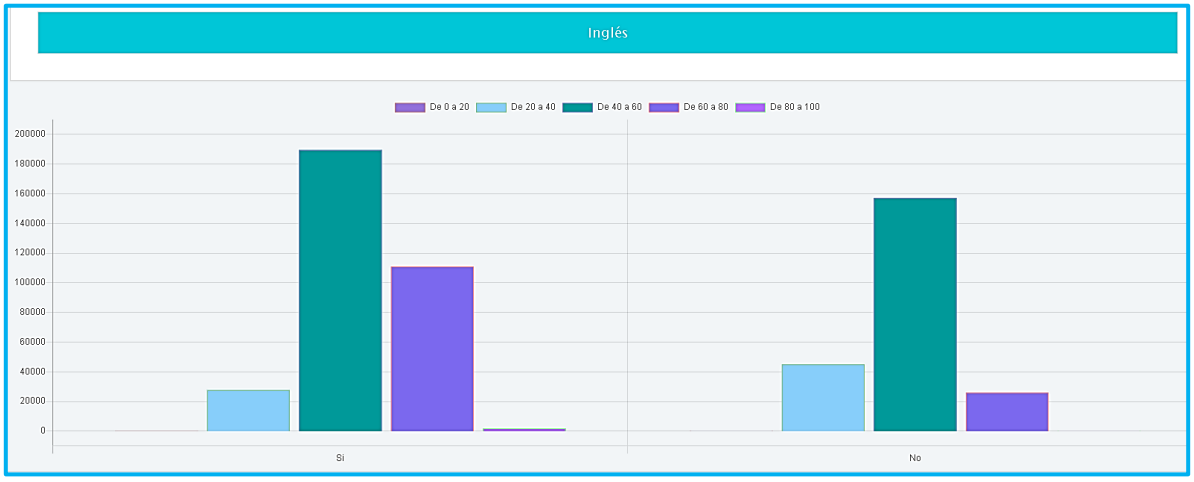


Figura 97. Inglés vs Tiene computador, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

d) Área de ubicación del colegio

Desde el punto de vista social, se tiene en esta prueba el indicador del área de ubicación de las instituciones educativas, el cual al relacionarlo con el género de sus estudiantes nos muestra que: las mujeres son mayoría en las dos categorías que presenta, es decir en el sector rural y urbano, siendo más significativa esta diferencia en el área urbana.

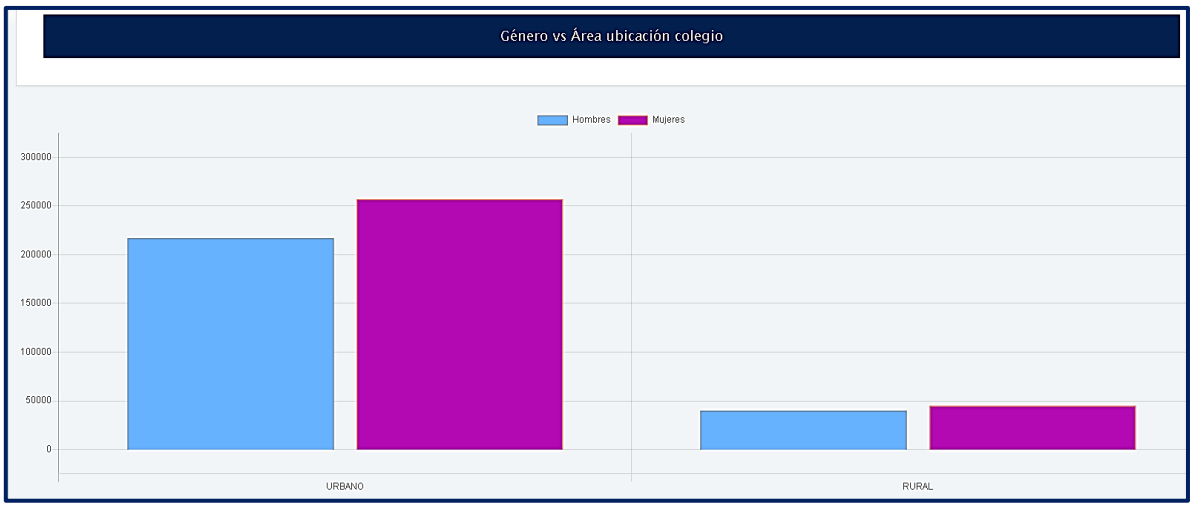


Figura 98. Género vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Con respecto a los puntajes obtenidos, se logra visualizar en los núcleos Lectura crítica y Matemáticas que el rango de 40 a 60 es el más frecuente en ambas respuestas. Pero, aun así, se puede ver una mayor presencia de puntajes superiores en la zona urbana.

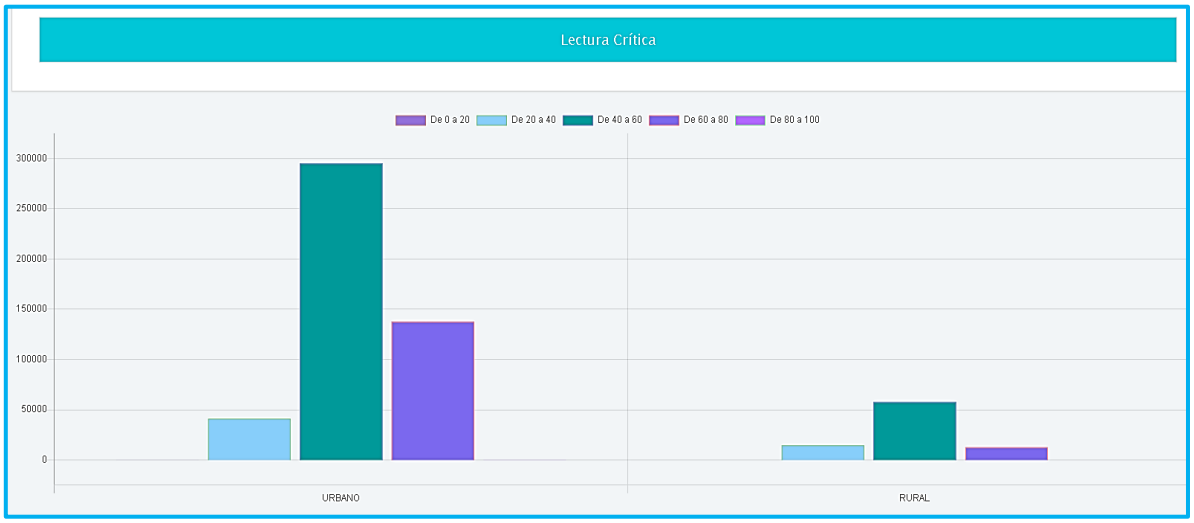


Figura 99. Lectura crítica vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)



Figura 100. Matemáticas vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo Ciencias Naturales, para la zona urbana se observa una mayor cantidad de estudiantes que presentan un promedio de 60 a 80 en sus puntajes.

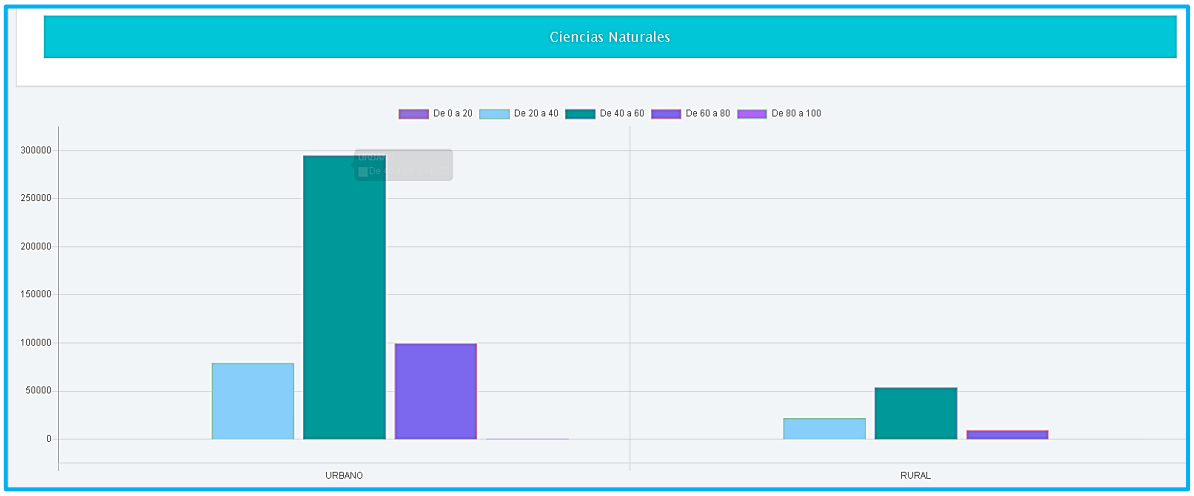


Figura 101. Ciencias Naturales vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Para Ciencias Sociales e Inglés el rango de 40 a 60 es el que presenta con mayor frecuencia en ambas categorías. Sin embargo, aun así, se pueden ver más puntajes superiores en la zona urbana.

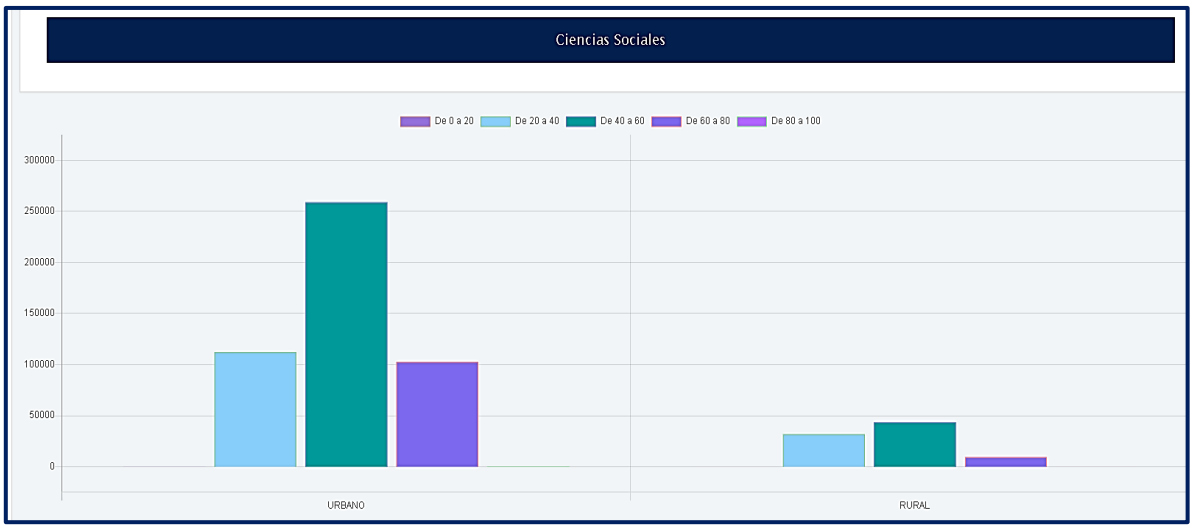


Figura 102. Ciencias sociales vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

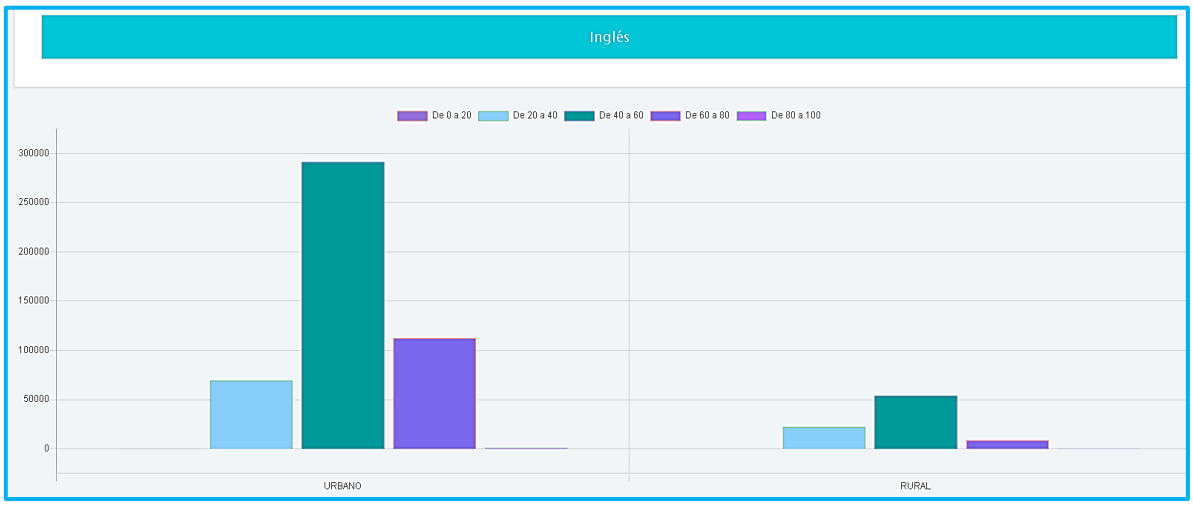


Figura 103. Inglés vs Área ubicación colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

e) Género del colegio

Este indicador presenta una cantidad de estudiantes significativamente superior en la categoría de colegios mixtos, asimismo en dicha categoría la cantidad de mujeres es mayor que la de hombres, mientras que, en los colegios masculinos la mayoría son de su género y de igual forma sucede con los colegios del género femenino.

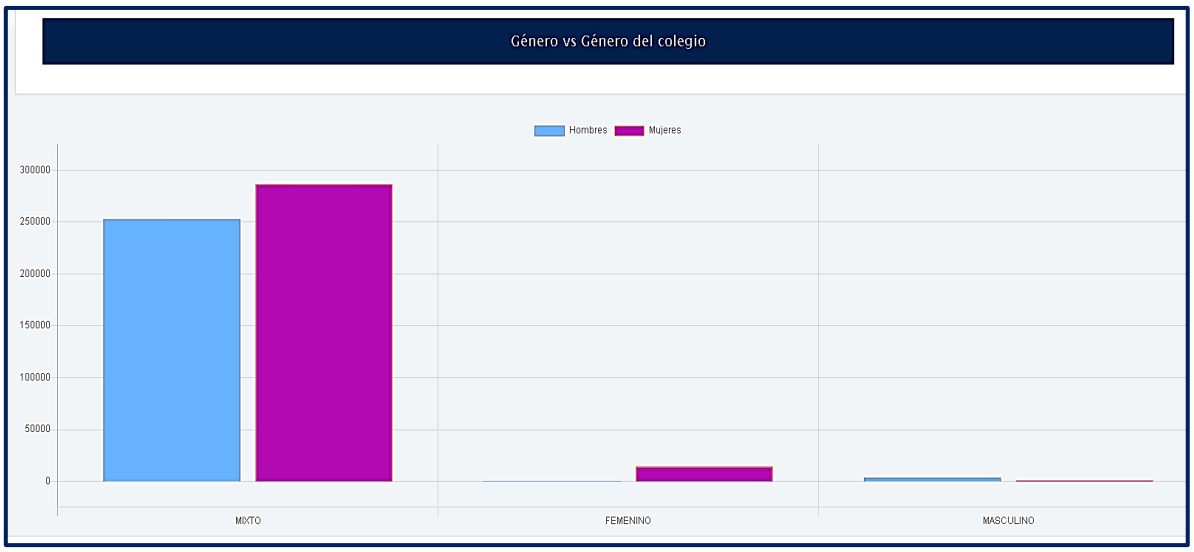


Figura 104. Género vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Analizando los resultados de Lectura crítica, se encuentra que en los colegios de género mixto el rango de puntajes más representativo es entre 40 y 60, otro grupo importante de resultados se presenta entre 60 y 80. En la categoría de colegios femeninos, el rango más frecuente es de 60 y 80, seguido por el de 40 y 60. Para los de género masculino se evidencia resultados entre 60 y 80 mayoritariamente.



Figura 105. Lectura crítica vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Matemáticas, la categoría de colegios mixtos tiene más presencia entre 40 y 60, seguido de cerca por los de 60 y 80. El género femenino muestra el mismo comportamiento.



Figura 106. Matemáticas vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias naturales, los resultados se asemejan a los presentados en Lectura crítica.

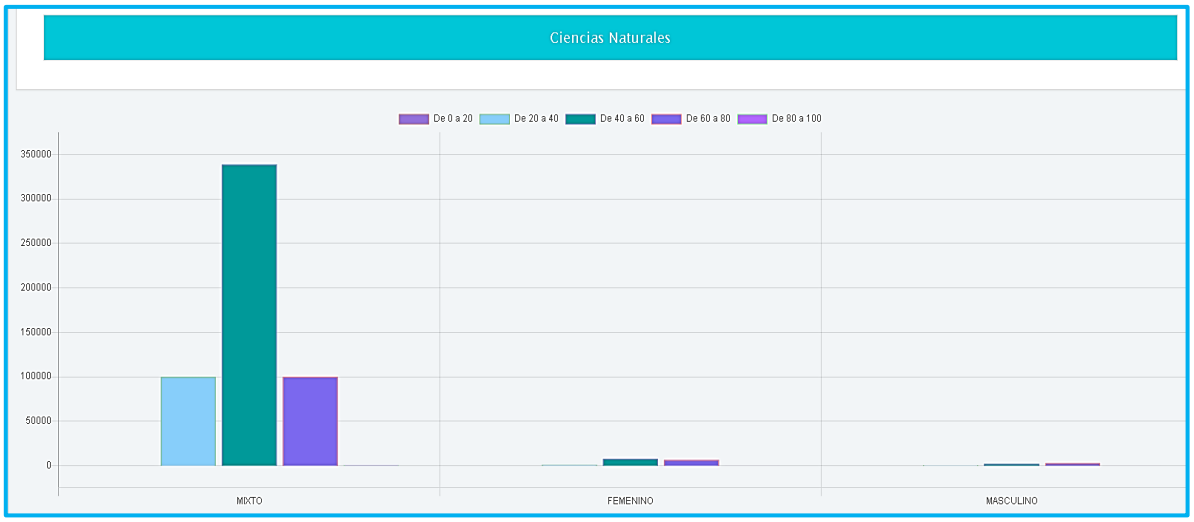


Figura 107. Ciencias naturales vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Para la materia de Sociales, se observa que el promedio de los resultados es predominante en el rango de 40 a 60 para diferentes categorías.



Figura 108. Ciencias sociales vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se destaca que el género femenino presenta mejores puntajes que los estudiantes de colegios mixtos.

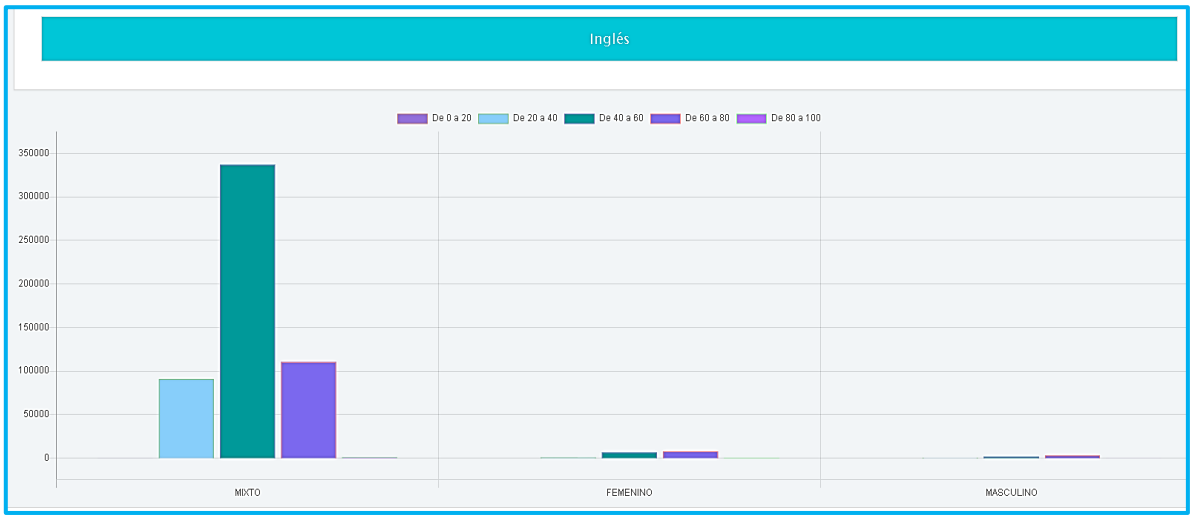


Figura 109. Inglés vs Género del colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En este indicador, existe una brecha positiva para las mujeres en los núcleos de Lectura, Ciencias naturales e inglés. Mientras que, es negativa para los 2 restantes, matemáticas y sociales.

f) Jornada de colegio

Este indicador muestra que la mayoría de alumnos estudian en la jornada de la mañana, seguidos por la completa y tarde. Hay una menor proporción en las jornadas sabatina y noche. Las mujeres son mayoría en cada una de las categorías.

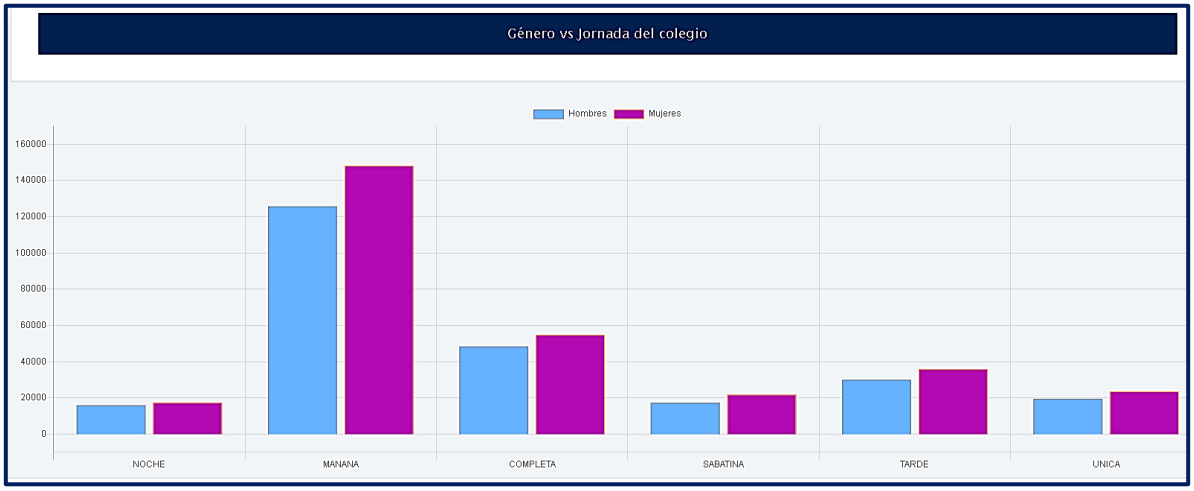


Figura 110. Género vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Los puntajes presentados en Lectura crítica, muestran que el promedio general se encuentra entre 40 y 60, en las jornadas mañana y completa se presentan resultados levemente significativos entre 60 y 80. Y en noche, se alcanzan a presentar resultados mayoritarios en el rango de 20 y 40 mostrando menor rendimiento con respecto a las otras categorías.

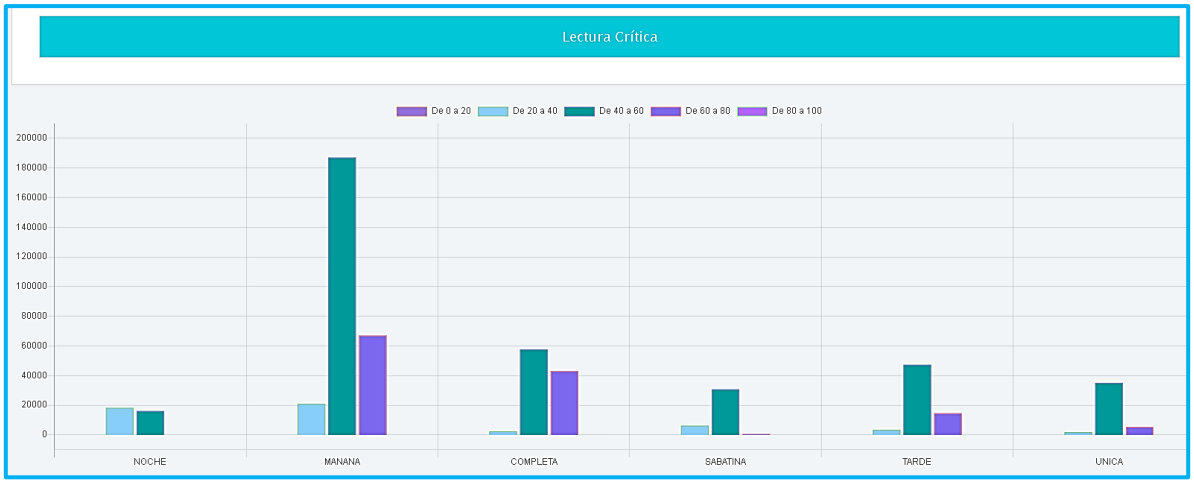


Figura 111. Lectura crítica vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Matemáticas muestra en sus puntajes resultados más bajos en relación con Lectura; puesto que, jornadas como mañana, sabatina y noche incrementan su proporción de estudiantes con promedio entre 20 y 40.

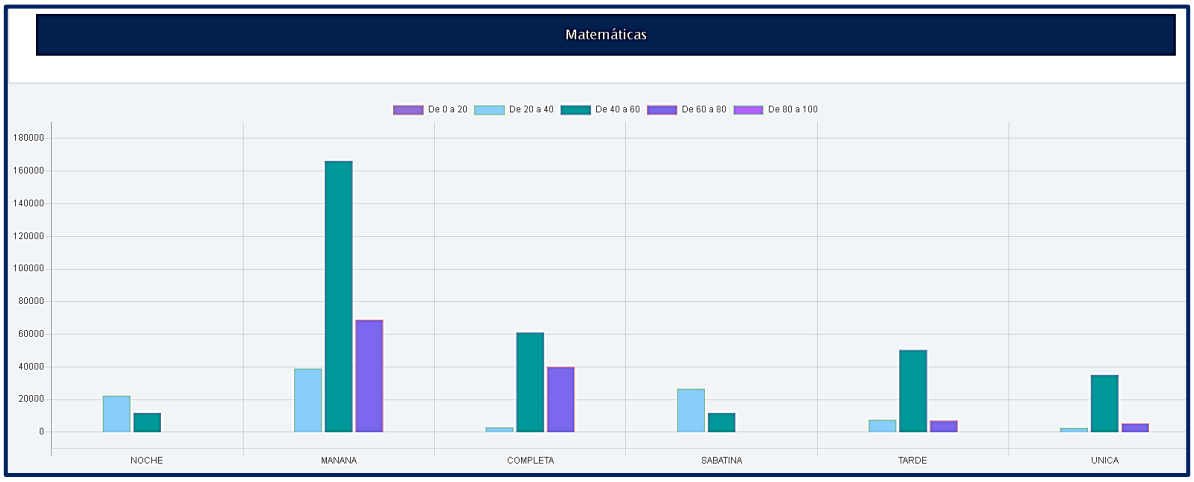


Figura 112. Matemáticas vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

La jornada sabatina y noche en Ciencias Naturales son las que presentan peores resultados, mientras que la jornada completa les gana significativamente a las otras jornadas presentando gran cantidad de estudiantes con promedio entre 60 y 80.

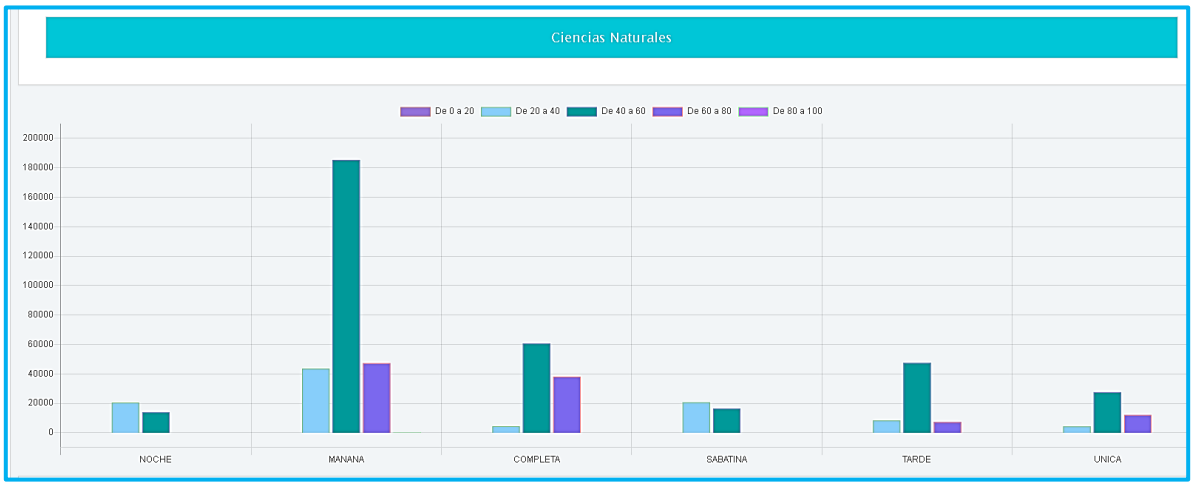


Figura 113. Ciencias naturales vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias sociales las jornadas tarde, noche y sabatina, obtienen resultados representativos entre 20 y 40 puntos, las demás categorías el rango comprendido

entre 40 y 60 es el que se presenta con mayor frecuencia. También en la mañana se puede ver que los resultados son más bajos que en otros núcleos.

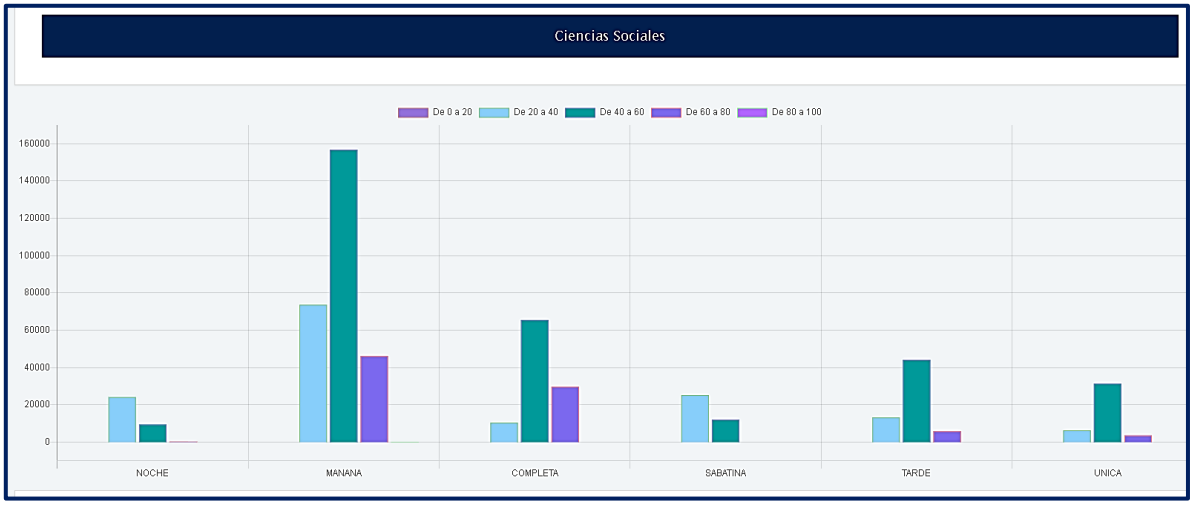


Figura 114. Ciencias sociales vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo temático inglés, se destaca de manera negativa la jornada noche por presentar de manera mayoritaria puntajes entre 20 y 40. Mientras que, las demás categorías presentan resultados aceptables, incluso mejor que otros núcleos.



Figura 115. Inglés vs Jornada de colegio, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En términos generales, se puede decir que la jornada completa es la categoría con mejores resultados. Mientras que, según el núcleo las jornadas tarde, noche y sabatina tienden a presentar resultados bajos.

g) Educación madre

En el caso del nivel educativo de la madre de los estudiantes, se logra observar una mayor representación en los niveles de educación primaria y secundaria, teniendo en cuenta quienes culminan con sus estudios o no. Seguidos en menor proporción están los niveles de Técnica y Profesional.

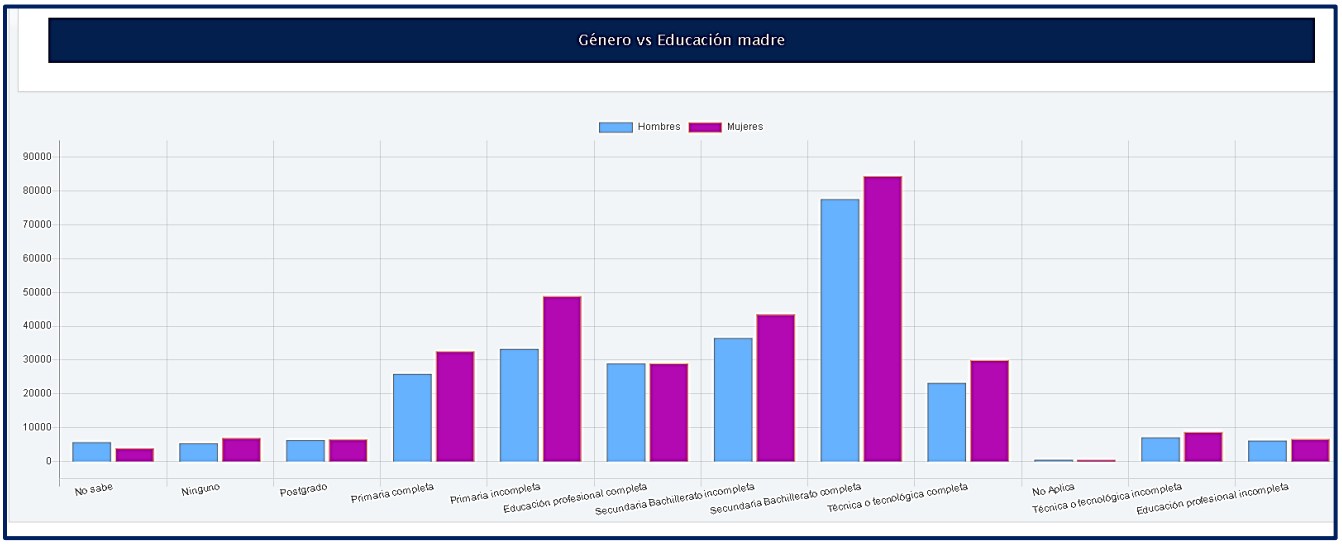


Figura 116. Género vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Por otra parte, se evidencian algunas brechas entre ambos géneros en las categorías de educación secundaria, técnica y profesional. En esta última, el género masculino tiende a tener mayor representación de madres entre los estudios culminados y sin terminar.

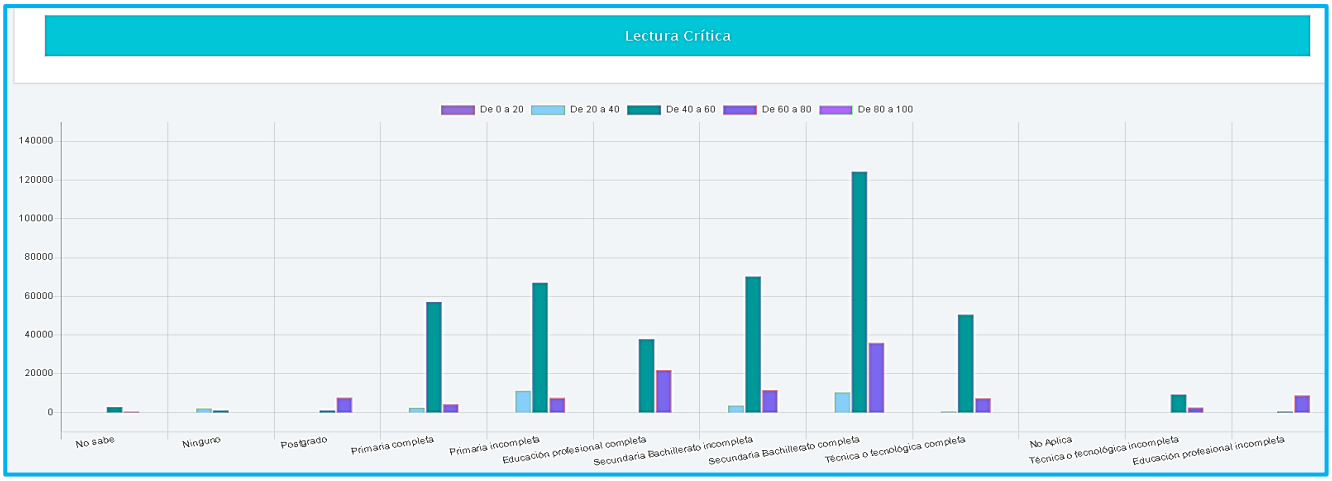


Figura 117. Lectura crítica vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Al analizar el indicador con los puntajes obtenidos por los estudiantes en el núcleo de Lectura crítica, se evidencia puntajes entre 40 y 60 como los más frecuentes. Los estudiantes con madres con rangos educativos más altos, presentan puntajes más altos.

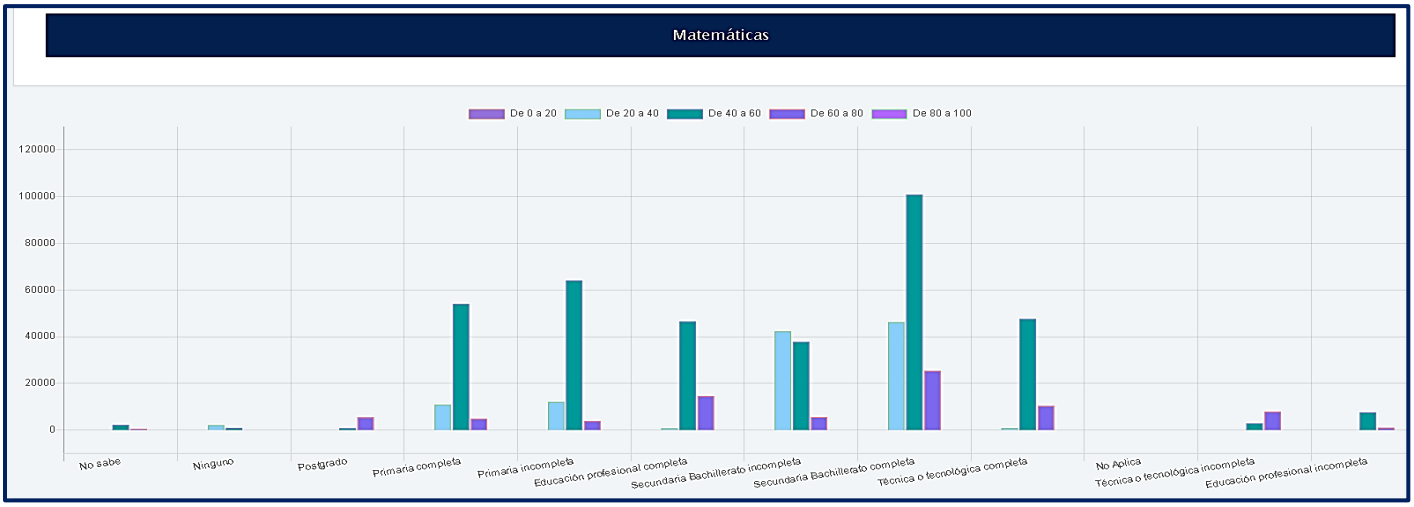


Figura 118. Matemáticas vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de matemáticas, se puede evidenciar como aquellos estudiantes que tienen una madre con estudios en postgrado y técnicos lograron tener resultados más sobresalientes. Sin embargo, en el nivel de secundaria, los resultados son negativos con respecto a las demás categorías y en términos generales el rango que se presenta con más frecuencia es entre 40 a 60.

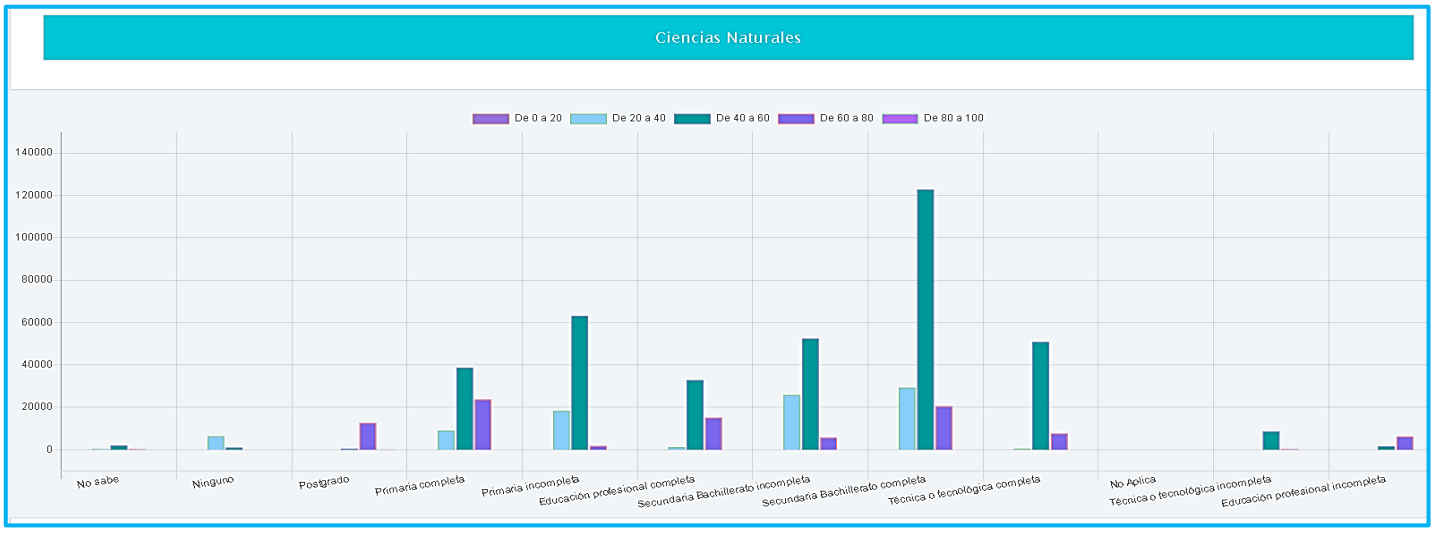


Figura 119. Ciencias naturales vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el área de ciencias naturales, el rango más frecuente es entre 40 a 60. En la categoria de postgrado los puntajes son mayores con respecto a las demás categorías, caso contrario sucede con ninguno, donde se evidencian los puntajes más bajos.

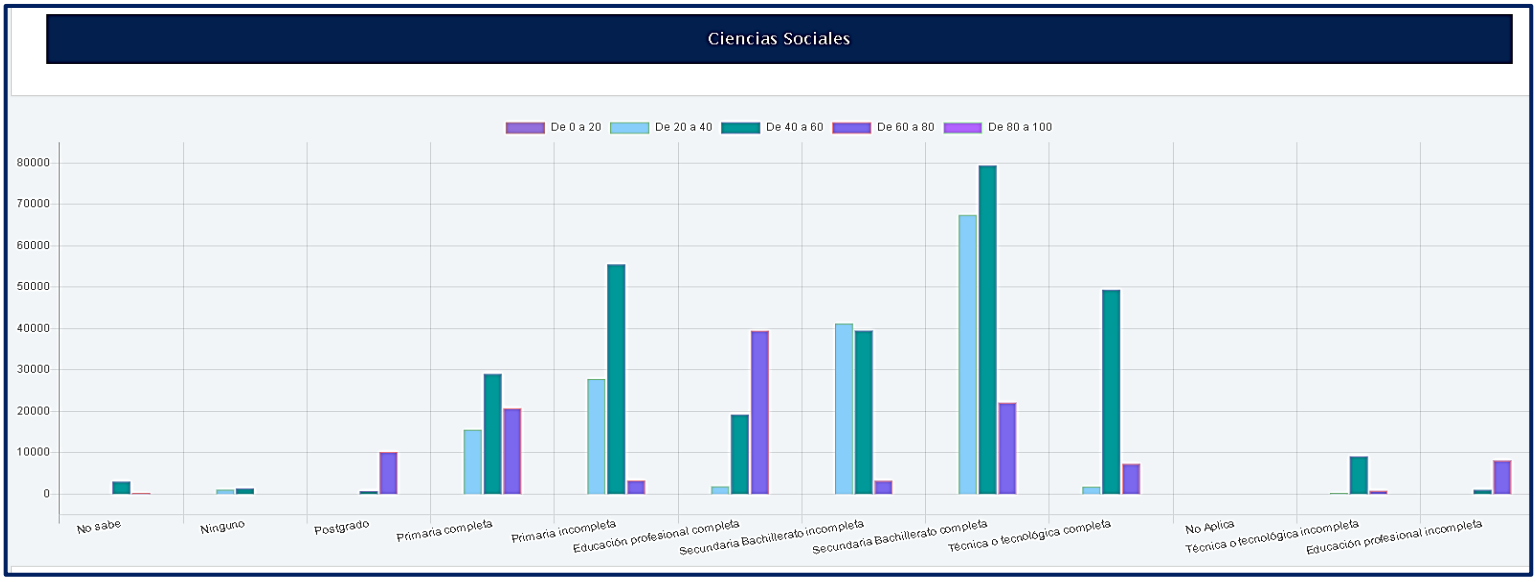


Figura 120. Ciencias sociales vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias Sociales, los resultados más preocupantes se evidencian en secundaria incompleta, al presentar una gran cantidad de resultados ubicados entre 20 y 40. Por otra parte, se destaca de manera positiva los buenos puntajes en la categoría de educación profesional incompleta al igual que postgrado.

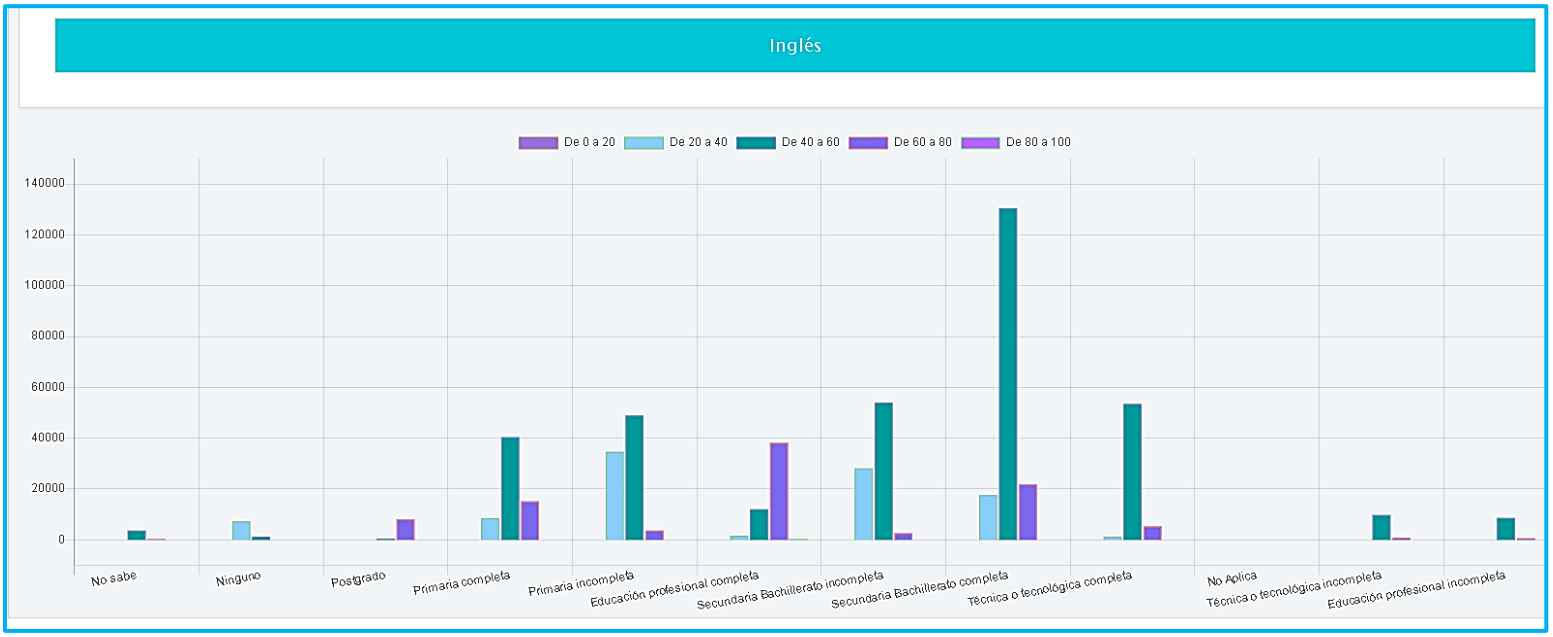


Figura 121. Inglés vs Educación madre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando con este indicador, se puede decir que los estudiantes en el núcleo de inglés tienen en promedio puntajes entre 40 y 60; pero en primaria incompleta se ve una afectación negativa en los resultados, caso contrario para los de postgrado

h) Educación padre

En este indicador es pareja la cantidad de padres presentada entre hombres y mujeres en las diferentes categorías; sin embargo, en primaria completa se evidencia que son en gran proporción la mayoría padres de mujeres; mientras que, de forma contraria sucede en educación profesional incompleta donde los padres de estudiantes hombres representan la mayoría.

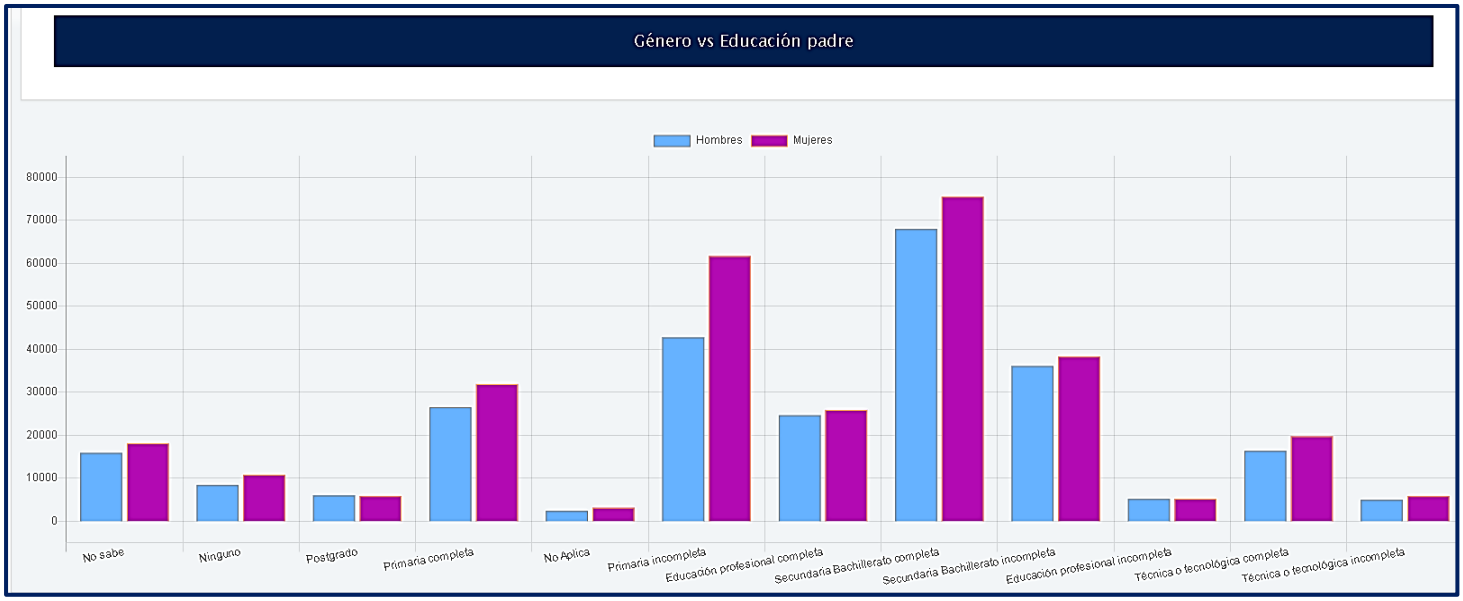


Figura 122. Género vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis de Lectura crítica, se encuentra que la mayoría de estudiantes obtienen puntajes dentro del rango de 40 a 60, evidenciando algunos resultados más bajos en secundaria y primaria incompleta, y más marcados en la categoría no sabe. Por el contrario, más altos, en postgrado, educación profesional completa e incompleta y secundaria completa.

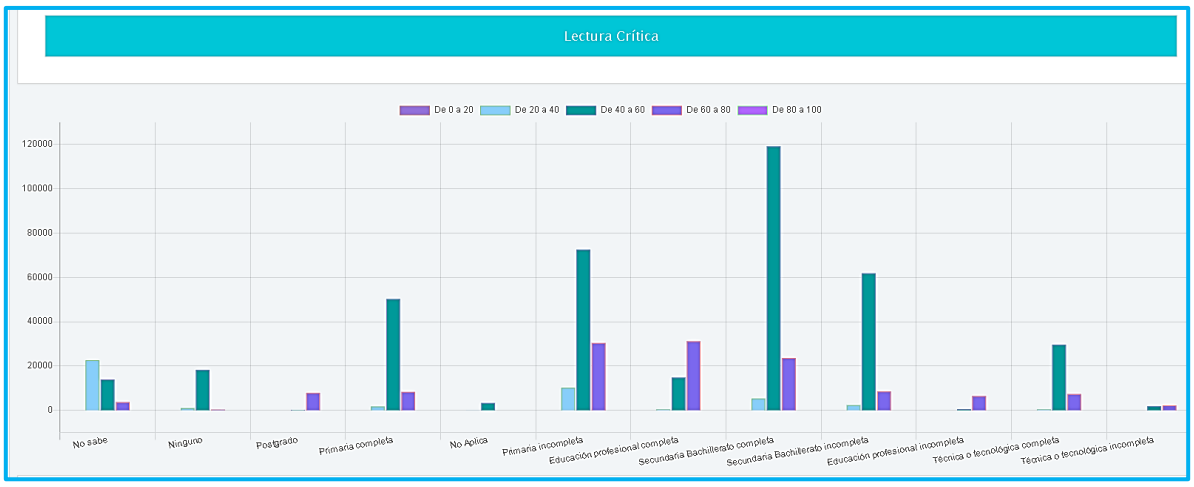


Figura 123. Lectura crítica vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Matemáticas la calificación más frecuente es en el rango de 40 a 60; pero, a pesar de ello, en las categorías primaria completa y no sabe los resultados son en su mayoría negativos. Caso opuesto sucede con educación profesional completa y Postgrado donde son evidentes los puntajes entre 60 y 80 puntos.

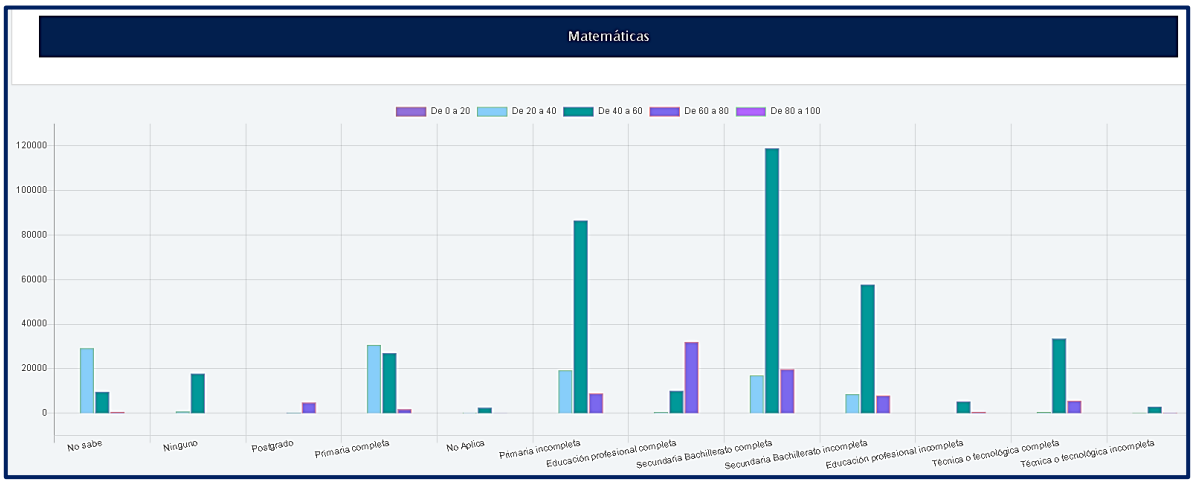


Figura 124. Matemáticas vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Comparando con el núcleo anterior, se observa diferencia en la categoría de ninguno donde se evidencia puntajes menores, mientras que en primaria completa mejoran.

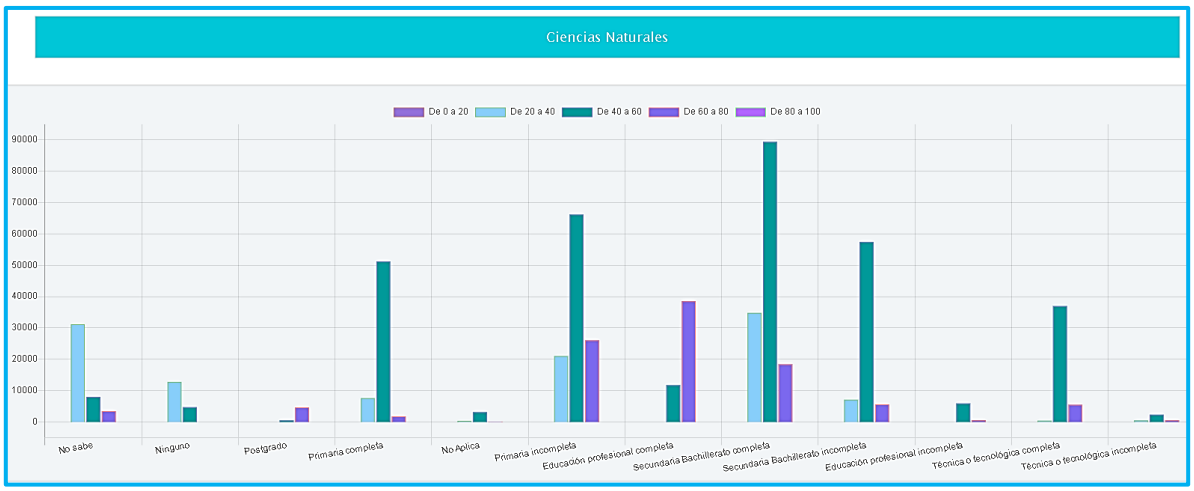


Figura 125. Ciencias naturales vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Ciencias sociales, se logra visualizar buenos resultados en educación profesional completa e incompleta y postgrado, puntajes deficientes en los niveles de primaria incompleta, no sabe y ninguno.

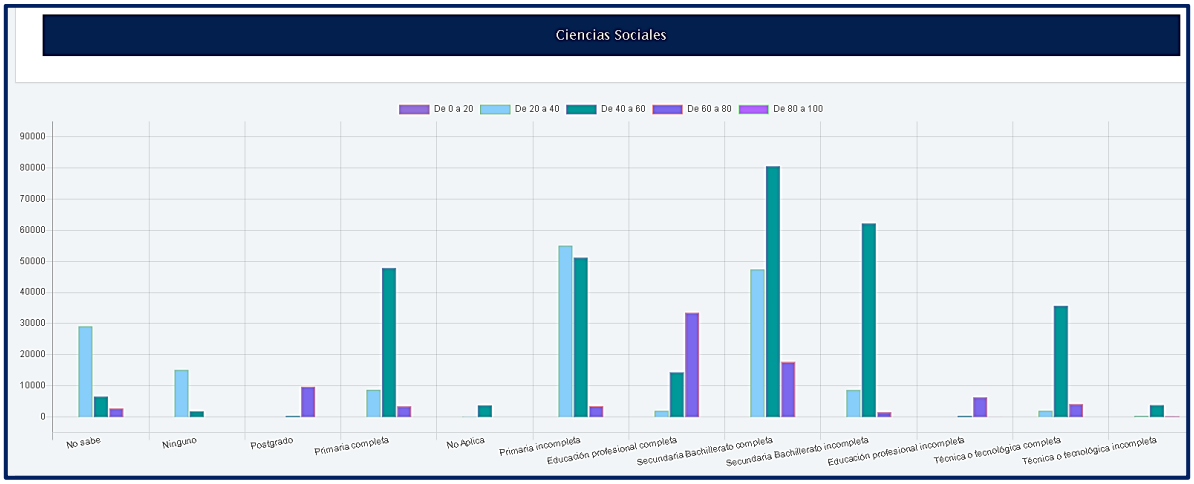


Figura 126. Ciencias sociales vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Por último, las categorías que tienen en su mayoría resultados destacados es Educación profesional completa y Postgrado. Mientras que, es preocupante los resultados clasificados como No sabe tanto en este núcleo como en los analizados anteriormente.

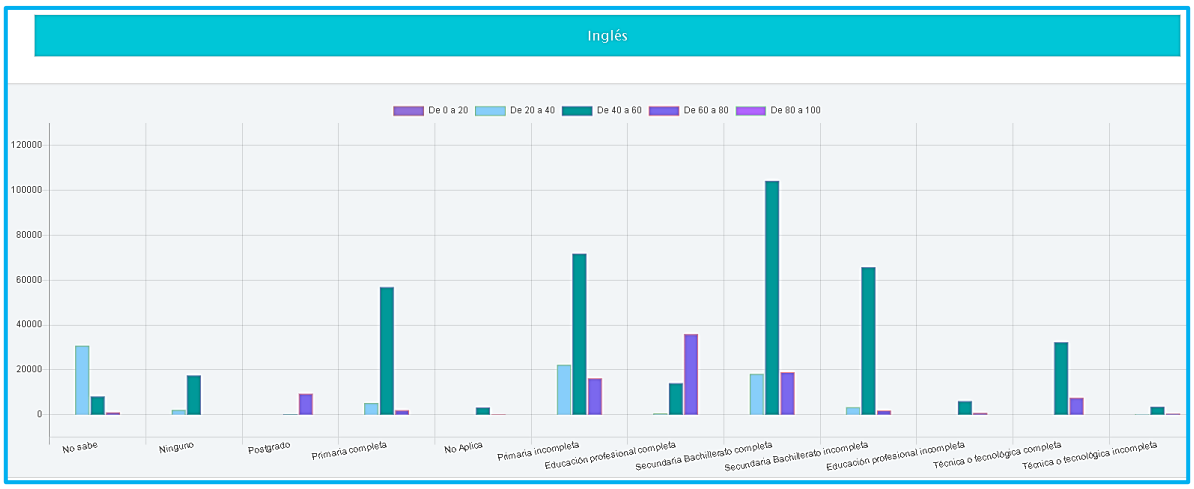


Figura 127. Inglés vs Educación padre, Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

4.2. PRUEBAS SABER TYT

En las pruebas Saber TyT según los indicadores de brecha de género que afectan los resultados se realizó la aplicación del algoritmo K-Means, encontrando los clústeres más representativos en cada indicador, los cuales son mostrados en las gráficas, a continuación.

a) Estrato de vivienda

Para el indicador de Estrato de vivienda, en relación con el género de los estudiantes que presentaron la prueba se puede observar que las mujeres son mayoría en cada uno de los niveles, menos en el Estrato 4. Asimismo, se evidencia que en los estratos 1, 2 y 3 es donde se concentra la gran mayoría de estudiantes, esto nos indica que la población estudiantil del país vive en condiciones socioeconómicas bajas.

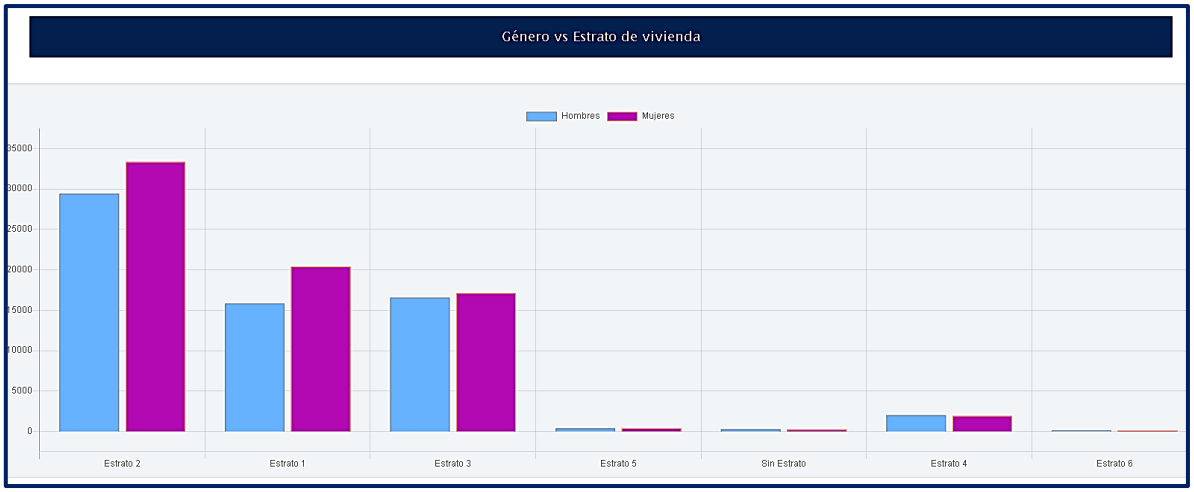


Figura 128. Género vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Haciendo el análisis descriptivo para este indicador con respecto al núcleo de Lectura Crítica, se muestra que el rango de 90 a 120 es el más frecuente en los resultados de los estudiantes, presentándose en mayor proporción en los Estratos 1, 2, 3 y 4 que son los que alcanzan a obtener clústeres significativos.

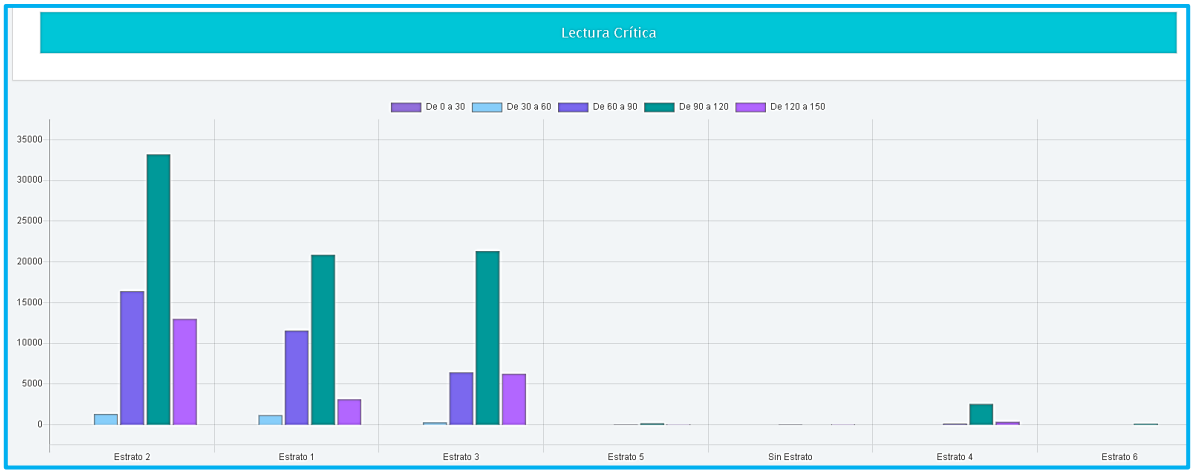


Figura 129. Lectura crítica vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Razonamiento cuantitativo, los resultados de los estratos 2 y 3 presentan el rango entre 90 y 120 como el más representativo; mientras que, en los estratos 1 y 4, los resultados son más bajos, siendo en su mayoría entre 60 y 90 puntos.

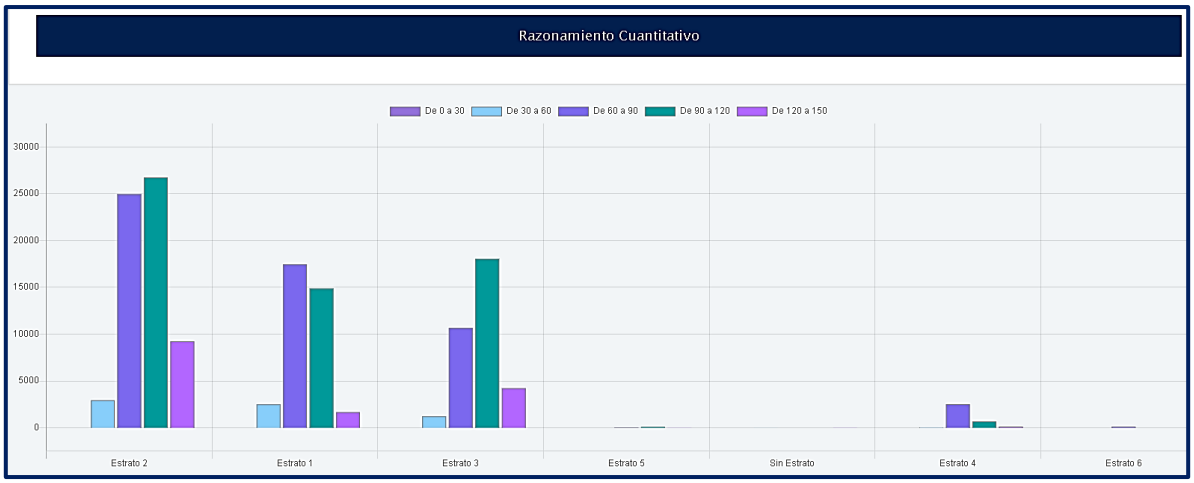


Figura 130. Razonamiento cuantitativo vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el caso de comunicación escrita, el mismo rango de puntajes mencionado anteriormente sigue siendo el más frecuente. Sin embargo, se destaca que en el Estrato 2 y 3 se alcanza a mostrar cierta cantidad de puntajes dentro del rango más alto que es el comprendido entre 120 y 150. Por otra parte, el estrato 1 se ve

afectado de forma negativa por el rango entre 60 y 90 y éste mismo alcanza a tener presencia en los estratos 2 y 3 también.

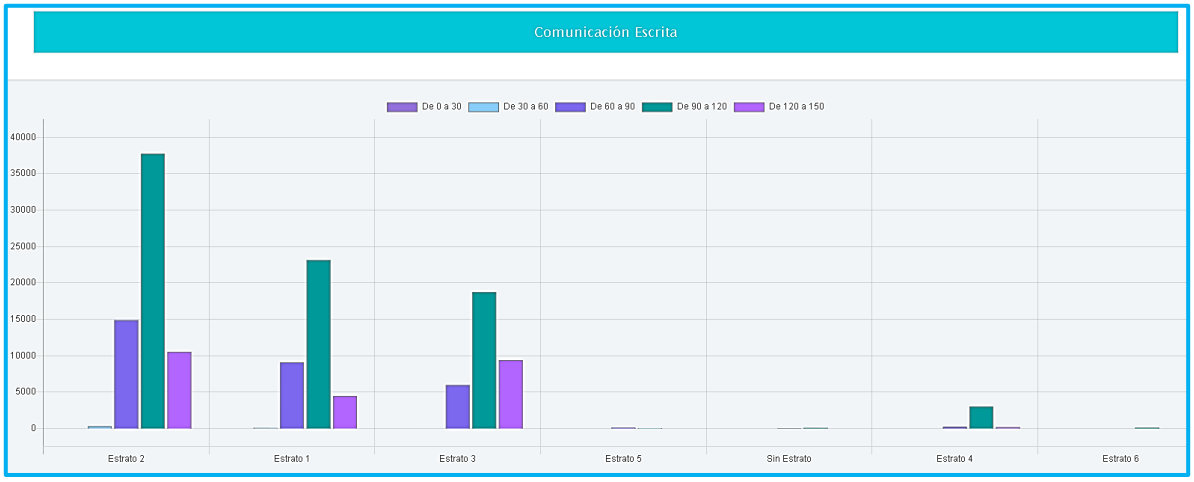


Figura 131. Comunicación escrita vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas, siguen siendo los Estratos 1 y 4 los que presentan los resultados más bajos, sin dejar a un lado que en general los puntajes en los diferentes niveles se ubican en el mismo rango promedio.

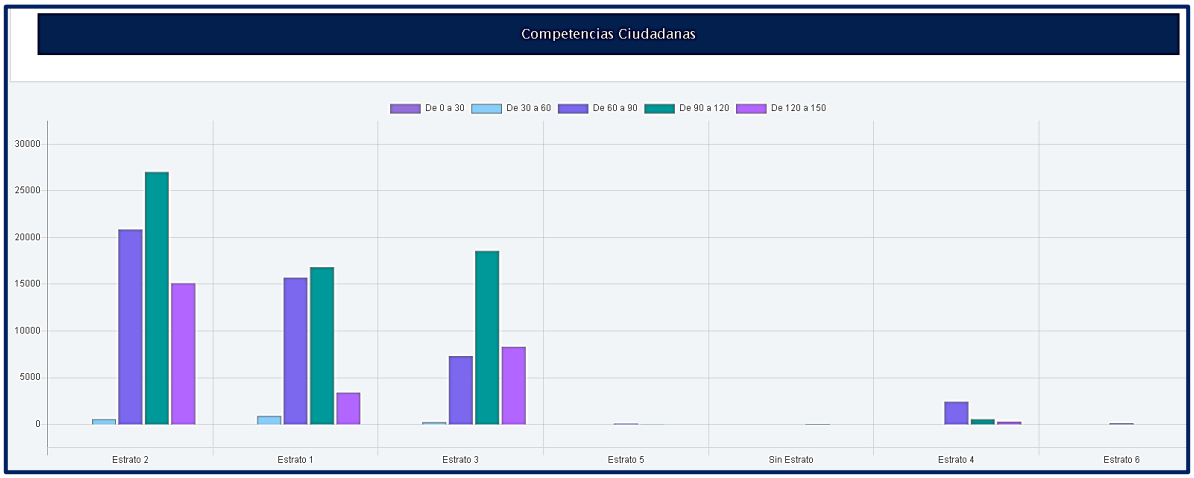


Figura 132. Competencias ciudadanas vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Por su parte, en inglés, los puntajes tienden a ser más bajos, puesto que hay mayor representación del rango entre 60 y 90, especialmente en el Estrato 1 donde es el que mayor representación tiene con una alta proporción respecto a los otros.



Figura 133. Inglés vs Estrato de vivienda, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

b) Pago de matrícula a crédito

La mayoría de estudiantes no pagan la matrícula a crédito, es resultado en este sentido es parejo entre hombres y mujeres.

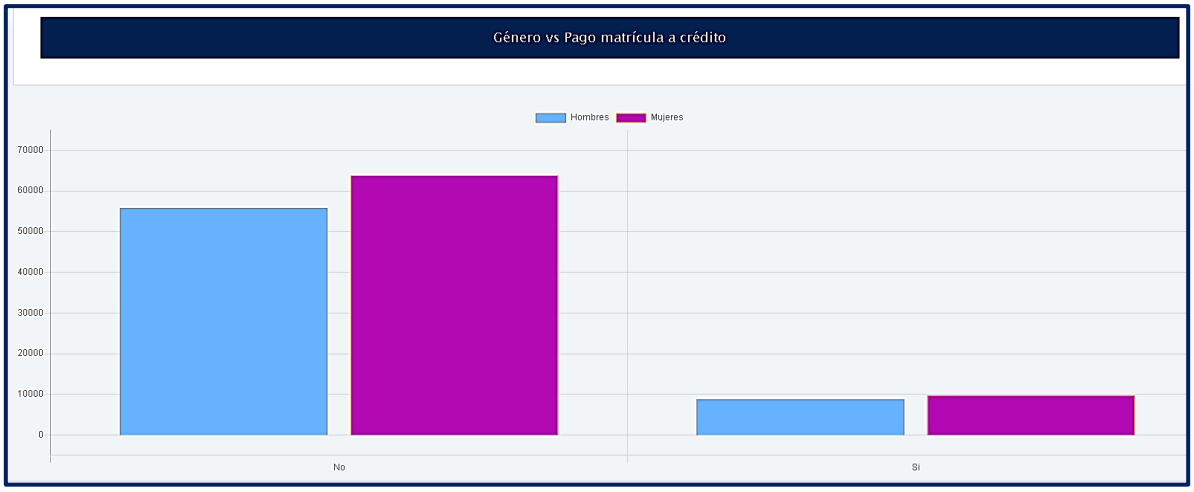


Figura 134. Género vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Con respecto a Lectura Crítica, ambas categorías de respuesta tienen resultados promedio dentro del rango de 90 a 120 con mayor proporción, seguidos del rango entre 60 y 90, y en tercer lugar el rango entre 120 y 150.



Figura 135. Lectura crítica vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Analizando el núcleo de Razonamiento cuantitativo, ambas categorías de respuesta tienen la mayoría de sus resultados dentro del rango de 90 a 120, pero también tienen alta proporción en el rango de 60 y 90, y alcanzan a tener presencia en el rango de 30 y 60.

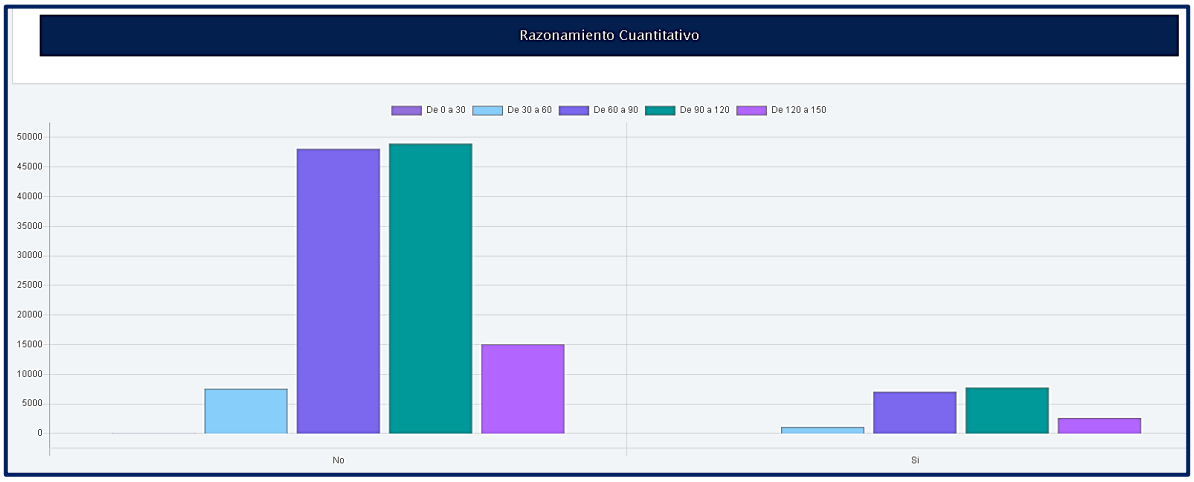


Figura 136. Razonamiento cuantitativo vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Con respecto a Comunicación escrita, el clúster con mayor representación es el de 90 a 120, en la categoría de No también se muestra una cantidad importante de resultados en el rango entre 60 y 90, y en tercer lugar el rango entre 120 y 150.



Figura 137. Comunicación escrita vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas, ambas categorías de respuesta tienen resultados promedio dentro del rango de 90 a 120 con mayor proporción, seguidos del rango entre 60 y 90, y en tercer lugar el rango entre 120 y 150.

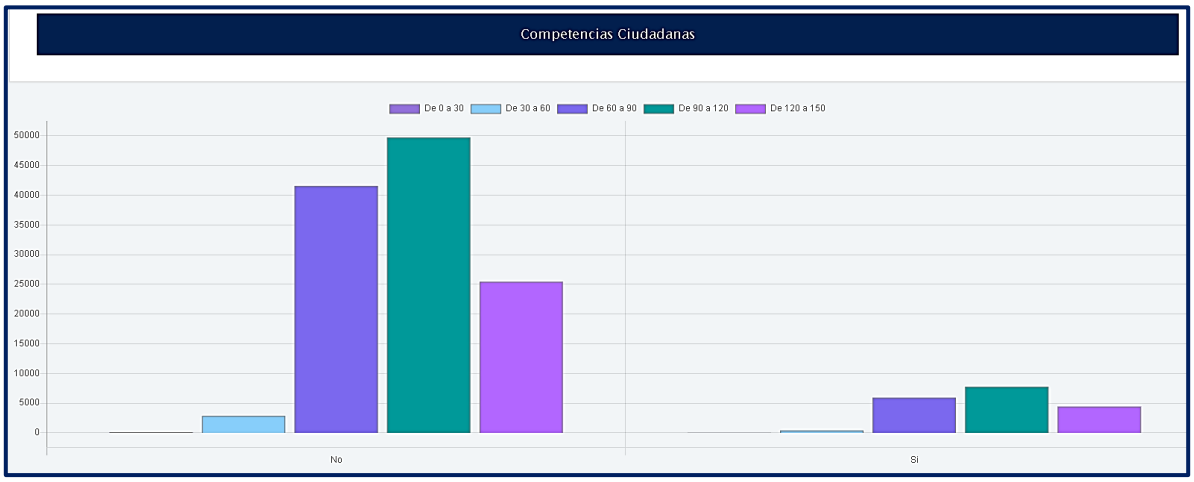


Figura 138. Competencias ciudadanas vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se destaca a diferencia de los otros núcleos que en ambas categorías se alcanza a evidenciar una cantidad de resultados en el rango más bajo que se ubica dentro de 0 y 30.



Figura 139. Inglés vs Pago matrícula a crédito, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

c) Tiene internet

En este indicador se evidencia que la mayoría de estudiantes, tanto mujeres como hombres, cuentan con servicio de internet en sus hogares.



Figura 140. Género vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo Lectura crítica se visualiza que la gran mayoría de estudiantes sacan un puntaje entre 90 y 120, sin depender de la categoría del indicador. Sin embargo, en la categoría Si, se puede observar una mínima representación de puntajes de 0 a 30.

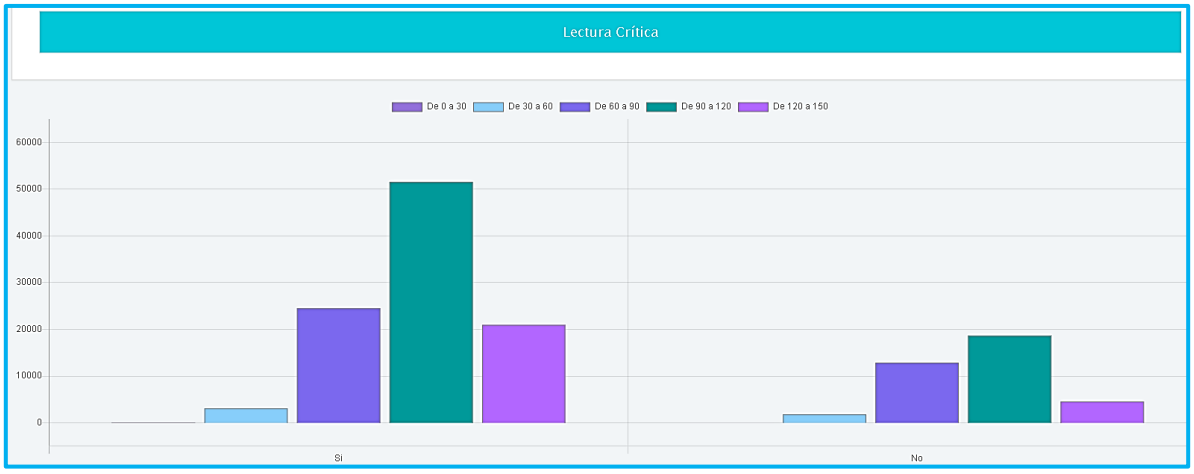


Figura 141. Lectura crítica vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo Razonamiento cuantitativo se logra evidenciar una diferencia bastante pronunciada entre las dos categorías, debido al promedio más bajo en el puntaje obtenido por los estudiantes que no poseen internet, con diferencia del promedio entre 90 y 120 puntos que se visualiza en aquellos que poseen este servicio.

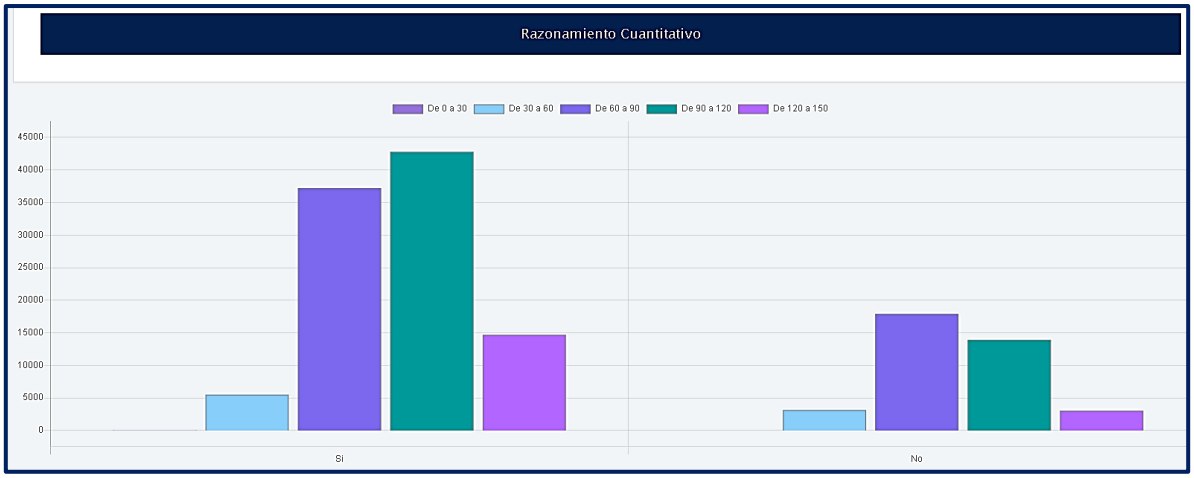


Figura 142. Razonamiento cuantitativo vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, se visualiza que la gran mayoría de estudiantes sacan un puntaje entre 90 y 120, sin depender de la categoría del indicador. Consecutivamente se encuentra el rango de 60 a 120 y por último la representación en el rango 120 a 150.



Figura 143. Comunicación escrita vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo Competencias ciudadanas se observa que la gran mayoría de estudiantes en la categoría Si sacan un puntaje entre 90 y 120, también se presenta una cantidad considerable de resultados entre 120 y 150, y una mínima representación de puntajes de 30 a 60. Por otra parte, los estudiantes que no cuenta con servicio de internet presentan puntajes más bajos al ser el rango entre 60 y 90 el más significativo. A diferencia de los análisis anteriores, se puede visualizar la presencia de cluster en los puntajes entre 0 y 30 para ambas categorías.

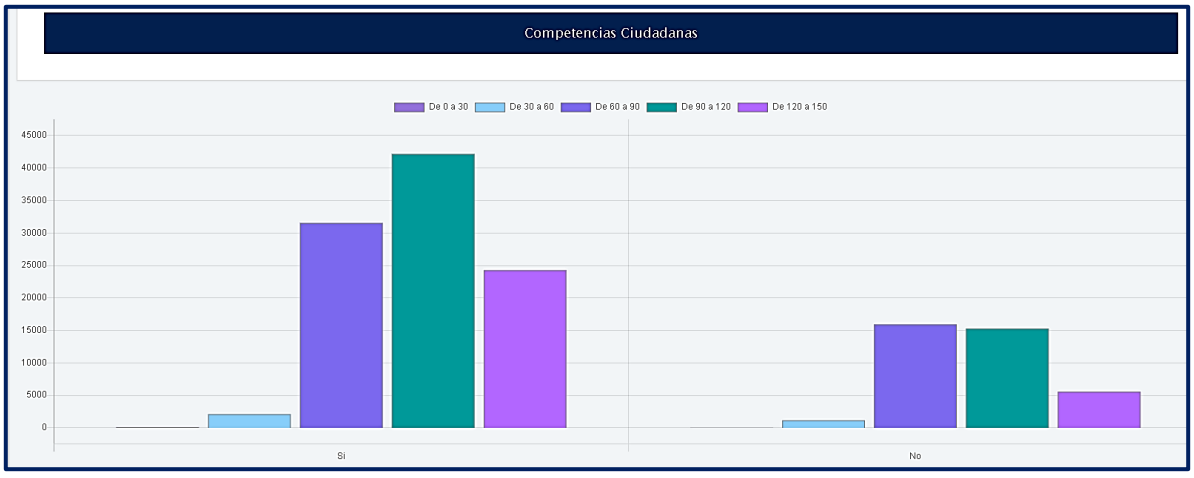


Figura 144. Competencias ciudadanas vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se evidencia que sin importar la categoria el puntaje más destacado es en el rango de 90 a 120, además, se alcanza a presentar resultados en el nivel más bajo que es entre 0 y 30. Para el caso de los estudiantes que no cuentan con servicio de internet se observa que obtienen peores resultados, puesto que su segundo clúster más significativo es el comprendido entre 60 y 90.



Figura 145. Inglés vs Tiene internet, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

d) Área de residencia

Con el análisis de este indicador es notable la diferencia entre ambas respuestas; puesto que, la gran mayoría de estudiantes viven en las áreas urbanas del país, o como se llama en las pruebas: en las cabeceras municipales.

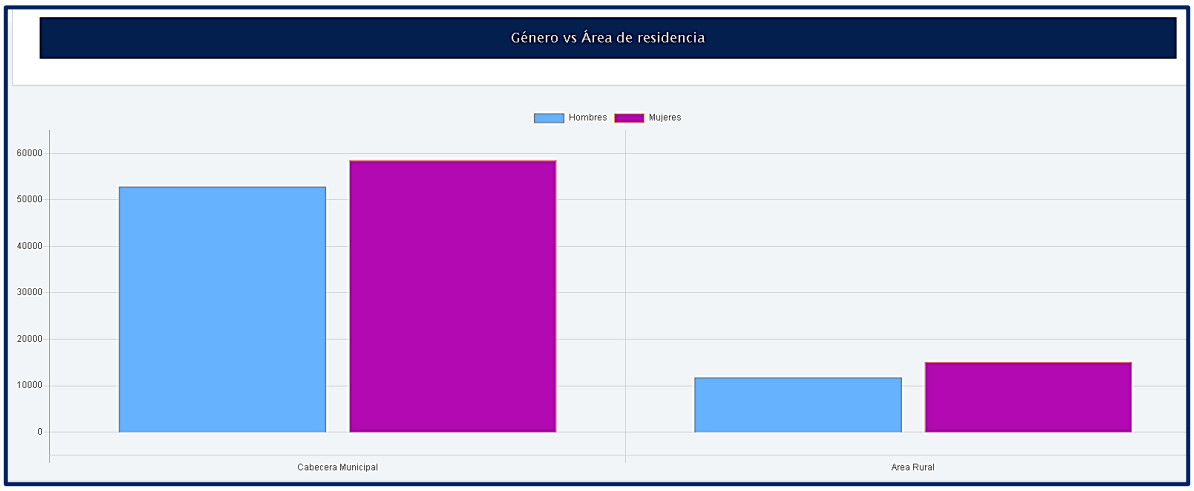


Figura 146. Género vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis de Lectura crítica, se muestra que en ambas categorías el rango más representativo es el de 90 y 120. A pesar de ello, en los resultados de los estudiantes de cabecera municipal se presentan puntajes en el rango de 0 a 30.

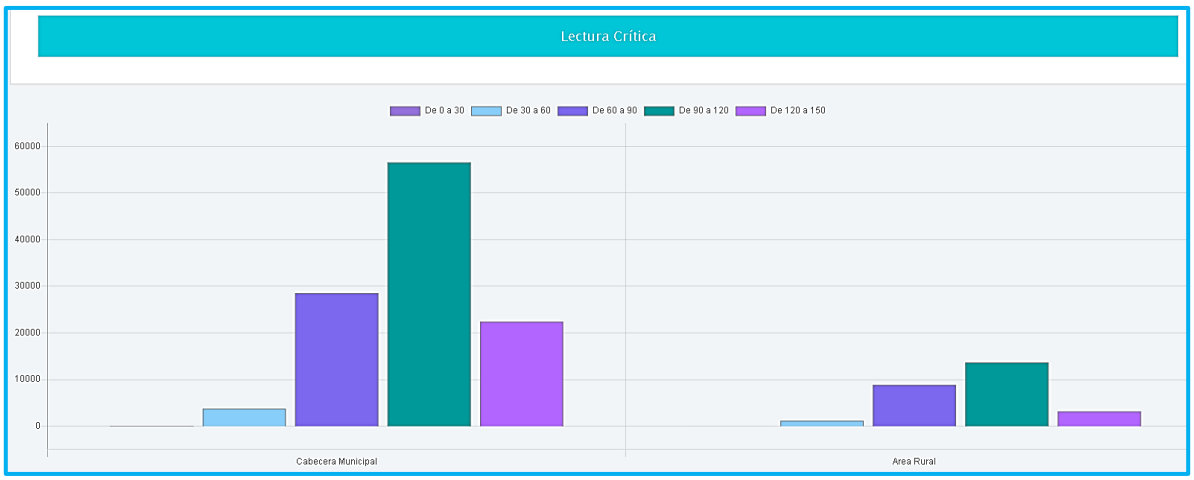


Figura 147. Lectura crítica vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Razonamiento cuantitativo, se evidencia de forma significativa que los puntajes de los estudiantes de cabeceras municipales son mejores. El clúster más representativo en el área rural es el comprendido entre 60 y 90.

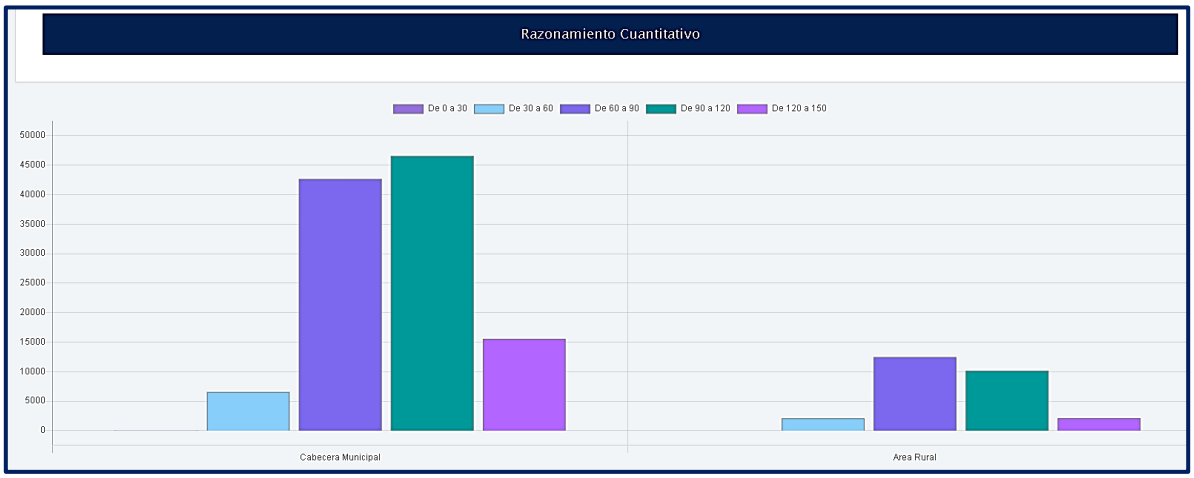


Figura 148. Razonamiento cuantitativo vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Comunicación escrita, se destaca que en ambas categorías el rango más representativo es el de 90 y 120 y presentan un comportamiento bastante similar.

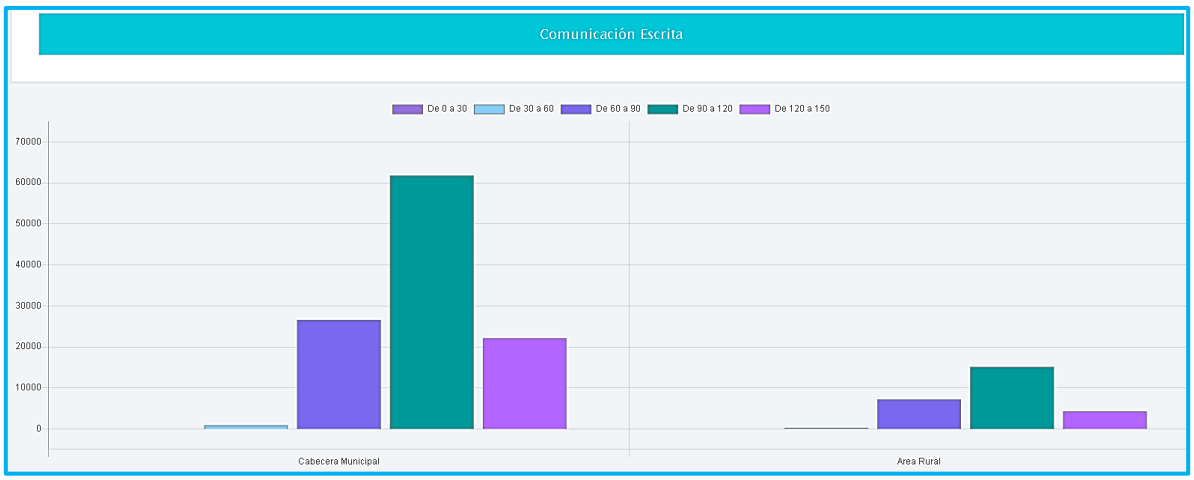


Figura 149. Comunicación escrita vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para Competencias ciudadanas, el comportamiento que se presenta es similar al de Lectura crítica, diferenciándose por la alta presencia de puntajes bajos.

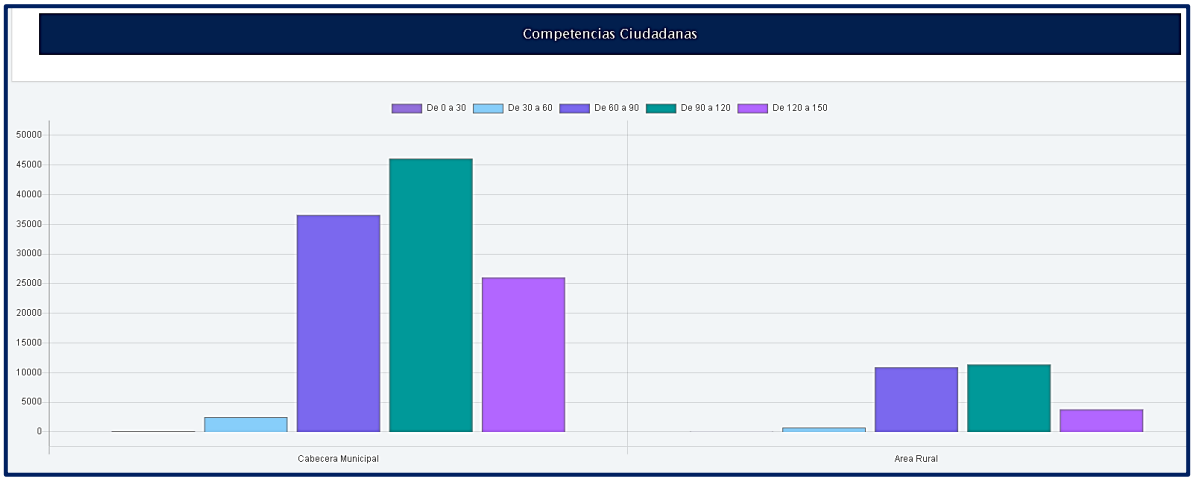


Figura 150. Competencias ciudadanas vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando con este indicador, se analiza el núcleo de inglés encontrando que los estudiantes alcanzan a presentar cierta cantidad de resultados dentro del rango de 0 a 30 en ambas categorías.



Figura 151. Inglés vs Área de residencia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

e) Estado civil

En este indicador se observa que la mayoría de estudiantes se encuentran solteros, la segunda categoría más representativa es la de unión libre, seguidos de cerca por los casados, y por último los separados y/o viudos. Como sucede en varios indicadores las mujeres tienen más presencia en cada una de las opciones de respuesta, pero la diferencia se relaciona con la cantidad de mujeres que presentan la prueba.

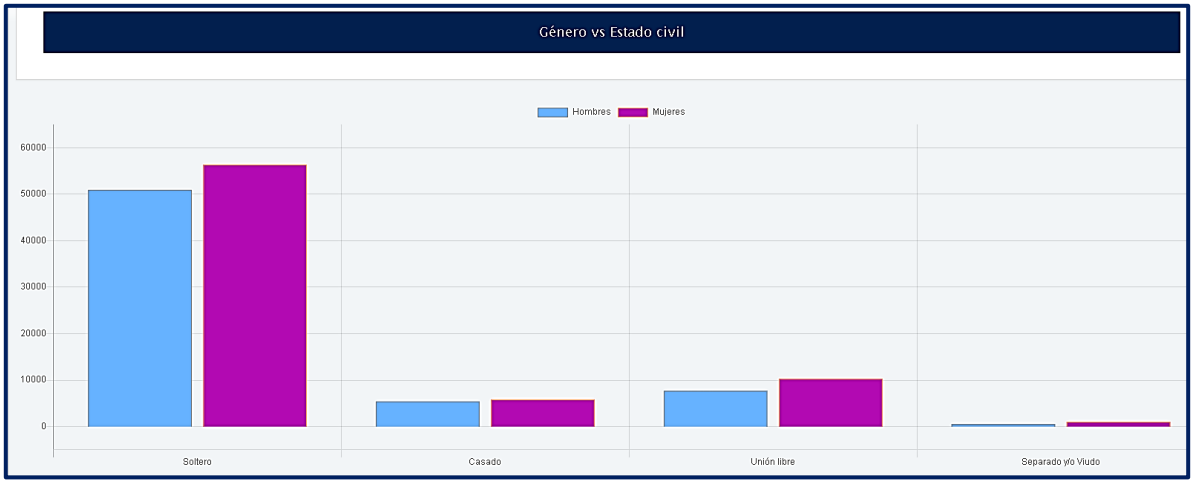


Figura 152. Género vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, se logra evidenciar que en la mayoría de las categorías el rango más presentado es entre 90 y 120, seguido por el rango entre 60 y 90, excluyendo los estudiantes separados quienes presentan puntajes más bajos.

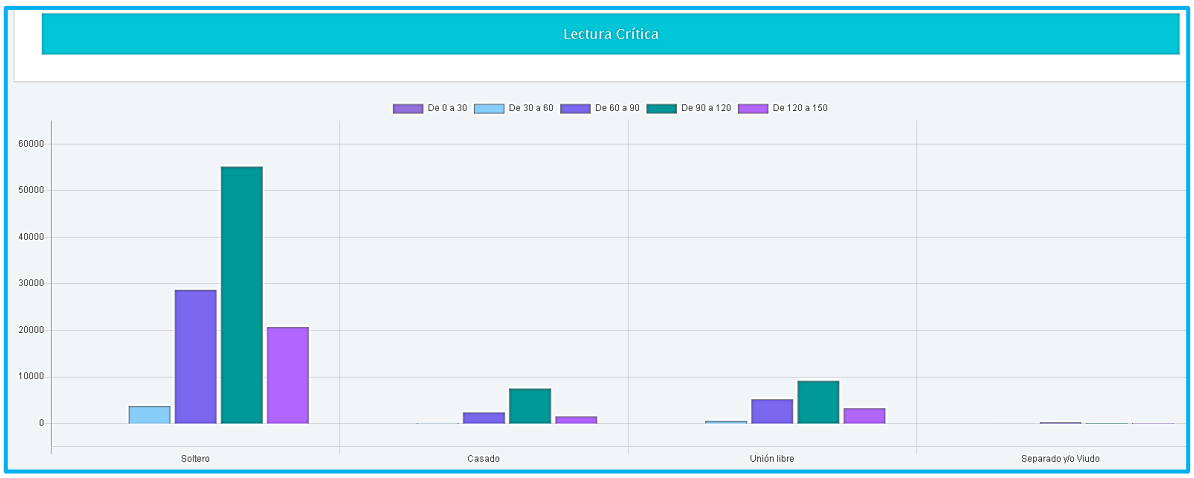


Figura 153. Lectura crítica vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo Razonamiento cuantitativo se observa un comportamiento similar al nucleo anterior, con cierta diferencia en la categoria Soltero, donde se evidencia cierta cantidad de puntajes más bajos.

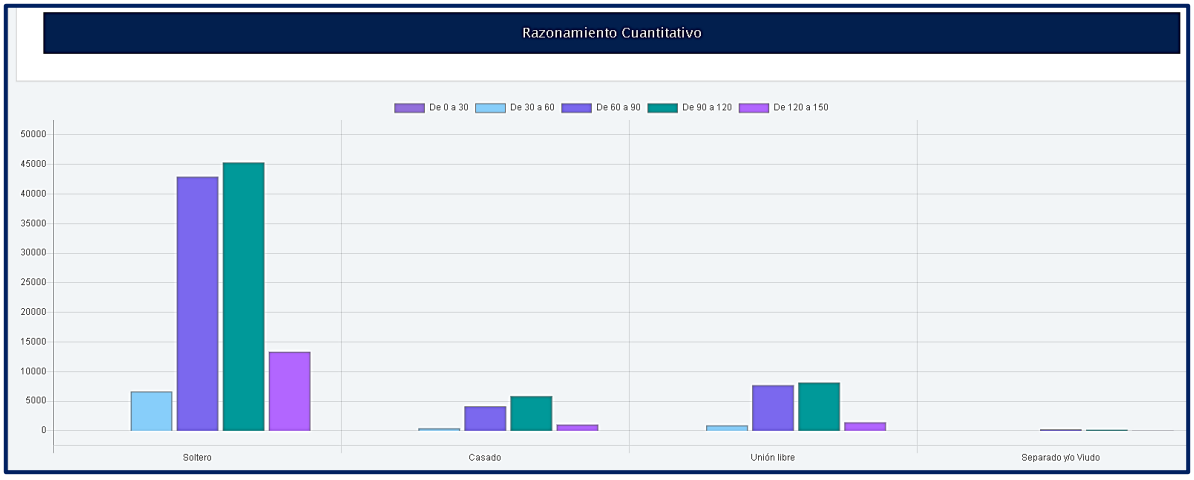


Figura 154. Razonamiento cuantitativo vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo Comunicación escrita, se logra evidenciar que en las 4 categorías el rango más presentado es entre 90 y 120, presentando una variación en la categoría de soltero donde se muestran resultados en el rango de 30 a 60.

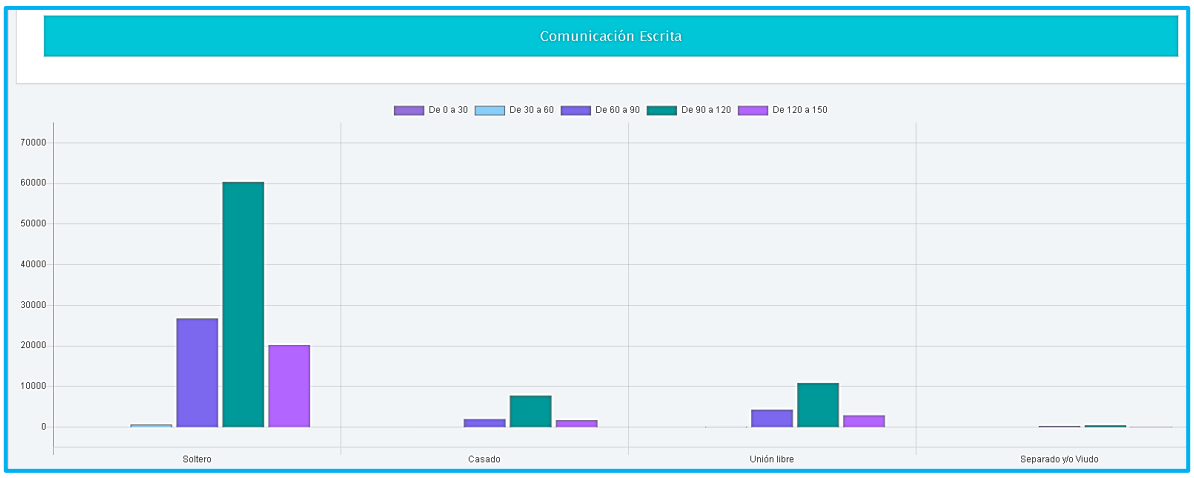


Figura 155. Comunicación escrita vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis de Competencias ciudadanas se presenta un comportamiento semejante al del núcleo Razonamiento cuantitativo.

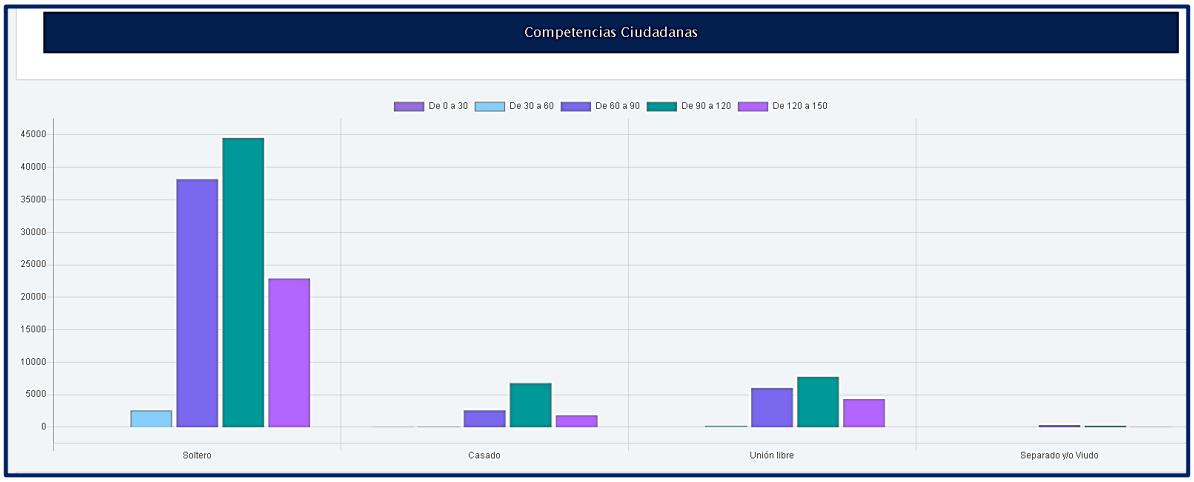


Figura 156. Competencias ciudadanas vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

El núcleo inglés presenta puntajes más bajos en 3 de las categorías del indicador, a diferencia de los demás núcleos temáticos. Se destaca que en Soltero se alcanza a evidenciar resultados dentro del rango más bajo.



Figura 157. Inglés vs Estado civil, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

f) Cabeza de familia

En este indicador, los estudiantes en términos generales no son cabeza de familia; en el caso de la categoría No las mujeres son mayoría; mientras que, en la otra categoría pasa lo contrario. Esto indica que los hombres son quienes tienen más responsabilidad en el hogar.

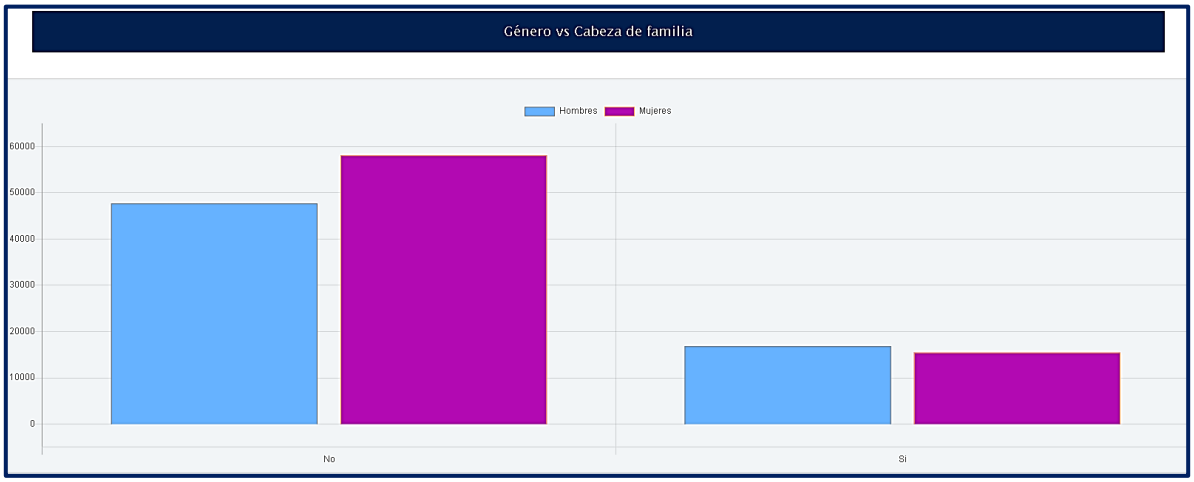


Figura 158. Género vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis de Lectura crítica, se muestra que los resultados de los estudiantes en ambas categorías se encuentran en su mayoría dentro del rango de 90 y 120, seguidos por el rango 60 y 90.

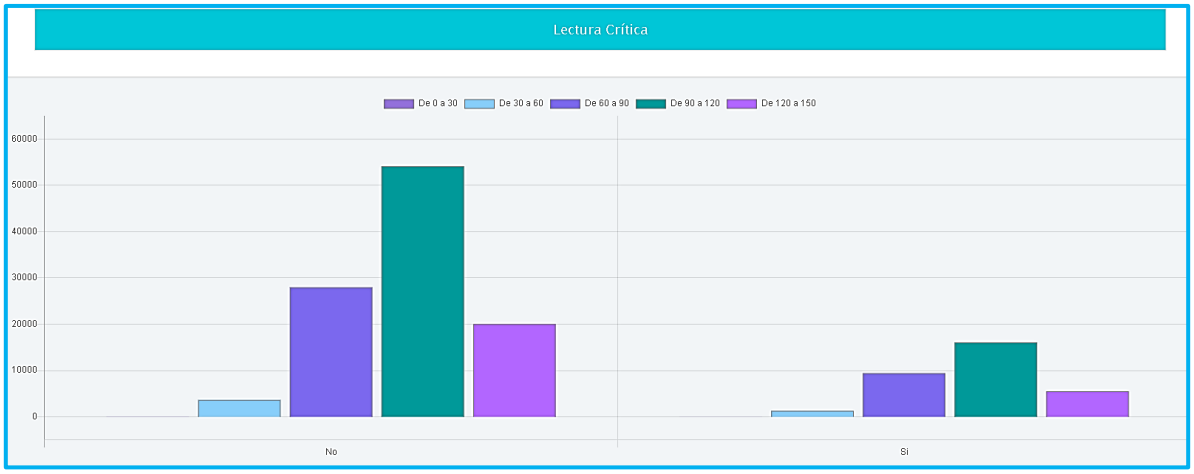


Figura 159. Lectura crítica vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Razonamiento cuantitativo, los estudiantes que son cabeza de familia tienen una cantidad semejante entre el clúster comprendido entre 60 y 90 y el de 90 y 120; mientras que, los que No tienen mejores resultados, siendo su promedio entre 90 y 120.

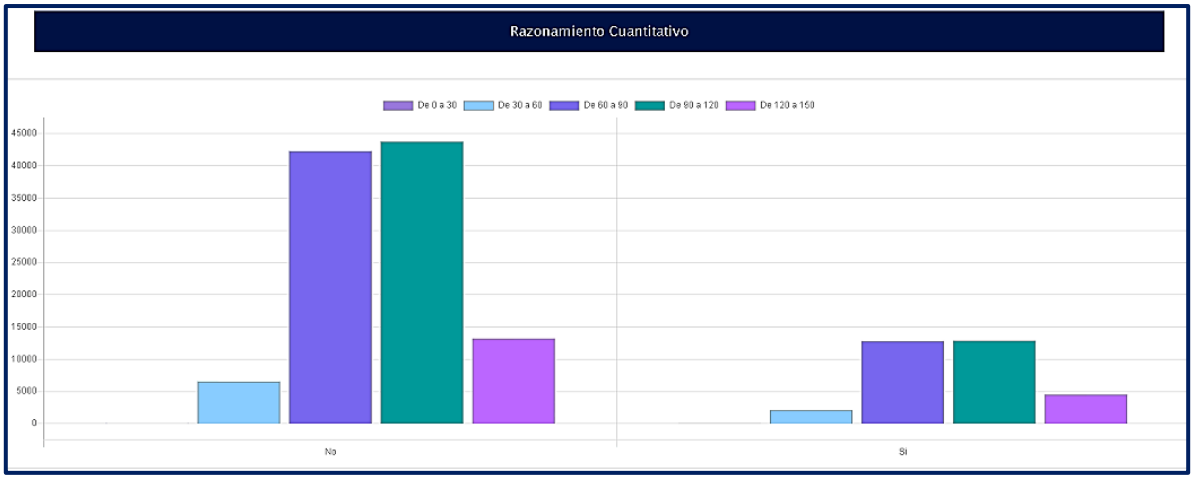


Figura 160. Razonamiento cuantitativo vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Comunicación escrita, se destaca que en ambas categorías el primer puntaje más representativo es el de 90 y 120, la segunda puntuación con más alta proporción es de 60 a 90 y por ultimo los puntajes entre 120 y 150.

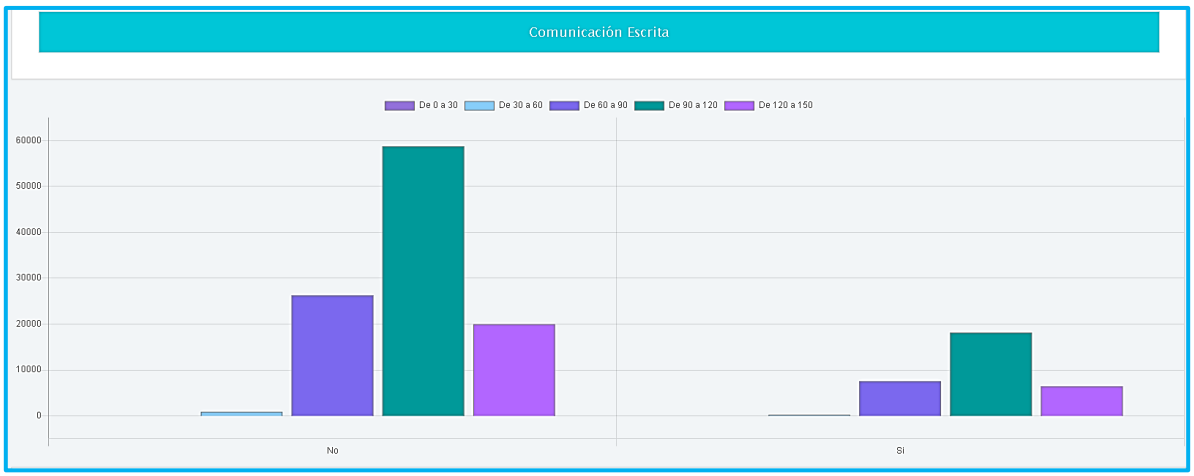


Figura 161. Comunicación escrita vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas se puede evidenciar que en ambas categorías el promedio de puntajes se encuentra entre 90 a 120; sin embargo, los estudiantes que no son cabeza de familia presentan puntajes más bajos.

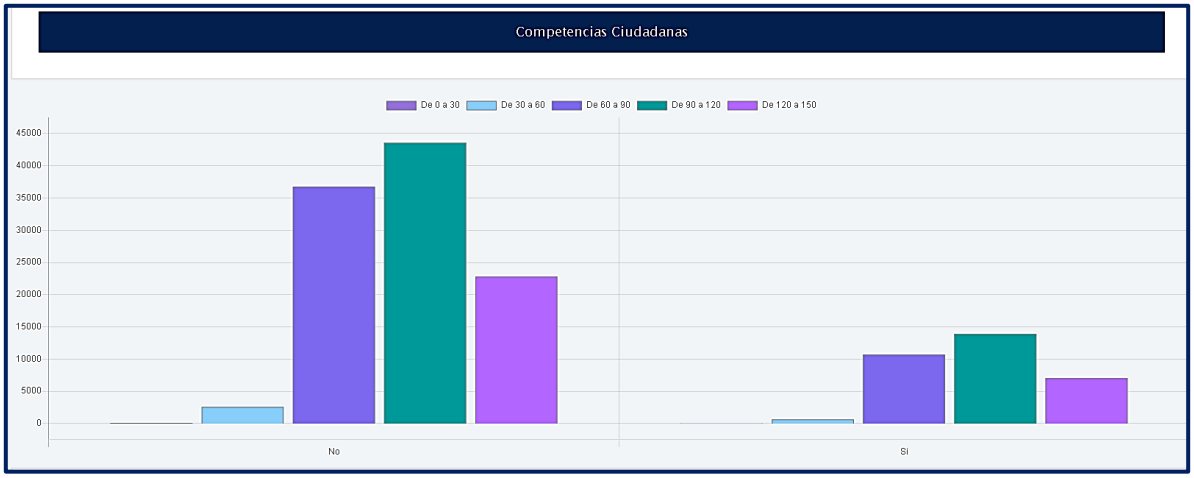


Figura 162. Competencias ciudadanas vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis de inglés se logran observar clústeres para todas las categorías, siendo el rango principal de 90 a 120 y una representación mínima en los rangos más bajos.

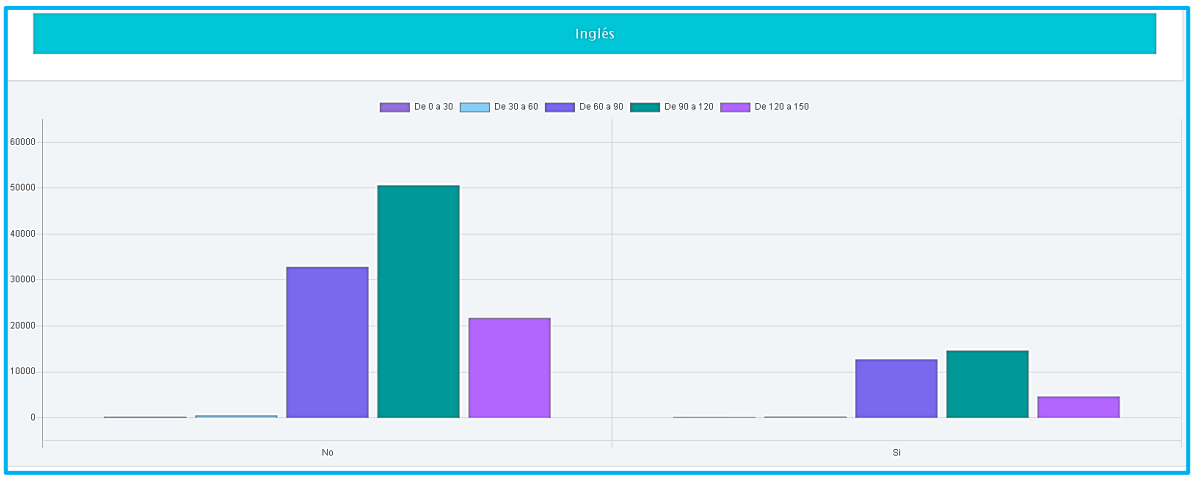


Figura 163. Inglés vs Cabeza de familia, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

g) Número de personas a cargo

Para el caso del indicador Número de personas a cargo, en su mayoría los estudiantes no tienen personas a cargo. Además, se puede evidenciar que a medida que aumenta la cantidad de personas a cargo, va disminuyendo la cantidad de estudiantes.

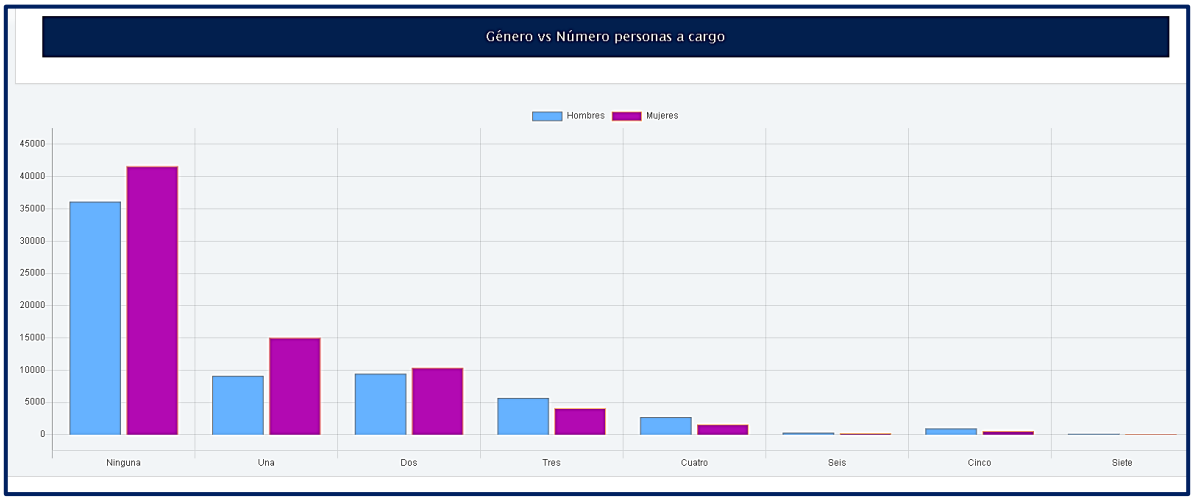


Figura 164. Género vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, se observa que todas las categorías presentan el mismo comportamiento, dónde el clúster más representativo es el comprendido entre 90 y 120; en segundo lugar, el rango entre 60 y 90, seguido de cerca por el de 120 y 150.

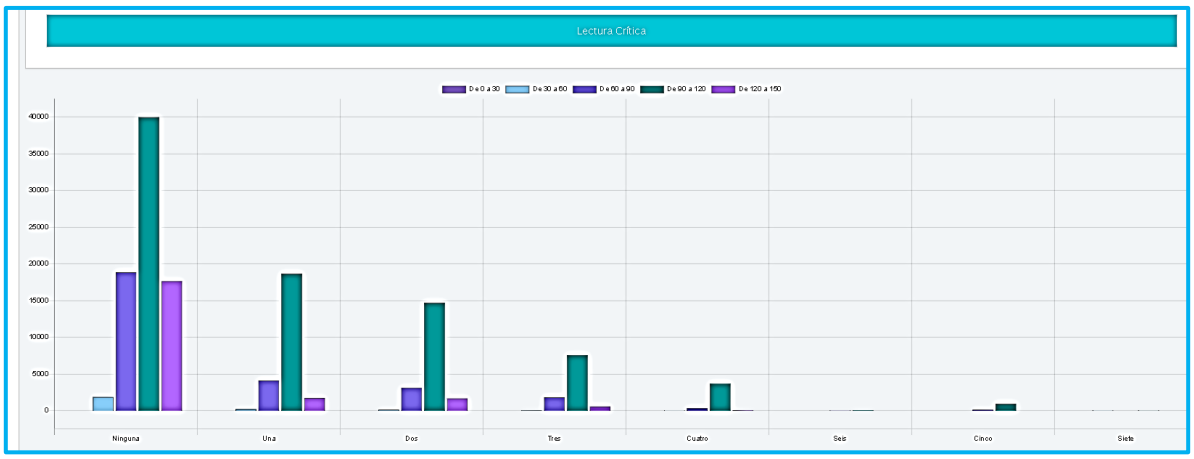


Figura 165. Lectura crítica vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo Razonamiento cuantitativo, se logra observar cómo los estudiantes que tienen 2, 4 y 5 personas a cargo presentan un promedio más bajo en sus resultados, a comparación de las otras categorías donde el promedio se encuentra en el rango de 90 a 120.

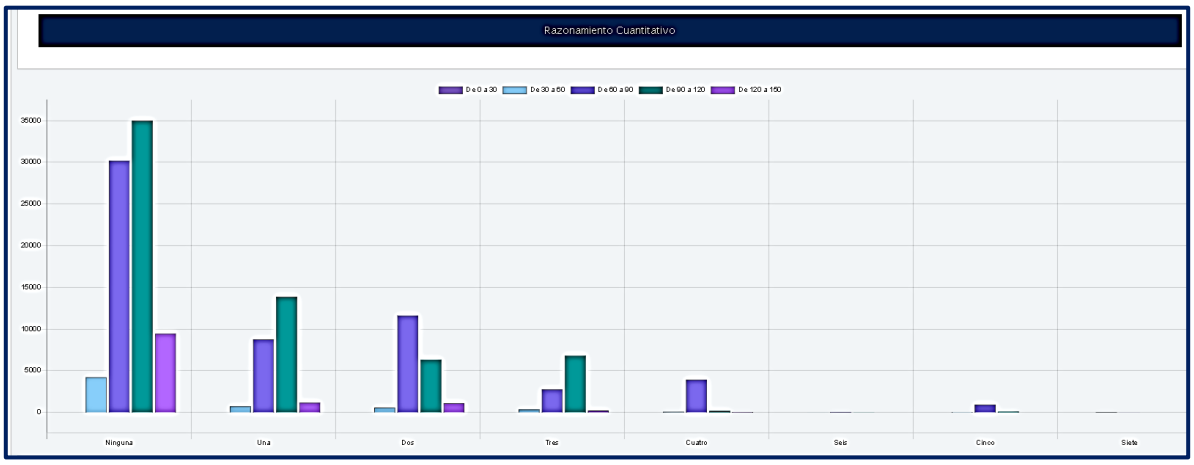


Figura 166. Razonamiento cuantitativo vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Los resultados en el núcleo de Comunicación escrita, presentan un comportamiento similar a los presentados en Lectura crítica.

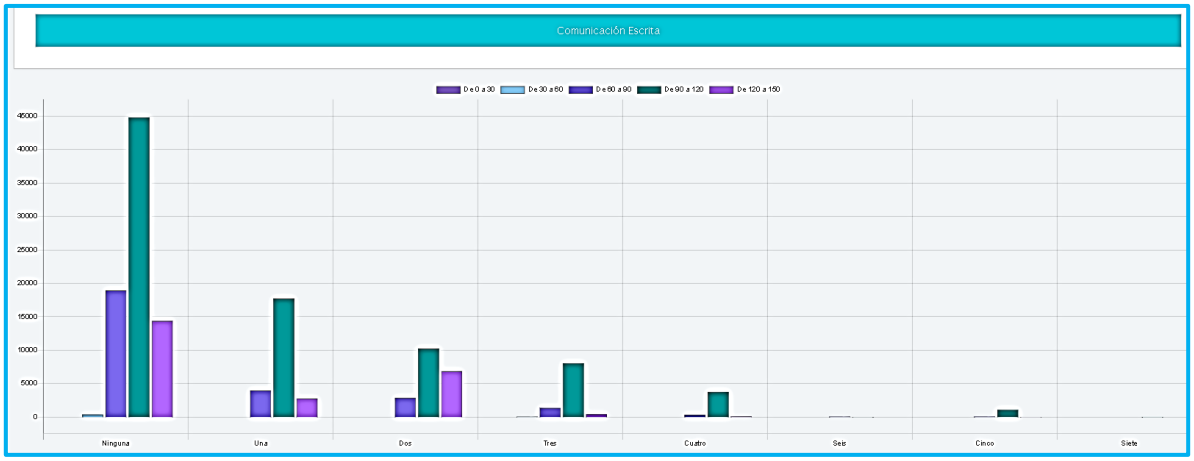


Figura 167. Comunicación escrita vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, no se evidencian grandes diferencias entre las diferentes categorías, sólo se destaca que los estudiantes que no tienen ninguna persona a cargo aumentan su representación en el rango de 60 a 90, con respecto a los demás.

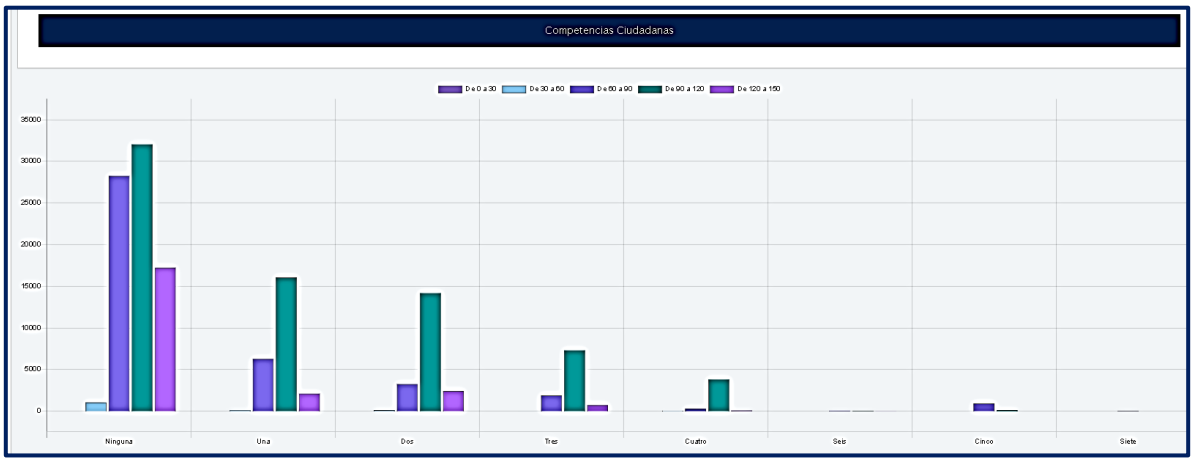


Figura 168. Competencias ciudadanas vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo de Competencias ciudadanas, se logra observar que los estudiantes con 1, 2 y 5 personas a cargo presentan un promedio más bajo en sus resultados, puesto que su rango más representativo es entre 60 y 90; mientras que, las otras categorías tienen como promedio el rango de 90 a 120.



Figura 169. Inglés vs Número personas a cargo, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Analizando el núcleo inglés, se evidencia resultados más representativos en los puntajes comprendidos entre 90 y 120 en la mayoría de las categorías, excluyendo a aquellos estudiantes que tienen 1, 2 y 5 personas a su cargo, quienes presentan como promedio puntajes más bajos.

h) Educación madre

Para este indicador se registra mayor presencia de estudiantes del género masculino en las categorías No aplica, No sabe, Postgrado y Educación profesional completa. En la categoría Secundaria bachillerato completa se evidencia como la cantidad de mujeres y de hombres es equitativa.

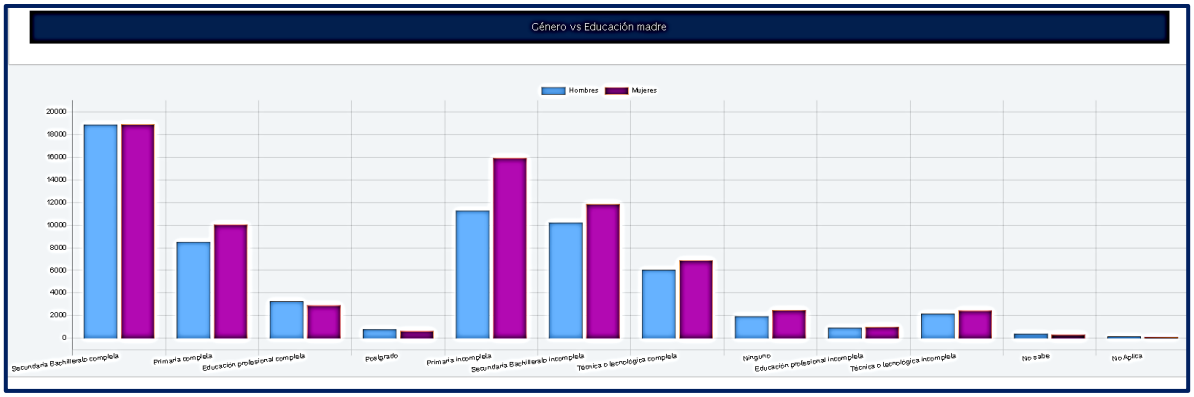


Figura 170. Género vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Lectura crítica presenta dentro de sus categorías, puntajes más altos en Postgrado, Educación profesional incompleta, Primaria completa y Secundaria incompleta dónde se encuentra con mayoría el rango más alto de la prueba. Mientras que, en Primaria incompleta el clúster más representativo es entre 60 y 90 siendo el de resultados más bajos.

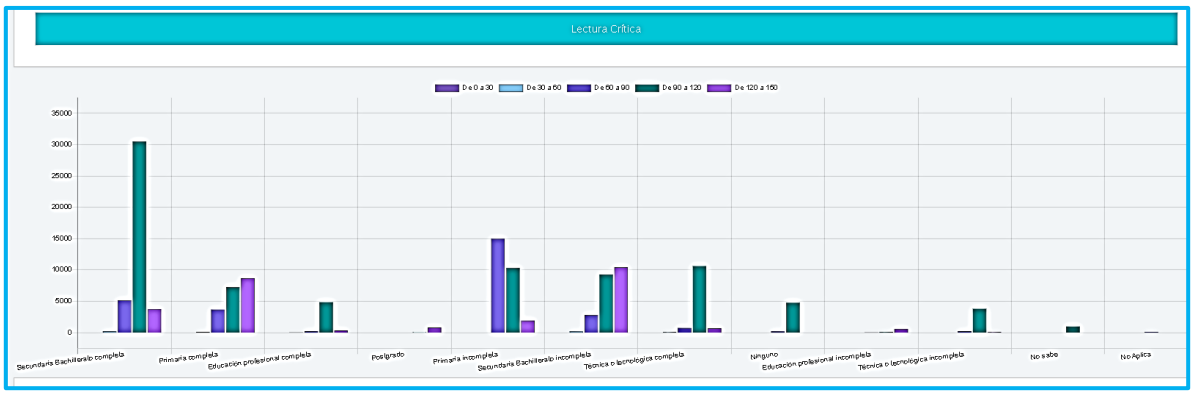


Figura 171. Lectura crítica vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Los resultados en Razonamiento cuantitativo, muestran que los estudiantes con madre en Primaria completa y Técnica completa obtienen puntajes más altos. Por otra parte, en las categorías de No sabe, Educación profesional completa y Técnica incompleta son quienes obtienen puntajes más bajos.

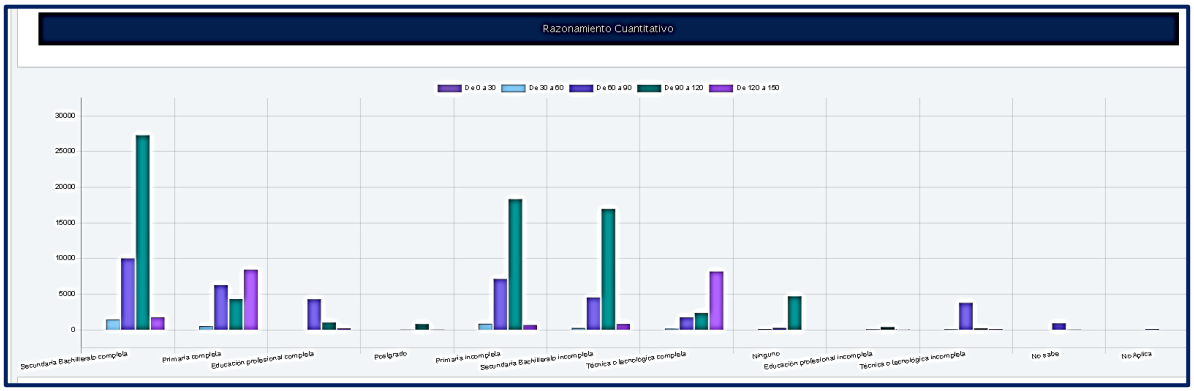


Figura 172. Razonamiento cuantitativo vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita todas las categorías tienen como clúster más representativo el comprendido entre 90 y 120, sólo se destaca positivamente la categoría Educación profesional incompleta que obtiene resultados más altos.

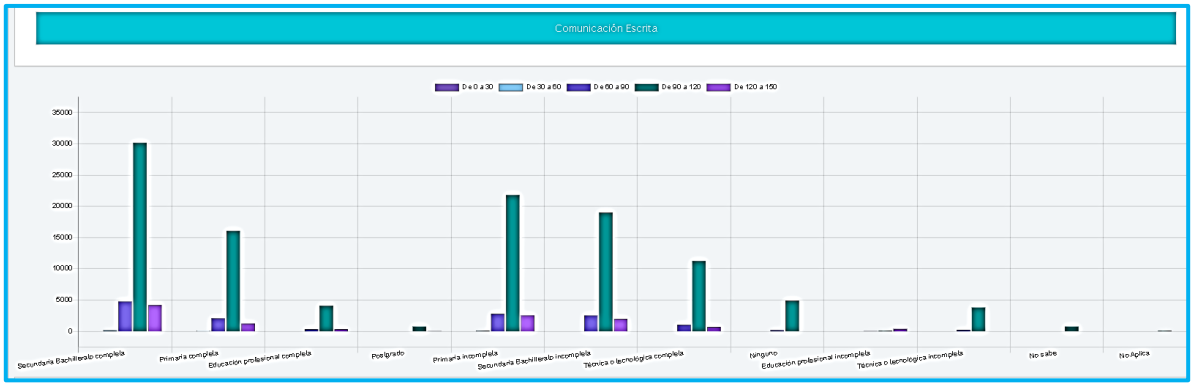


Figura 173. Comunicación escrita vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, se presentan puntajes más bajos que los núcleos analizados anteriormente puesto que, categorías como Secundaria completa e incompleta, Primaria incompleta y Técnica incompleta tienen en promedio resultados entre 60 y 90.

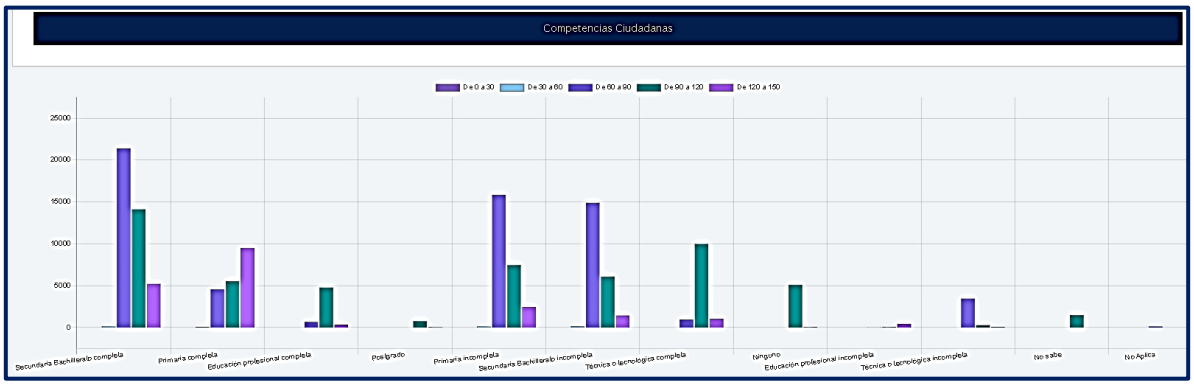


Figura 174. Competencias ciudadanas vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Por último, en inglés se observa que todas las categorías tienen como clúster más representativo el comprendido entre 90 y 120, sólo se destaca positivamente la categoría de Postgrado al obtener resultados más altos.

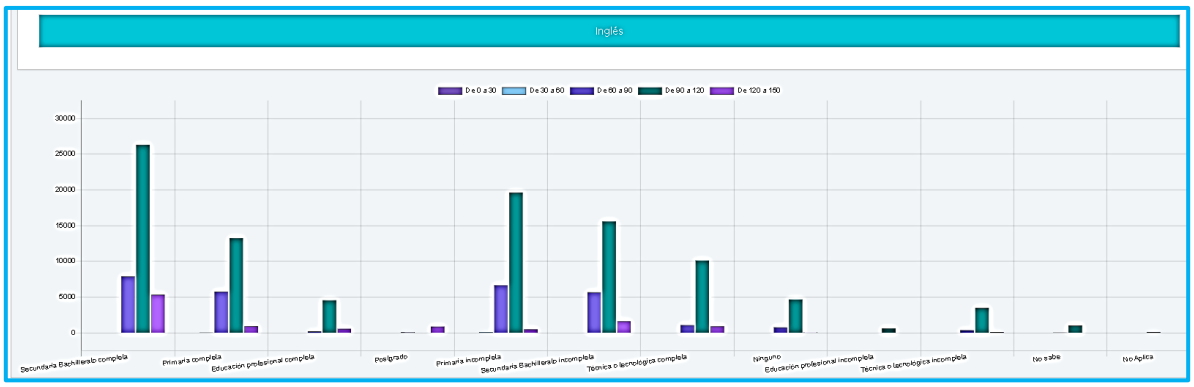


Figura 175. Inglés vs Educación madre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

i) Educación padre

En el indicador de Educación padre se muestra que, los estudiantes del género masculino son mayoría en las categorías Educación profesional completa e incompleta, Técnica completa, Postgrado y No aplica. En las demás categorías son mayoría las mujeres, pero en Primaria incompleta es donde se evidencia significativamente esa diferencia.

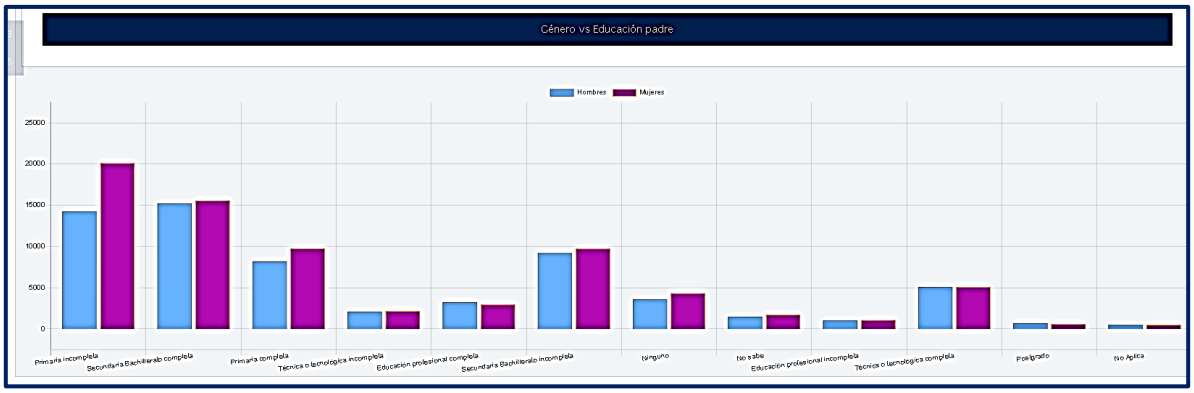


Figura 176. Género vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica las categorías presentan en promedio un puntaje comprendido entre 90 y 120, pero se evidencia diferencia positiva en Secundaria bachillerato incompleta y negativa en los estudiantes donde el padre no cuenta con ningún nivel de estudios.

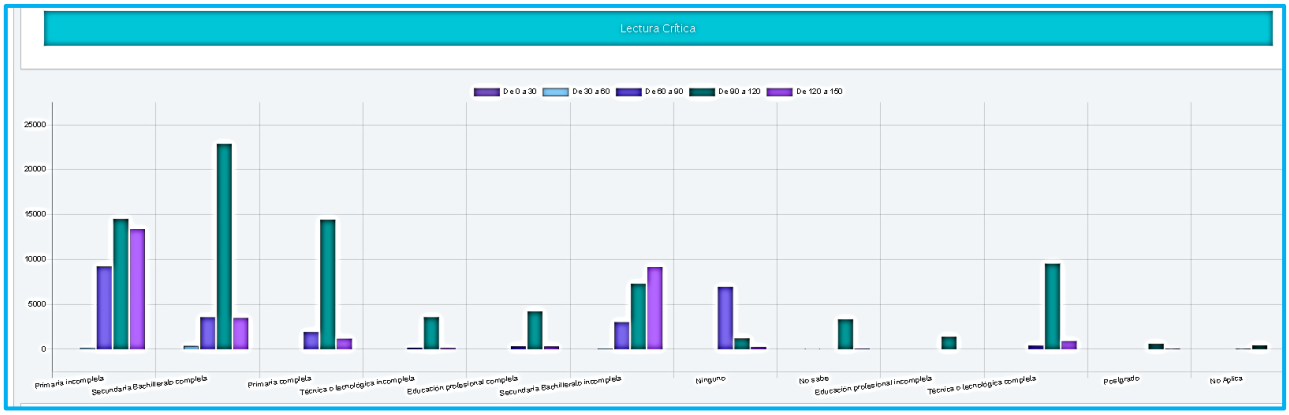


Figura 177. Lectura crítica vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, varias categorías presentan resultados más bajos que el promedio ubicado entre 90 y 120, estas son: Técnica incompleta, Ninguno y No sabe.

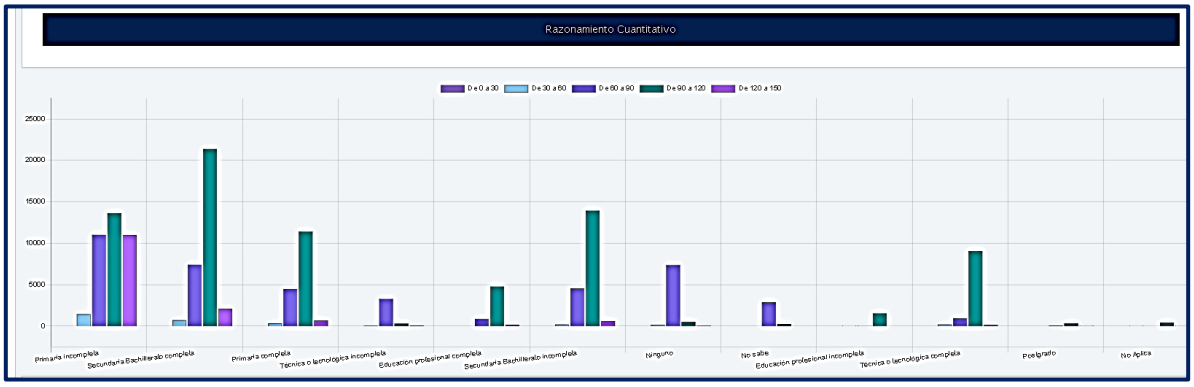


Figura 178. Razonamiento cuantitativo vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Comunicación escrita, se muestra que las categorías que se destacan por encima del promedio son Postgrado y Secundaria bachillerato incompleta. Mientras que, el caso contrario sucede para las categorías Ninguno, Primaria completa y Secundaria bachillerato completa.

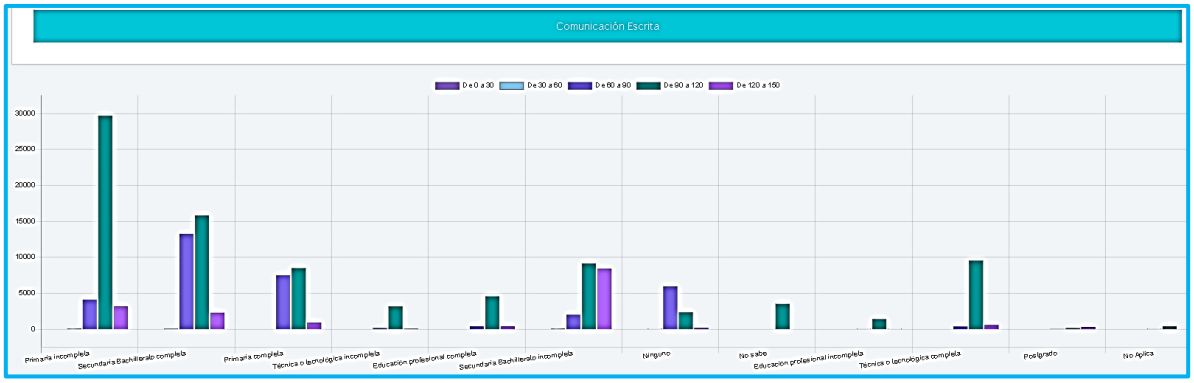


Figura 179. Comunicación escrita vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas, se destaca positivamente la categoría de Primaria incompleta con puntajes entre 120 y 150, por el contrario, los niveles de estudio Técnico incompleto y Secundaria incompleta con resultados en el rango de 60 a 90.

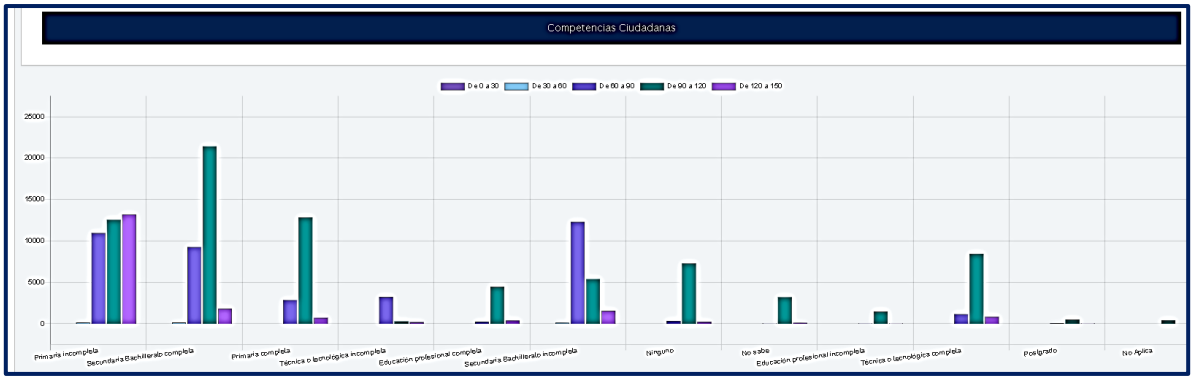


Figura 180. Competencias ciudadanas vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se observa que los estudiantes con padre con Postgrado obtienen mejores resultados; mientras que, los puntajes de las categorías Secundaria bachillerato incompleta, Ninguno y Primaria incompleta son más bajos, en esta última se alcanza a presentar un clúster para el rango de 0 a 30.

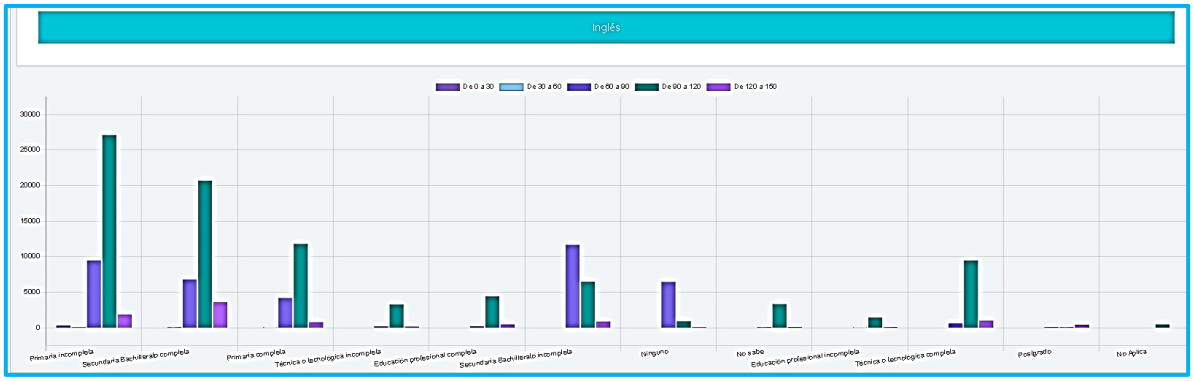


Figura 181. Inglés vs Educación padre, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

j) Dedicación de lectura diaria

Para este indicador se evidencia mayor cantidad de estudiantes del género femenino, salvo en la categoría No leo por entretenimiento. Cabe destacar que la mayoría de estudiantes de ambos géneros tienden a leer menos de una hora.

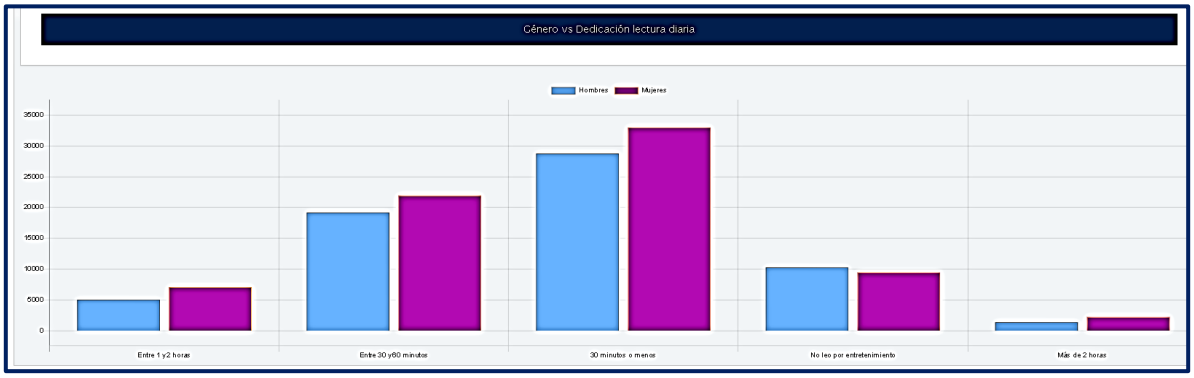


Figura 182. Género vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo Lectura crítica, el análisis muestra un comportamiento similar para todas las categorías, donde el promedio de puntajes abarca el rango de 90 a 120, seguido del rango entre 60 a 90 y por último el rango entre 120 y 150.

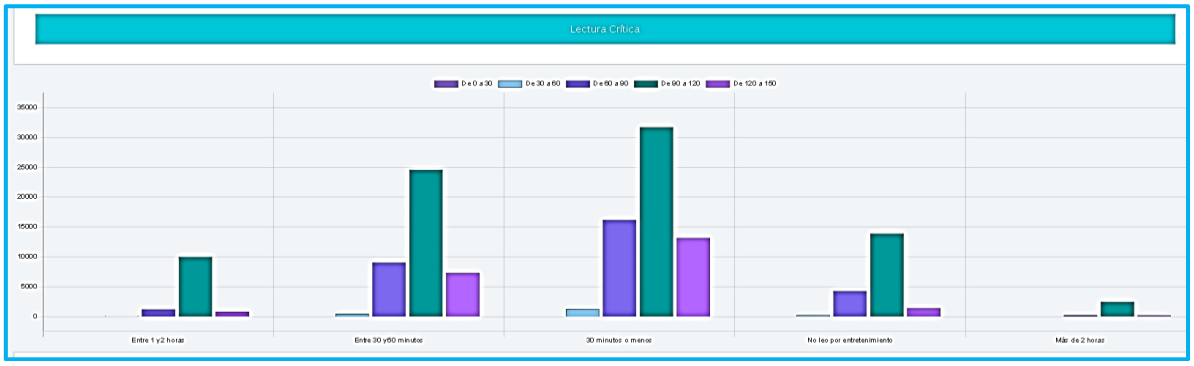


Figura 183. Lectura crítica vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

El análisis realizado a Razonamiento cuantitativo evidencia que, en las categorías Entre 1 y 2 horas, 30 minutos o menos y Más de 2 horas los estudiantes tienen puntajes con un promedio entre 60 a 90. Las otras categorías centran sus puntajes en el rango de 90 a 120.

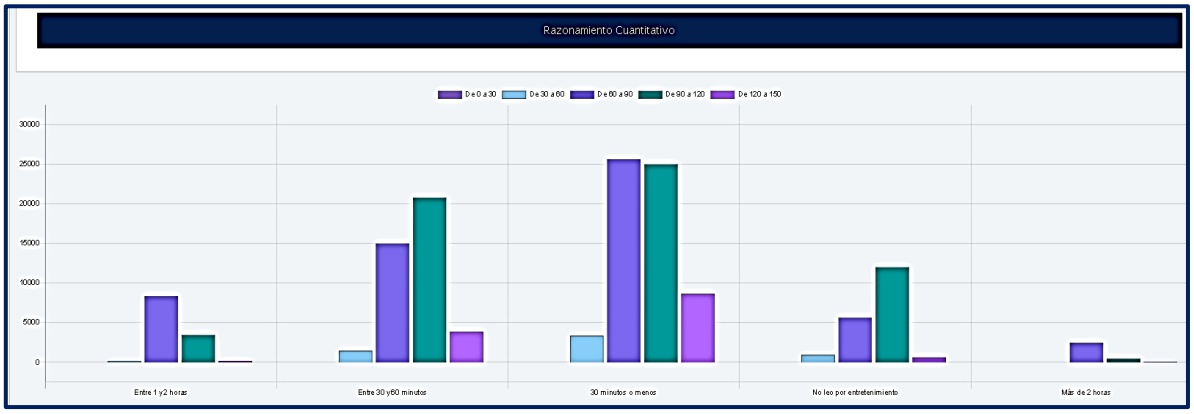


Figura 184. Razonamiento cuantitativo vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita se observa un comportamiento similar al presentado en el núcleo de Lectura crítica.

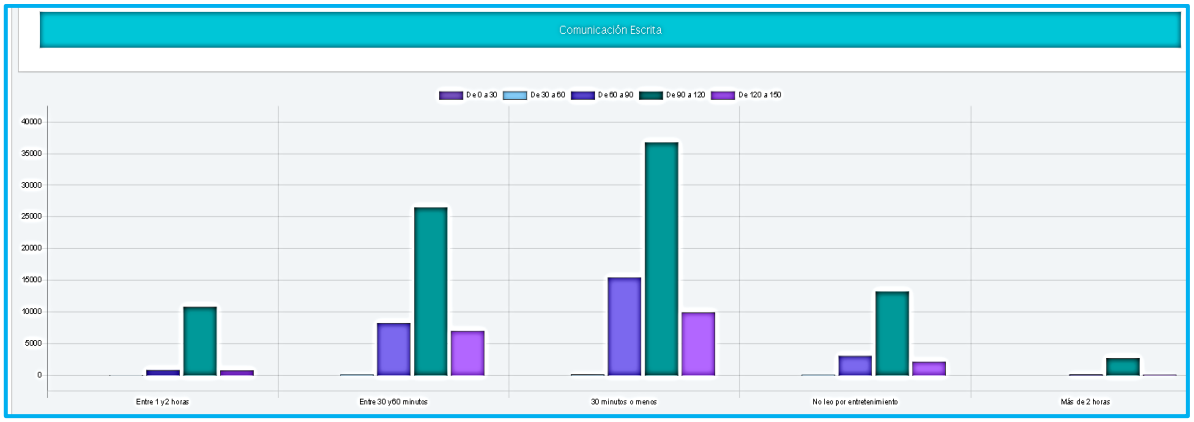


Figura 185. Comunicación escrita vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, las categorías de Entre 1 y 2 horas y No leo por entrenamiento presentan puntajes más bajos que las demás categorías donde el clúster más representativo es el comprendido entre 90 y 120.

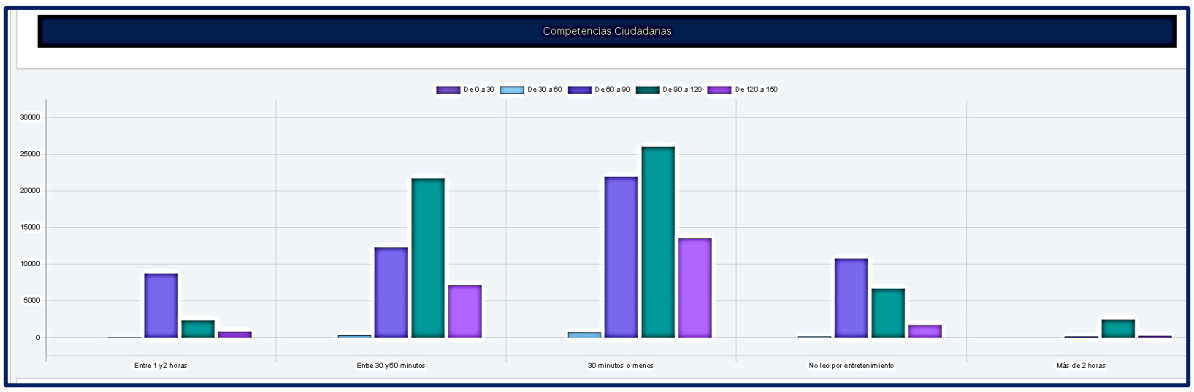


Figura 186. Competencias ciudadanas vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo inglés, el cluster más representativo en todas las categorías es el comprendido entre 90 y 120. Vale la pena destacar que en la categoria 30 minutos o menos, se alcanza a presentar una cantidad de estudiantes en el rango de 0 a 30.

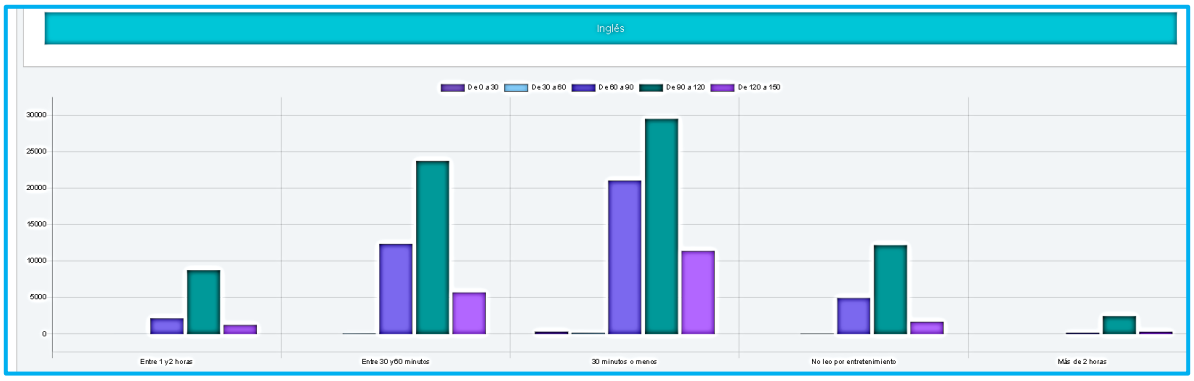


Figura 187. Inglés vs Dedicación lectura diaria, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

k) Área de la carrera

En este indicador se presenta mayor cantidad de hombres en las áreas de: Técnico y tecnológico en ingeniería, industria y minas, Técnico y tecnológico en TIC, Tecnológico en ciencias agropecuarias, Recreación y deportes y Tecnológico en militar y policial. Mientras que, en los técnicos y tecnológicos de administración y turismo es evidente la mayoría de mujeres, al igual que sucede en el tecnológico de salud y normales superiores.

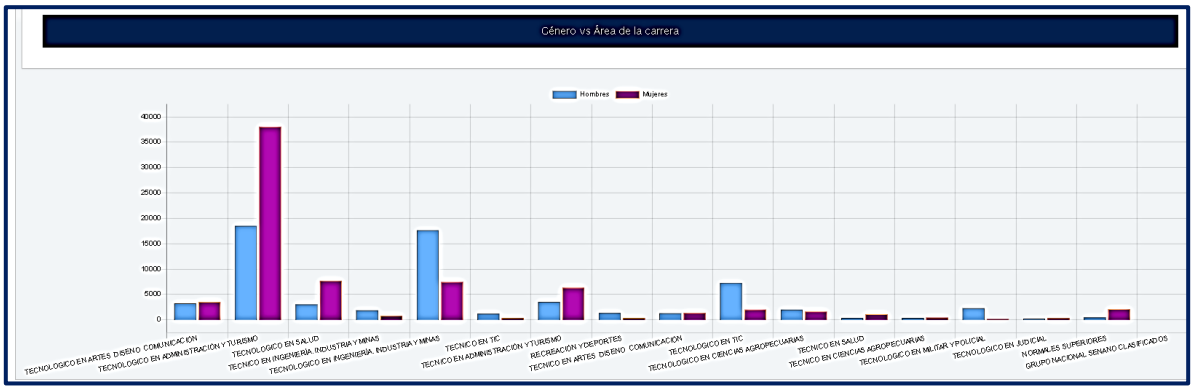


Figura 188. Género vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

El clúster promedio en Lectura crítica, es el comprendido entre 90 y 120. Sin embargo, en Recreación y deportes, Tecnológico en TIC y Tecnológico en ciencias agropecuarias obtienen resultados más bajos. Por otra parte, los estudiantes de Tecnológico en judicial y Normales superiores son quienes mejores puntajes obtienen.

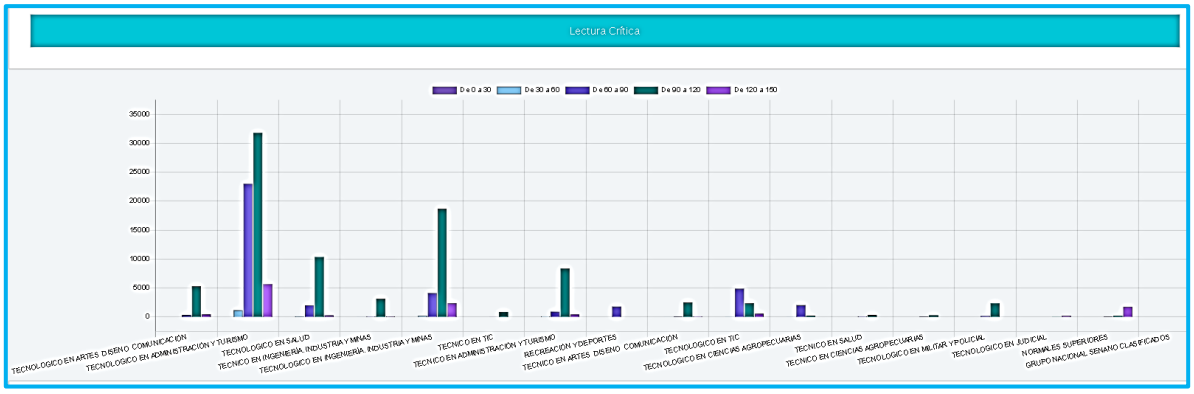


Figura 189. Lectura crítica vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el nucleo de Razonamiento cuantitativo, los estudiantes de Tecnológico en ingeniería, industria y minas, son quienes mejores puntajes obtienen con un clúster representativo para el rango comprendido entre 120 y 150. Mientras que, los estudiantes de Recreación y deportes, Tecnológico en militar y policial y Normales superiores obtienen resultados más bajos que el promedio.

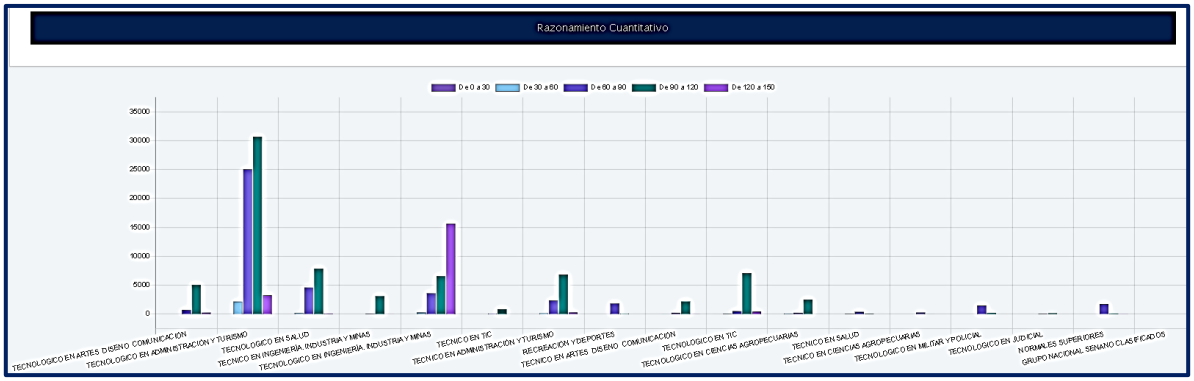


Figura 190. Razonamiento cuantitativo vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Comunicación escrita se destaca positivamente los estudiantes del Técnico en artes, diseño y comunicación, quienes obtienen resultados dentro del rango de 120 a 150. Las demás categorías obtienen puntajes promedio entre 90 y 120.

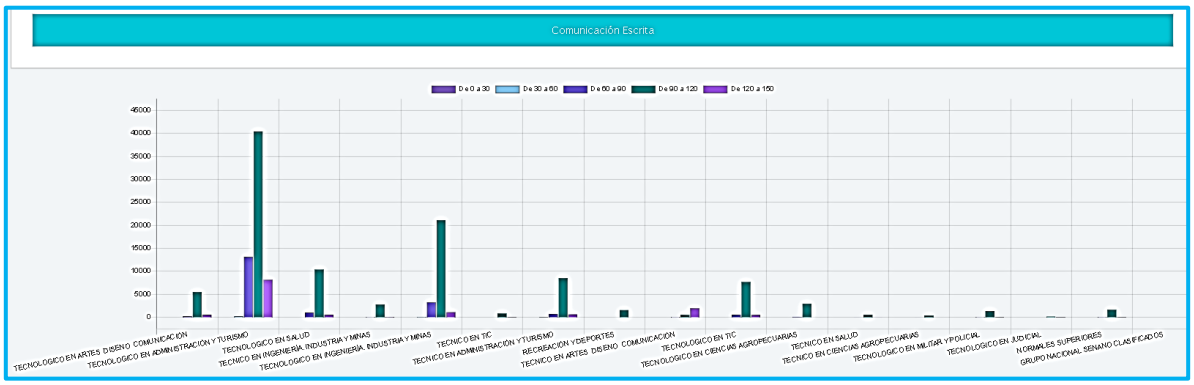


Figura 191. Comunicación escrita vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, los estudiantes de Tecnológico en ingeniería, industria y minas, Técnico en TIC y Tecnológico en judicial son quienes se destacan positivamente, obteniendo puntajes comprendidos dentro del rango de 120 a 150. Por otra parte, en las áreas de Tecnológico y técnico en administración y turismo, Tecnológico en salud, Tecnológico en TIC, Tecnológico en ciencias agropecuarias, Tecnológico en militar y policial y Recreación y deportes presentan resultados dentro del rango de 60 a 90.

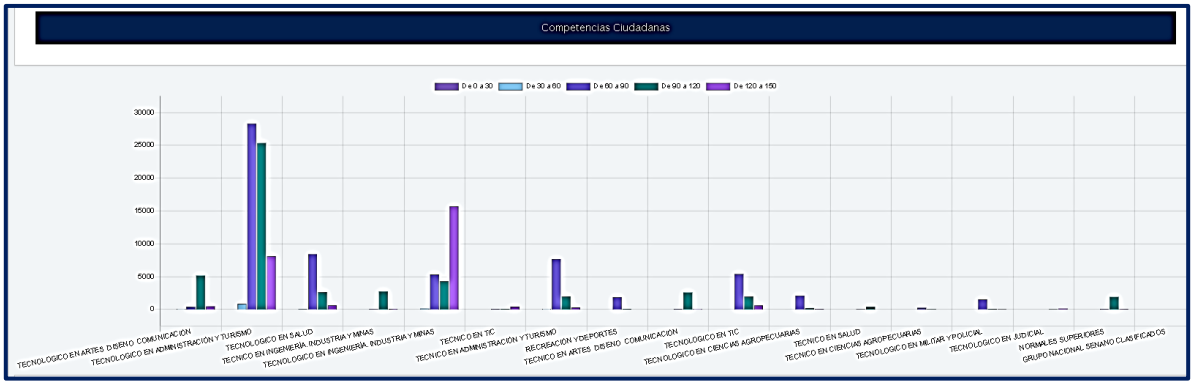


Figura 192. Competencias ciudadanas vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Por último, en el núcleo de inglés, los estudiantes del Técnico en artes, diseño y comunicación son quienes mejores puntajes obtienen. Mientras que, los que presentan resultados más bajos son: Tecnológico y técnico en administración y turismo y Tecnológico en ciencias agropecuarias.

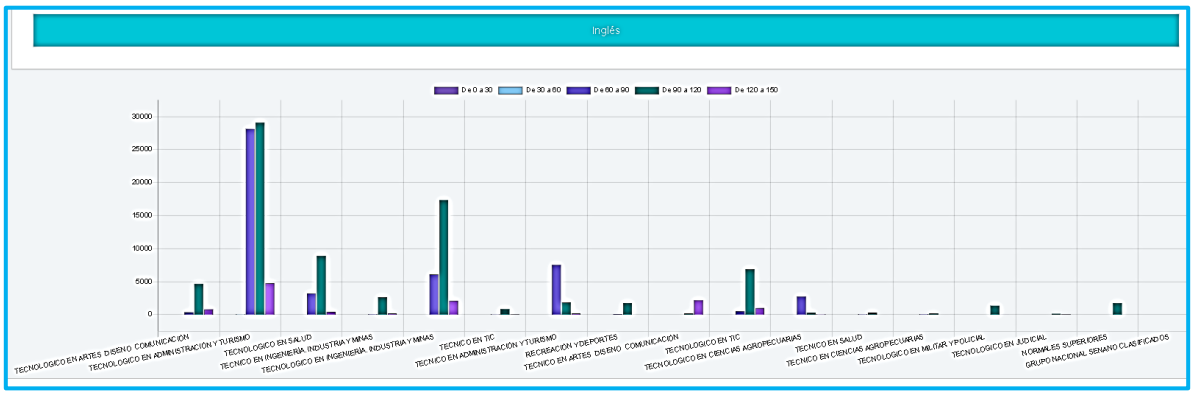


Figura 193. Inglés vs Área de la carrera, Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

4.3. PRUEBAS SABER PRO

a) Valor matricula de universidad

Los resultados del análisis descriptivo realizado al indicador Valor de matrícula de universidad, reflejan que la mayoría de estudiantes en Colombia, pagan en promedio una cifra de dinero comprendida entre 1 millón y 4 millones de pesos. Además, en cada una de las categorías las mujeres son mayoría.

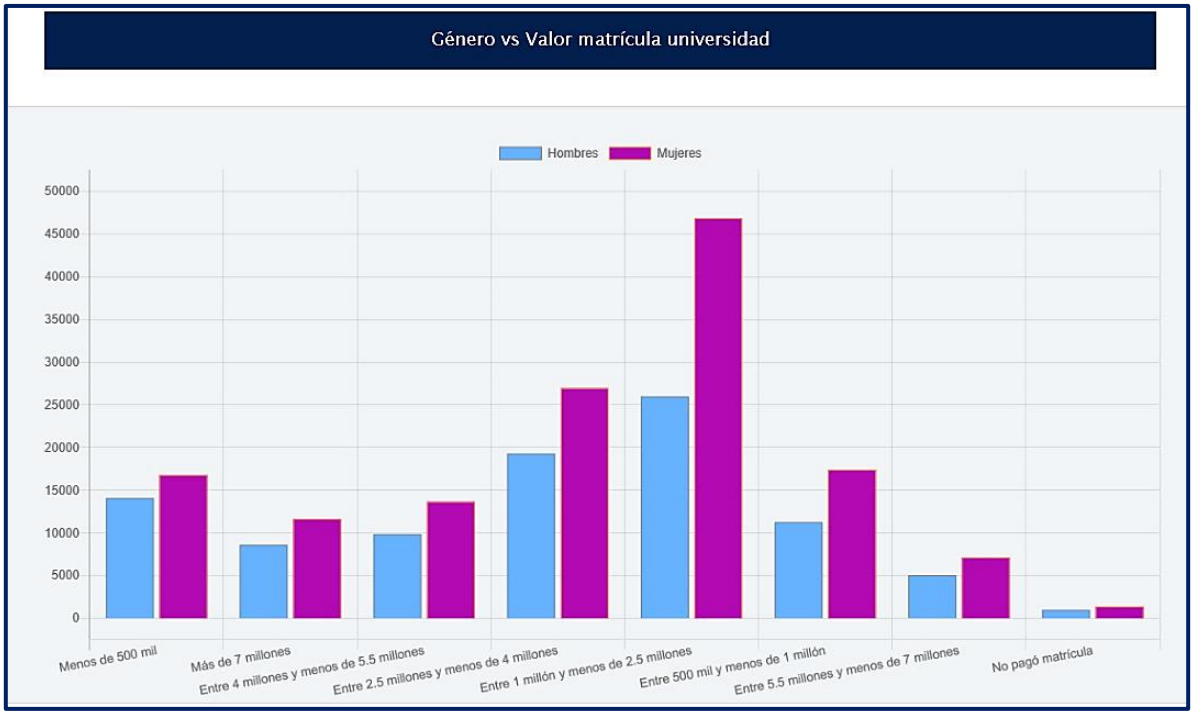


Figura 194. Género vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, se observa que los estudiantes que pagan Menos de 500 mil, Más de 7 millones y Entre 4 y 5.5 millones son quienes mejores resultados obtienes estando ubicados en un rango promedio de 150 a 180. Mientras que, la categoría Entre 500 mil y menos de un millón es la que presenta puntajes más bajos, con un clúster mayoritario en el rango de 90 a 120.

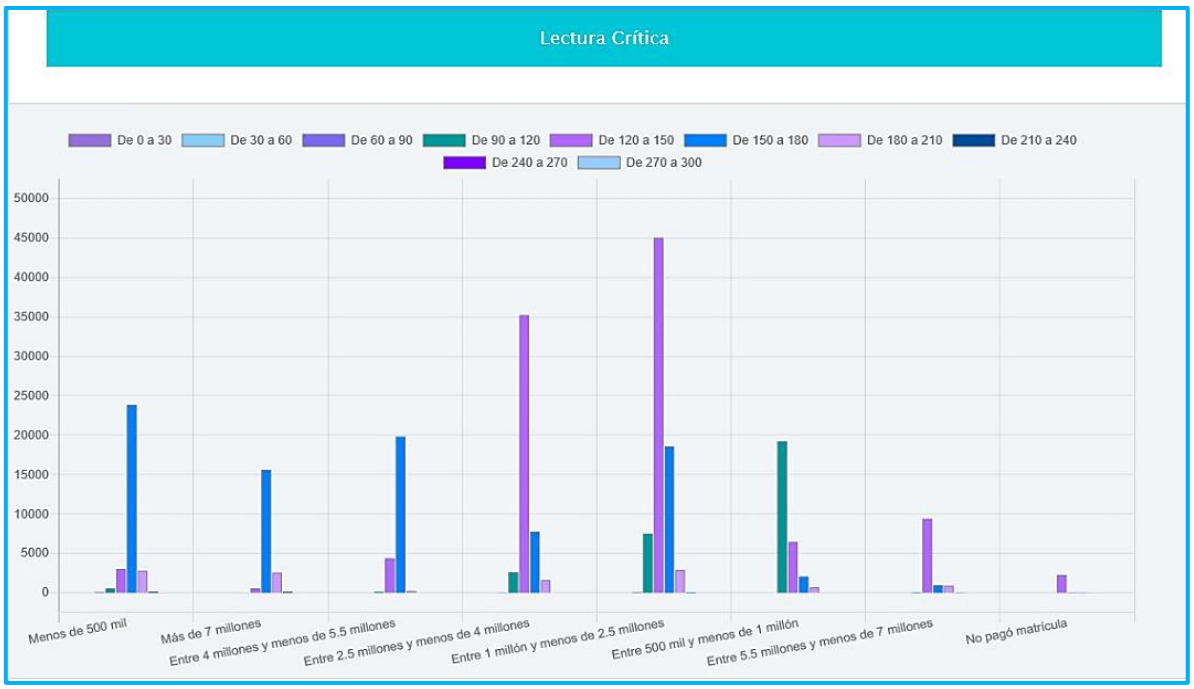


Figura 195. Lectura crítica vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, se presentan puntajes más altos que el núcleo anterior puesto que el clúster más representativo es entre 150 y 180, incluso en la categoría de estudiantes que no pagaron matrícula se evidencia mejores resultados.

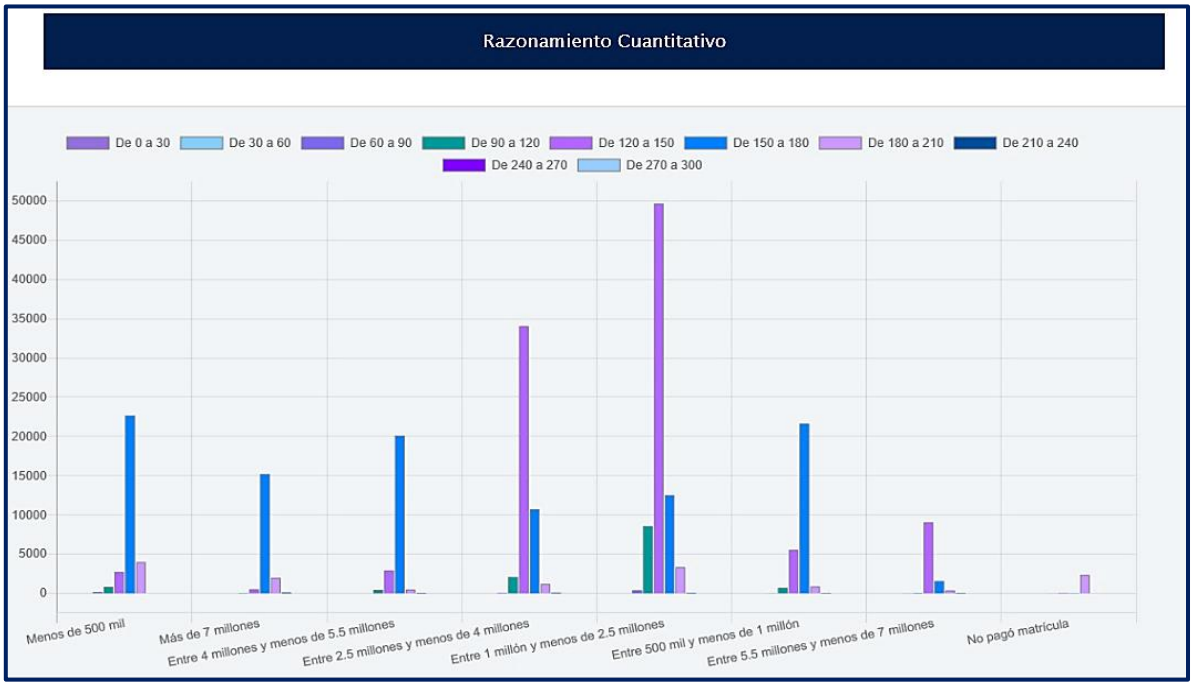


Figura 196. Razonamiento cuantitativo vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, se encuentra que el clúster promedio es el comprendido entre 120 y 150 puntos. Asimismo, vale la pena destacar que, las categorías de pago menor a 500 mil y entre 4 millones y menos de 5.5 millones obtienen resultados más altos.

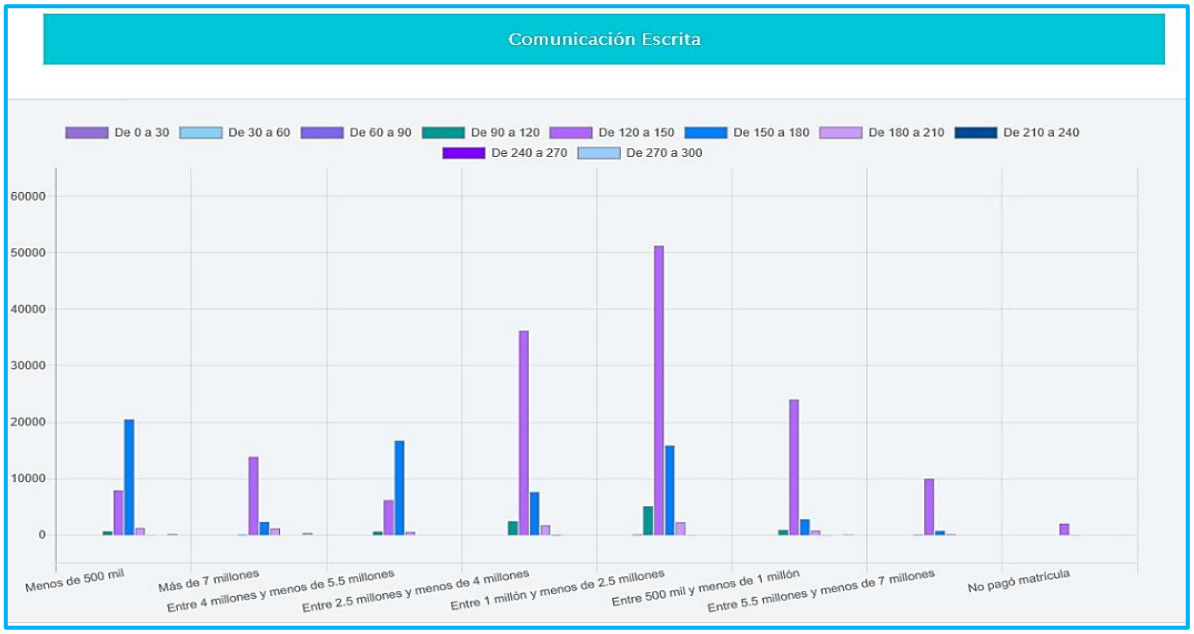


Figura 197. Comunicación escrita vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas, la mayoría de categorías obtienen resultados comprendidos dentro del rango de 150 y 180. Sin embargo, la categoría Entre 2.5 millones y menos de 4 millones presenta puntajes en un rango inferior, y con resultados aún más bajos se encuentran los estudiantes que pagan entre 500 mil y menos de 2.5 millones.

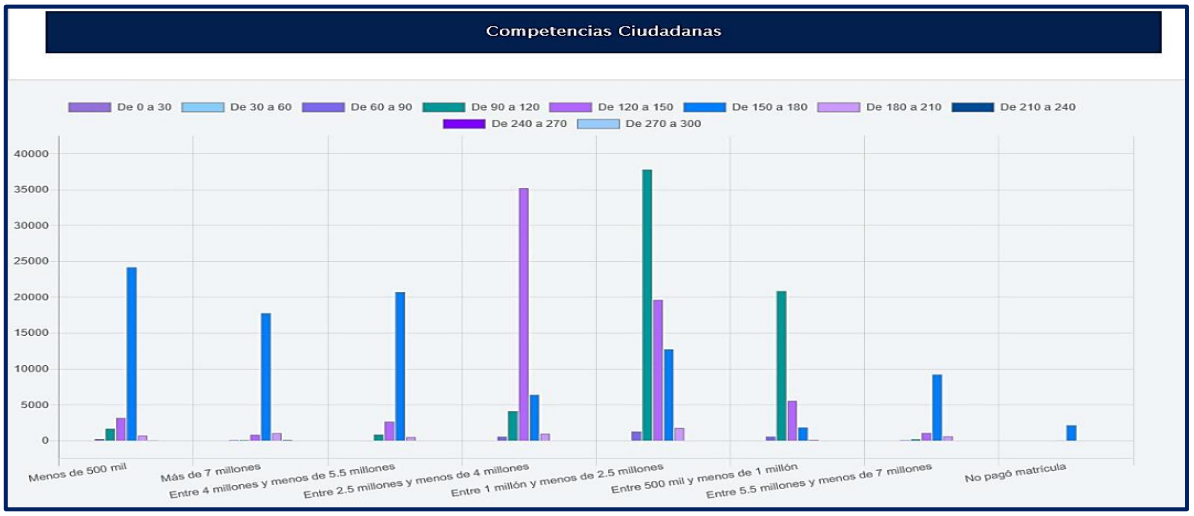


Figura 198. Competencias ciudadanas vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Por último, en el núcleo de inglés, se muestra que los resultados son mejores que los demás núcleos, al ser el clúster más significativo el comprendido entre 150 y 180 y además que la categoría Más de 7 millones presenta una mayoría de puntajes en el rango de 180 a 210 y alcanza a presentar en el rango de 210 a 240.

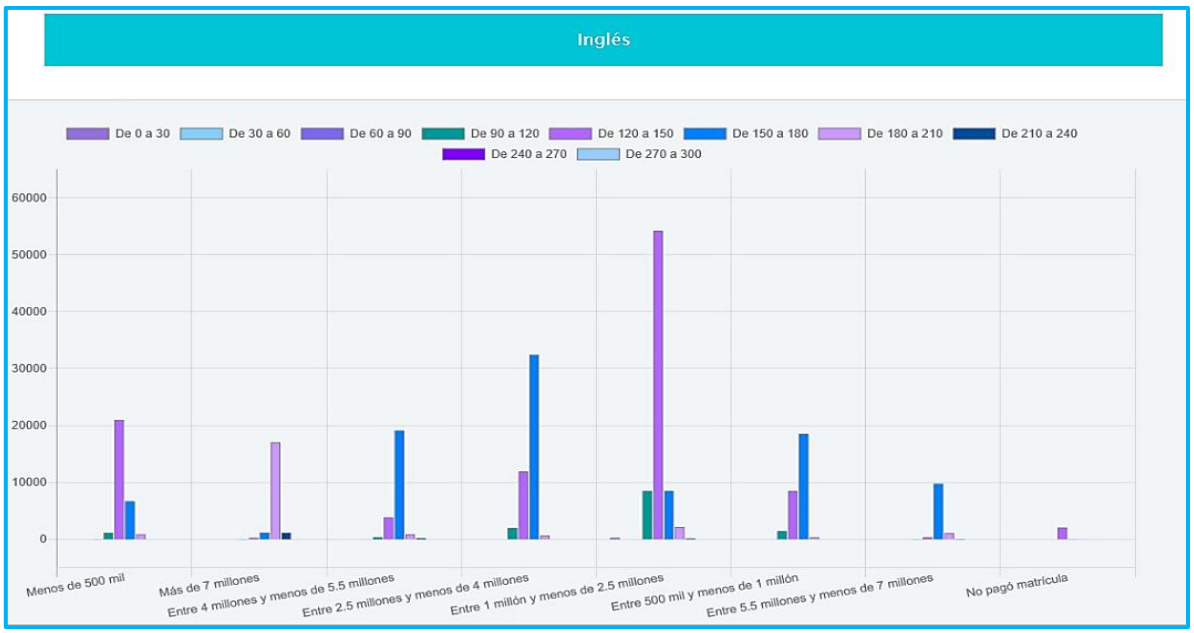


Figura 199. Inglés vs Valor matrícula universidad, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

b) Pago matrícula a crédito

En este indicador, se observa gracias al análisis que en ambas categorías el género femenino constituye la mayoría y asimismo que los estudiantes en términos generales no pagan la matrícula a crédito.

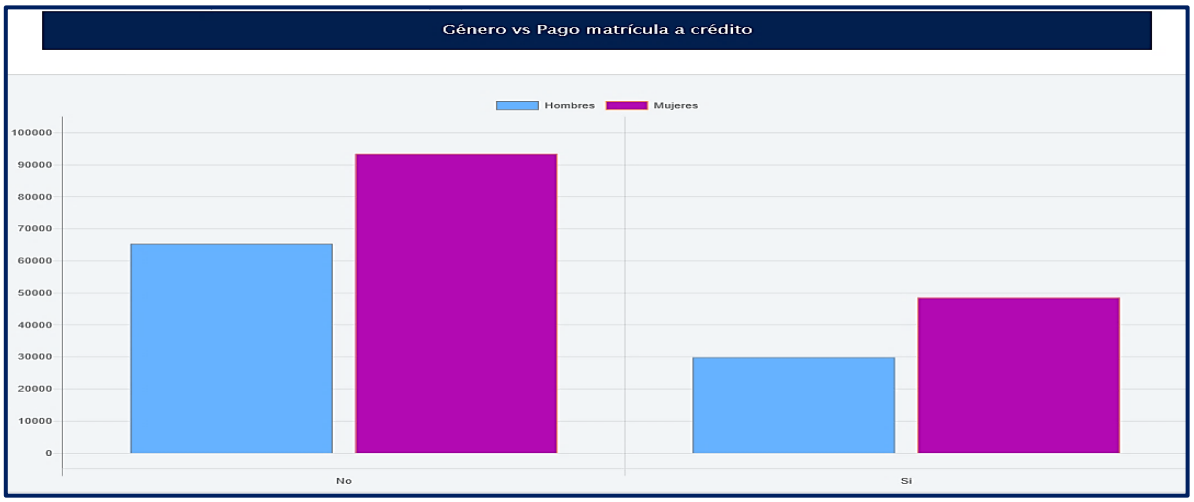


Figura 200. Género vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, ambas categorías presentan resultados muy similares donde los 2 clústeres más representativos son en su respectivo orden entre 120 a 150 y de 150 a 180 puntos.

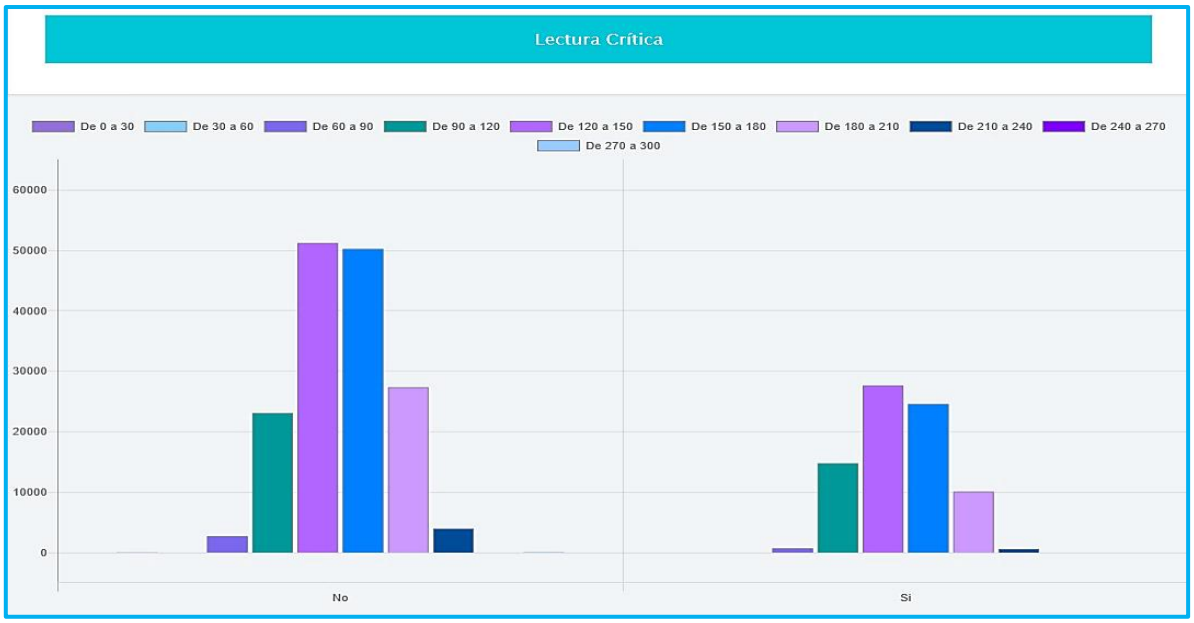


Figura 201. Lectura crítica vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Razonamiento cuantitativo, ambas categorías presentan resultados muy similares donde los 2 clústeres más representativos son en su respectivo orden entre 120 a 150 y de 150 a 180 puntos. Además, los estudiantes que no pagan su matrícula a crédito alcanzan a obtener puntajes en el rango de calificación más alto.

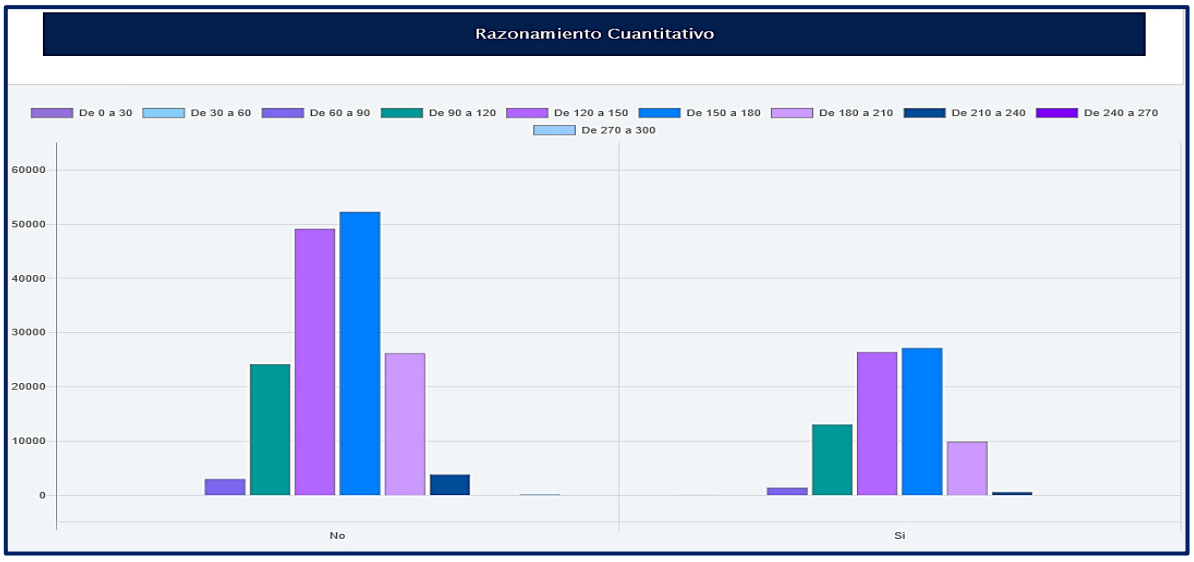


Figura 202. Razonamiento cuantitativo vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Comunicación escrita, ambas categorías presentan resultados muy similares donde los 2 clústeres más representativos son en su respectivo orden entre 120 a 150 y de 150 a 180 puntos. Además, los estudiantes alcanzan a obtener puntajes en el rango de calificación más alto, lo cual es destacable.

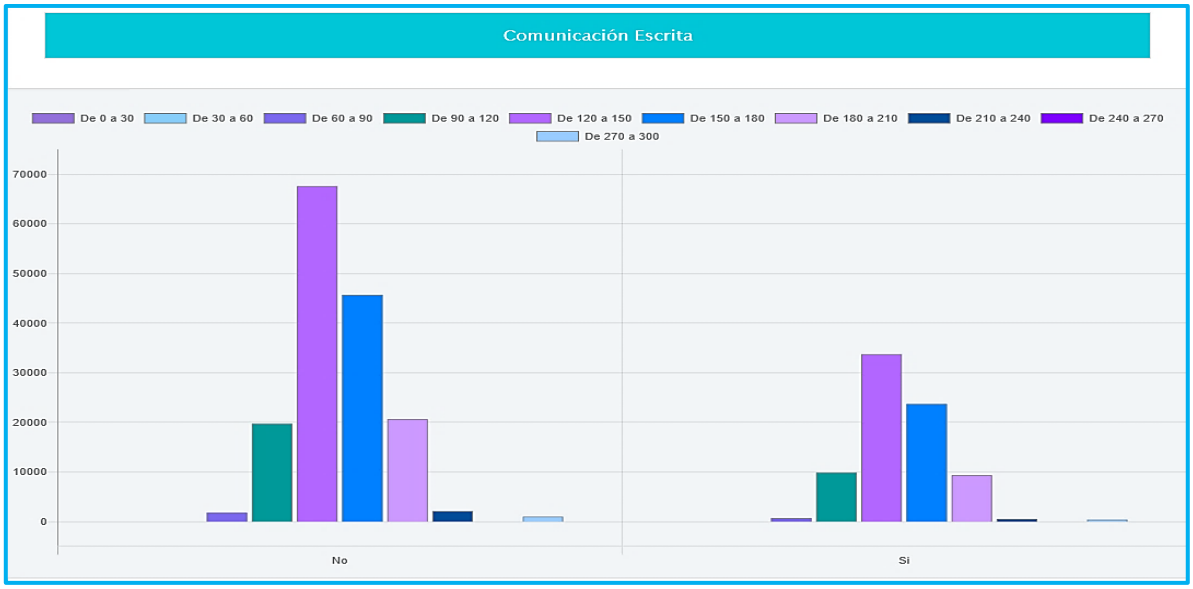


Figura 203. Comunicación escrita vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, se evidencia que los estudiantes que no pagan su matrícula a crédito alcanzan a obtener puntajes en el rango de calificación comprendido entre 60 a 90, siendo estos resultados más bajos que en los núcleos anteriores.

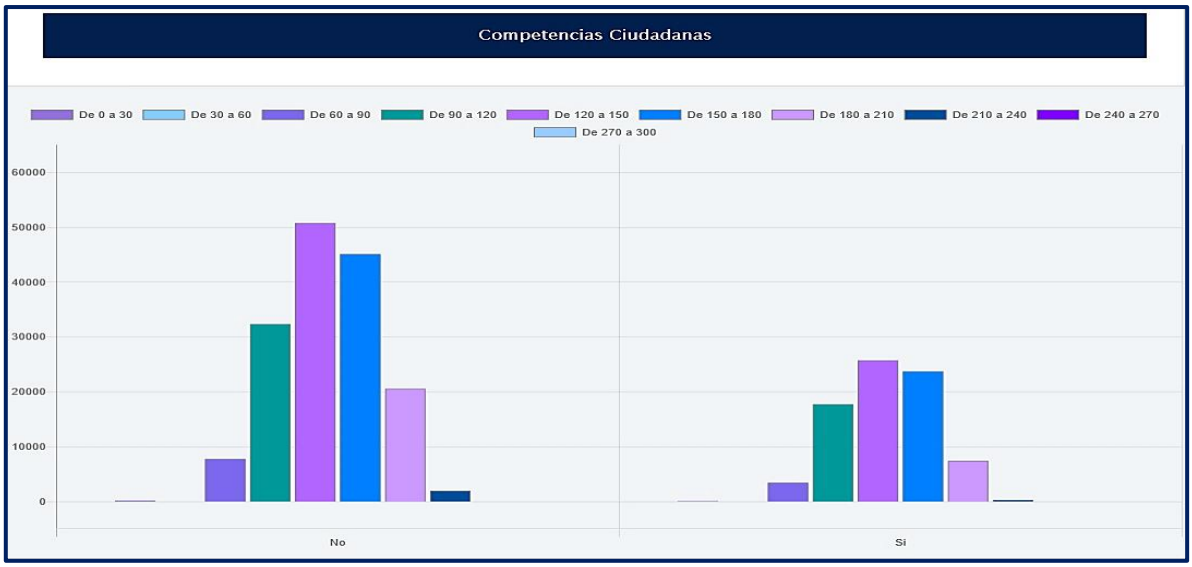


Figura 204. Competencias ciudadanas vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando el análisis de este indicador, se encuentra que, en inglés, los estudiantes de ambas categorías alcanzan a obtener puntajes en el rango de calificación más alto, pero asimismo en el más bajo y su clúster más representativo es el comprendido entre 120 a 150.

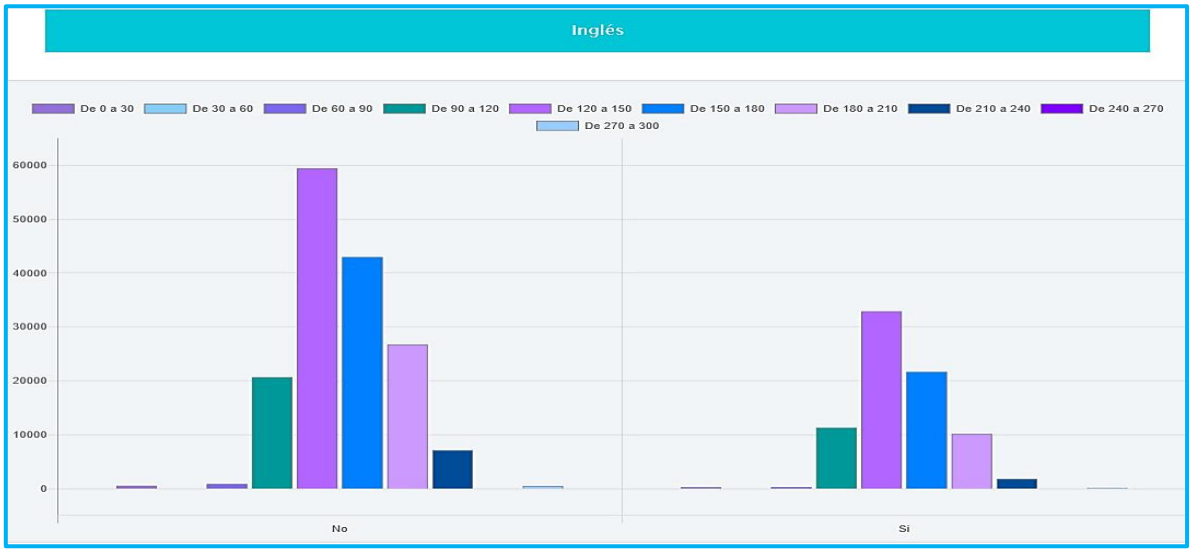


Figura 205. Inglés vs Pago matrícula a crédito, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

c) Estrato de vivienda

Gracias al análisis realizado con este indicador se observa que la mayoría de estudiantes se encuentran ubicados en los estratos 2 y 3 del país. Además, en las diferentes categorías las mujeres son mayoría a excepción de los estudiantes categorizados como Sin estrato.

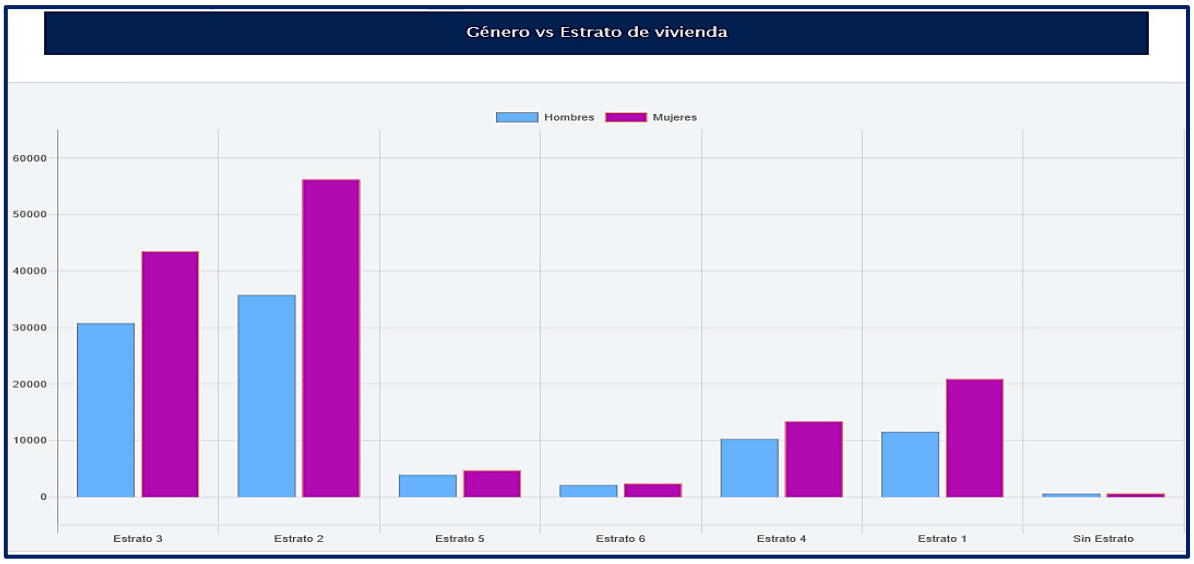


Figura 206. Género vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, para los estratos 5 y 6 el clúster más representativo es el rango entre 180 y 210. Mientras que, los estratos 1, 2 y 3 tienen un promedio entre los puntajes 120 y 150. Los estudiantes de estrato 4 tienen puntajes que abarcan el rango de 150 a 180.

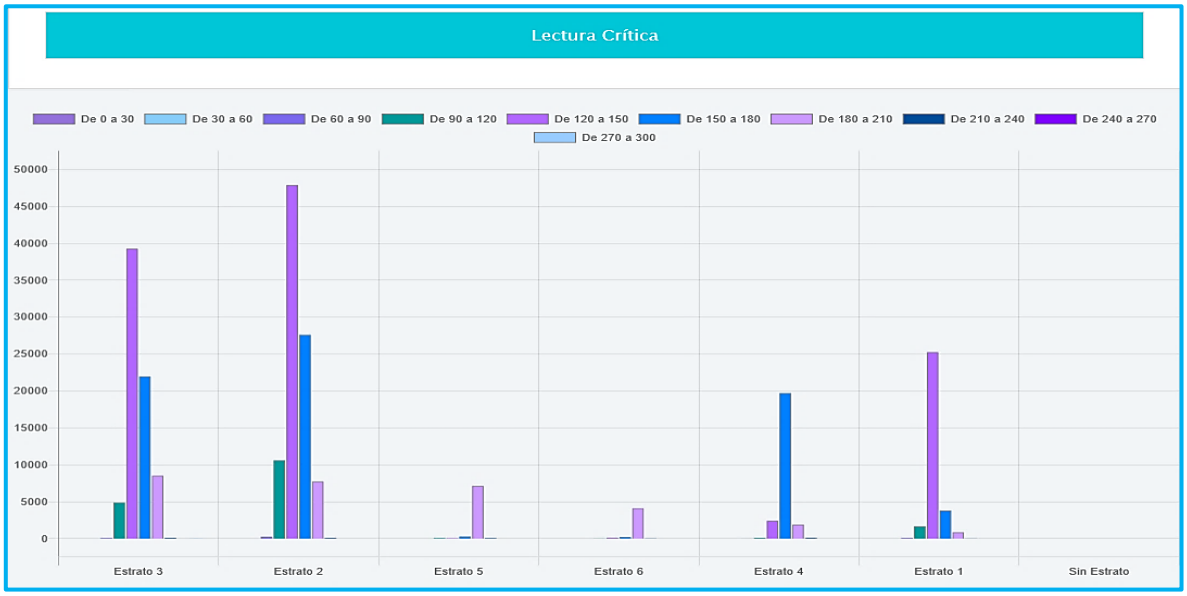


Figura 207. Lectura crítica vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el análisis del núcleo Razonamiento cuantitativo se evidencia que el estrato 6 es la categoria que mejores puntajes presenta, teniendo como clúster más representativo el rango de 180 a 210. Caso contrario sucede con los estratos 1, 2 y 3, donde el promedio de los puntajes se encuentra entre 120 a 150.

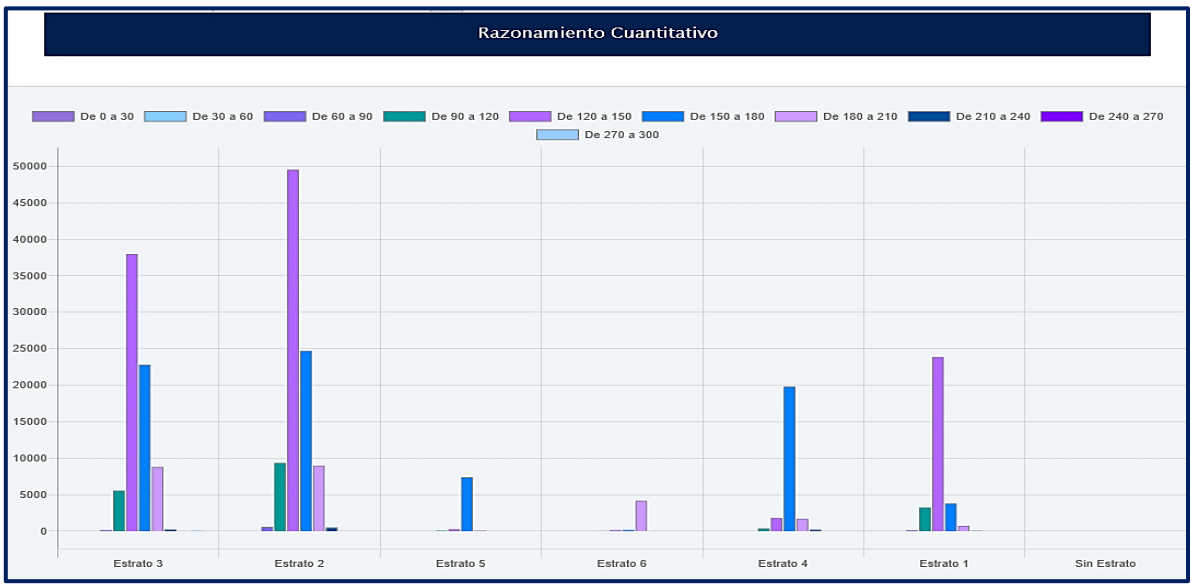


Figura 208. Razonamiento cuantitativo vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, la mayoría de las categorías presentan un comportamiento similar, teniendo como puntaje más representativo el rango de 120 a 150, seguido del clúster de 150 a 180. El estrato 4 por su parte, presenta como promedio este último rango mencionado.

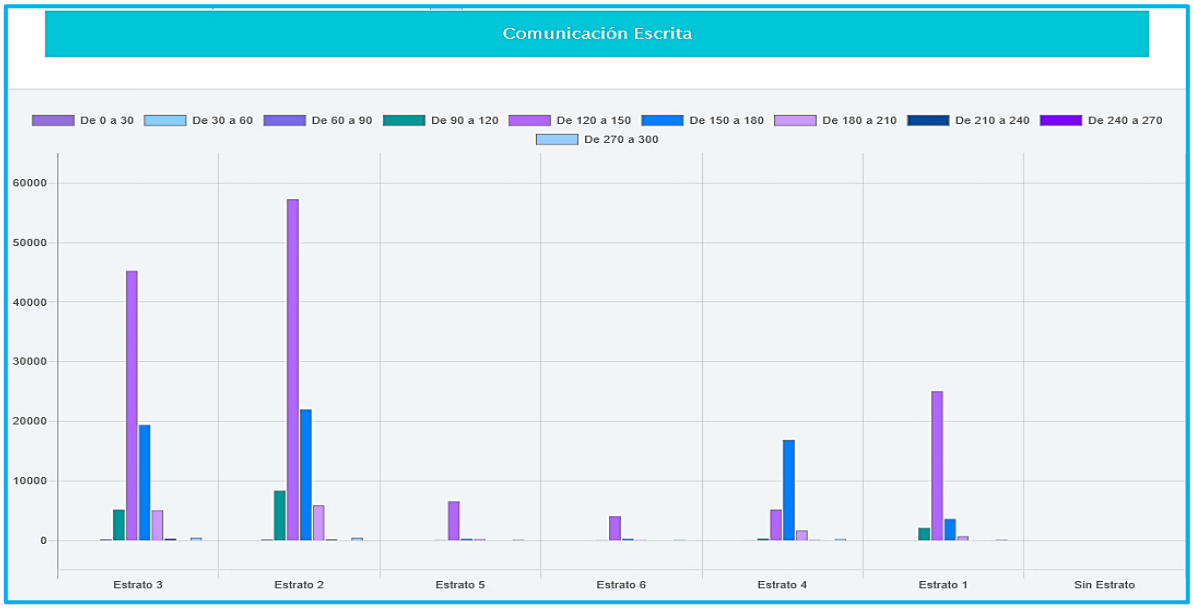


Figura 209. Comunicación escrita vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Los estratos 3, 4 y 5 son las categorías que mejores puntajes presentan en el núcleo Competencias ciudadanas, teniendo como clúster más representativo el rango entre 150 a 180. Los otros estratos presentan como promedio el rango de 120 a 150.

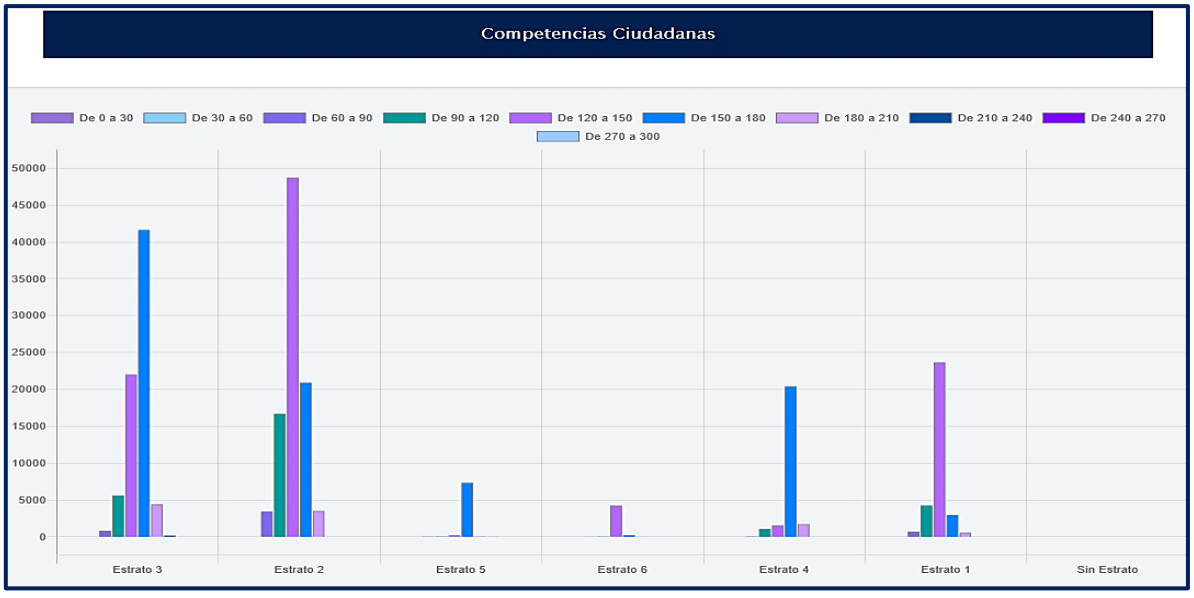


Figura 210. Competencias ciudadanas vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se observa que los estudiantes de estratos 1 y 3 presentan el promedio de puntajes más bajo, teniendo como rango más representativo de 120 a 150. Estrato 2 por su parte, a pesar de tener como clúster representativo el rango entre 150 a 180, presenta estudiantes que obtuvieron puntajes entre 0 y 30. El estrato 6 tiene como promedio el rango de calificación más alto, el rango 180 a 210 es el clúster más común en el estrato 5 y en estrato 4 los estudiantes obtienen puntajes con un promedio entre 150 y 180.

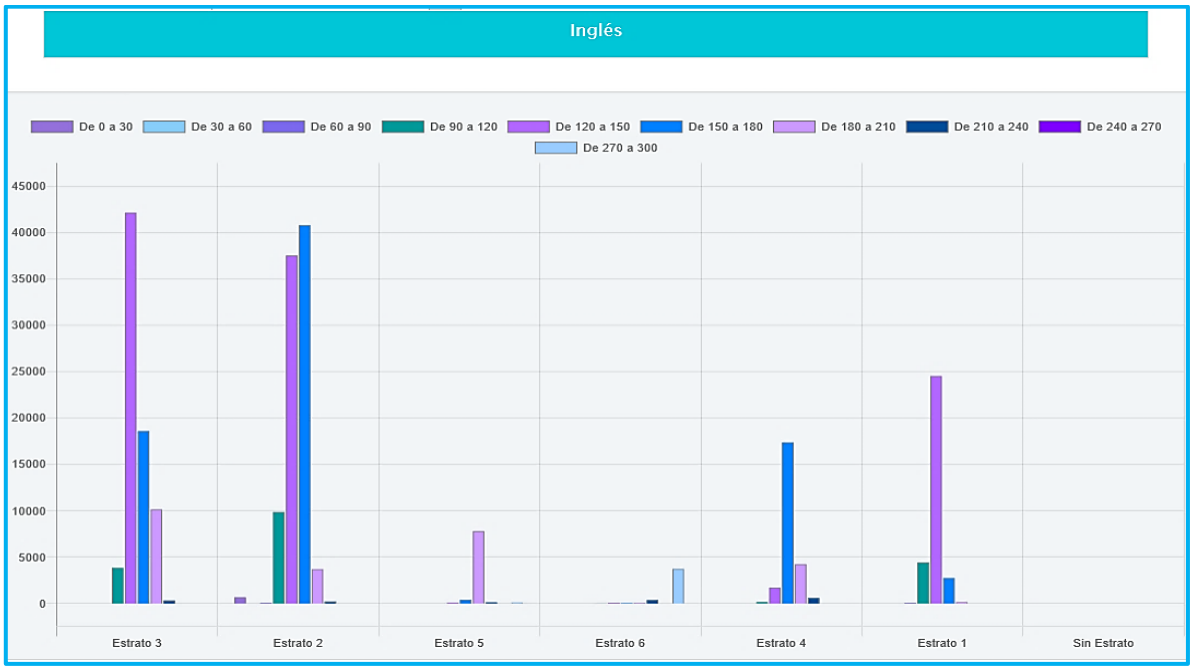


Figura 211. Inglés vs Estrato de vivienda, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

d) Tiene internet

En este indicador se observa que la mayoría de los estudiantes del país cuenta con servicio de internet en sus hogares, asimismo el género femenino constituye la mayoría en las dos categorías de clasificación.

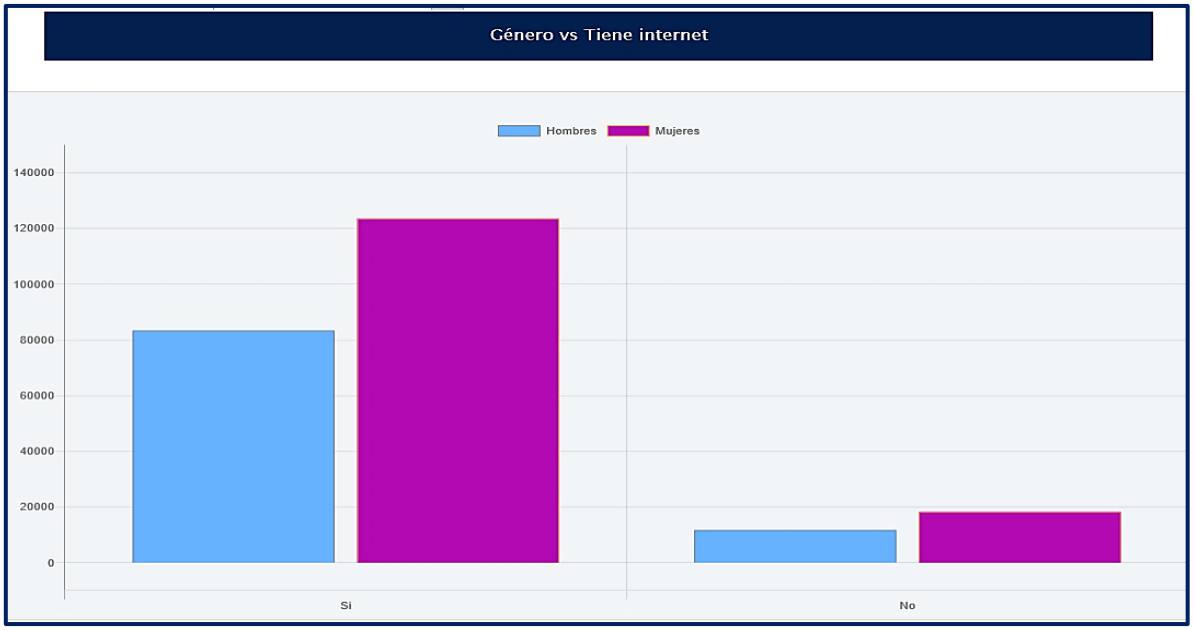


Figura 212. Género vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, se observa que los estudiantes que cuentan con servicio de internet obtienen mejores resultados que no los que no cuentan con este servicio. Además, se destaca que para los de la categoría Si, hay presencia de resultados dentro del rango entre 210 a 240.

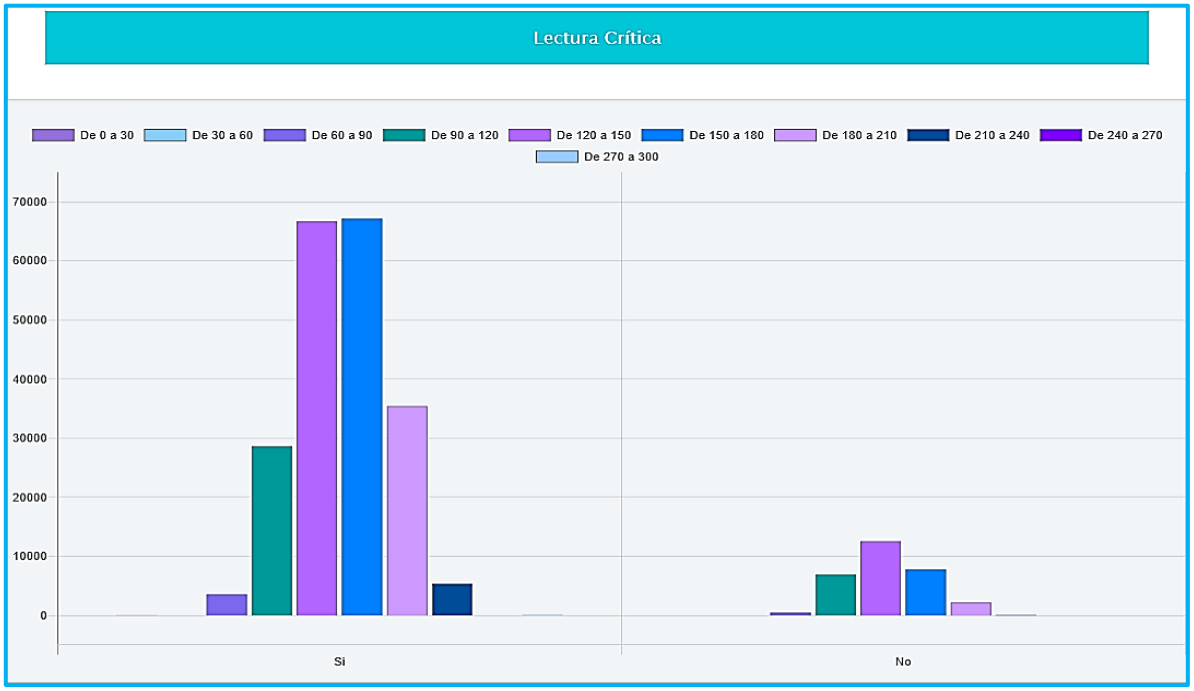


Figura 213. Lectura crítica vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, se muestra que el clúster más representativo para la categoría Si es el comprendido entre 150 a 180, mientras que para los de No es entre 120 a 150. Asimismo, los que No tienen internet alcanzar a tener resultados en el rango de 270 a 300.

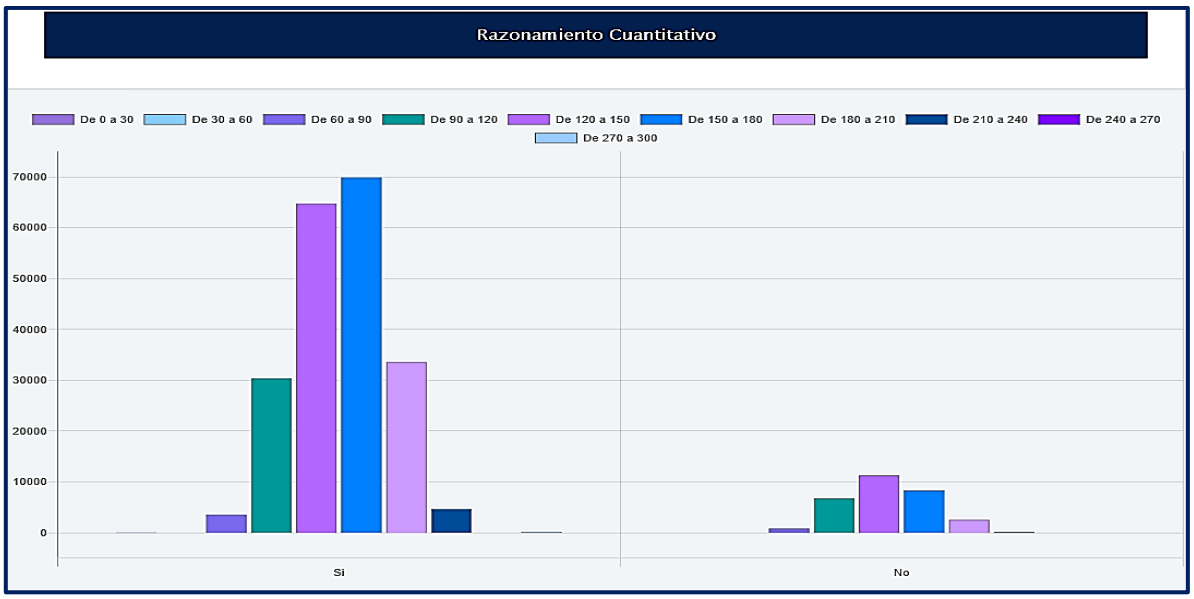


Figura 214. Razonamiento cuantitativo vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Comunicación escrita ambas categorías tienen como rango más representativo el comprendido entre 120 a 150. Sin embargo, se destaca de manera positiva que en la categoría de Si hay una cantidad de resultados en el rango más alto.

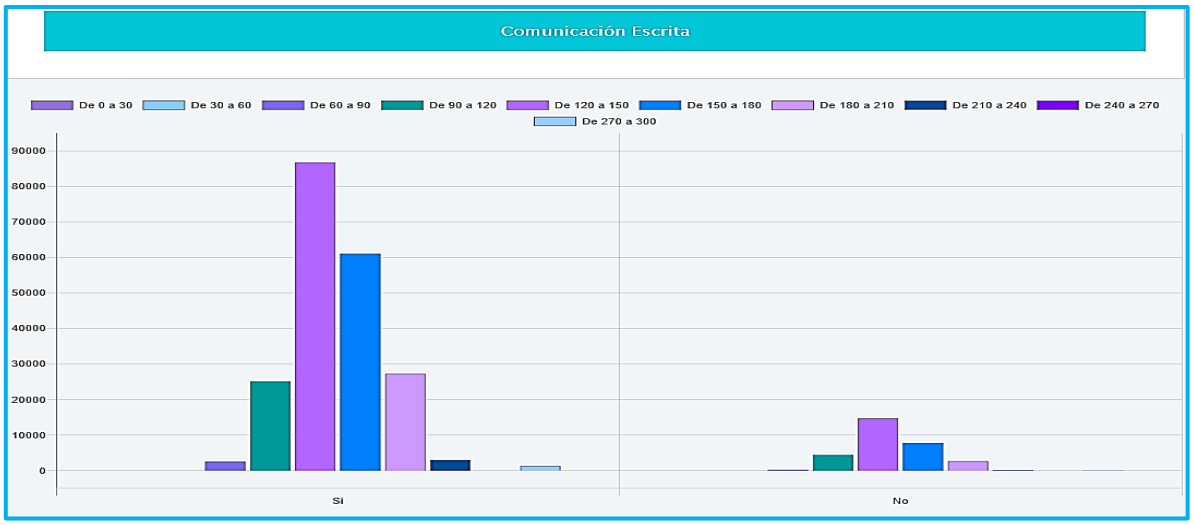


Figura 215. Comunicación escrita vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En competencias ciudadanas, el comportamiento es similar para las dos categorías, pero se destaca que en la categoría de Si hay una cantidad importante de resultados dentro del rango de 210 a 240.

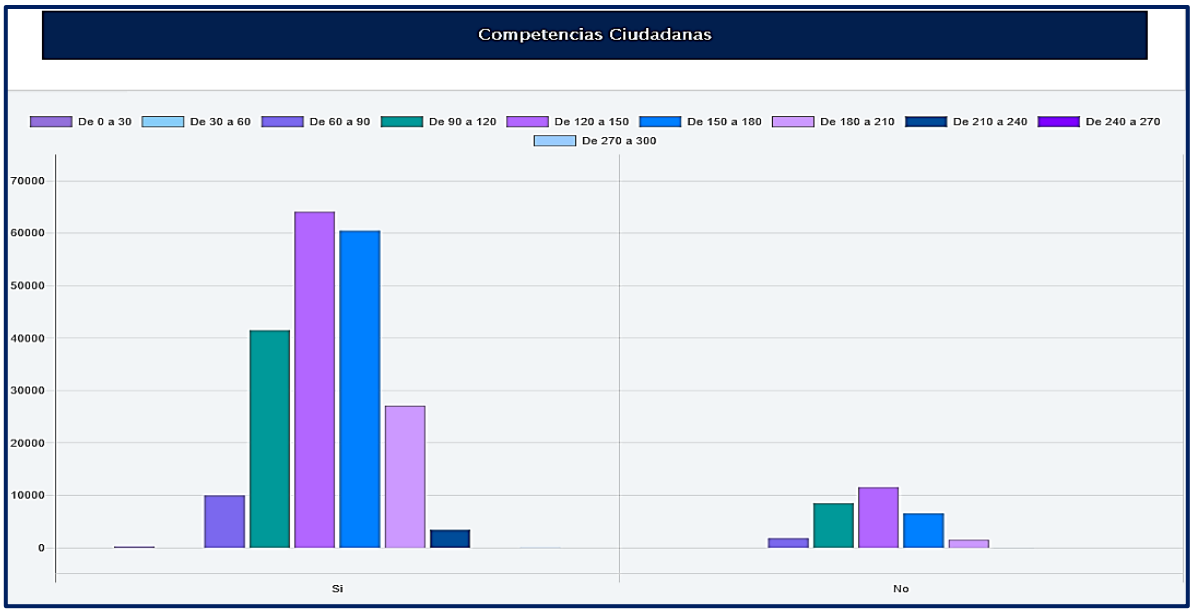


Figura 216. Competencias ciudadanas vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Por último, en el núcleo de inglés se destaca que en la categoría de Si se evidencian resultados en los rangos extremos (de 0 a 30 y de 210 a 300), pero en términos generales el comportamiento es similar para ambas categorías.

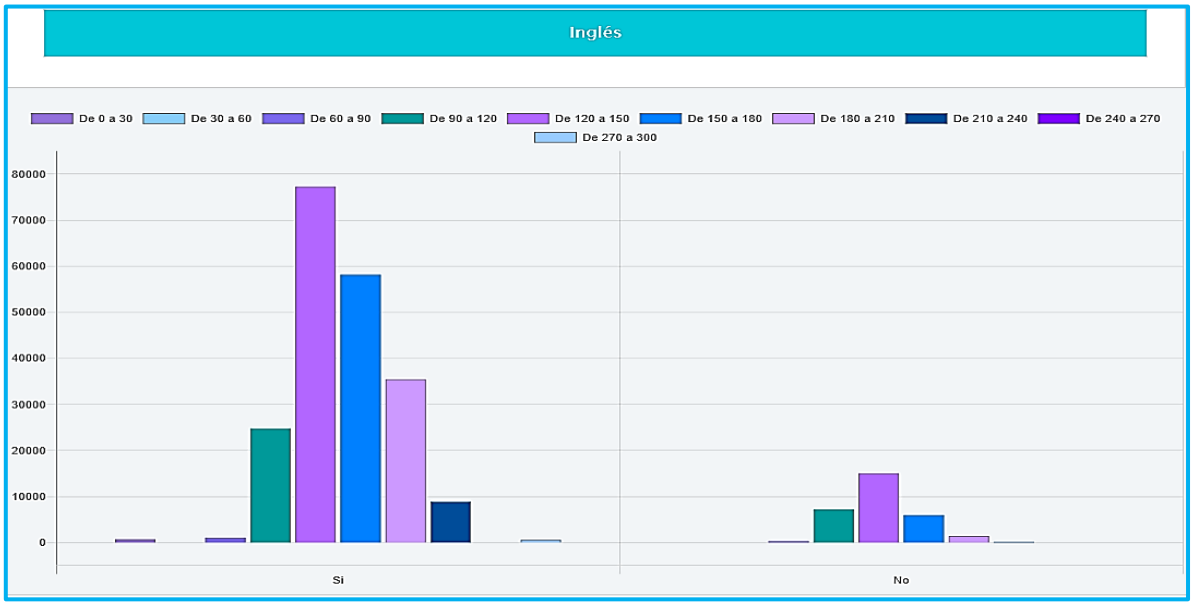


Figura 217. Inglés vs Tiene internet, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

e) Área de residencia

En el indicador de Área de residencia, se evidencia que la gran mayoría de estudiantes viven en las cabeceras municipales del país. Además, en las dos categorías de clasificación hay mayor cantidad de mujeres.

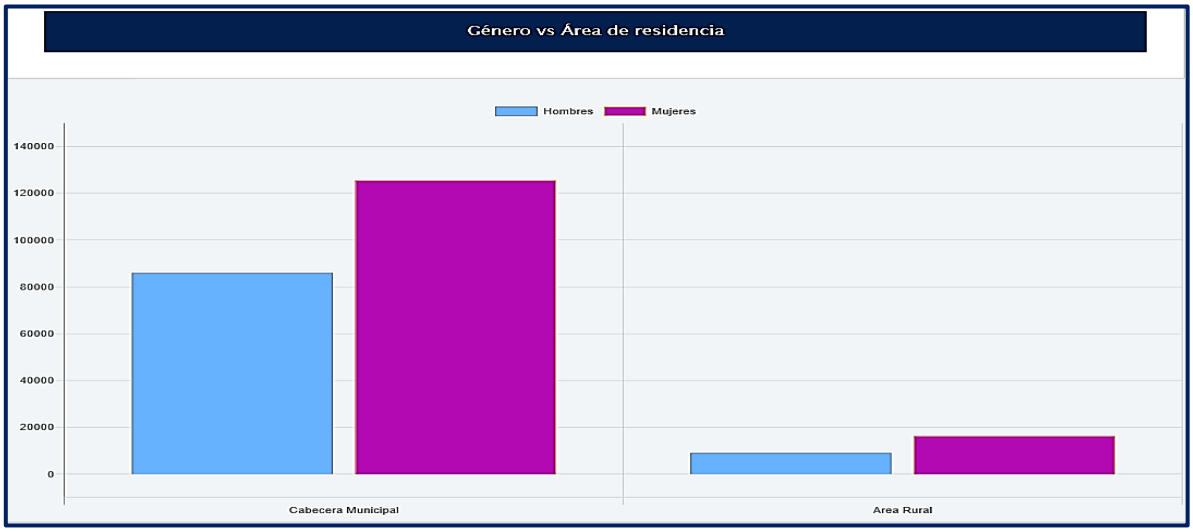


Figura 218. Género vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, los estudiantes de Cabecera municipal obtienen mejores resultados del área rural, quienes tienen como clúster más representativo el comprendido entre 120 a 150.

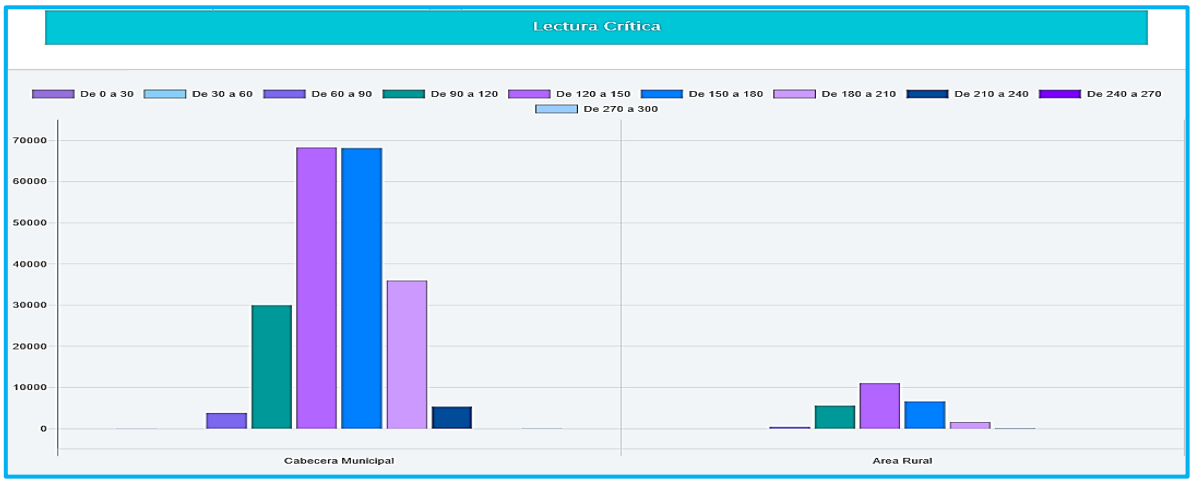


Figura 219. Lectura crítica vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Razonamiento cuantitativo, los estudiantes de Cabecera municipal tienen como rango más representativo el comprendido entre 150 a 180. Mientras que, en el Área rural el rango con mayor presencia es el de 120 a 150.

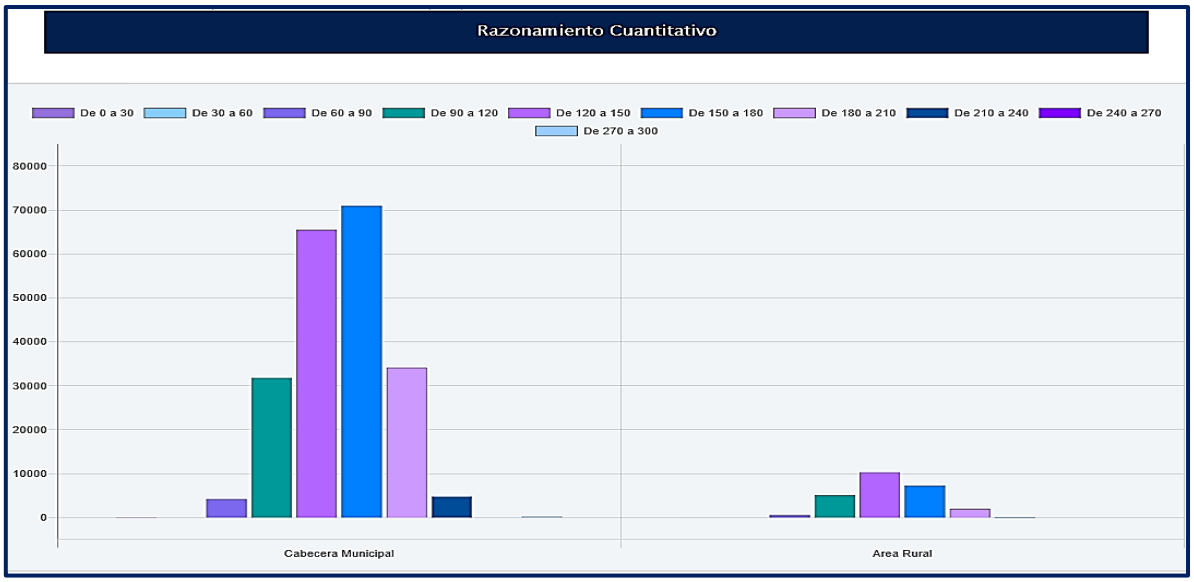


Figura 220. Razonamiento cuantitativo vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, ambas categorías obtienen como rango más representativo el comprendido entre 120 a 150. Por su parte, los estudiantes de cabecera municipal superan a los de área rural al tener más resultados en rango altos.

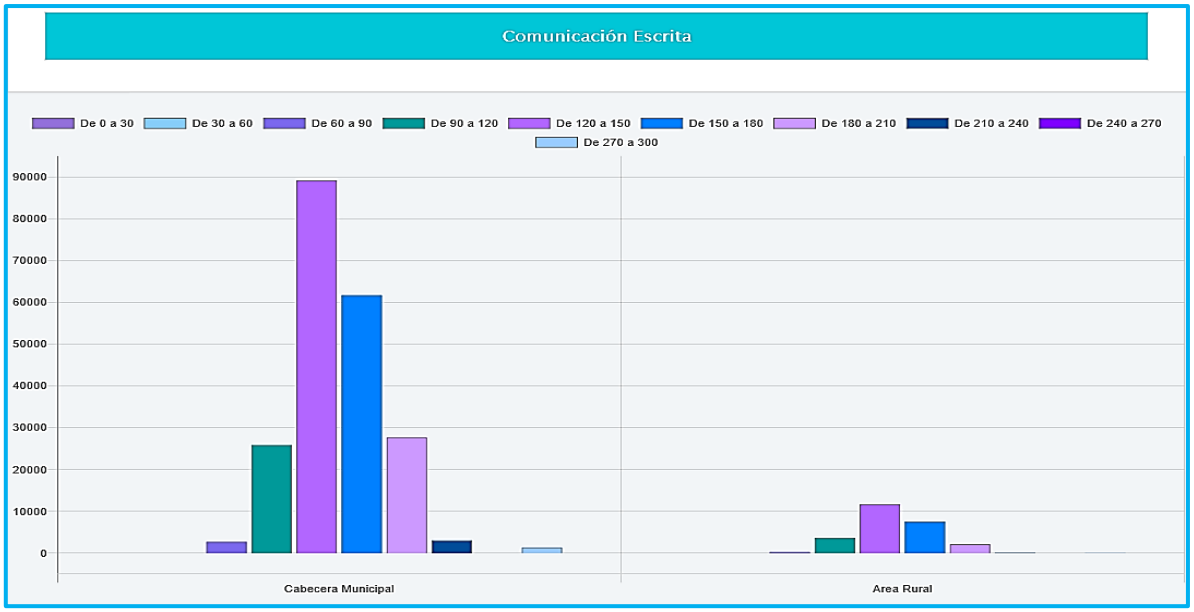


Figura 221. Comunicación escrita vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Competencias ciudadanas, ambas categorías obtienen como rango más representativo el comprendido entre 120 a 150. Por otra parte, se destaca los resultados en el rango de 210 a 240 de la categoría de Cabecera municipal.

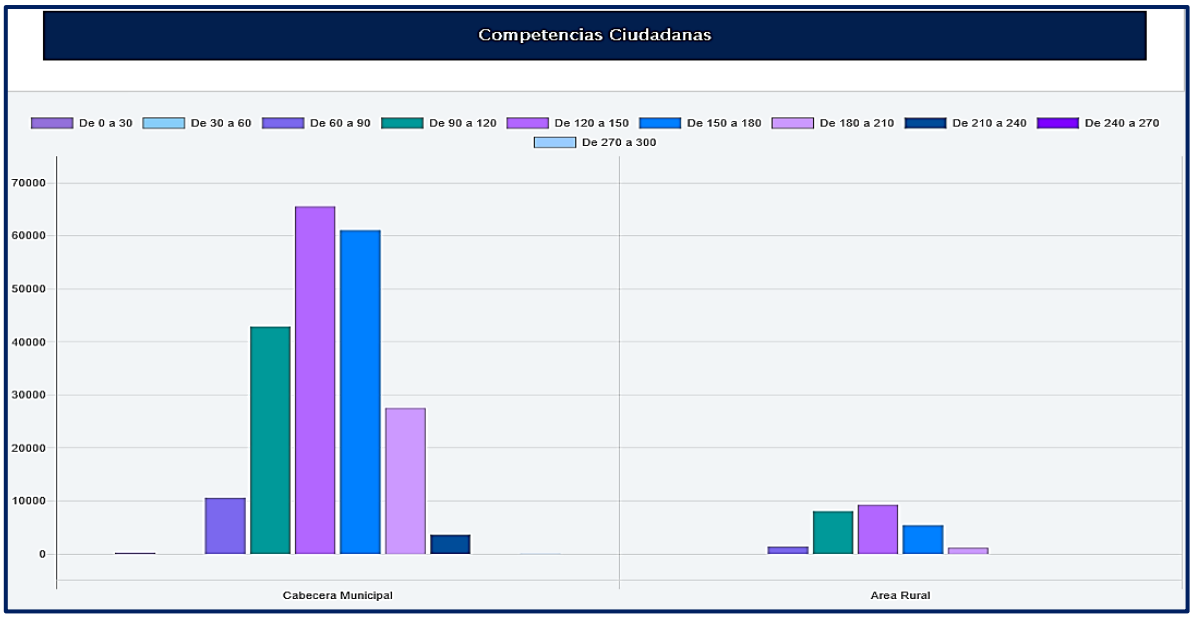


Figura 222. Competencias ciudadanas vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se destaca que en la categoría de Si se evidencian resultados en los rangos extremos, pero en términos generales el comportamiento es similar para ambas categorías, donde el clúster más significativo es el de 120 a 150.

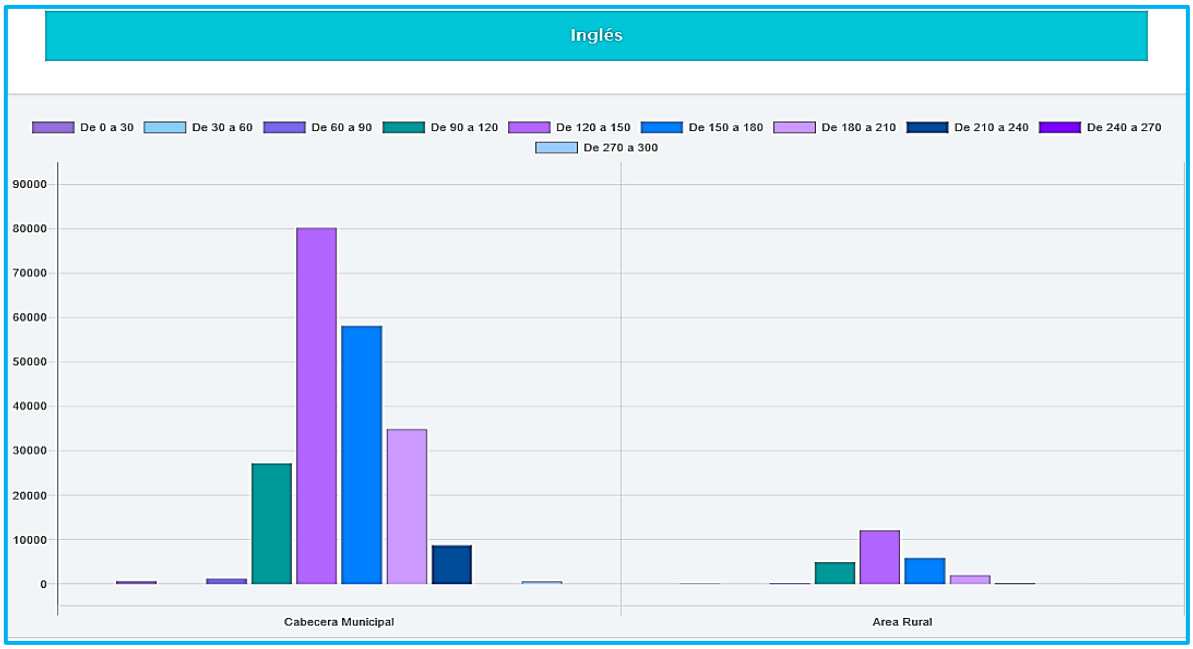


Figura 223. Inglés vs Área de residencia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

f) Estado civil

El indicador de Estado civil, muestra que la mayoría de estudiantes son solteros, en segundo lugar, se encuentran los estudiantes en unión libre, seguidos de cerca por los casados y finalmente los que están separados y/o viudos.

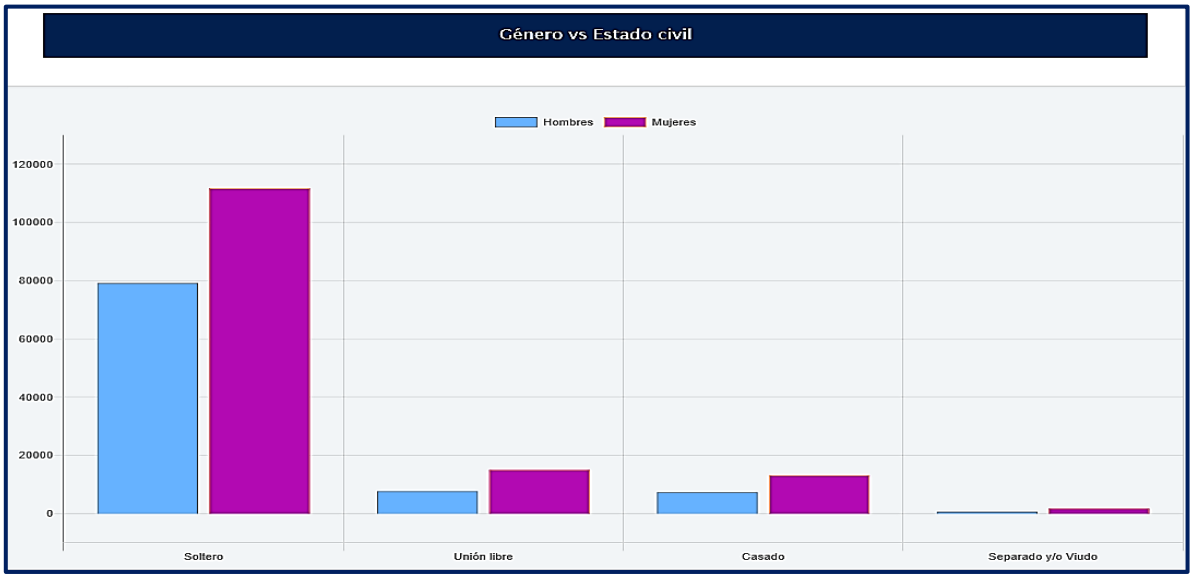


Figura 224. Género vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, se evidencia que el clúster más representativo es el comprendido entre 120 y 150. Sin embargo, los estudiantes de la categoría Separado y/o viudo tienen mejores resultados, ubicados en promedio en el rango de 150 a 180.

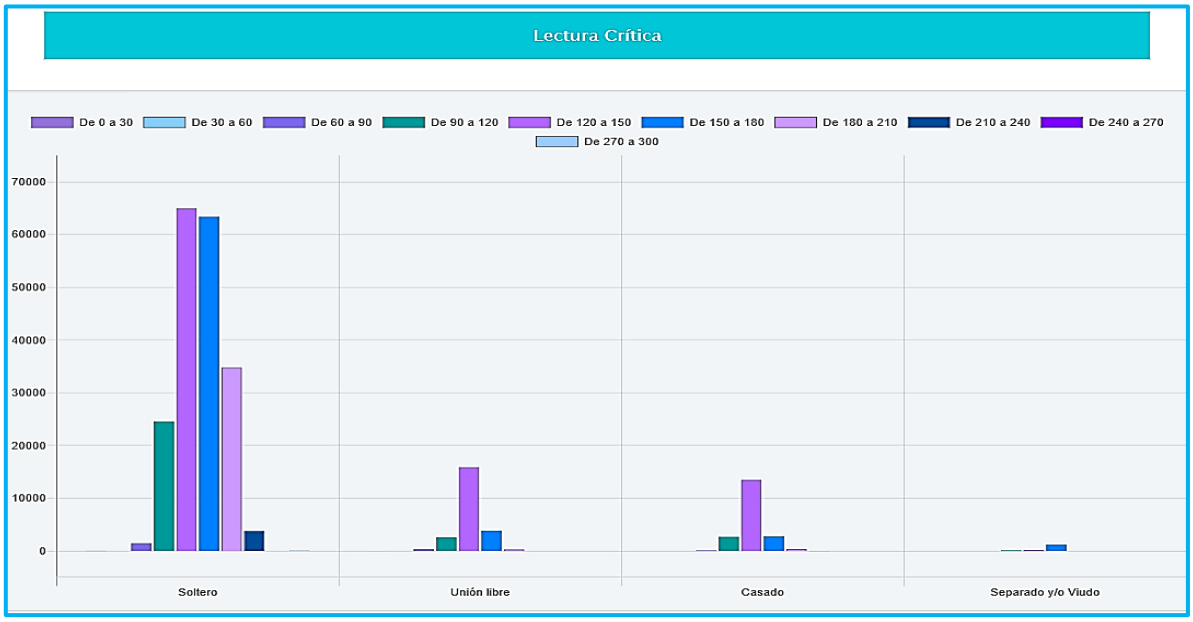


Figura 225. Lectura crítica vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, quienes presentan mejores resultados en este núcleo temático son los estudiantes que se encuentran solteros puesto que tiene mayor presencia en los rangos altos de calificación.

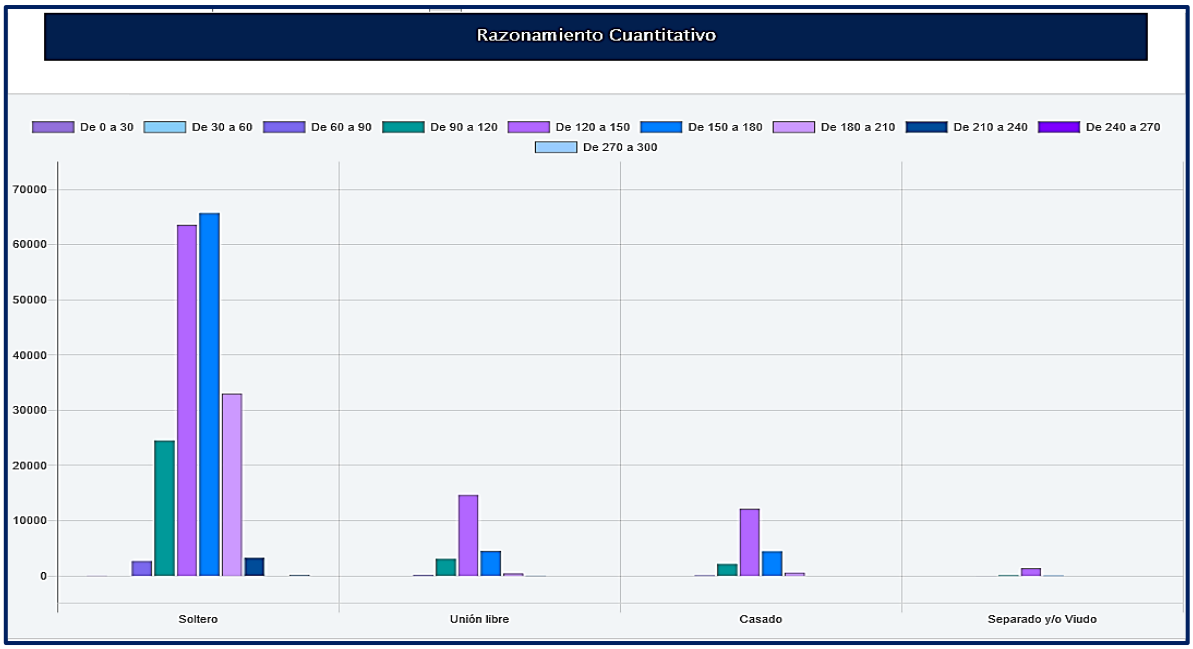


Figura 226. Razonamiento cuantitativo vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, todas las categorías presentan un comportamiento similar teniendo como clúster más representativo es el comprendido entre 120 y 150.

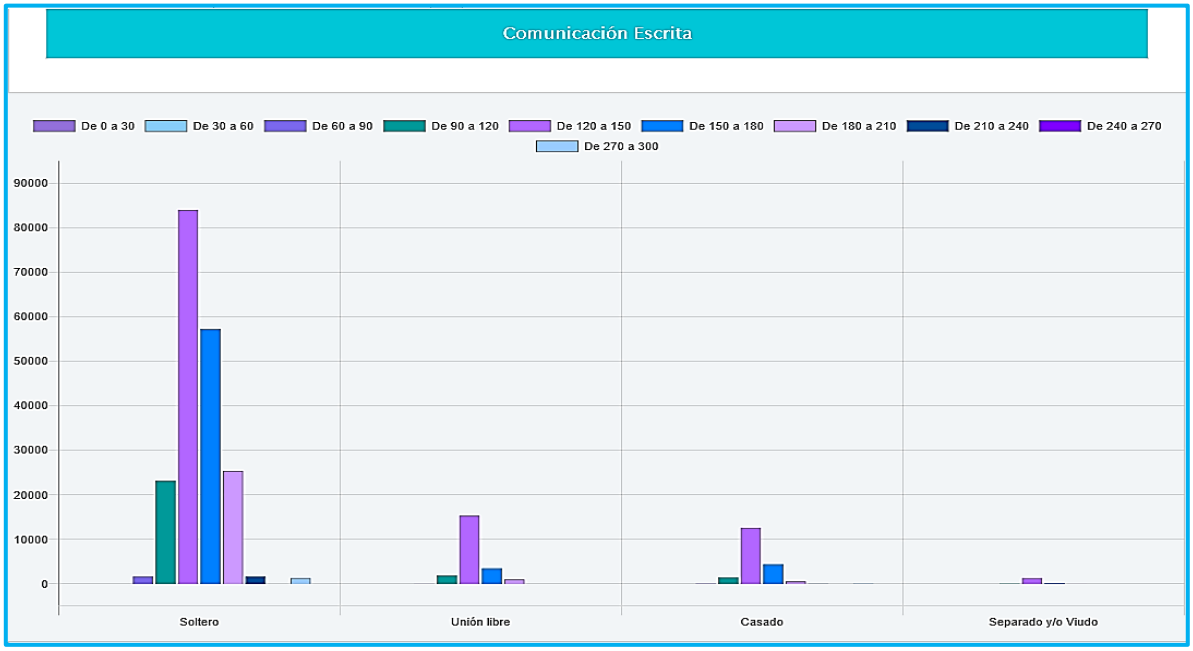


Figura 227. Comunicación escrita vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, todas las categorías presentan como clúster más representativo el comprendido entre 120 y 150. En la categoría Soltero, se evidencian resultados entre 60 y 240 puntos.

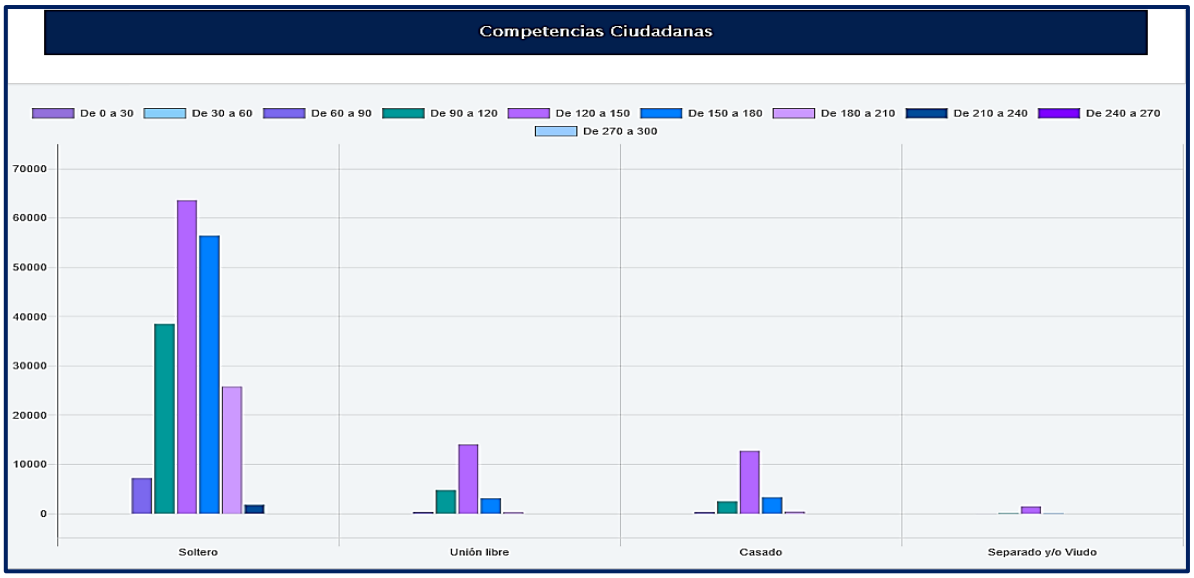


Figura 228. Competencias ciudadanas vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando este análisis, el núcleo de inglés el comportamiento es similar que en Competencias ciudadanas. Sin embargo, en la categoría Soltero se alcanzan a presentar resultados en el rango más bajo de calificación.

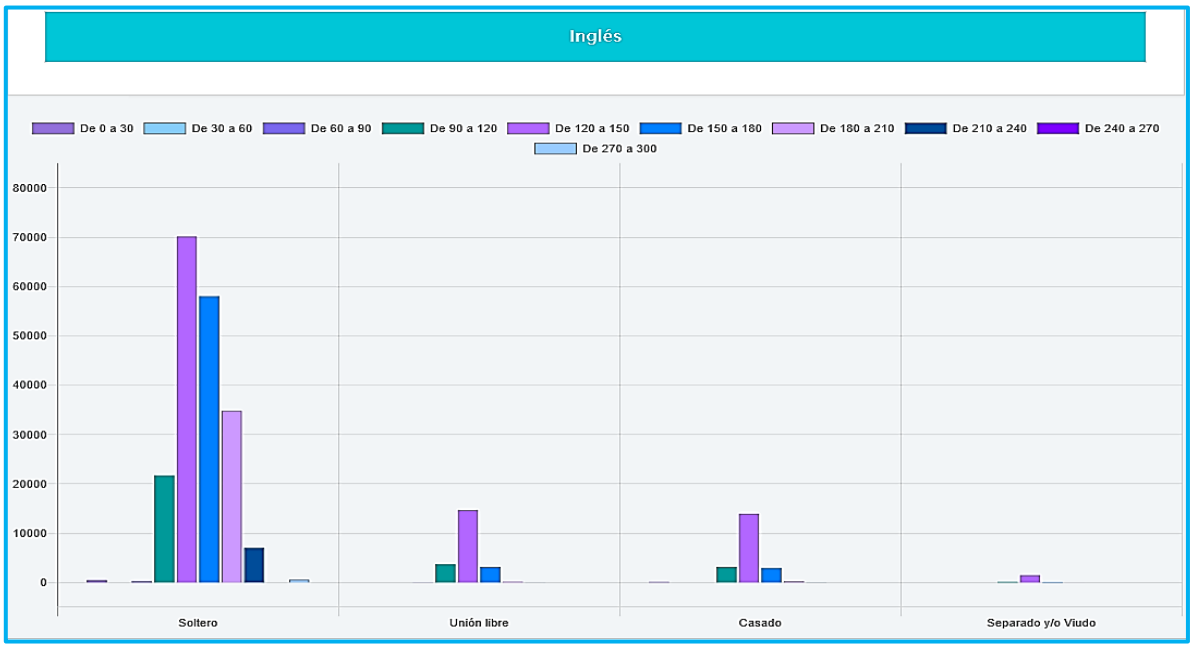


Figura 229. Inglés vs Estado civil, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

g) Cabeza de familia

En este indicador se encuentra que, por lo general los estudiantes no son cabeza de familia en sus hogares, esta diferencia se ve marcada especialmente en las mujeres.

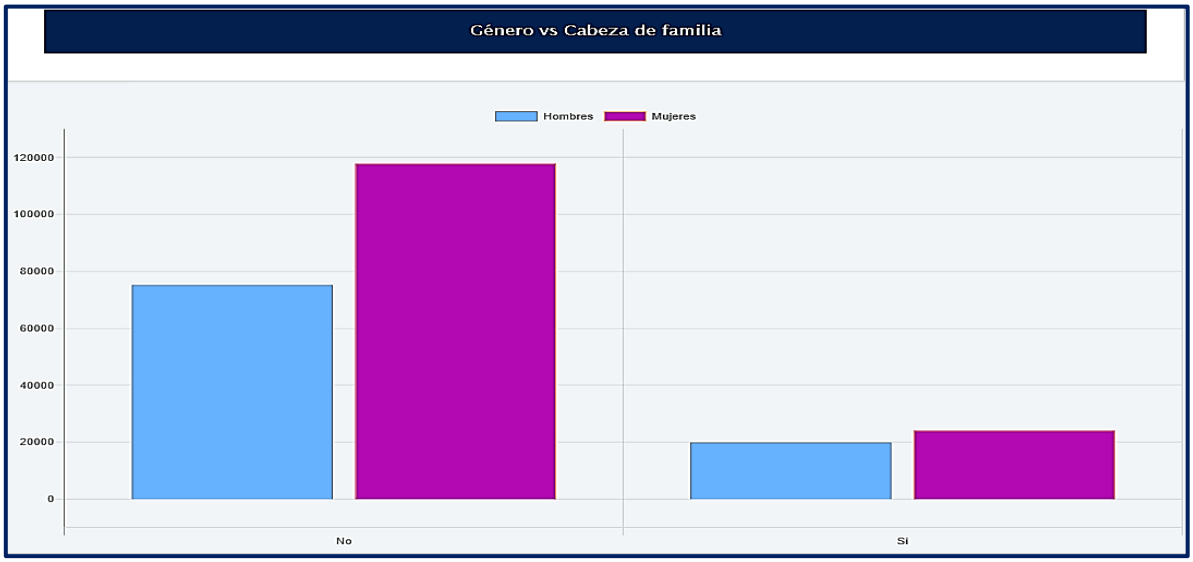


Figura 230. Género vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Lectura crítica, se encontró gracias al análisis que los estudiantes que no son cabeza de familia obtienen mejores resultados puesto que hay mayor cantidad de puntajes altos en esta categoría.

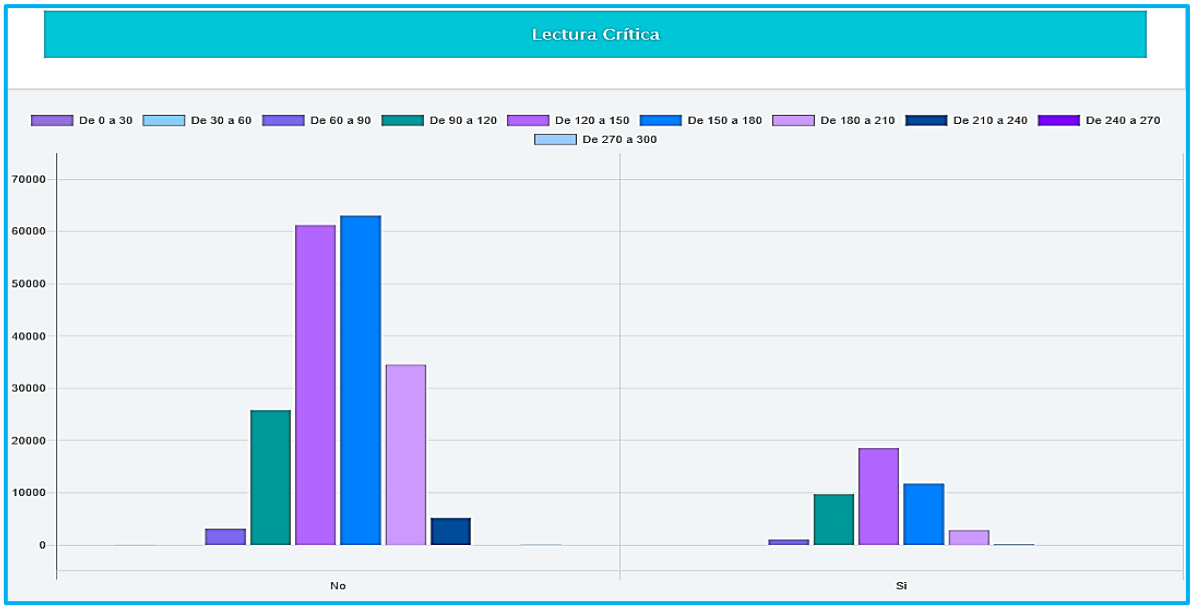


Figura 231. Lectura crítica vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, sucede un comportamiento similar al de Lectura crítica donde tienen mejores resultados los estudiantes que no son cabeza de familia, donde el rango más presentado es el de 150 a 180.

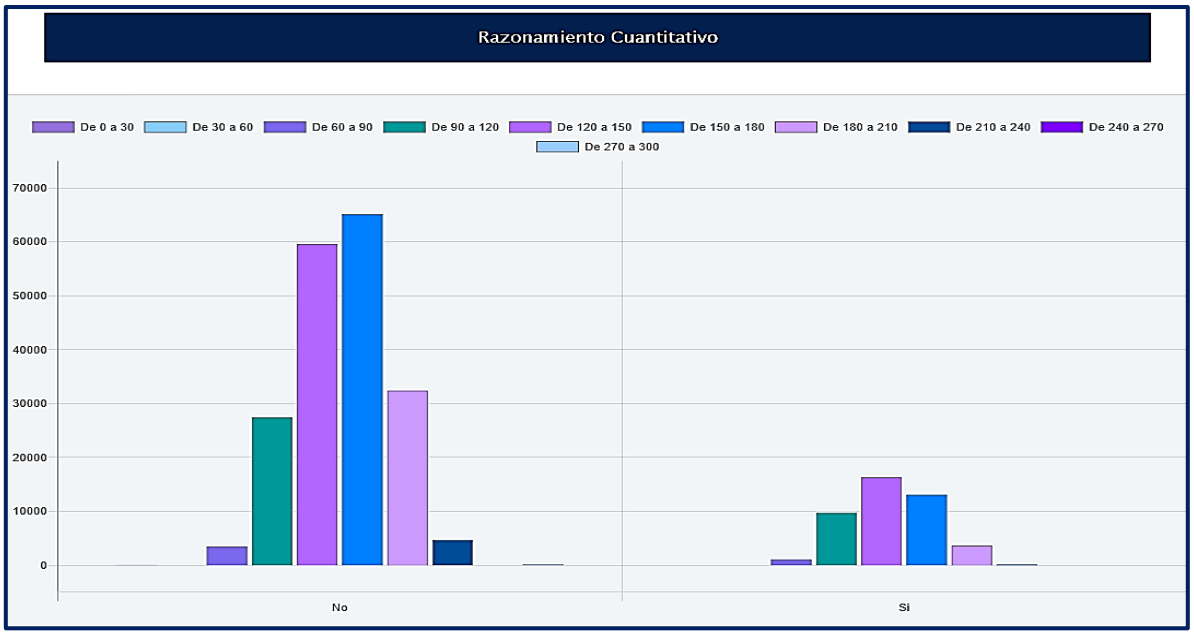


Figura 232. Razonamiento cuantitativo vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, el comportamiento es similar para ambas categorías donde el clúster más representativo es el de 120 a 150 y en segundo lugar el de 150 a 180.

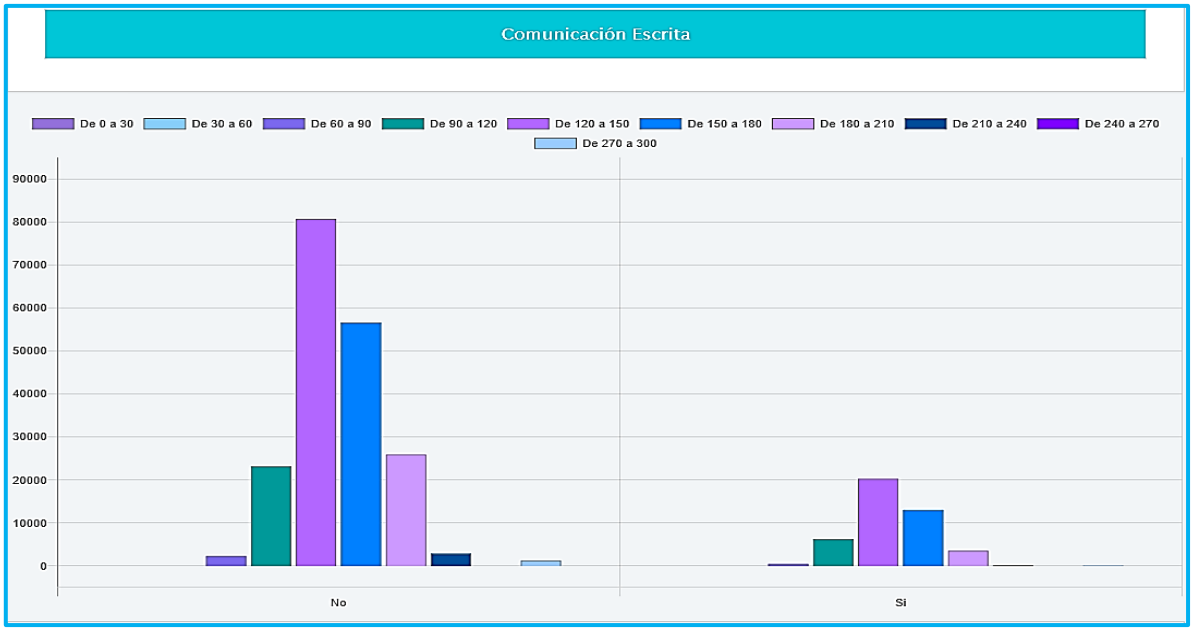


Figura 233. Comunicación escrita vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, el comportamiento es similar al de Comunicación escrita, sino que en este núcleo temático la categoría No alcanza a obtener resultados en el clúster más bajo.

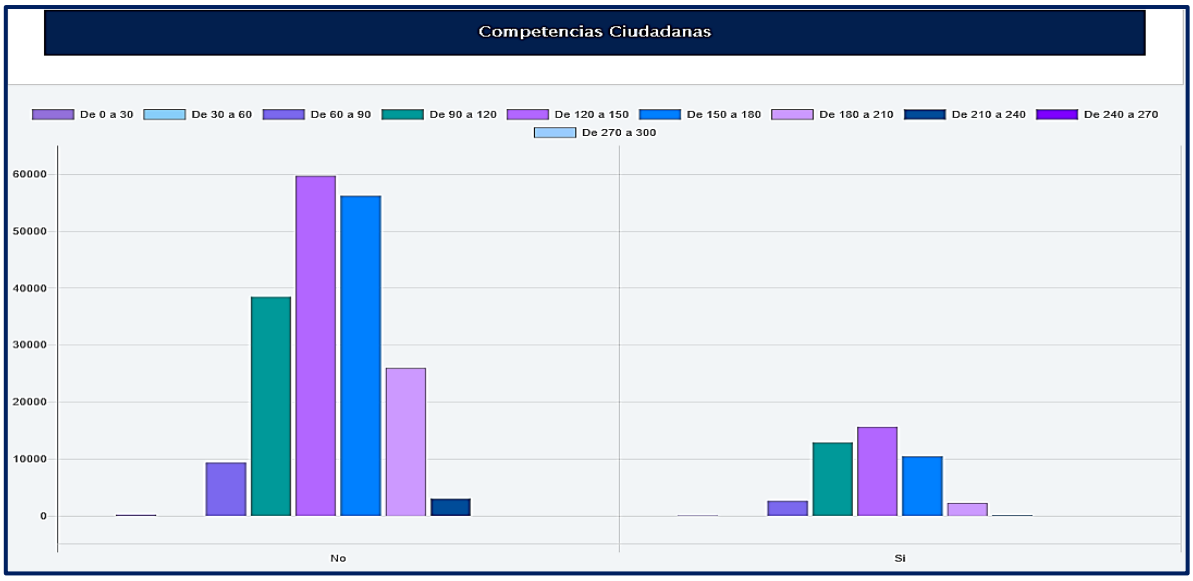


Figura 234. Competencias ciudadanas vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, el comportamiento es similar para ambas categorías donde el clúster más representativo es el de 120 a 150 y en segundo lugar el de 150 a 180. Sin embargo, en la categoría No se evidencian puntajes en el rango más alto de calificación.

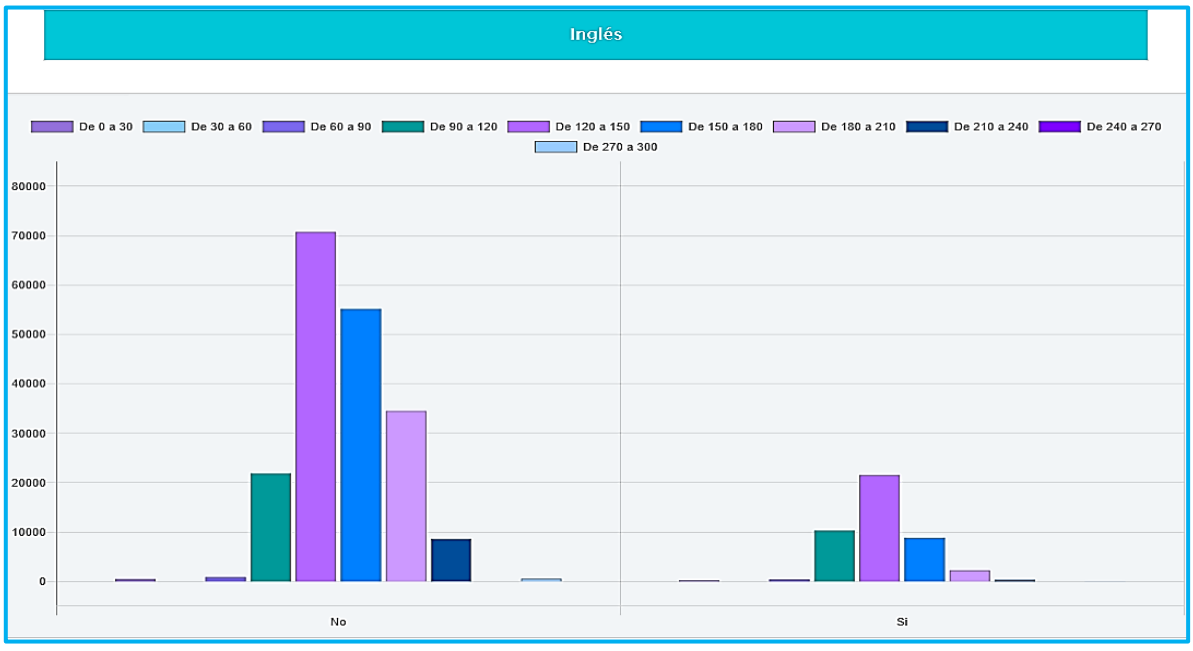


Figura 235. Inglés vs Cabeza de familia, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

h) Número de personas a cargo

En este indicador, las mujeres son mayoría en cada una de las categorías menos en la de cinco personas a cargo. Además, se evidencia que los estudiantes en su mayoría no tienen ninguna persona a cargo.

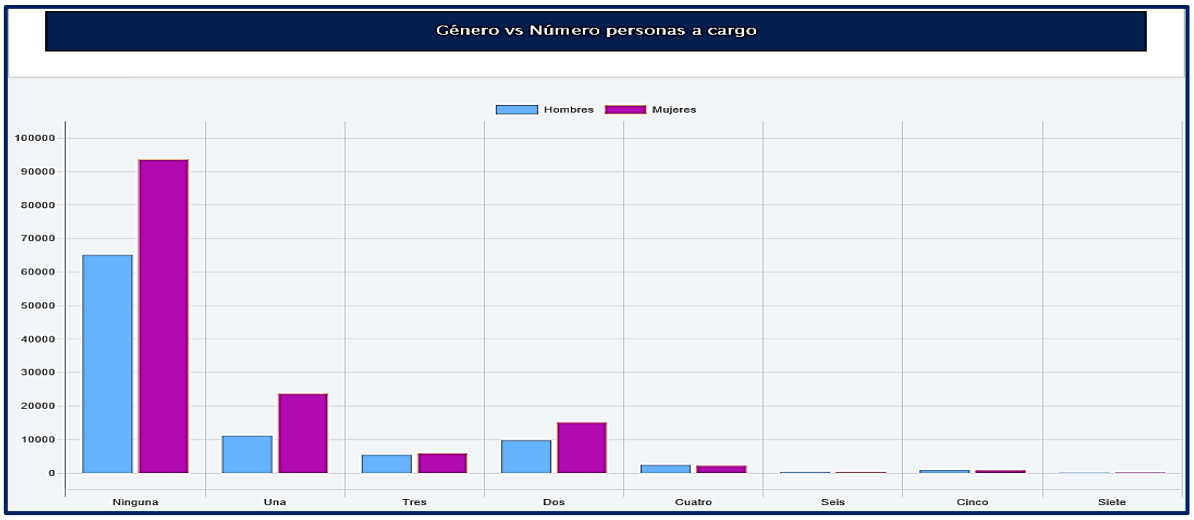


Figura 236. Género vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, en promedio las diferentes categorías presentan resultados entre 120 a 150, con excepción de los estudiantes que tienen 4 personas a cargo quienes obtienen puntajes dentro del rango de 150 a 180.

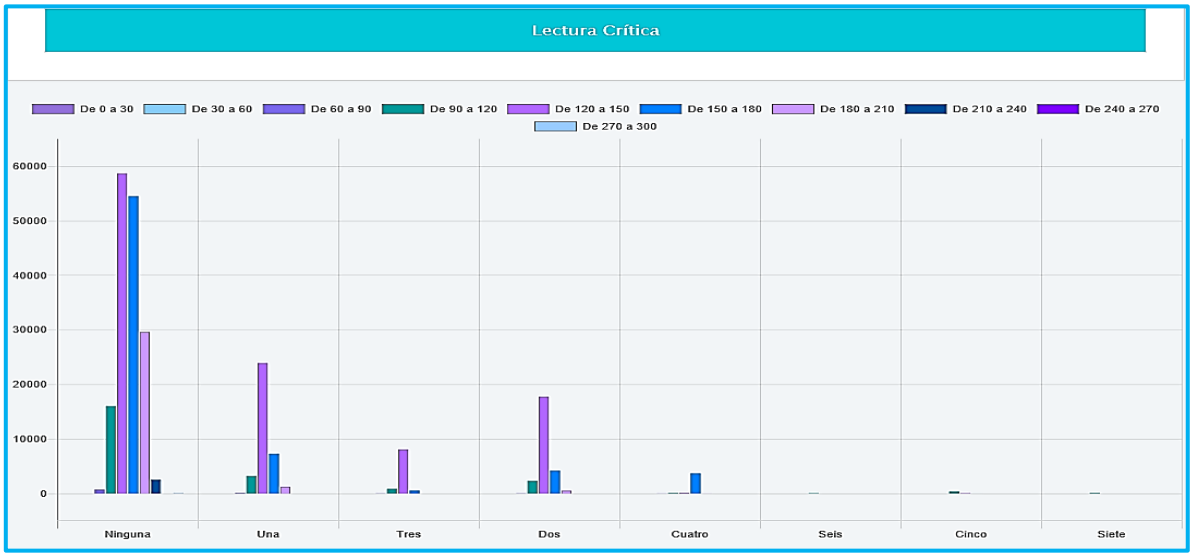


Figura 237, Lectura crítica vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, las categorías Ninguna, Cuatro y Siete obtienen como clúster más significativo el comprendido entre 150 a 180. Mientras que, las demás categorías tienen un promedio entre 120 a 150.

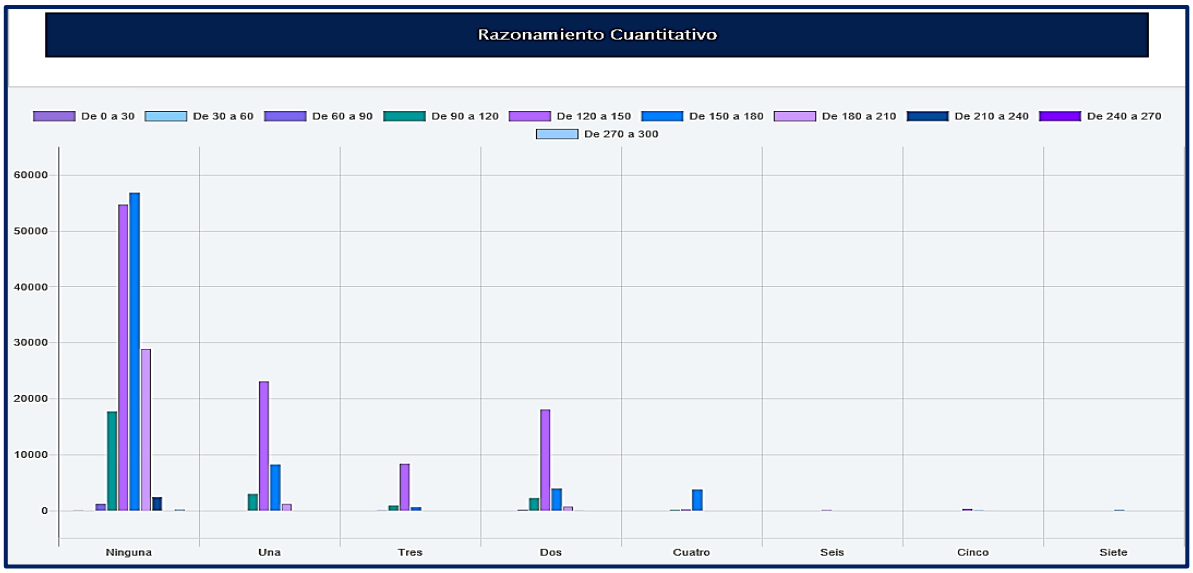


Figura 238. Razonamiento cuantitativo vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, las categorías en promedio registran puntajes entre 120 a 150. Por otra parte, en la categoría Dos, presenta mejores resultados puesto su clúster representativo es el comprendido ente 150 y 180.

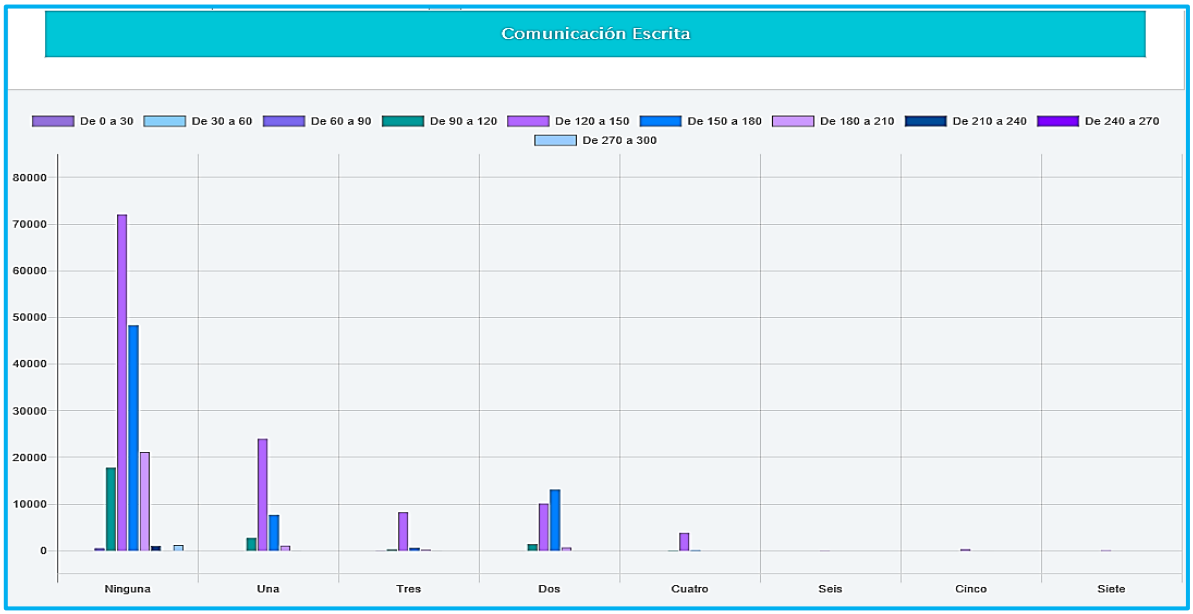


Figura 239. Comunicación escrita vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, las categorías con mejores resultados son Ninguna y Cuatro; por otra parte, quienes presentan puntajes más bajos son Dos y Cinco al obtener como rango más significativo el de 90 a 120.

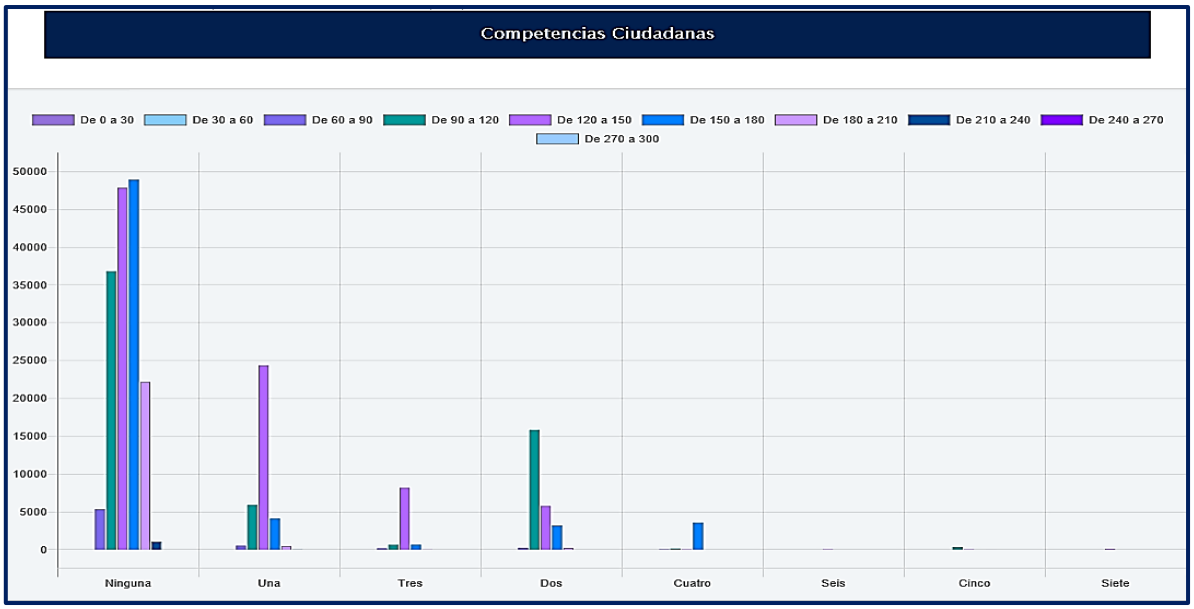


Figura 240. Competencias ciudadanas vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se destaca positivamente la categoría Cuatro al obtener mejores resultados que las demás con un clúster representativo comprendido entre 150 y 180, y de manera negativa Siete, estando por debajo del promedio.

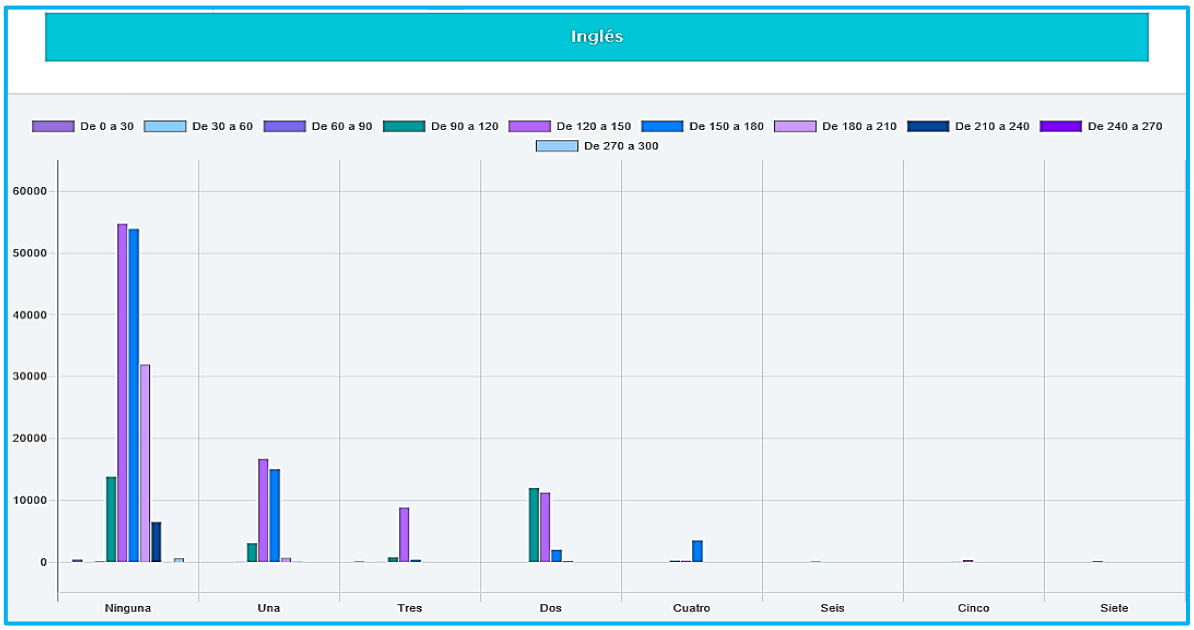


Figura 241. Inglés vs Número personas a cargo, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

i) Educación madre

En este indicador hay más estudiantes que cuentan con una madre con estudios de Secundaria completa, seguidos por los de Primaria incompleta, Técnica completa y Secundaria completa. Además, en la categoría No sabe hay más presencia del género masculino.

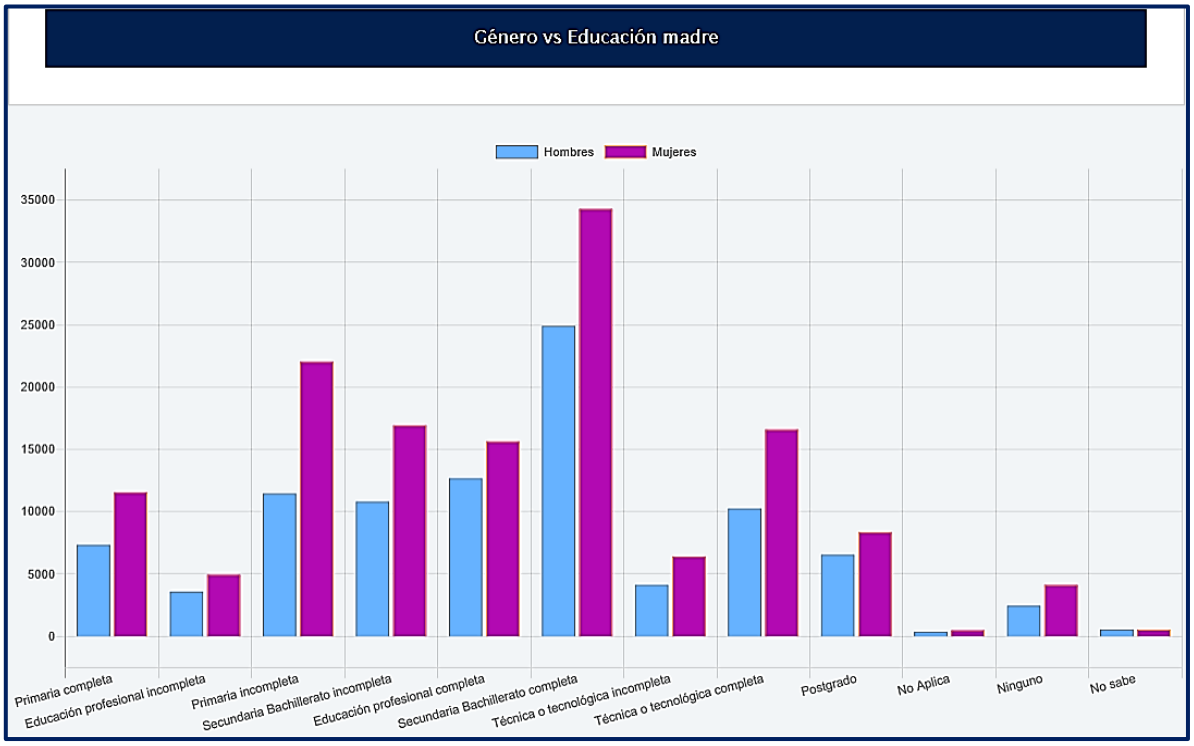


Figura 242. Género vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, se observa que el rango promedio para las diferentes categorías es el comprendido entre 120 y 150. Sin embargo, la categoría Secundaria incompleta presenta como clúster más significativo el rango de 90 a 120. Caso contrario sucede con las categorías Educación profesional completa, Secundaria completa y Técnica y tecnológica incompleta donde se destacan se manera positiva con respecto a las demás.

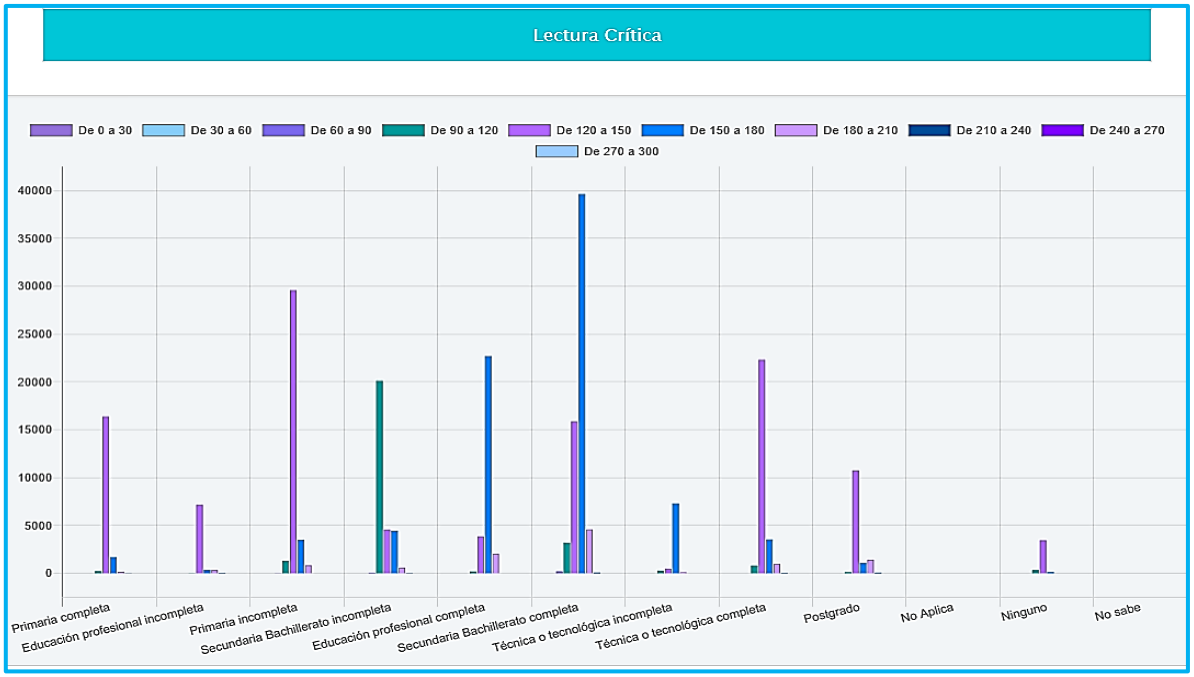


Figura 243. Lectura crítica vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, las categorías presentan en promedio un puntaje entre 120 y 150. Sin embargo, la categoría Secundaria incompleta obtiene resultados más bajos, con clúster más representativo entre 90 y 120; y, por otra parte, se destaca positivamente las categorías Educación profesional completa, Secundaria completa y Técnica o tecnológica incompleta.

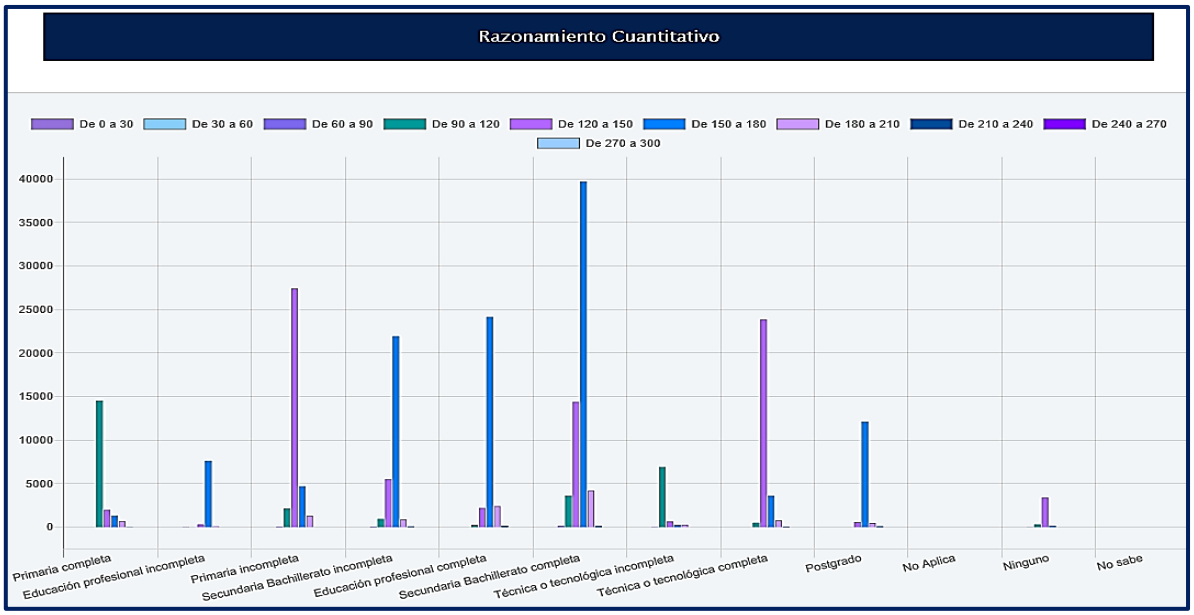


Figura 244. Razonamiento cuantitativo vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento Cuantitativo, el puntaje promedio es de 150 y 180 para las diferentes categorías. Sin embargo, las categorías Primaria incompleta, Técnica o tecnológica completa y Ninguno presentan resultados promedio entre 120 y 150. Además, Primaria completa y Técnica o tecnológica incompleta obtienen puntajes entre 90 y 120.

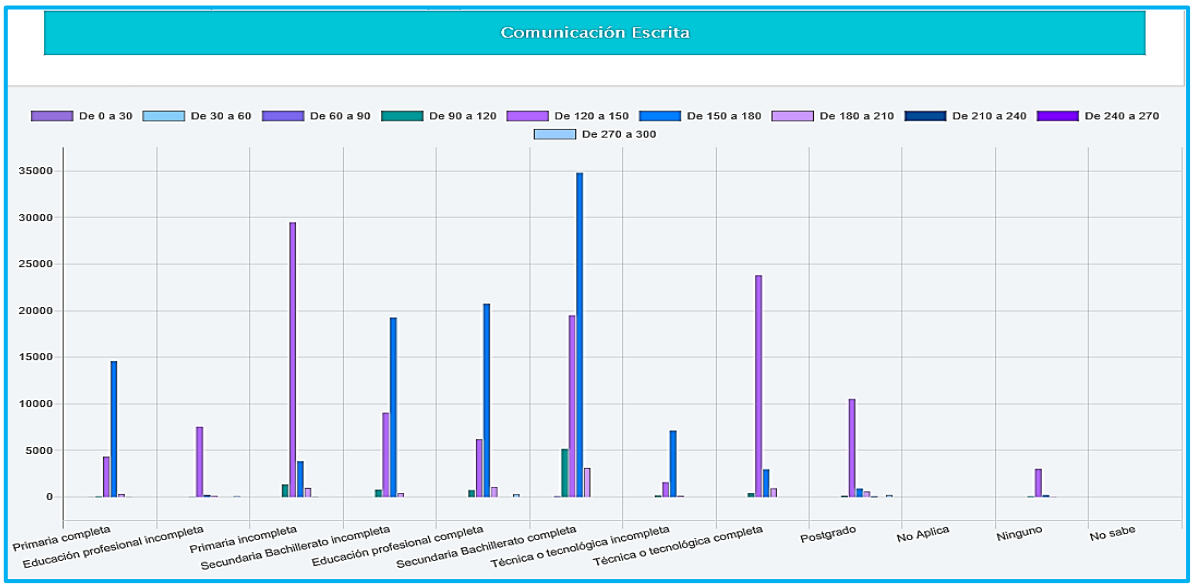


Figura 245. Comunicación escrita vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En competencias ciudadanas, se destaca positivamente la categoría Postgrado al obtener resultados comprendidos entre 180 y 120. Mientras que, de forma negativa se presentan resultados bajos en las categorías Técnica o tecnológica completa y Ninguno.

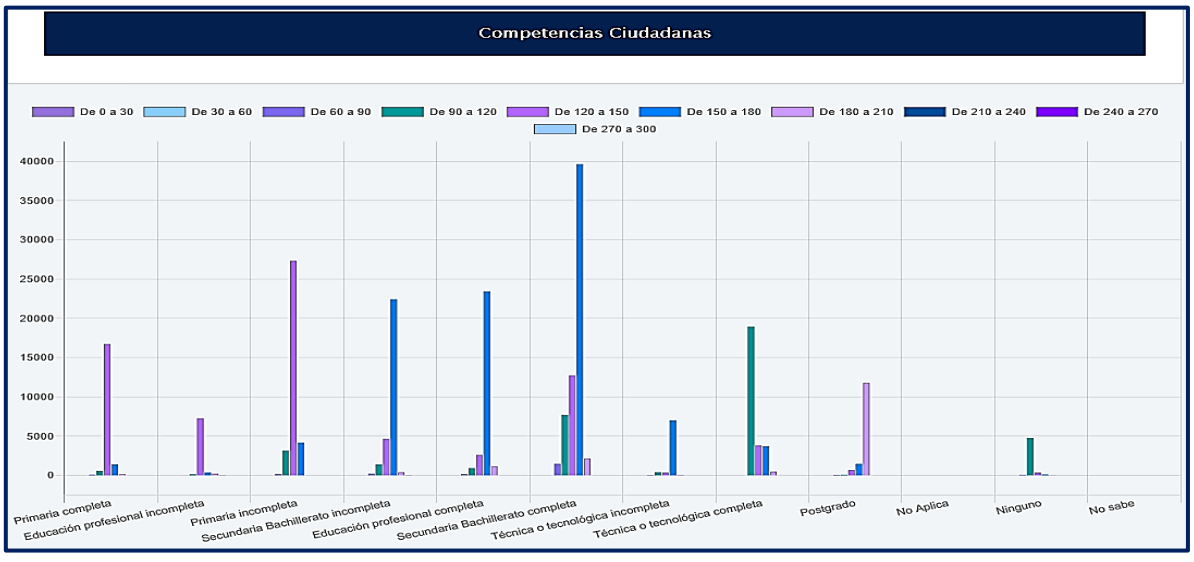


Figura 246. Competencias ciudadanas vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando el análisis para este indicador, se encuentra que en inglés la categoría Educación profesional completa se destaca por encima de las demás categorías teniendo sus resultados mayoritariamente entre 180 y 210. Además, el promedio de puntajes en este núcleo temática es más alto que en los otros presentados en esta prueba.

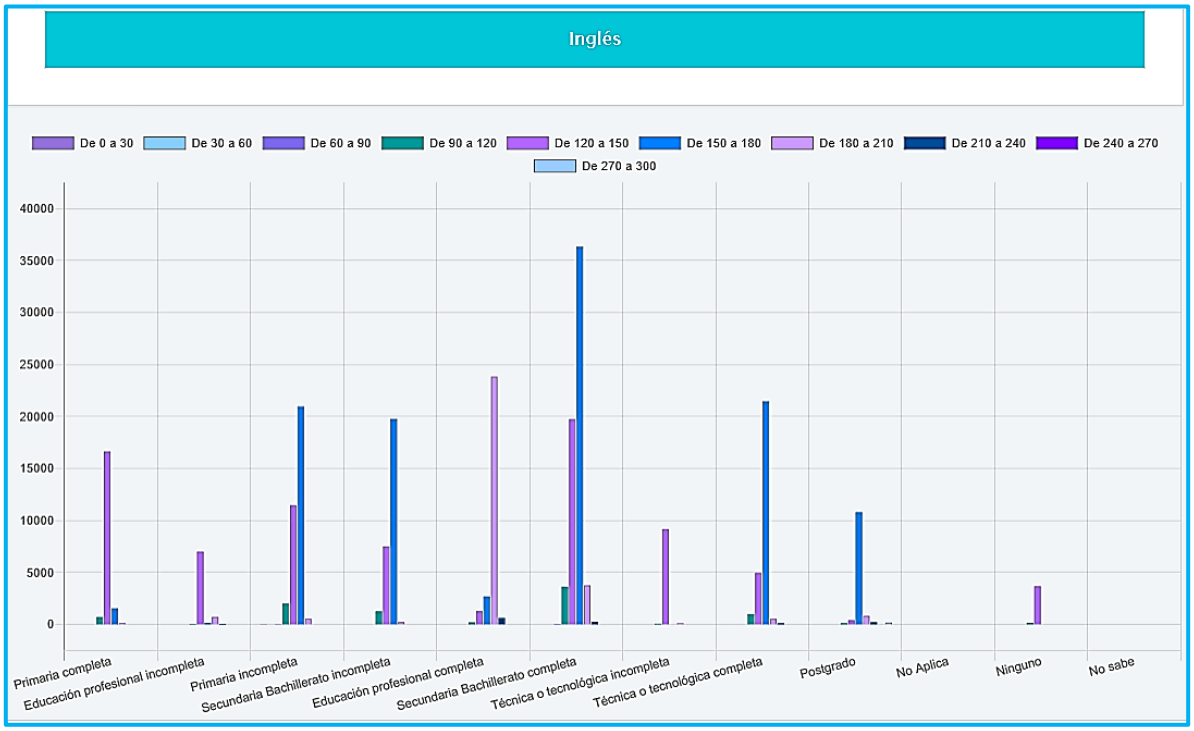


Figura 247. Inglés vs Educación madre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

j) Educación padre

En este indicador hay más estudiantes que cuentan con un padre con estudios de Secundaria completa, seguidos por los de Primaria incompleta, Educación profesional completa y Secundaria completa. Vale la pena aclarar, que en todas las categorías el género femenino es mayoría.

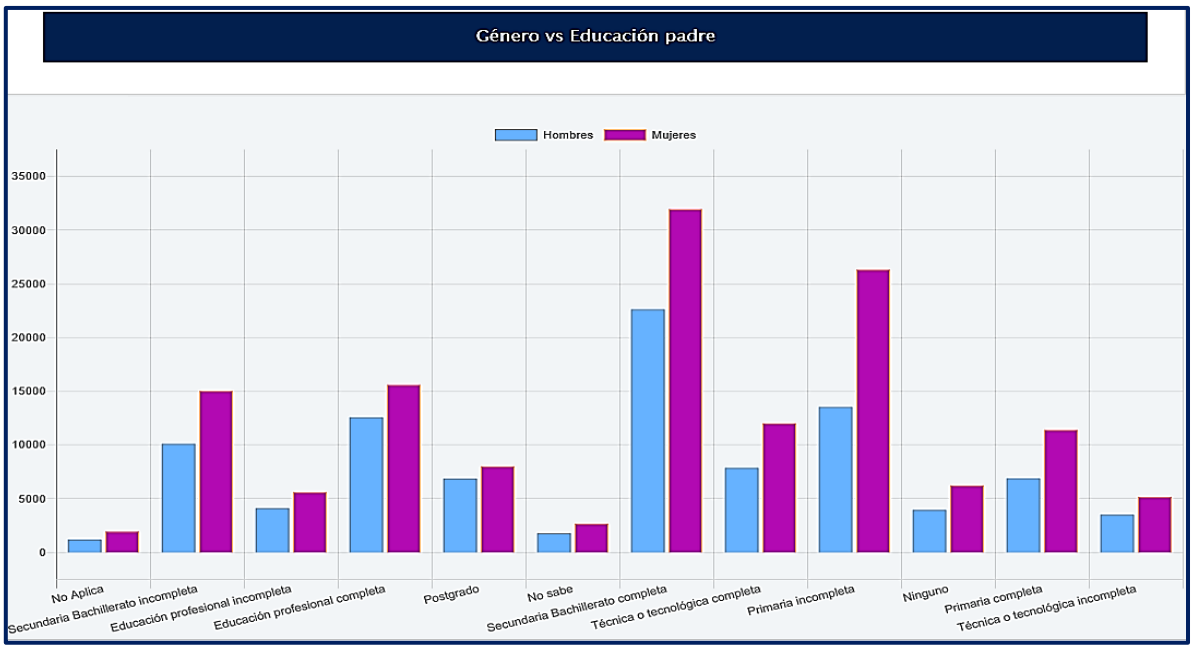


Figura 248. Género vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, los estudiantes que mejores puntajes obtienen son los de la categoría Postgrado. Por otro lado, los que menores resultados presentan son los de Secundaria completa e incompleta y Ninguno, estando bajo el promedio.

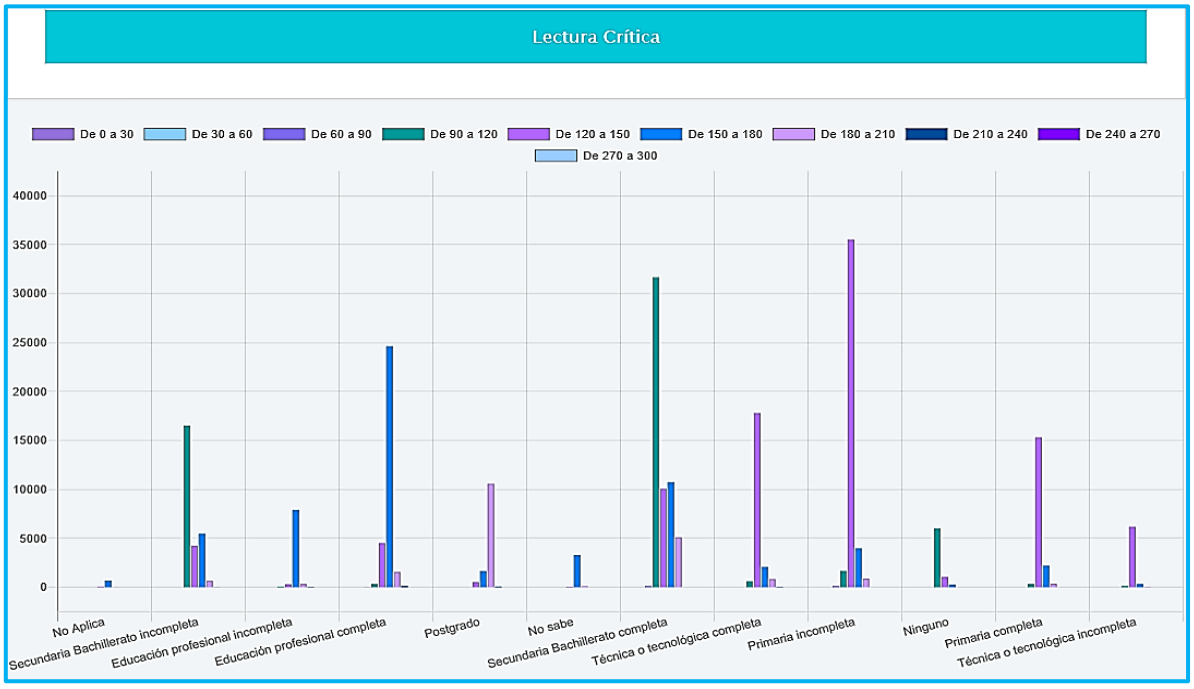


Figura 249. Lectura crítica vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, se encuentra que las categorías Secundaria incompleta y Ninguno presentan un rango de puntajes por debajo del promedio, donde las demás categorías obtienen entre 120 y 180.

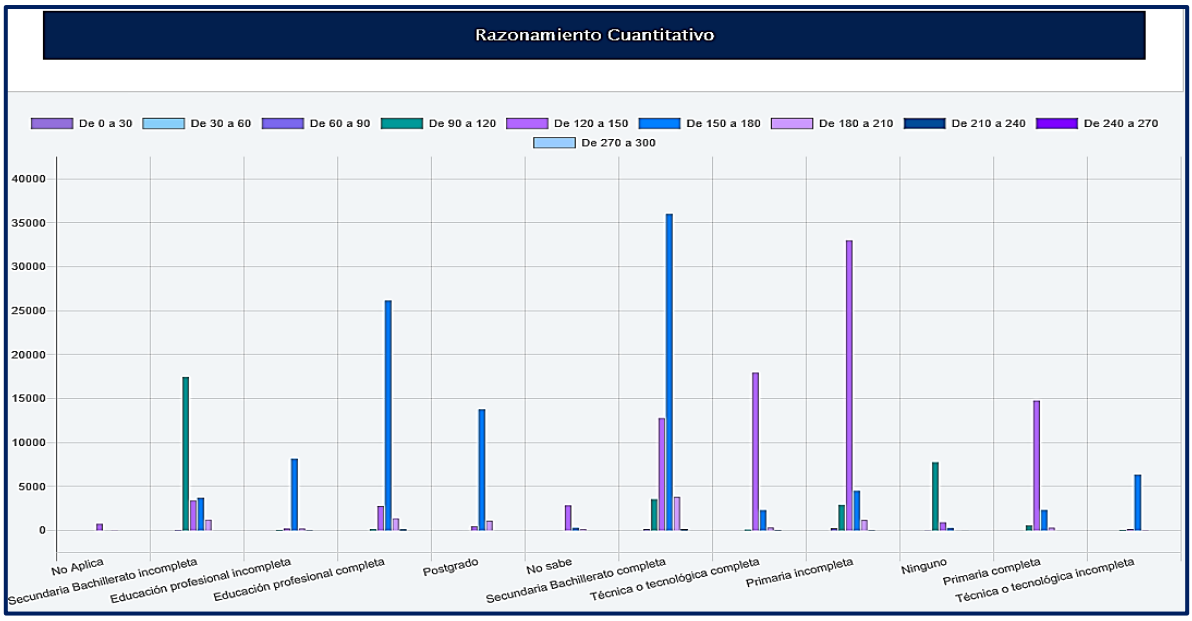


Figura 250. Razonamiento cuantitativo vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Por su parte, Comunicación escrita presenta un comportamiento similar entre las diferentes categorías donde obtienen entre 120 y 180 como rango representativo.

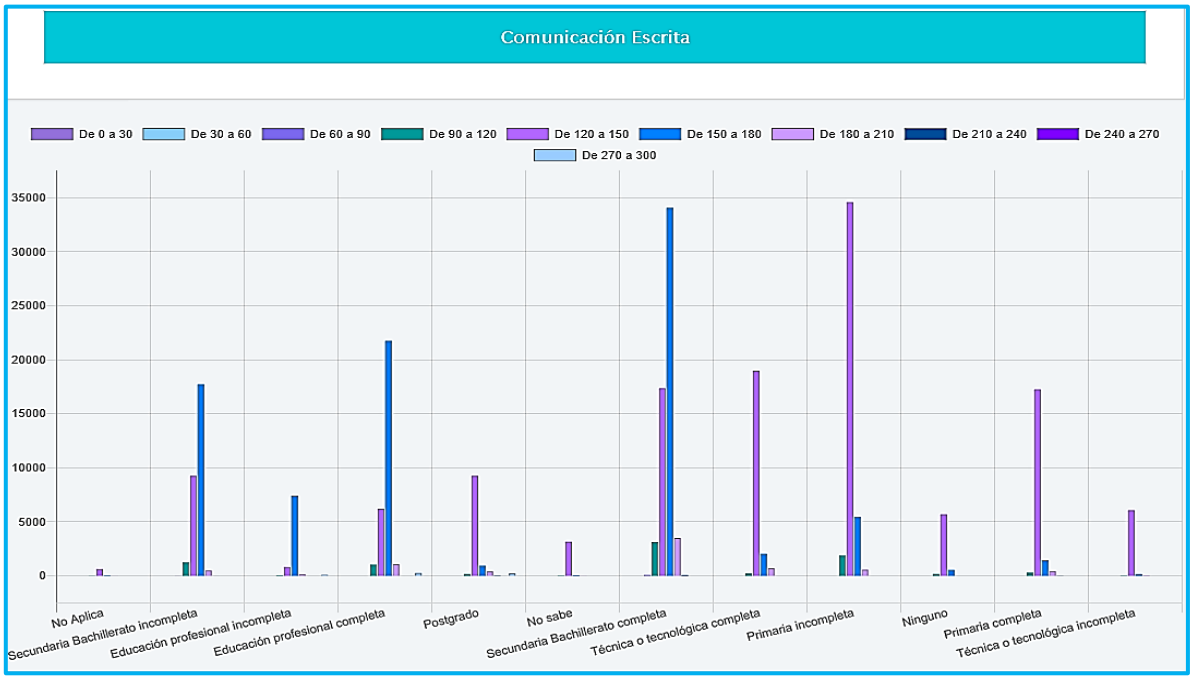


Figura 251. Comunicación escrita vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, la categoría Técnica o tecnológica completa presenta como clúster más significativo el comprendido entre 90 y 120 encontrándose por debajo del promedio de las demás categorías.

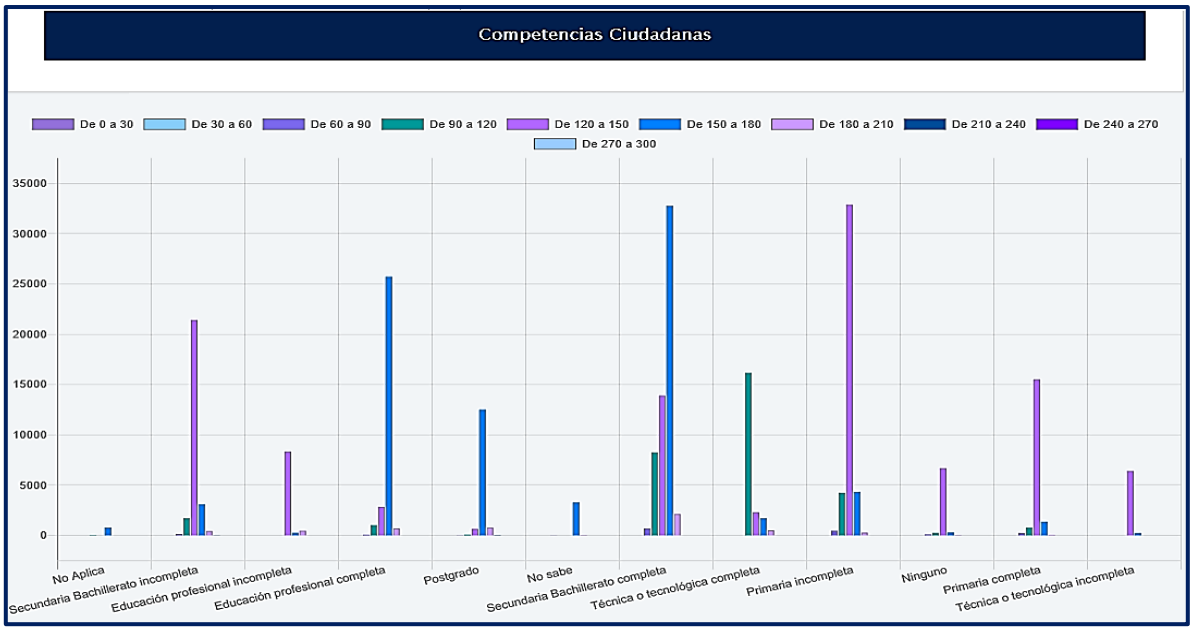


Figura 252. Competencias ciudadanas vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se observa que Postgrado obtiene resultados mucho más altos que las demás categorías puesto que se encuentra con un clúster promedio de 210 a 240.

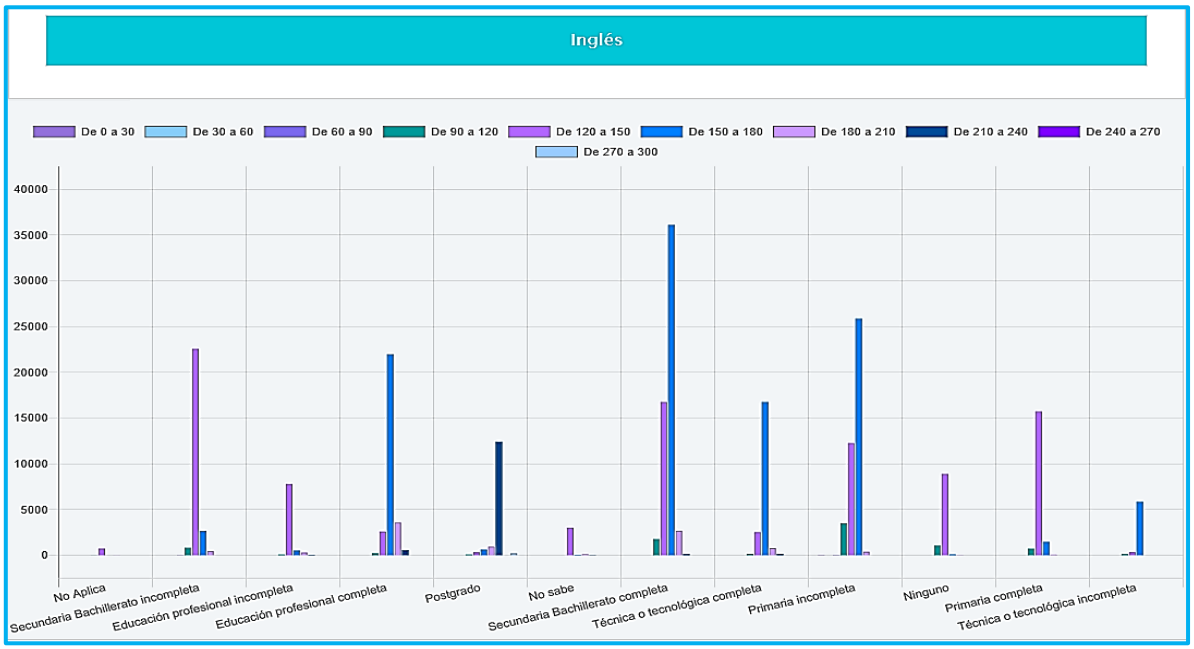


Figura 253. Inglés vs Educación padre, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

k) Dedicación lectura diaria

En el indicador de Dedicación de lectura diaria se muestra que la mayoría de estudiantes leen menos de 60 minutos en promedio. Asimismo, en cada una de las categorías son mayoría el género femenino.



Figura 254. Género vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, se puede ver que los estudiantes que más leen son los que mejores resultados obtienen, es el caso de la categoría de Más de 2 horas donde su cluster más representativo es el comprendido entre 180 a 210, seguido por Entre 1 y 2 horas con un rango de 150 a 180, y finalmente las demás categorías presentan resultados entre 120 y 150.

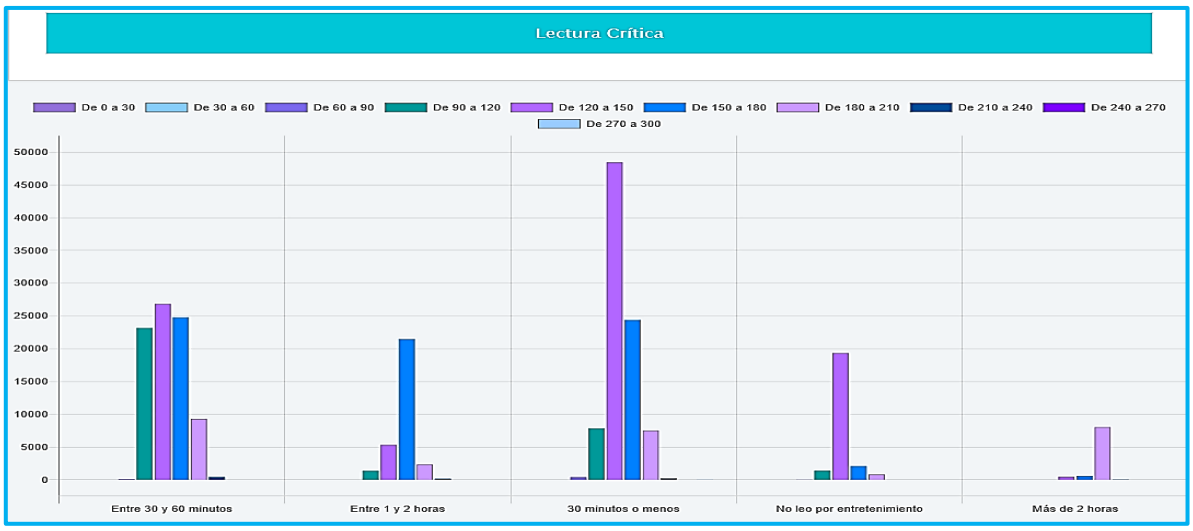


Figura 255. Lectura crítica vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, las categorías de 30 minutos o menos y Más de 2 horas son quienes obtienen resultados más bajos al tener como clúster representativo el comprendido entre 120 y 150. Mientras que, las demás de 150 a 180.

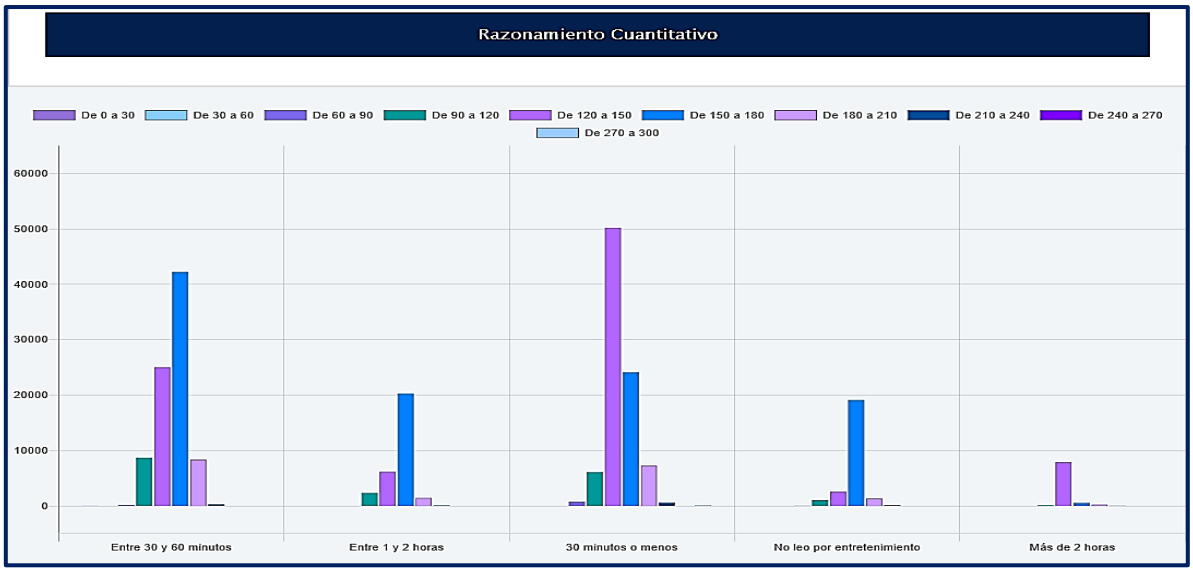


Figura 256. Razonamiento cuantitativo vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En comunicación escrita, se destaca Entre 1 y 2 horas por encima de las otras categorías al obtener resultados de 150 a 180, es decir un clúster más alto que las demás.

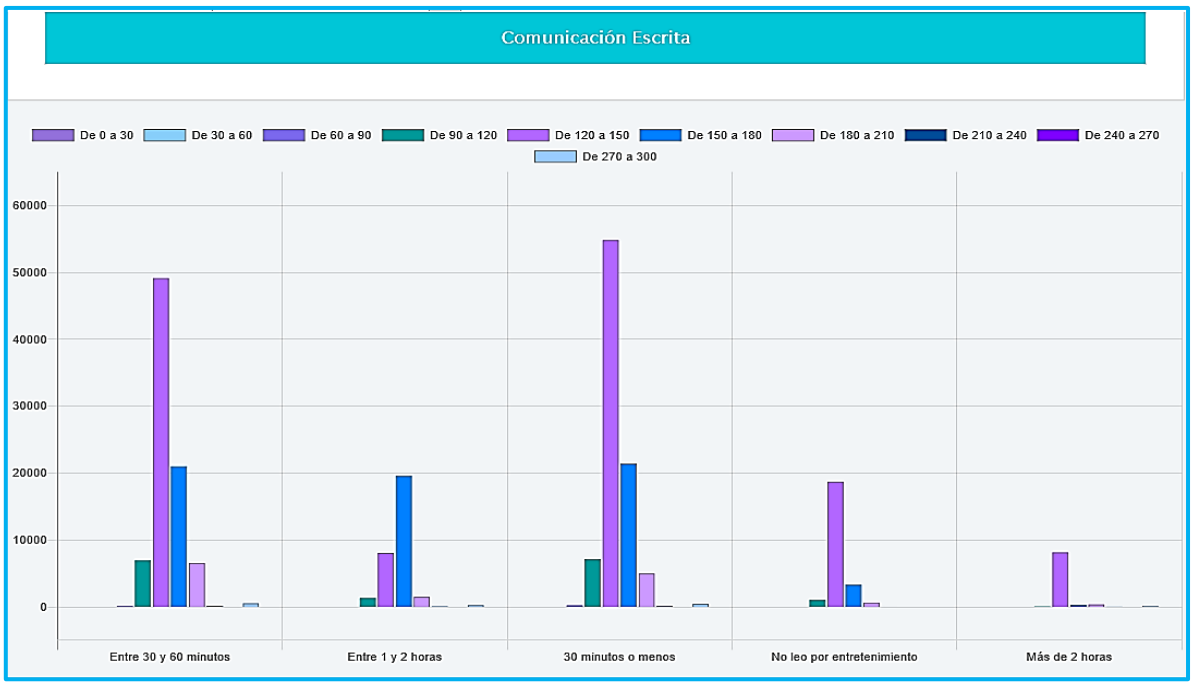


Figura 257. Comunicación escrita vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En competencias ciudadanas, el rango promedio es entre 150 y 180. Pero, las categorías 30 minutos o menos y No leo por entretenimiento presentan resultados de 120 a 150, siendo estos más bajos.

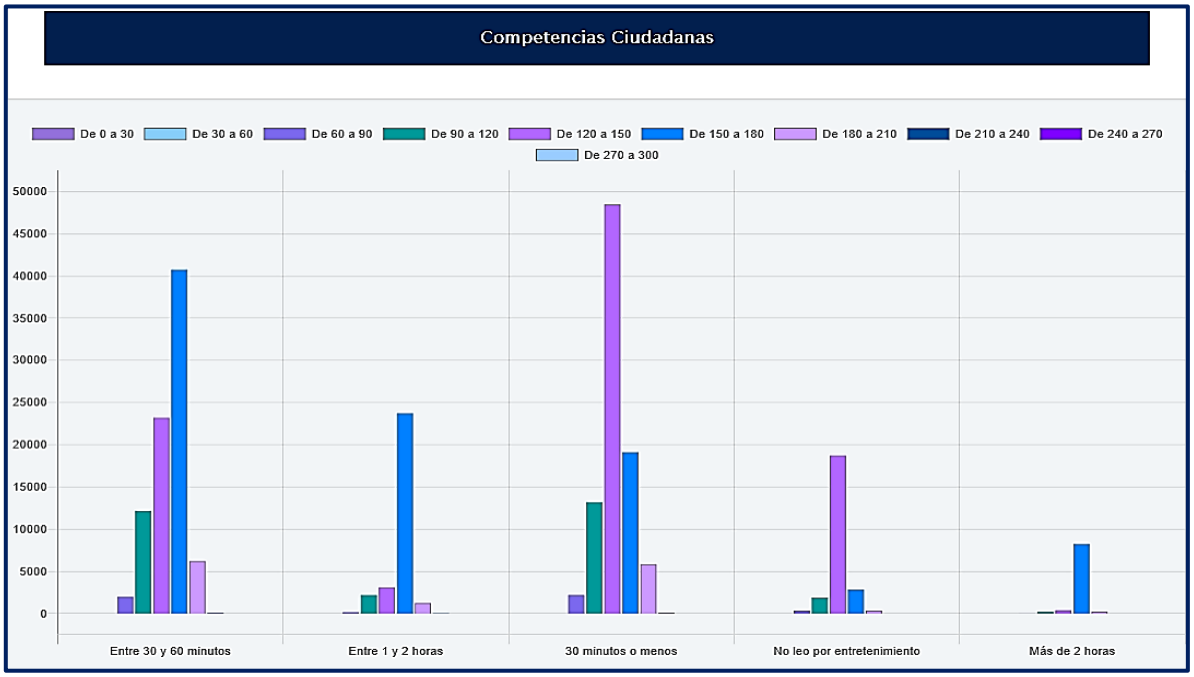


Figura 258. Competencias ciudadanas vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se evidencia que la categoría de 30 minutos o menos es la que mejor resultados obtiene, estando un rango por encima de las demás categorías evaluadas.

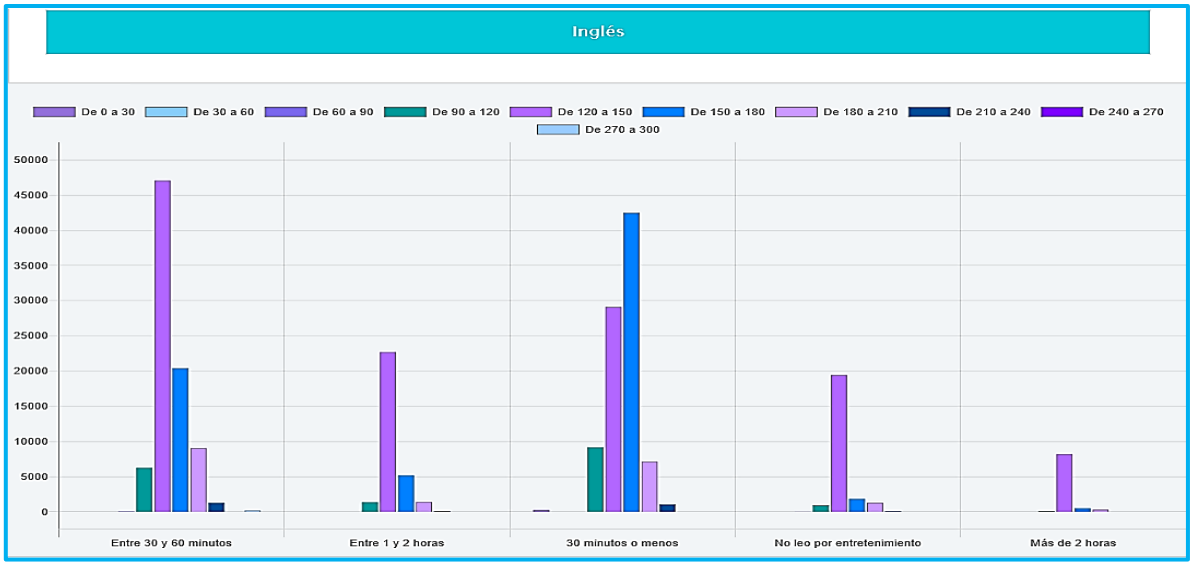


Figura 259. Inglés vs Dedicación lectura diaria, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

l) Área de la carrera

En este indicador, se muestra que las áreas con mayor número de estudiantes son: Administración y afines, Ingeniería, Contaduría y afines, Psicología, Educación y Derecho. Asimismo, en Ingeniería, Humanidades, Economía, Recreación y deportes y Ciencias militares y navales hay mayor presencia del género masculino que del femenino.

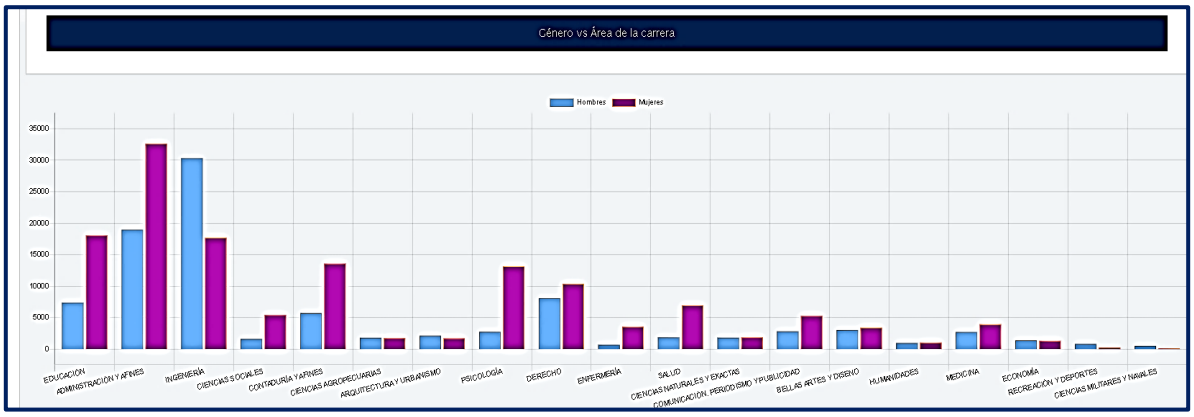


Figura 260. Género vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Lectura crítica, las áreas que tienen mejores resultados son: Ciencias naturales y exactas, Comunicación, periodismo y publicidad, Medicina, Economía y Ciencias militares y navales. Mientras que, el área que obtiene los puntajes más bajos es Ingeniería. Las demás categorías se encuentran dentro del promedio.

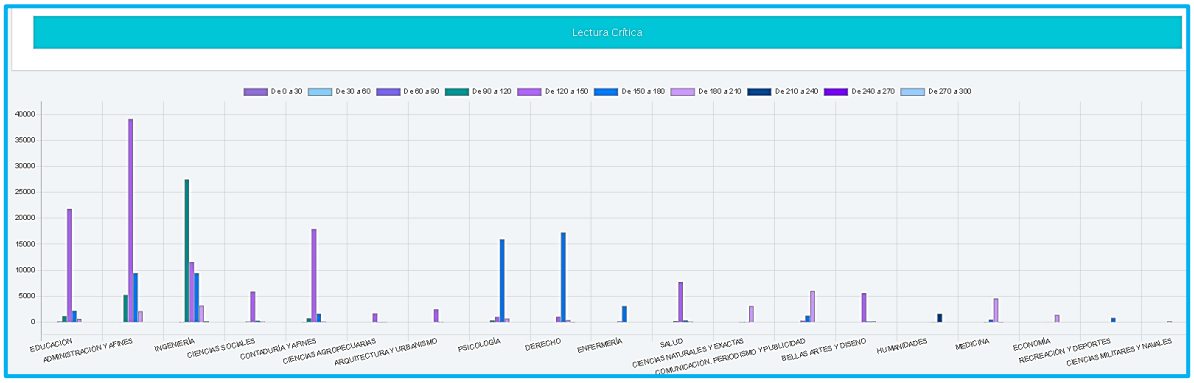


Figura 261. Lectura crítica vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, se destacan positivamente las áreas de Ciencias naturales y exactas, Medicina y Economía. Caso contrario sucede con Derecho y Enfermería, quienes tienen un clúster comprendido entre 90 y 120 puntos.

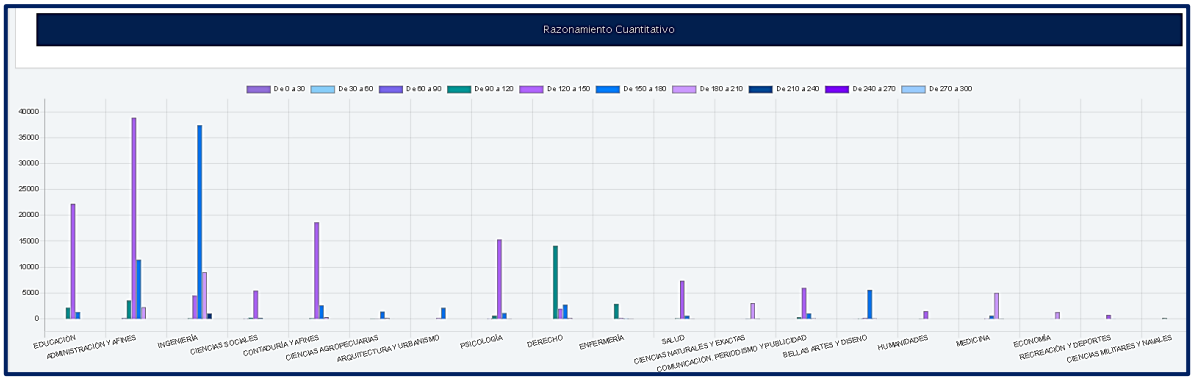


Figura 262. Razonamiento cuantitativo vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Comunicación escrita, las diferentes categorías obtienen resultados comprendidos entre 120 y 180. Sin embargo, las áreas de Ciencias agropecuarias, Ciencias naturales y exactas y Bellas artes y diseño, son quienes presentan puntajes por debajo del promedio.

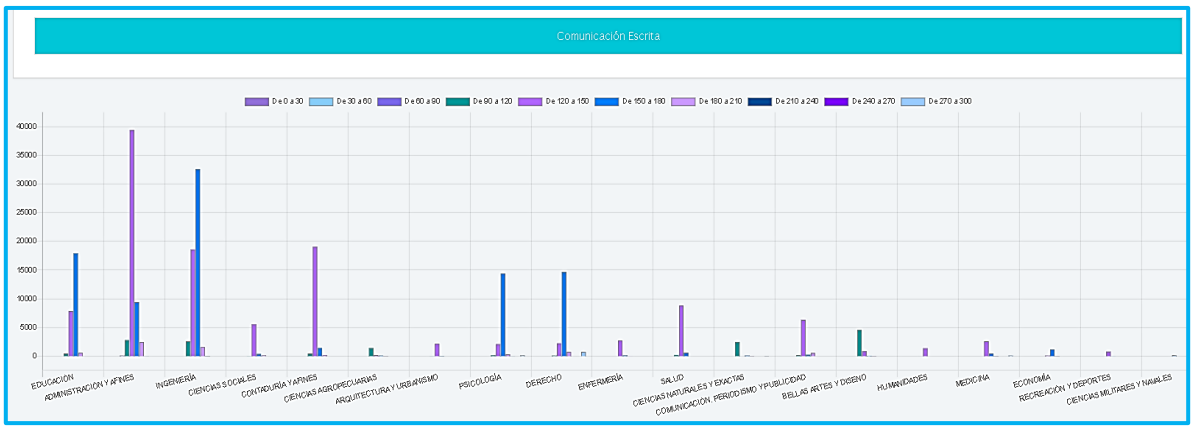


Figura 263. Comunicación escrita vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se observa que las áreas de Ciencias naturales y exactas, Humanidades, Medicina, Ciencias militares y navales y Economía obtienen un clúster entre 180 y 210. Por otra parte, el rendimiento es mucho más bajo para las áreas de Educación, Administración y afines, Ingeniería y Salud.

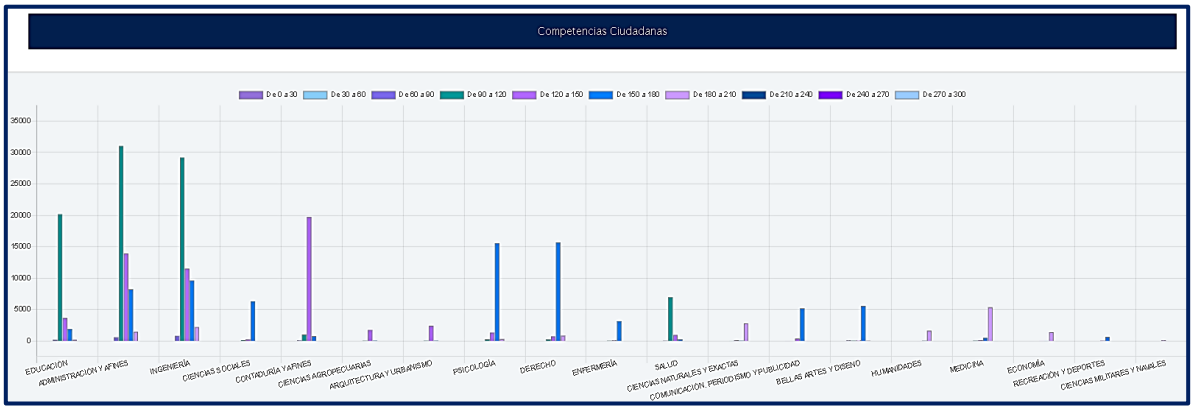


Figura 264. Competencias ciudadanas vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando con este indicador se encuentra que, para el núcleo de inglés, las áreas de Ciencias naturales y exactas y Bellas artes y diseño son las más destacadas con un rango promedio de 180 a 210. Por el contrario, las áreas de Educación y Ciencias sociales presentan resultados entre 90 y 120, siendo estos los más bajos.

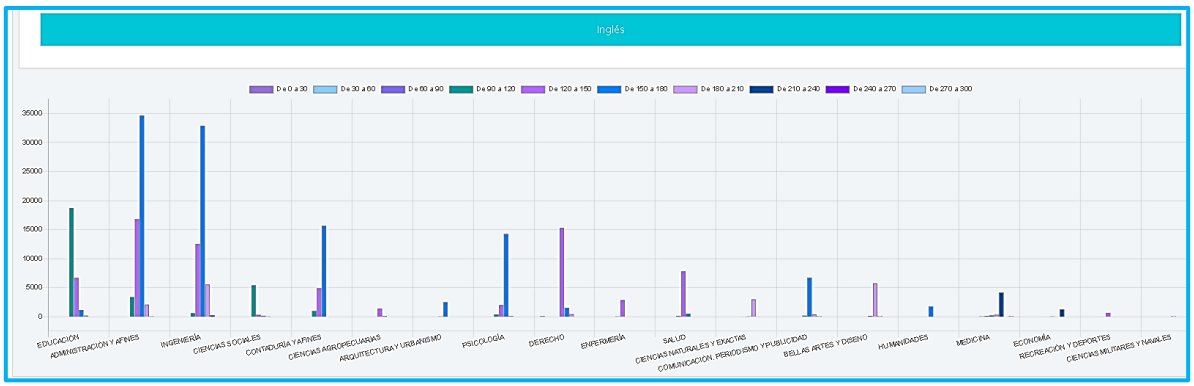


Figura 265. Inglés vs Área de la carrera, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

m) Método programa

En el indicador de Método de programa, se observa que la mayoría de estudiantes asisten a clases de forma presencial, seguidos por la categoria Distancia y Distancia virtual.

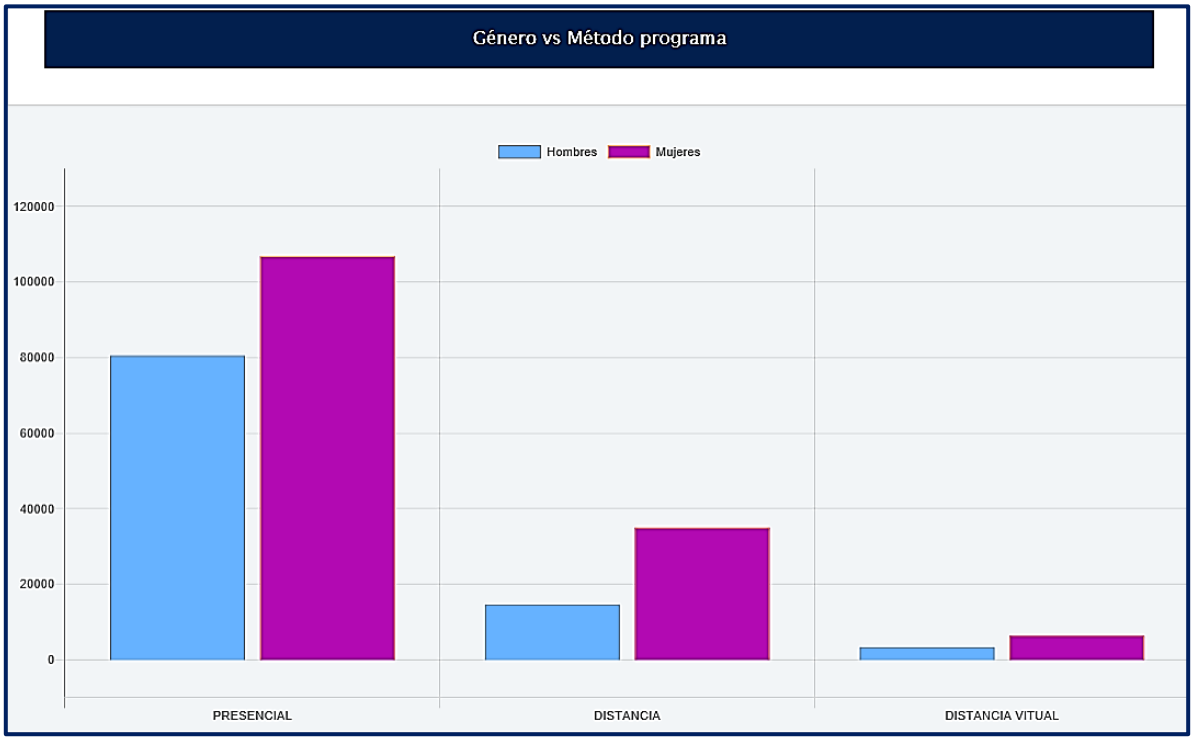


Figura 266. Género vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Lectura crítica, a los estudiantes de forma presencial les va mejor en los resultados de la prueba que las otras dos categorías, obteniendo puntajes dentro del rango de 150 y 180. Mientras que, los demás tienen como clúster más representativo el comprendido entre 120 y 150.

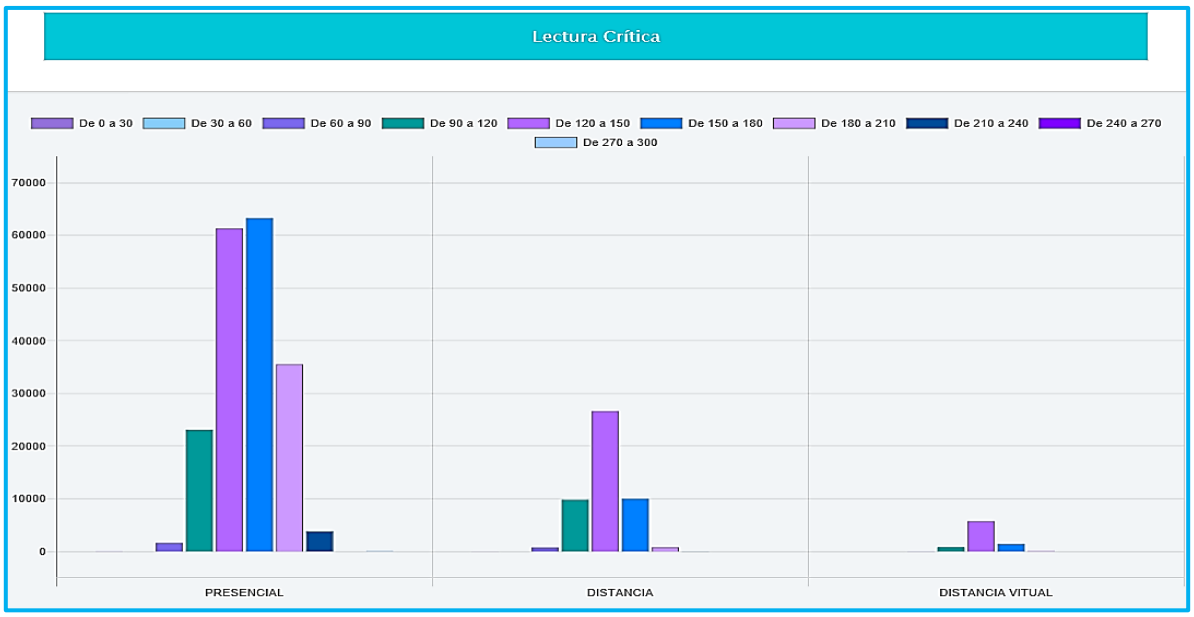


Figura 267. Lectura crítica vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, el comportamiento presentado es similar al de Lectura crítica donde quienes se destacan positivamente son los estudiantes que estudian de forma presencial.

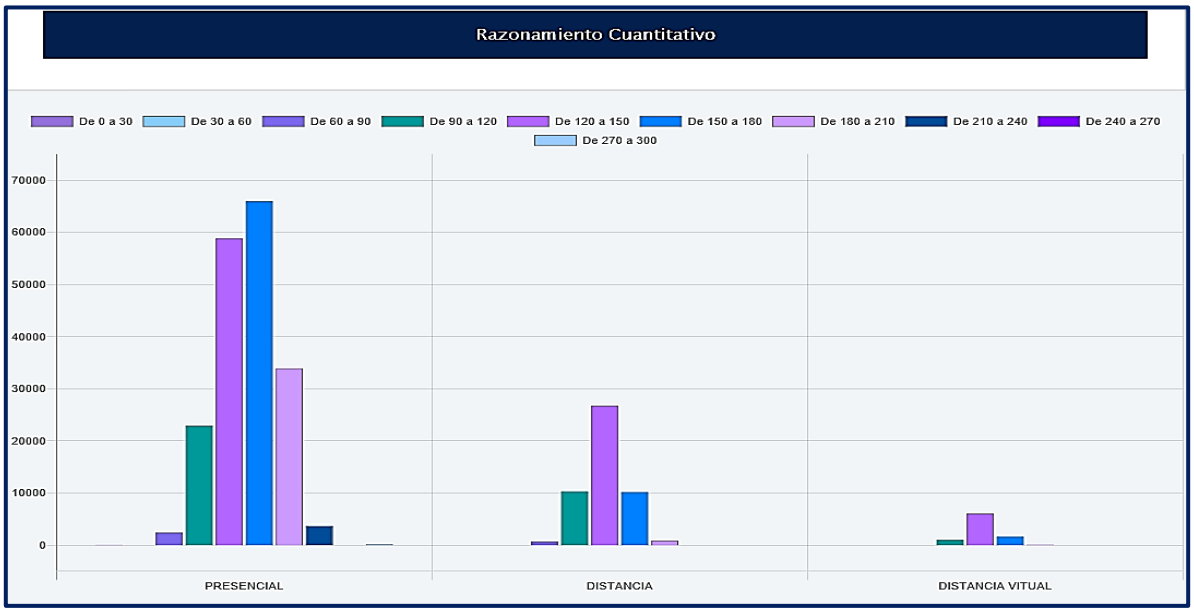


Figura 268. Razonamiento cuantitativo vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, las tres categorías tienen como clúster representativo el comprendido entre 120 y 150. Sin embargo, los estudiantes de la categoria presencial alcanzan a obtener resultados en rangos más altos.

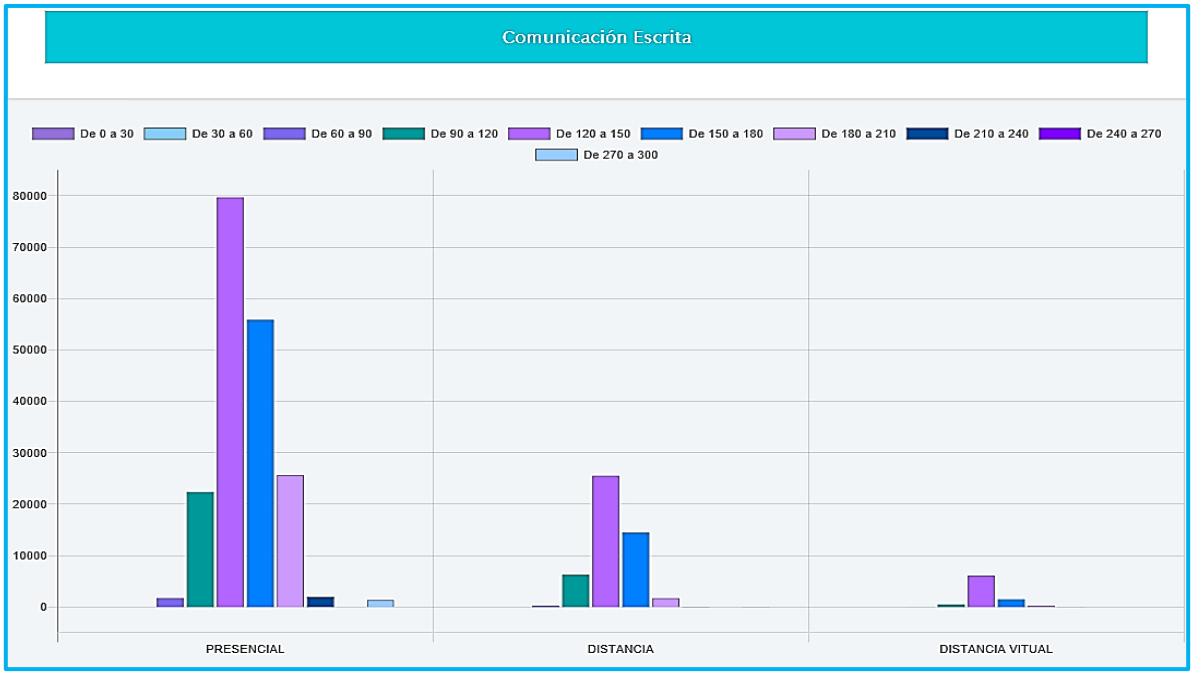


Figura 269. Comunicación escrita vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se presenta el mismo comportamiento que en el núcleo de Comunicación escrita, con la diferencia de que en la categoria Presencial disminuyen los puntajes altos.

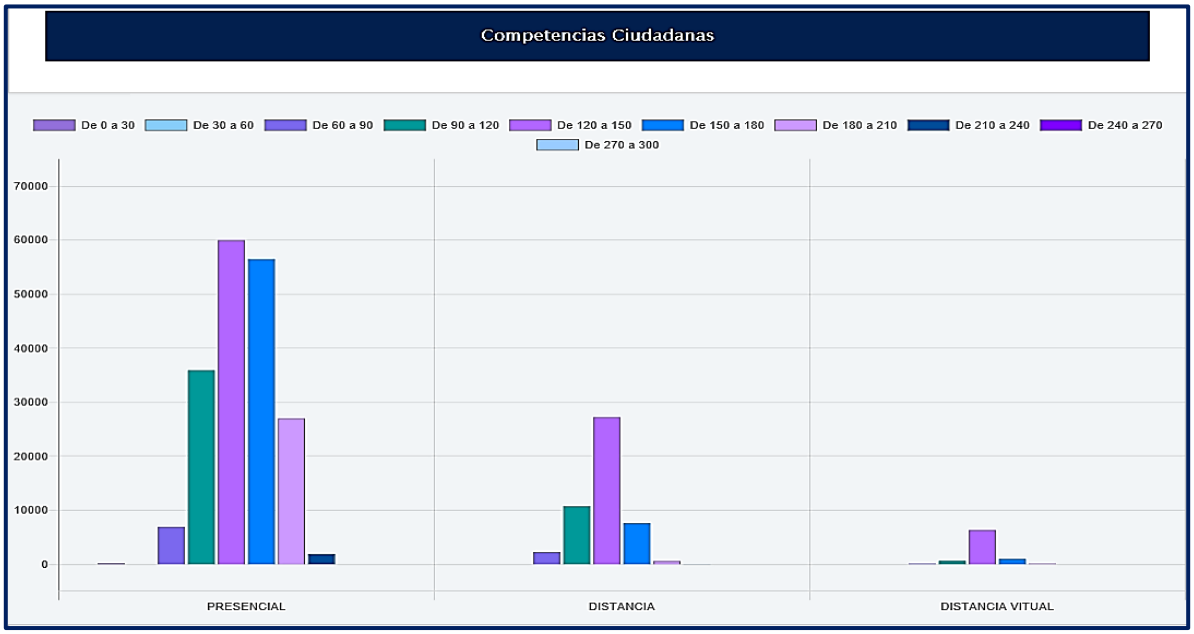


Figura 270. Competencias ciudadanas vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Finalizando con el análisis descriptivo de esta prueba, se puede decir que el núcleo temático de inglés los estudiantes tienden a presentar puntajes promedio de 120 a 150 puntos. Además, se alcanzan a presentar resultados en los rangos extremos de calificación.

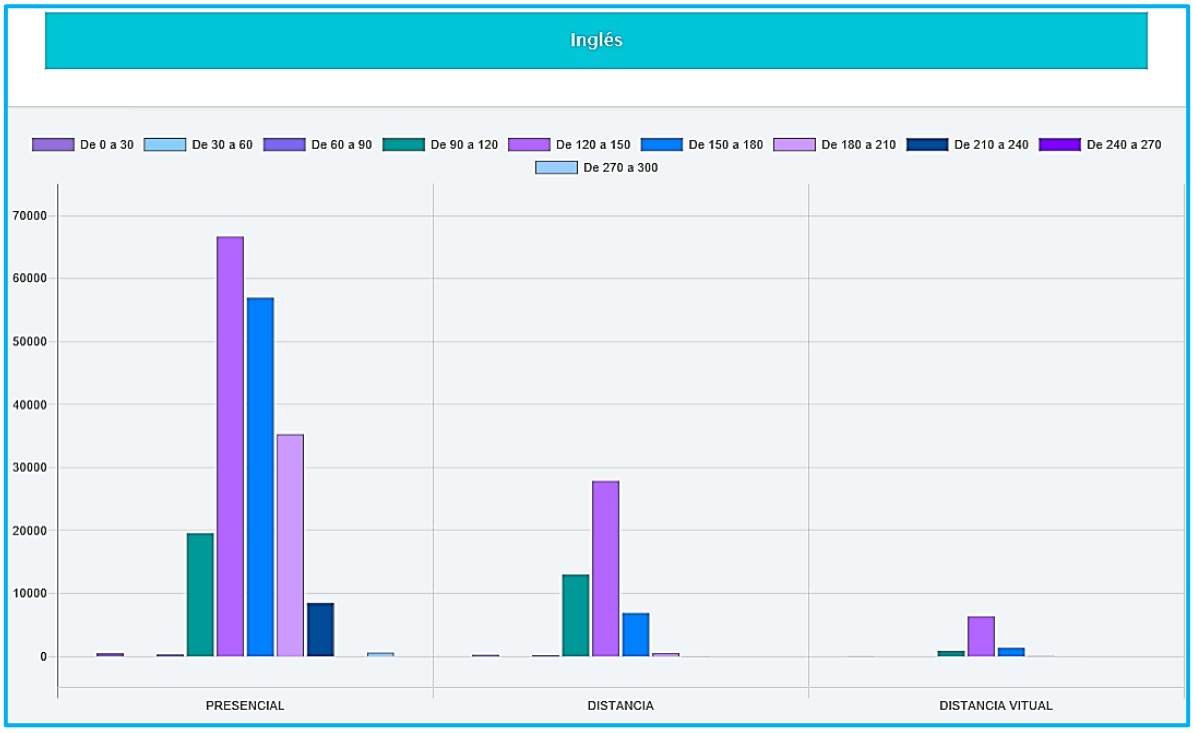


Figura 271. Inglés vs Método programa, Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

V. MODELO PREDICTIVO

El primer algoritmo evaluado de árboles de decisión fue Random Forest. Este algoritmo logra clasificar la información en una gran cantidad de árboles generados por vectores totalmente aleatorios, asimismo sus reglas varían según el árbol que haya sido creado. Debido a la poca probabilidad de que una regla vuelva a presentarse y a las limitaciones encontradas en este algoritmo, como lo son la poca posibilidad de cambio de las variables y la baja adaptabilidad en la cantidad de niveles del árbol, no es útil para el desarrollo del proyecto.

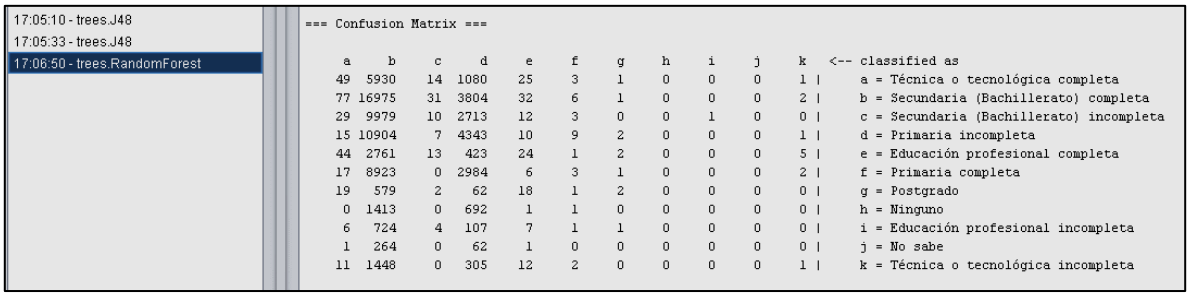


Figura 272. Matriz de confusión Random Forest
Autores: (Manrique, Celeita, Barahona, 2020)

En la figura 273 se muestra una prueba realizada de este algoritmo, donde se hace uso de las variables Género, Educación padre y puntaje en el núcleo razonamiento cuantitativo, en la cual se evidencia que, a pesar de que las variables son pocas, el árbol generado es de una gran extensión, lo cual complica su manejo.

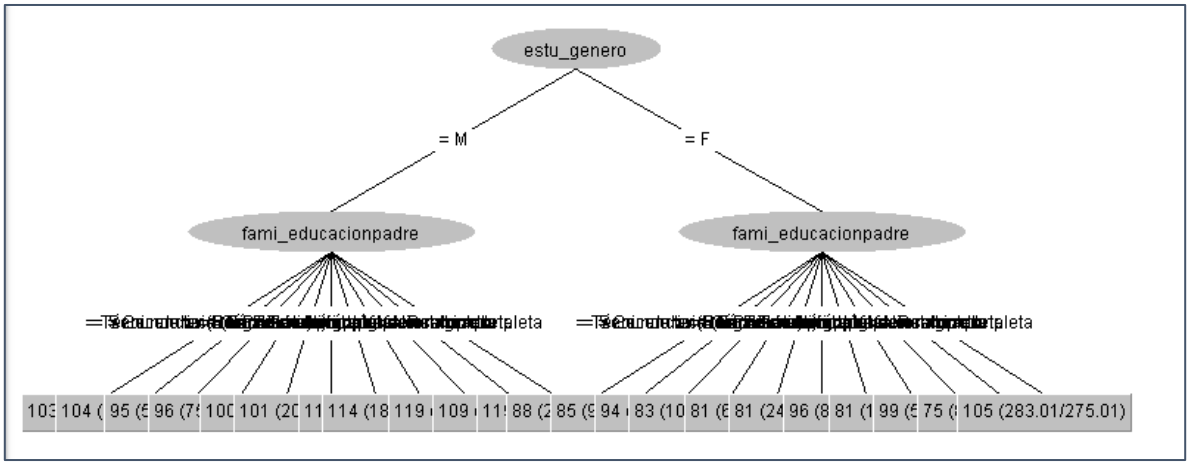


Figura 273. Random Tree WEKA
Autores: (Manrique, Celeita, Barahona, 2020)

Random Tree es el segundo algoritmo evaluado, el cual presenta características similares al algoritmo Random Forest, con la diferencia de que sus reglas son más comprensibles y de más fácil manejo.

```
RandomTree
==========

estu_genero = M
|   fami_educacionmadre = Técnica o tecnológica completa : 93 (3399/3326)
|   fami_educacionmadre = Secundaria (Bachillerato) completa : 100 (10625.94/10449.94)
|   fami_educacionmadre = Secundaria (Bachillerato) incompleta : 95 (5950.94/5841.94)
|   fami_educacionmadre = Primaria incompleta : 101 (6318.47/6204.47)
|   fami_educacionmadre = Educación profesional completa : 107 (1776/1742)
|   fami_educacionmadre = Primaria completa : 100 (5487.47/5383.47)
|   fami_educacionmadre = Postgrado : 93 (365/354)
|   fami_educacionmadre = Ninguno : 79 (929.47/907.47)
|   fami_educacionmadre = Educación profesional incompleta : 128 (412/402)
|   fami_educacionmadre = No sabe : 73 (197/191)
|   fami_educacionmadre = Técnica o tecnológica incompleta : 96 (834/812)
estu_genero = F
|   fami_educacionmadre = Técnica o tecnológica completa : 88 (3704.09/3621.09)
|   fami_educacionmadre = Secundaria (Bachillerato) completa : 93 (10302.31/10086.31)
|   fami_educacionmadre = Secundaria (Bachillerato) incompleta : 85 (6796.22/6649.22)
|   fami_educacionmadre = Primaria incompleta : 83 (8972.75/8768.75)
|   fami_educacionmadre = Educación profesional completa : 85 (1497.04/1464.04)
|   fami_educacionmadre = Primaria completa : 81 (6448.69/6306.69)
|   fami_educacionmadre = Postgrado : 106 (318.01/307.01)
|   fami_educacionmadre = Ninguno : 81 (1177.56/1142.56)
|   fami_educacionmadre = Educación profesional incompleta : 100 (438.01/427.01)
|   fami_educacionmadre = No sabe : 75 (131/125)
|   fami_educacionmadre = Técnica o tecnológica incompleta : 88 (945.02/915.02)

Size of the tree : 25
```

Figura 274. Reglas Random Tree
Autores: (Manrique, Celeita, Barahona, 2020)

Pero, a pesar de lo anterior, aunque sean seleccionados los mismos indicadores es poco frecuente que una regla vuelva a presentarse. A su vez. Se encuentra bastante limitada la documentación sobre su uso, lo cual complica la implementación en el proyecto.

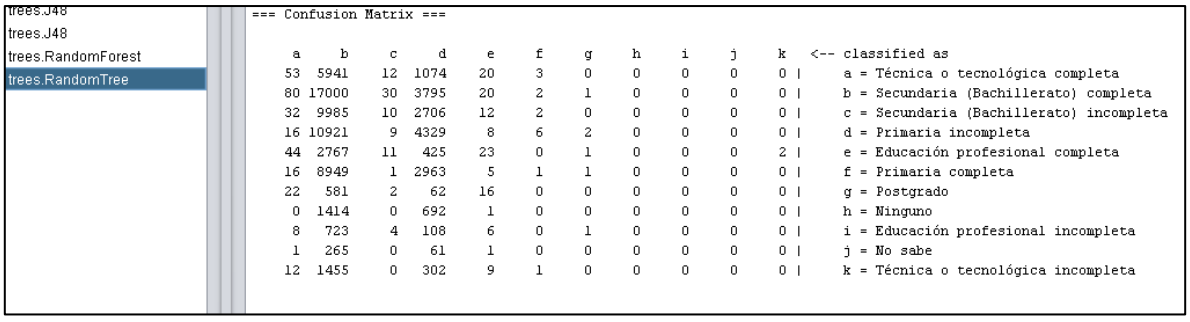


Figura 275. Matriz de confusión Random Tree
Autores: (Manrique, Celeita, Barahona, 2020)

En la documentación sobre análisis predictivo, el árbol J-48 es el algoritmo más recomendado para este procedimiento. Esto se debe a la alta precisión que ha obtenido en los diversos comparativos y a su fácil uso. Para la primera prueba con este algoritmo se manejaron las variables Género, educación madre y puntaje en el núcleo razonamiento cuantitativo. Como resultado obtuvimos un árbol de pocos niveles, pero con gran cantidad de ramas.

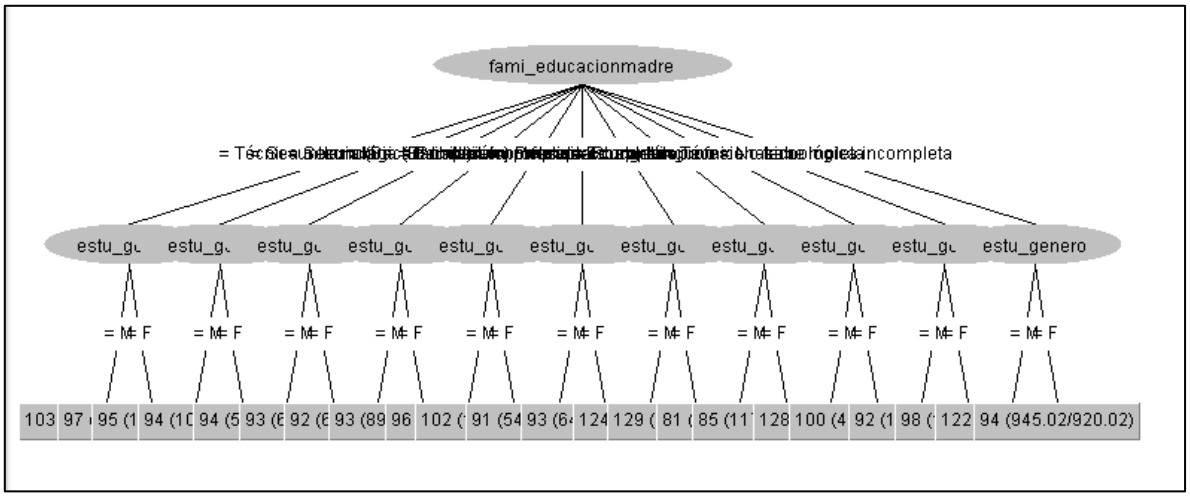


Figura 276. Árbol J48 WEKA
Autores: (Manrique, Celeita, Barahona, 2020)

Pese a su gran extensión, es muy sencilla la comprensión de sus reglas, pero, al aumentar las variables que se van a analizar, su error y su extensión aumentan considerablemente, como se muestra a continuación.

En la matriz de confusión se puede evidenciar la cantidad de registros clasificados según las categorías. Los registros que se encuentren en la diagonal de la matriz son aquellos que tienen una clasificación correcta y como se puede observar solo 20928 de todas las instancias están debidamente clasificadas. Esto nos indica que deben ser controladas las variables que van a ser analizadas, para que de esta forma disminuya el error presentado.

```
=== Confusion Matrix ===

    a     b     c     d     e     f     g     h     i     j     k   <-- classified as
    0  7103     0     0     0     0     0     0     0     0     0 |     a = Técnica o tecnológica completa
    0 20928     0     0     0     0     0     0     0     0     0 |     b = Secundaria (Bachillerato) completa
    0 12747     0     0     0     0     0     0     0     0     0 |     c = Secundaria (Bachillerato) incompleta
    0 15291     0     0     0     0     0     0     0     0     0 |     d = Primaria incompleta
    0  3273     0     0     0     0     0     0     0     0     0 |     e = Educación profesional completa
    0 11936     0     0     0     0     0     0     0     0     0 |     f = Primaria completa
    0   683     0     0     0     0     0     0     0     0     0 |     g = Postgrado
    0  2107     0     0     0     0     0     0     0     0     0 |     h = Ninguno
    0   850     0     0     0     0     0     0     0     0     0 |     i = Educación profesional incompleta
    0   328     0     0     0     0     0     0     0     0     0 |     j = No sabe
    0  1779     0     0     0     0     0     0     0     0     0 |     k = Técnica o tecnológica incompleta
```

Figura 277. Matriz de confusión J48
Autores: (Manrique, Celeita, Barahona, 2020)

Al realizar un análisis manejando como variables: un indicador, los puntajes obtenidos en alguno de los núcleos temáticos y una de las opciones de género de los estudiantes, se disminuye considerablemente el error generado, y como resultado se obtiene un árbol bastante fácil de comprender y de recorrer.

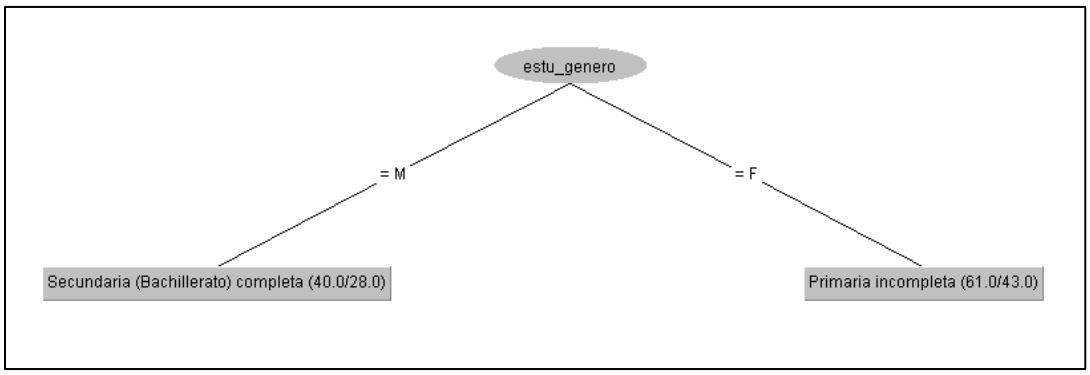


Figura 278. Árbol filtrado J48
Autores: (Manrique, Celeita, Barahona, 2020)

Las reglas arrojadas en este árbol, además de ser siempre las mismas, permiten predecir cual es la probabilidad de que un estudiante obtenga un puntaje específico en el núcleo temático seleccionado, todo esto dependiendo de su género y de la categoría del indicador que haya elegido. Este proceso es el implementado en Java para el proyecto.

Como se puede visualizar cada una de las categorías tienen distintas reglas donde varían la cantidad de registros y el puntaje obtenido. Para este caso se seleccionó el patrón más representativo de cada una de las categorías, el cual contiene la mayor cantidad de registros.

```
punt_sociales = 92: Educación profesional completa (460.01/205.01)
punt_sociales = 80: Educación profesional completa (1051.02/575.02)
punt_sociales = 51: Secundaria Bachillerato completa (3058.06/2148.06)
punt_sociales = 49: Secundaria Bachillerato completa (3409.06/2412.06)
punt_sociales = 60: Secundaria Bachillerato completa (2233.04/1586.04)
punt_sociales = 65: Secundaria Bachillerato completa (1467.03/1059.03)
punt_sociales = 90: Educación profesional completa (573.01/286.01)
punt_sociales = 79: Educación profesional completa (1032.02/551.02)
punt_sociales = 67: Educación profesional completa (1368.02/1020.02)
punt_sociales = 87: Educación profesional completa (668.01/322.01)
punt_sociales = 59: Secundaria Bachillerato completa (2459.04/1769.04)
punt_sociales = 70: Educación profesional completa (1259.02/853.02)
punt_sociales = 56: Secundaria Bachillerato completa (2793.05/1965.05)
punt_sociales = 73: Educación profesional completa (1103.02/690.02)
punt_sociales = 95: Educación profesional completa (426.01/209.01)
punt_sociales = 97: Educación profesional completa (255.0/126.0)
punt_sociales = 85: Educación profesional completa (857.02/433.02)
punt_sociales = 61: Secundaria Bachillerato completa (2002.04/1446.04)
punt_sociales = 74: Educación profesional completa (1232.02/749.02)
punt_sociales = 35: Secundaria Bachillerato completa (1614.03/1195.03)
punt_sociales = 54: Secundaria Bachillerato completa (2664.05/1851.05)
punt_sociales = 50: Secundaria Bachillerato completa (2992.05/2082.05)
punt_sociales = 46: Secundaria Bachillerato completa (3252.06/2327.06)
punt_sociales = 53: Secundaria Bachillerato completa (2968.05/2075.05)
punt_sociales = 34: Secundaria Bachillerato completa (1429.03/1081.03)
punt_sociales = 45: Secundaria Bachillerato completa (3150.06/2285.06)
punt_sociales = 41: Secundaria Bachillerato completa (2577.05/1907.05)
punt_sociales = 36: Secundaria Bachillerato completa (1811.03/1374.03)
punt_sociales = 48: Secundaria Bachillerato completa (3080.06/2190.06)
punt_sociales = 44: Secundaria Bachillerato completa (2959.05/2101.05)
punt_sociales = 68: Educación profesional completa (1284.02/904.02)
punt_sociales = 37: Secundaria Bachillerato completa (1926.03/1438.03)
punt_sociales = 42: Secundaria Bachillerato completa (2839.05/2097.05)
punt_sociales = 62: Secundaria Bachillerato completa (1907.03/1391.03)
```

Figura 279. Reglas J48 Java
Autores: (Manrique, Celeita, Barahona, 2020)

Para saber con exactitud cuál es el porcentaje de probabilidad que existe con cierto patrón, se realiza una regla de 3 simple con el número de registros arrojados por la regla y el número de registros totales manejados por el árbol de decisión.

```
232             int cantR = 0;
233             for (PredictivoSuperior pred : predictivos) {
234                 cantR = cantR + Integer.parseInt(pred.getRegistro());
235             }
236             for (PredictivoSuperior pred : predictivos) {
237                 pred.setRegistro(String.valueOf(Integer.parseInt(pred.getRegistro()) * 100 / cantR));
238             }
239         }
```

Figura 280. Probabilidad de las reglas generadas
Autores: (Manrique, Celeita, Barahona, 2020)

VI. RESULTADOS MODELO PREDICTIVO

6.1. SABER 11

Los resultados encontrados a partir del análisis del algoritmo J48 se muestran a continuación representados en gráficas de torta donde se encuentran las principales reglas obtenidas de los nodos de los árboles de decisión, para cada uno de los indicadores evaluados y que son comunes para los diferentes años de presentación de la prueba Saber 11.

Las reglas obtenidas corresponden al cruce de categorías de cada indicador con los respectivos puntajes de calificación, de esta forma se encontraron los puntajes más probables que pueden llegar a presentarse para cada núcleo temático en cada una de las diferentes categorías.

a) Educación madre

Para el núcleo de Lectura crítica se encontró que los resultados de las mujeres se destacan en 4 categorías, las cuales son: Primaria incompleta (30 ptos), Secundaria bachillerato completa (54 ptos), Educación profesional completa (60 ptos) y Postgrado (86 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Educación profesional completa (64 ptos), Secundaria bachillerato completa (57 ptos) y Primaria incompleta (29 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria.

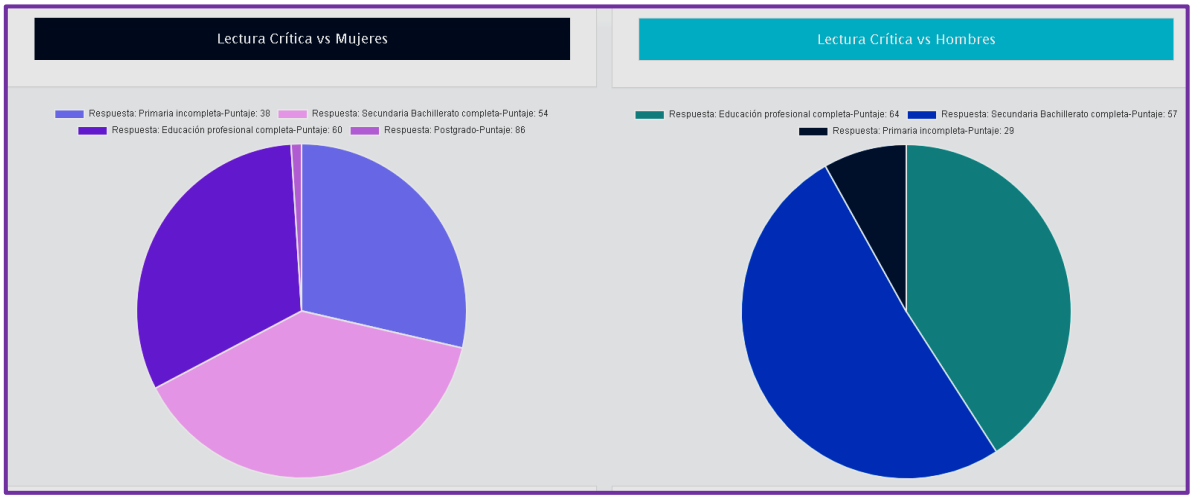


Figura 281. Lectura crítica vs Educación madre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Matemáticas, para las mujeres se encontraron probabilidades en Primaria incompleta (45 ptos), Secundaria bachillerato completa (55 ptos), Educación profesional completa (63 ptos) y Postgrado (81 ptos) donde los estudiantes que

tienen madre con este nivel educativo son quiénes tienen más probabilidad de tener un puntaje alto. Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (63 ptos), Secundaria bachillerato completa (55 ptos) y Primaria incompleta (30 ptos).



Figura 282.Matemáticas vs Educación madre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias naturales, para las mujeres se encontraron probabilidades en Primaria incompleta (39 ptos), Secundaria bachillerato completa (47 ptos), Educación profesional completa (60 ptos) y Postgrado (78 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (61 ptos), Secundaria bachillerato completa (52 ptos) y Postgrado (100 ptos).



Figura 283. Ciencias naturales vs Educación madre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias sociales, para las mujeres se encontraron probabilidades en Primaria incompleta (42 ptos), Secundaria bachillerato completa (49 ptos), Educación

profesional completa (65 ptos) y Postgrado (89 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (66 ptos), Secundaria bachillerato completa (49 ptos), Primaria incompleta (22 ptos) y Postgrado (91 ptos).

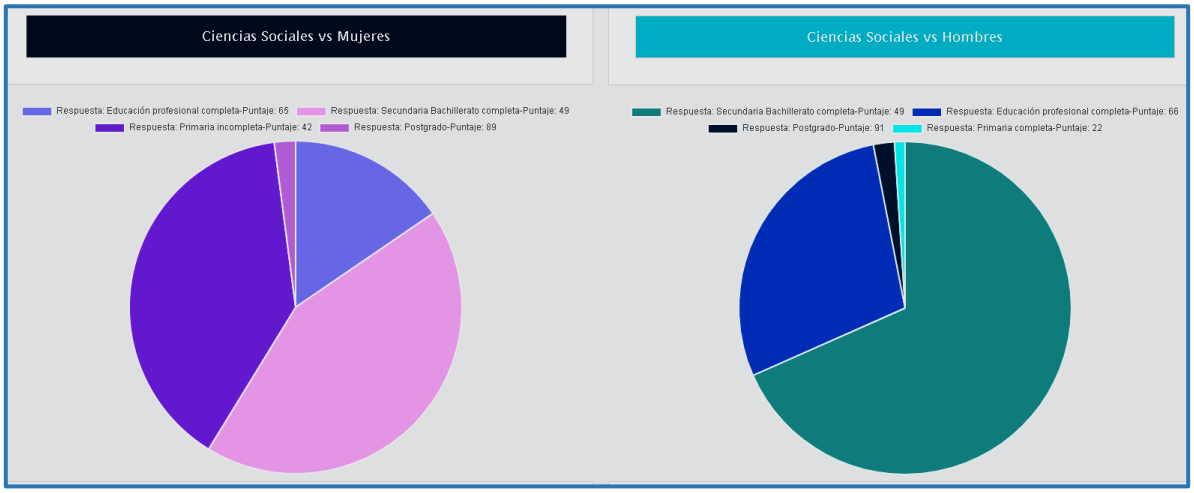


Figura 284. Ciencias sociales vs Educación madre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, para las mujeres se encontraron probabilidades en Primaria incompleta (42 ptos), Secundaria bachillerato completa (49 ptos), Educación profesional completa (61 ptos) y Postgrado (10 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (56 ptos) y Secundaria bachillerato completa (64 ptos).



Figura 285. Inglés vs Educación madre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

Las predicciones en este indicador muestran probabilidades en las mismas 4 categorías para las mujeres; mientras que, para los hombres varían sólo se mantiene 2 de ellas en todos los núcleos temáticos. Y se destaca que los

estudiantes que cuentan con madre con nivel educativo de Postgrado son quiénes más probabilidad tienen de obtener puntajes altos.

b) Educación padre

En Lectura crítica, para las mujeres se encontraron probabilidades en Primaria incompleta (47 ptos), Secundaria bachillerato completa (54 ptos), Educación profesional completa (63 ptos) y Postgrado (78 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (63 ptos), Secundaria bachillerato completa (55 ptos), Primaria incompleta (30 ptos) y Postgrado (100 ptos).

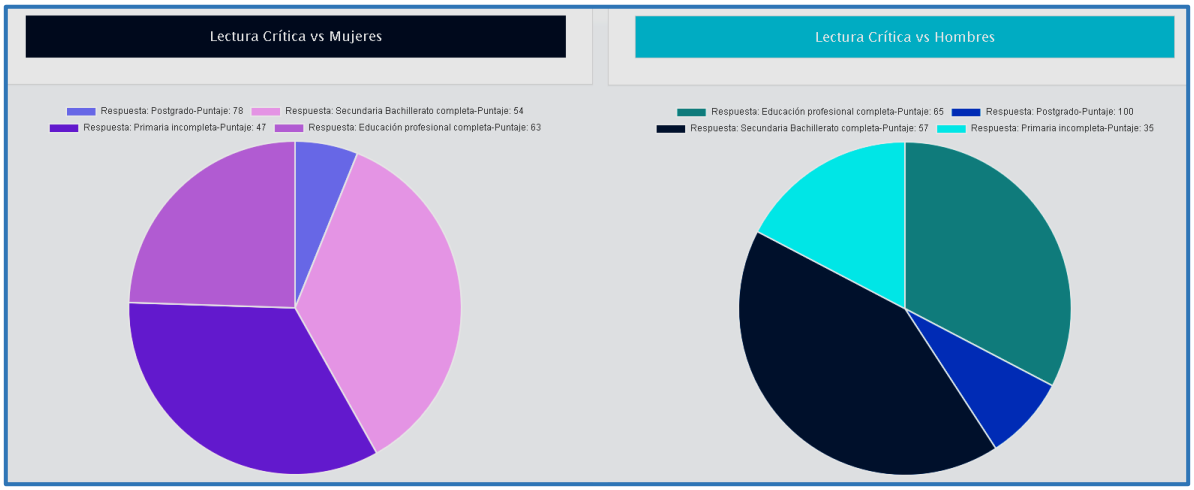


Figura 286. Lectura crítica vs Educación padre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Matemáticas, para las mujeres se encontraron probabilidades en Primaria incompleta (51 ptos), Secundaria bachillerato completa (55 ptos), Educación profesional completa (63 ptos) y Postgrado (81 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (63 ptos), Secundaria bachillerato completa (55 ptos), Primaria incompleta (43 ptos) y Postgrado (77 ptos).

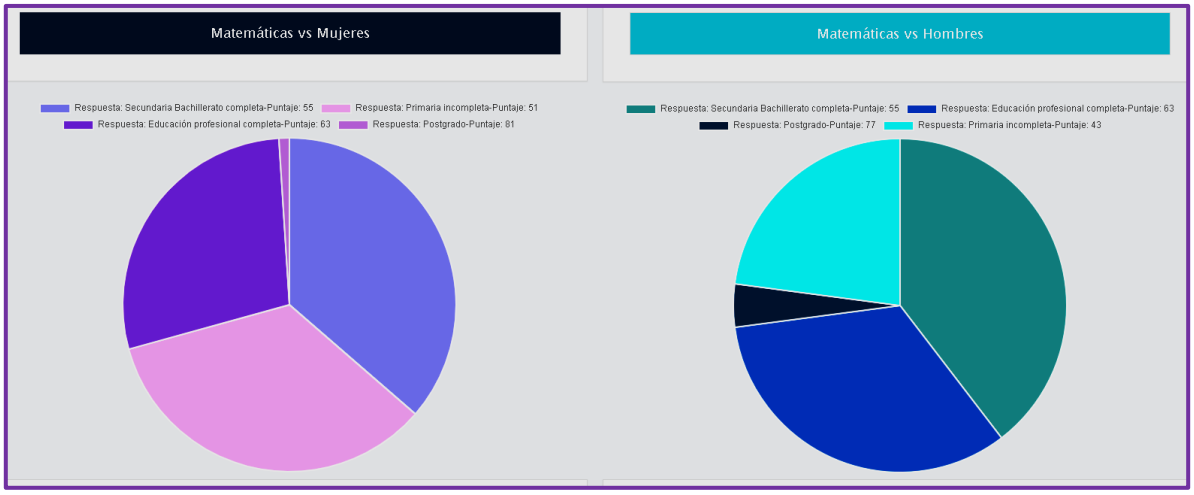


Figura 287. Matemáticas vs Educación padre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias naturales, para las mujeres se encontraron probabilidades en Primaria incompleta (47 ptos), Secundaria bachillerato completa (50 ptos), Educación profesional completa (61 ptos) y Postgrado (78 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (63 ptos), Secundaria bachillerato completa (52 ptos), Primaria incompleta (27 ptos), Secundaria bachillerato incompleta (24 ptos), y Postgrado (75 ptos).

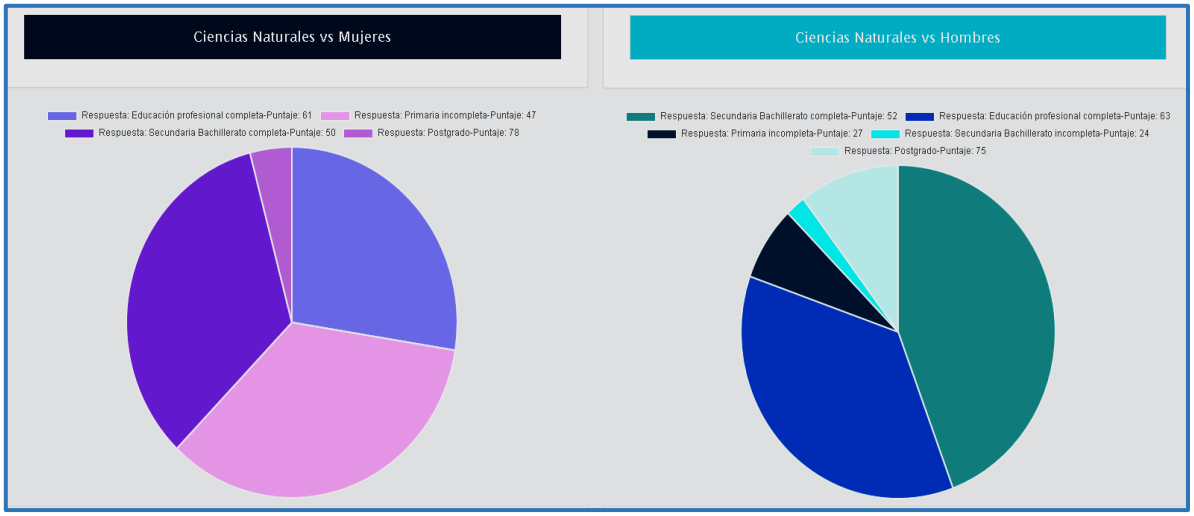


Figura 288. Ciencias naturales vs Educación padre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias sociales, para las mujeres se encontraron probabilidades en Primaria incompleta (46 ptos), Secundaria bachillerato completa (49 ptos), Educación profesional completa (66 ptos) y Postgrado (87 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (66 ptos), Secundaria bachillerato completa (49 ptos), Primaria incompleta (42 ptos), Secundaria bachillerato incompleta (25 ptos), y Postgrado (85 ptos).

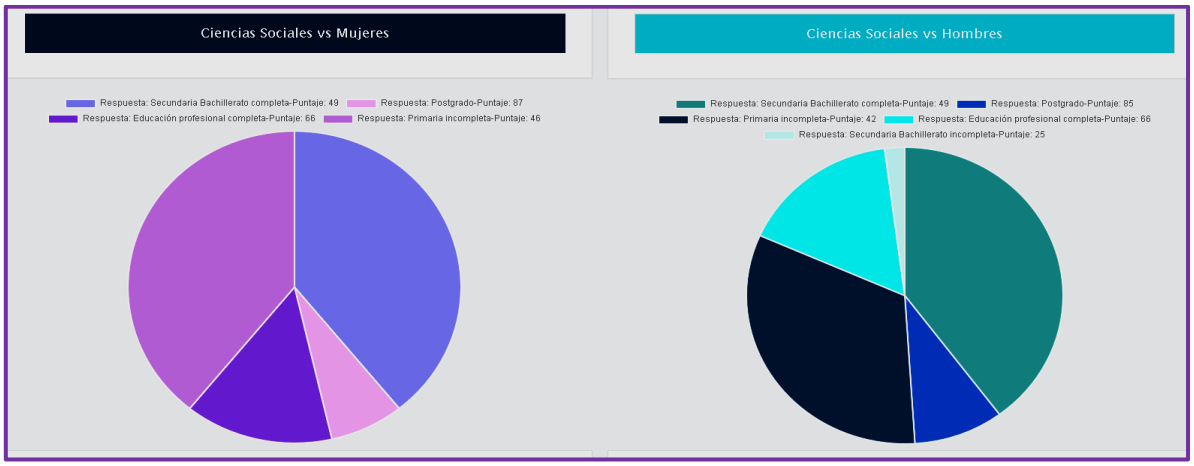


Figura 289. Ciencias sociales vs Educación padre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, para las mujeres se encontraron probabilidades en Primaria incompleta (48 ptos), Secundaria bachillerato completa (49 ptos), Educación profesional completa (61 ptos) y Postgrado (74 ptos). Además, para los hombres los puntajes con mayor proporción a presentarse son en las categorías Educación profesional completa (64 ptos), Secundaria bachillerato completa (56 ptos), Primaria incompleta (32 ptos) y Postgrado (78 ptos).



Figura 290. Inglés vs Educación padre, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

c) Estrato vivienda

En Lectura crítica, para las mujeres se encontraron probabilidades en Estrato 1 (49 ptos), en Estrato 2 (54 ptos), en Estrato 3 (68 ptos), en Estrato 4 (78 ptos), en Estrato 5 (73 ptos) y en Estrato 6 (74 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (46 ptos), en Estrato 2 (57 ptos), en Estrato 3 (68 ptos), en Estrato 4 (76 ptos), en Estrato 5 (79 ptos) y en Estrato 6 (78 ptos).



Figura 291. Lectura crítica vs Estrato de vivienda, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Matemáticas, para las mujeres se encontraron probabilidades en Estrato 1 (54 ptos), en Estrato 2 (55 ptos), en Estrato 3 (67 ptos), en Estrato 4 (74 ptos), en Estrato 5 (76 ptos) y en Estrato 6 (81 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (48 ptos), en Estrato 2 (55 ptos), en Estrato 3 (66 ptos), en Estrato 4 (75 ptos), en Estrato 5 (80 ptos) y en Estrato 6 (74 ptos).

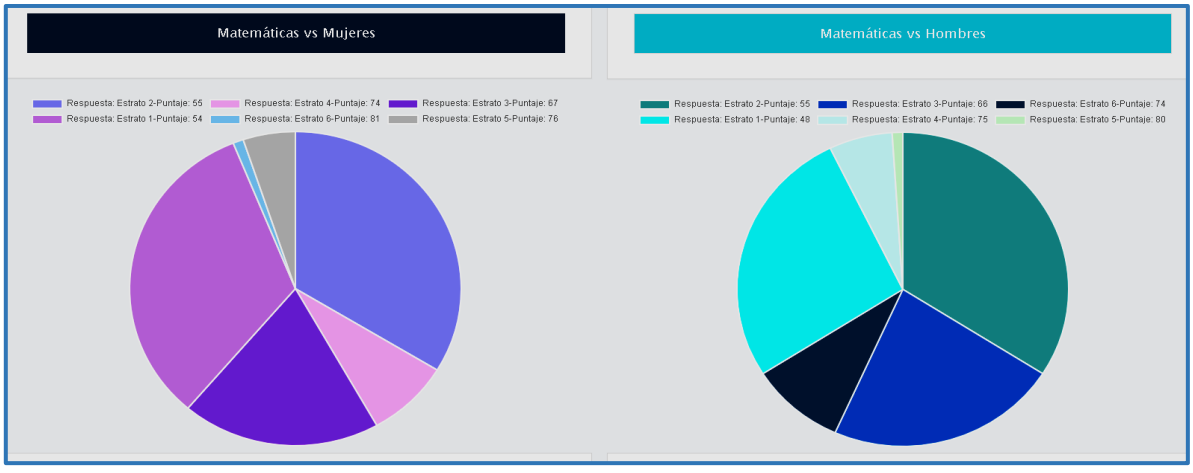


Figura 292. Matemáticas vs Estrato de vivienda, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias naturales, para las mujeres se encontraron probabilidades en Estrato 1 (47 ptos), en Estrato 2 (50 ptos), en Estrato 3 (65 ptos), en Estrato 4 (70 ptos), en Estrato 5 (80 ptos) y en Estrato 6 (74 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (43 ptos), en Estrato 2 (52 ptos), en Estrato 3 (65 ptos), en Estrato 4 (74 ptos), en Estrato 5 (80 ptos) y en Estrato 6 (76 ptos).

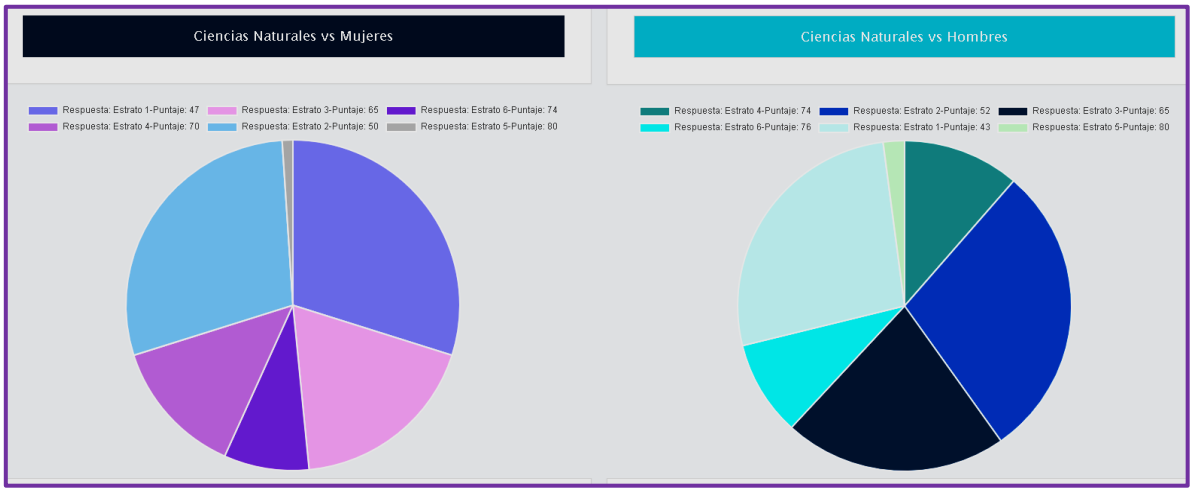


Figura 293. Ciencias naturales vs Estrato de vivienda, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En Ciencias sociales, para las mujeres se encontraron probabilidades en Estrato 1 (49 ptos), en Estrato 2 (52 ptos), en Estrato 3 (64 ptos), en Estrato 4 (77 ptos), en Estrato 5 (79 ptos) y en Estrato 6 (82 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (46 ptos), en Estrato 2 (49 ptos), en Estrato 3 (65 ptos), en Estrato 4 (79 ptos), en Estrato 5 (78 ptos) y en Estrato 6 (80 ptos).

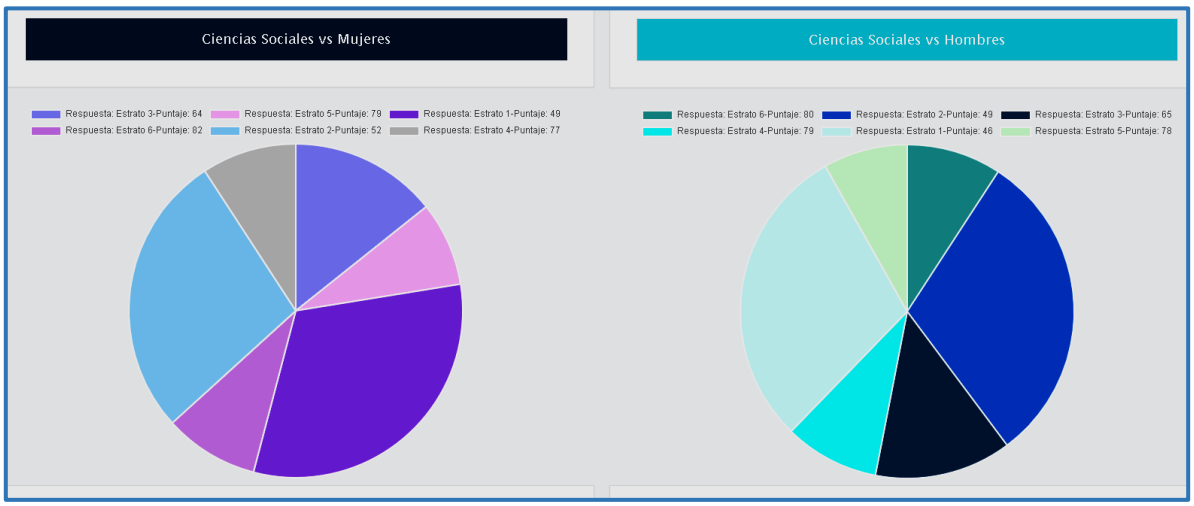


Figura 294. Ciencias sociales vs Estrato de vivienda, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, para las mujeres se encontraron probabilidades en Estrato 1 (49 ptos), en Estrato 2 (53 ptos), en Estrato 3 (64 ptos), en Estrato 4 (71 ptos), en Estrato 5 (74 ptos) y en Estrato 6 (72 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (51 ptos), en Estrato 2 (56 ptos), en Estrato 3 (66 ptos), en Estrato 4 (73 ptos), en Estrato 5 (100 ptos) y en Estrato 6 (74 ptos).

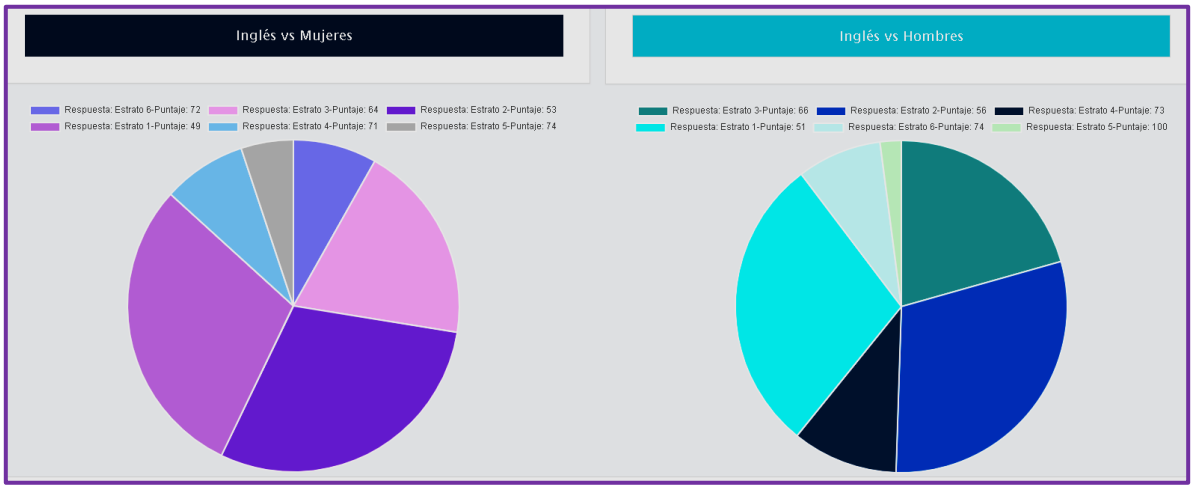


Figura 295. Inglés vs Estrato de vivienda, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

d) Tiene computador

En este indicador, para el núcleo Lectura crítica se encontró que, en el caso de las estudiantes del género femenino, tienen una probabilidad de obtener Si (54 ptos) y No (42 ptos), con un porcentaje más alto. Para los hombres la probabilidad es: Si (56 ptos) y No (40 ptos), siendo la categoría Si la más probable.



Figura 296. Lectura crítica vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Matemáticas se encontró que, en el caso de las estudiantes del género femenino, tienen una probabilidad de obtener Si (55 ptos) y No (46 ptos), con un porcentaje más alto. Para los hombres la probabilidad es: Si (55 ptos) y No (43 ptos), siendo la categoría Si la más probable.



Figura 297. Matemáticas vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Ciencias naturales se encontró que, en el caso de las estudiantes del género femenino, tienen una probabilidad de obtener Si (47 ptos) y No (41 ptos), con un porcentaje más alto. Para los hombres la probabilidad es: Si (52 ptos) y No (36 ptos), siendo la categoría Si la más probable. Los puntajes tienden a ser más que en los otros núcleos.

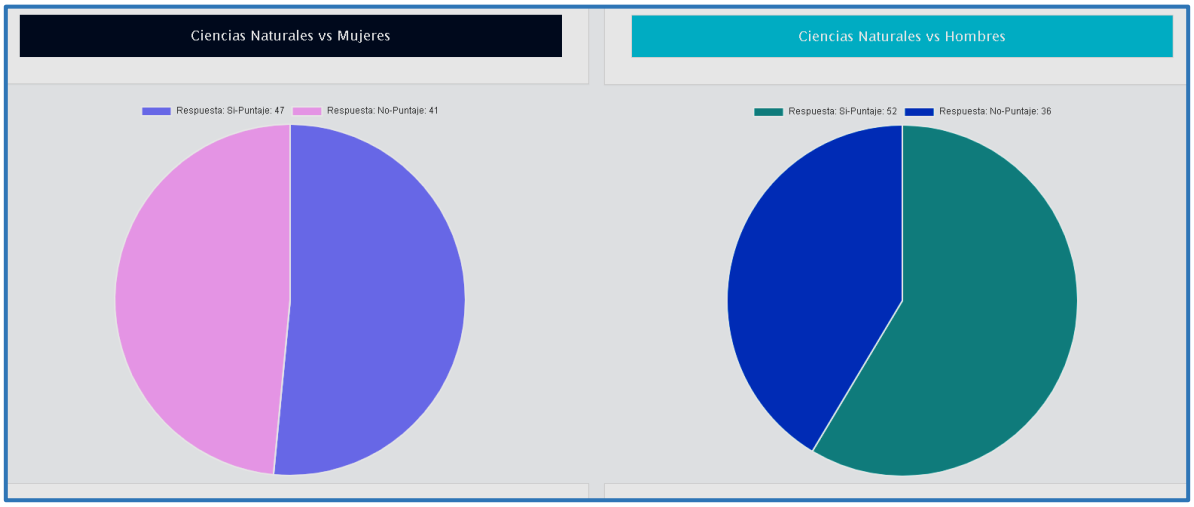


Figura 298. Ciencias naturales vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Ciencias sociales se encontró que, en el caso de las estudiantes del género femenino, tienen una probabilidad de obtener Si (49 ptos) y No (44 ptos) con un porcentaje bastante parejo entre ambas categorías. Para los hombres la probabilidad es: Si (49 ptos) y No (41 ptos), siendo la categoría Si la más probable.



Figura 299. Ciencias sociales vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de inglés se encontró que, en el caso de las estudiantes del género femenino, tienen una probabilidad de obtener Si (49 ptos) y No (45 ptos), con un porcentaje más alto. Para los hombres la probabilidad es: Si (52 ptos) y No (40 ptos), siendo la categoría Si la más probable.



Figura 300. Inglés vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

e) Jornada de colegio

Para este indicador en el núcleo de Lectura crítica se encontró que, en el caso de las mujeres, tienen probabilidades en las categorías Completa (64 ptos), Mañana (54 ptos) y Sabatina (46 ptos). Para los hombres la probabilidad es: Completa (68 ptos) y Mañana (56 ptos), siendo la categoría mañana la más probable para ambos géneros.



Figura 301. Lectura crítica vs Jornada de colegio, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Matemáticas se encontró que, en el caso de las mujeres, tienen probabilidades en las categorías Completa (63 ptos) y Mañana (55 ptos). De igual forma, para los hombres la probabilidad es: Completa (64 ptos) y Mañana (55 ptos), siendo la categoría mañana la más probable para ambos géneros.



Figura 302. Matemáticas vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Ciencias naturales se encontró que, en el caso de las mujeres, tienen probabilidades en las categorías Completa (63 ptos) y Mañana (47 ptos). De igual forma, para los hombres la probabilidad es: Completa (65 ptos) y Mañana (52 ptos), siendo la categoría mañana la más probable para ambos géneros.



Figura 303. Ciencias naturales vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de Ciencias sociales se encontró que, en el caso de las mujeres, tienen probabilidades en las categorías Completa (68 ptos), Mañana (49 ptos), Sabatina (28 ptos) y Noche (22 ptos). De igual forma, para los hombres la probabilidad es: Completa (71 ptos), Mañana (49 ptos) y Sabatina (22 ptos), siendo la categoría mañana la más probable para ambos géneros con una alta proporción.

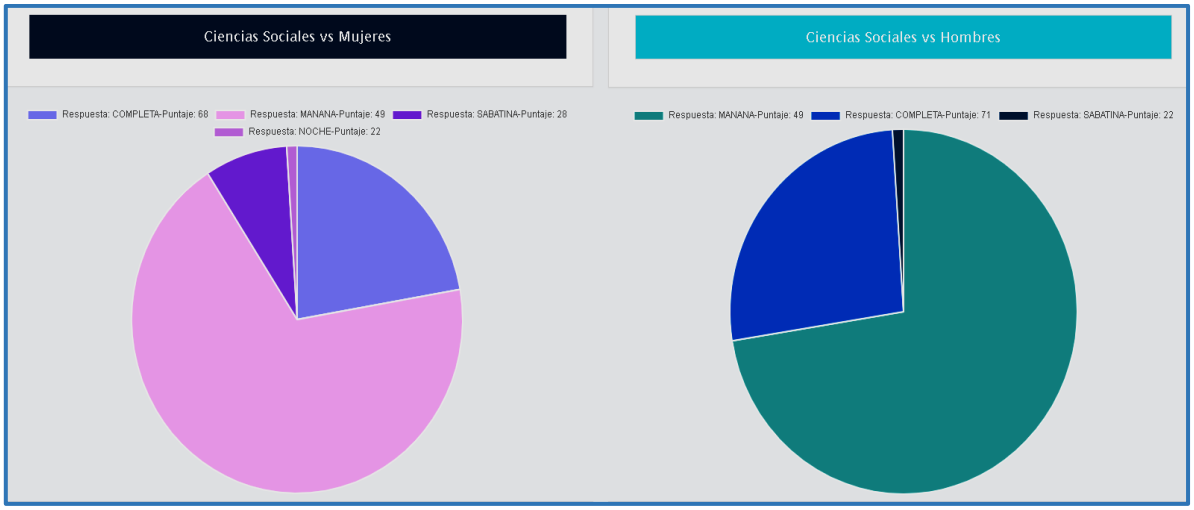


Figura 304. Ciencias sociales vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

En el núcleo de inglés se encontró que, en el caso de las mujeres, tienen probabilidades en las categorías Completa (62 ptos), Mañana (49 ptos) y Sabatina (25 ptos). De igual forma, para los hombres la probabilidad es: Completa (66 ptos) y Mañana (52 ptos), siendo la categoría mañana la más probable para ambos géneros.



Figura 305. Inglés vs Tiene computador, Predicción Saber 11
Autores: (Manrique, Celeita, Barahona, 2020)

6.2. SABER TYT

Los resultados encontrados a partir del análisis del algoritmo J48 se muestran a continuación representados en gráficas de torta donde se encuentran las principales reglas obtenidas de los nodos de los árboles de decisión, para cada uno de los indicadores evaluados y que son comunes para los diferentes años de presentación de la prueba Saber TyT.

Las reglas obtenidas corresponden al cruce de categorías de cada indicador con los respectivos puntajes de calificación, de esta forma se encontraron los puntajes más probables que pueden llegar a presentarse para cada núcleo temático en cada una de las diferentes categorías.

a) Educación madre

Para el núcleo de Lectura crítica se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (104 ptos), Secundaria bachillerato incompleta (114 ptos), Secundaria bachillerato completa (101 ptos), Primaria completa (61 ptos), Educación profesional completa (147 ptos) y Técnica y tecnológica completa (149 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (101 ptos), Secundaria bachillerato completa (150 ptos) y Primaria incompleta (66 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.

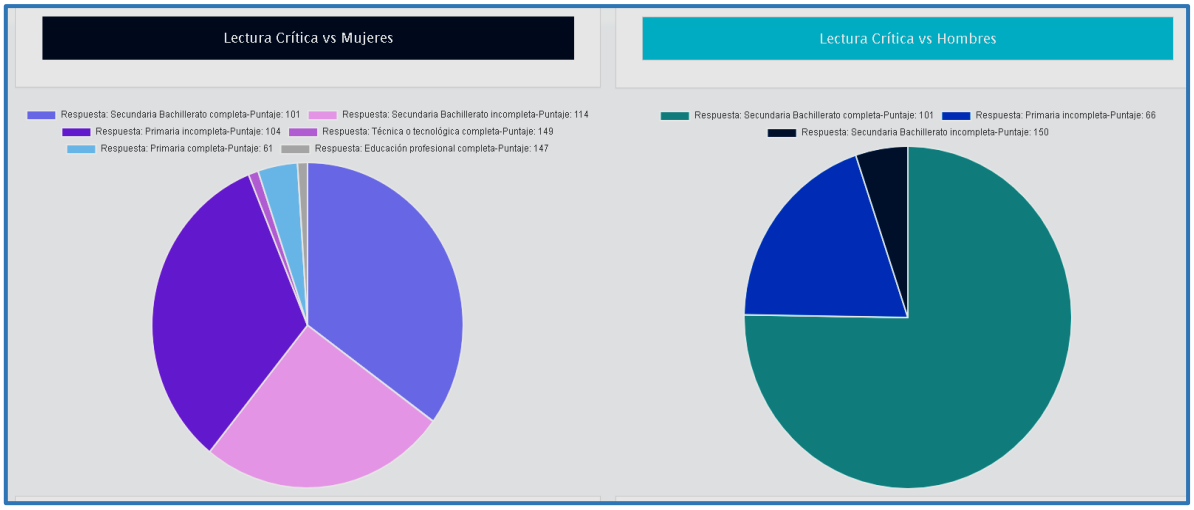


Figura 306. Lectura crítica vs Educación madre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (88 ptos) y Secundaria bachillerato completa (92 ptos), donde ambas tienen casi la misma probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (153 ptos), Secundaria bachillerato completa (102 ptos), Primaria incompleta (46 ptos) y Técnica y tecnológica completa (44 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.

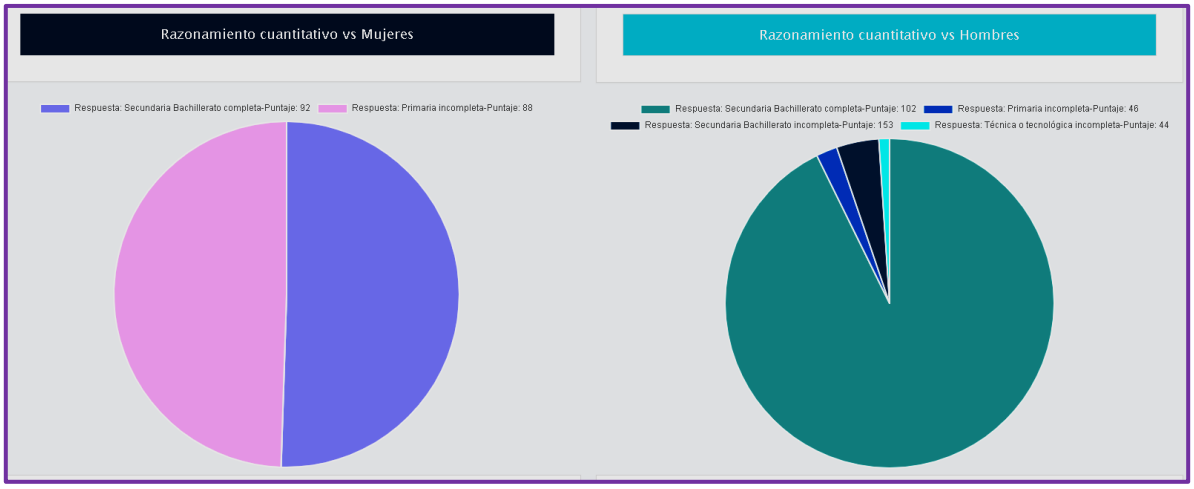


Figura 307. Razonamiento cuantitativo vs Educación madre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (106 ptos), Secundaria bachillerato incompleta (134 ptos) y Secundaria bachillerato completa (105 ptos), donde las categorías de Secundaria tienen las probabilidades más altas. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (135 ptos), Secundaria bachillerato completa (105 ptos) y Primaria incompleta (63 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.



Figura 308. Comunicación escrita vs Educación madre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (108 ptos), Primaria completa (54 ptos), Secundaria bachillerato incompleta (133 ptos) y Secundaria

bachillerato completa (98 ptos), donde las categorías de Secundaria tienen las probabilidades más altas. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (143 ptos), Secundaria bachillerato completa (92 ptos), Primaria incompleta (144 ptos) y Primaria completa (154 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.

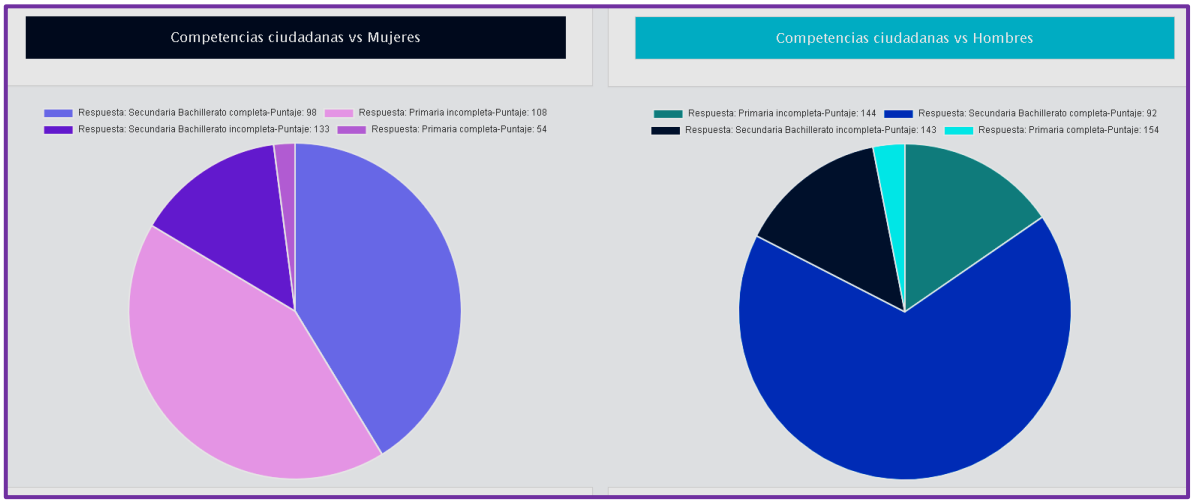


Figura 309. Competencias ciudadanas vs Educación madre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de inglés se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (100 ptos), Secundaria bachillerato completa (95 ptos), Secundaria bachillerato incompleta (59 ptos), Educación profesional completa (157 ptos) y Técnica y tecnológica completa (155 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (150 ptos), Secundaria bachillerato completa (99 ptos), Primaria incompleta (78 ptos), Primaria completa (82 ptos), Educación profesional completa (162 ptos) y Técnica y tecnológica completa (164 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria.

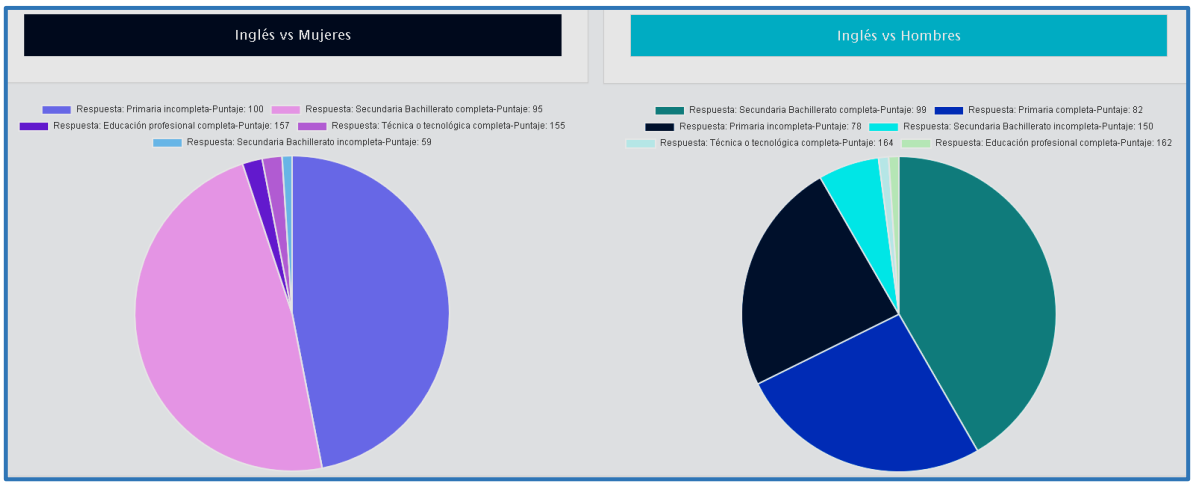


Figura 310. Inglés vs Educación madre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

b) Educación padre

Para el núcleo de Lectura crítica se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (101 ptos), Secundaria bachillerato completa (117 ptos), Secundaria bachillerato incompleta (135 ptos) y Educación profesional completa (146 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (64 ptos), Secundaria bachillerato completa (107 ptos), Primaria incompleta (101 ptos) y Primaria completa (48 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria incompleta.



Figura 311. Lectura crítica vs Educación padre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En razonamiento cuantitativo se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (92 ptos), Secundaria bachillerato completa (106 ptos), Secundaria bachillerato incompleta (141 ptos) y Primaria completa (142 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (149 ptos), Secundaria bachillerato completa (107 ptos), Primaria incompleta (102 ptos), Primaria completa (82 ptos), Primaria completa (48 ptos) y Técnica y tecnológica completa (151 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria incompleta.

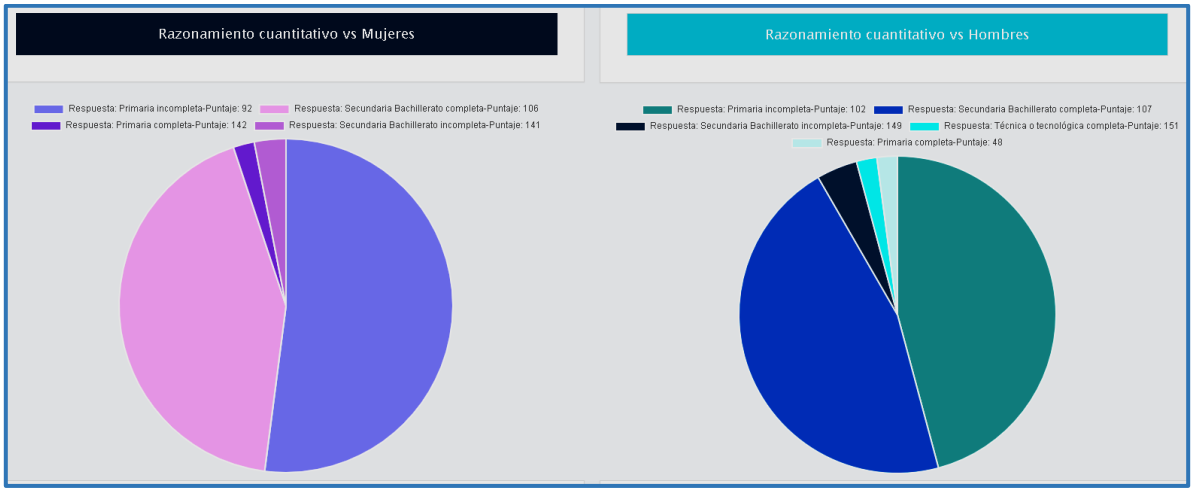


Figura 312. Razonamiento cuantitativo vs Educación padre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (105 ptos), Secundaria bachillerato completa (109 ptos) y Primaria completa (56 ptos), donde Primaria incompleta es la que tiene la probabilidad más alta. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (140 ptos), Secundaria bachillerato completa (105 ptos), Primaria incompleta (92 ptos) y Primaria completa (128 ptos), Primaria completa (48 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria incompleta.

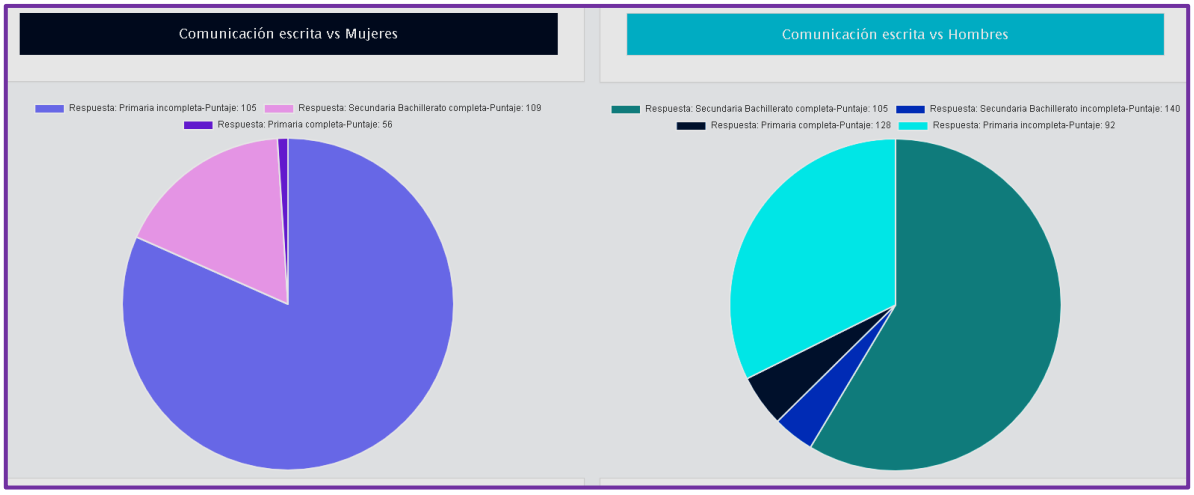


Figura 313. Comunicación escrita vs Educación padre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (108 ptos), Secundaria bachillerato completa (103 ptos), Secundaria bachillerato incompleta (149 ptos), Educación profesional completa (156 ptos) y Primaria completa (58 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (146 ptos), Secundaria bachillerato completa (92 ptos), Primaria incompleta (85 ptos), Primaria completa (49 ptos) y Técnica y tecnológica completa (156 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria incompleta.

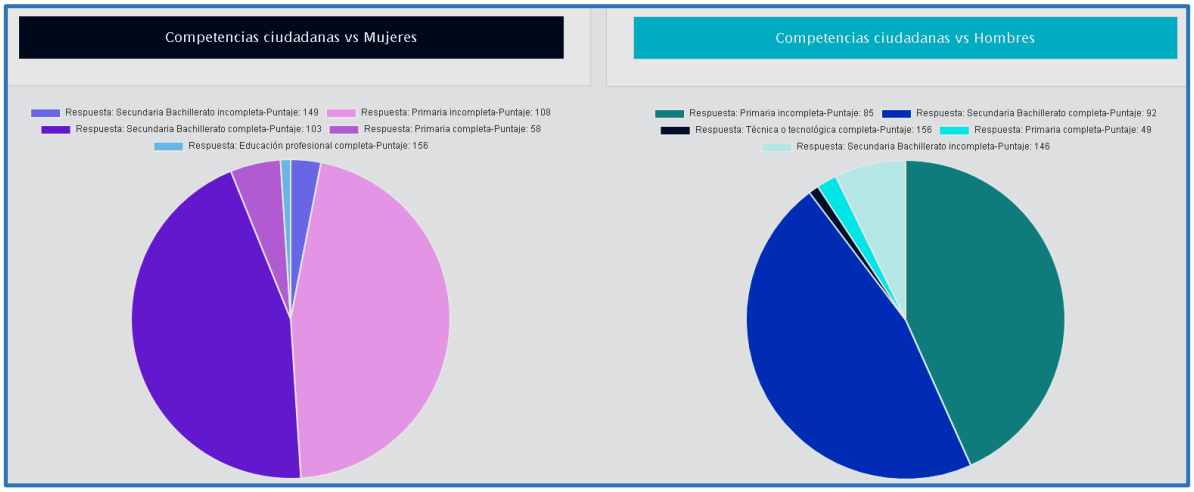


Figura 314. Competencias ciudadanas vs Educación padre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de inglés se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (95 ptos), Secundaria bachillerato completa (108 ptos), Educación profesional completa (152 ptos), Técnica y tecnológica completa (160 ptos) y Primaria completa (58 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (100 ptos), Primaria incompleta (99 ptos), Educación profesional completa (152 ptos) y Técnica y tecnológica completa (148 ptos), teniendo más porcentaje de probabilidad las categoría de Secundaria completa y Primaria incompleta.



Figura 315. Inglés vs Educación padre, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

c) Estrato vivienda

En Lectura crítica, para las mujeres se encontraron probabilidades en Estrato 1 (76 ptos), en Estrato 2 (101 ptos) y en Estrato 3 (135 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (74 ptos), en Estrato 2 (101 ptos) y en Estrato 3 (133 ptos). En ambas géneros es más probable le Estrato 2.

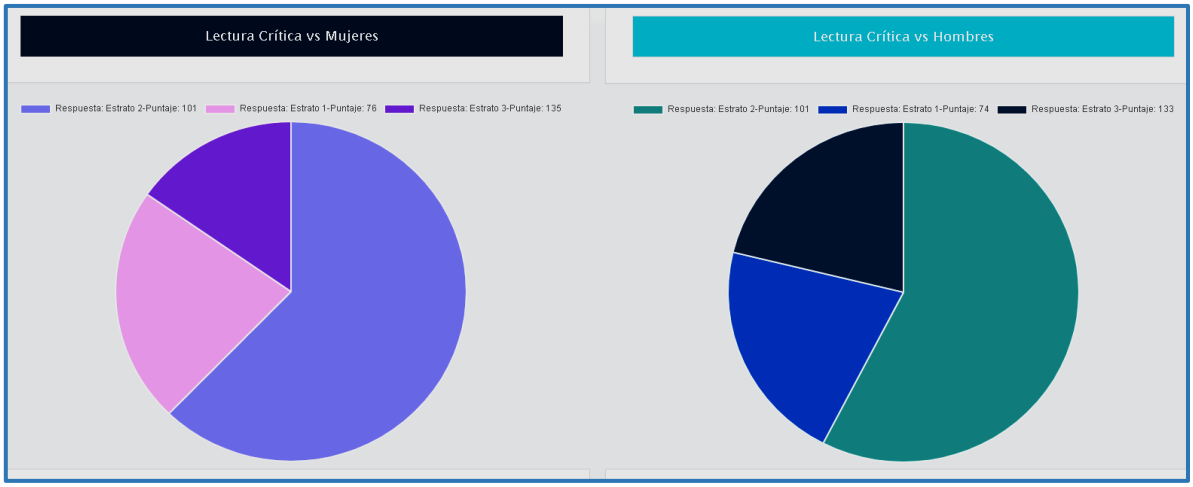


Figura 316. Lectura crítica vs Estrato vivienda, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, para las mujeres se encontraron probabilidades en Estrato 1 (63 ptos), en Estrato 2 (192 ptos) y en Estrato 3 (135 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (147 ptos), en Estrato 2 (102 ptos) y en Estrato 3 (67 ptos). En ambas géneros es más probable le Estrato 2.

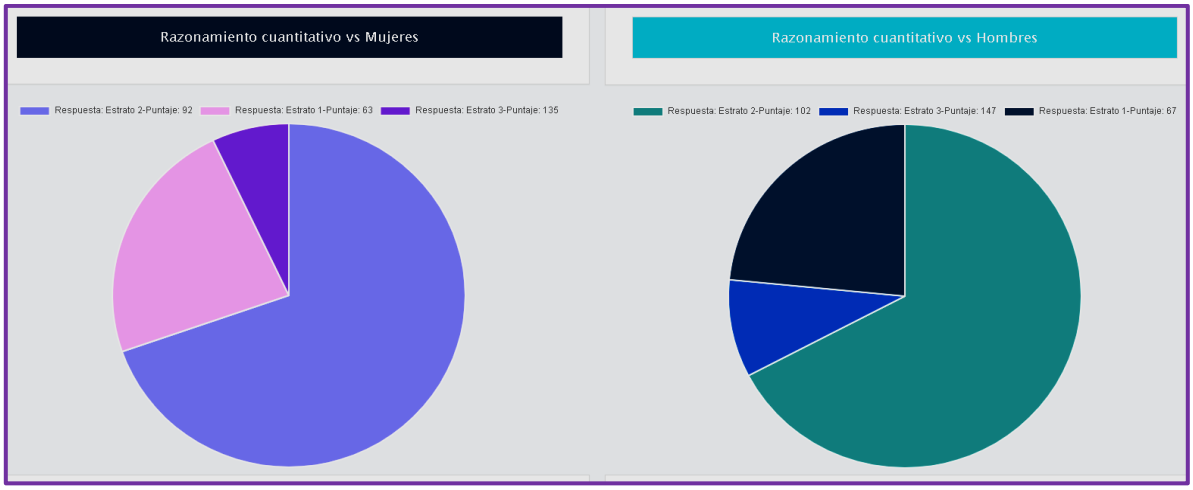


Figura 317. Razonamiento cuantitativo vs Estrato vivienda, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escritas, para las mujeres se encontraron probabilidades en Estrato 1 (70 ptos), en Estrato 2 (105 ptos) y en Estrato 3 (148 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (54 ptos), en Estrato 2 (105 ptos) y en Estrato 3 (132 ptos). En ambas géneros es más probable le Estrato 2.

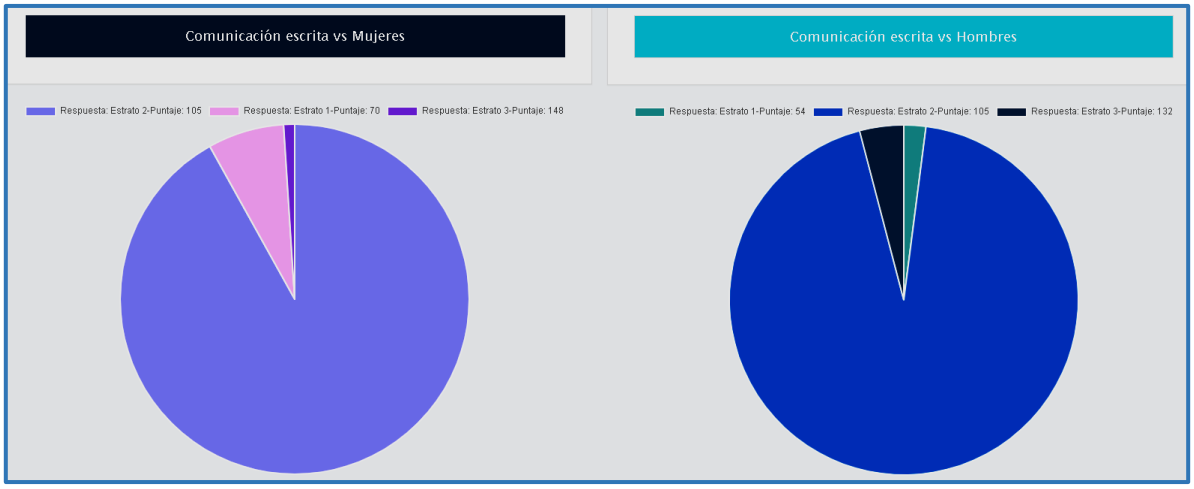


Figura 318. Comunicación escrita vs Estrato vivienda, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, para las mujeres se encontraron probabilidades en Estrato 1 (72 ptos), en Estrato 2 (108 ptos) y en Estrato 3 (144 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (70 ptos), en Estrato 2 (105 ptos) y en Estrato 3 (141 ptos). En ambas géneros es más probable le Estrato 2.

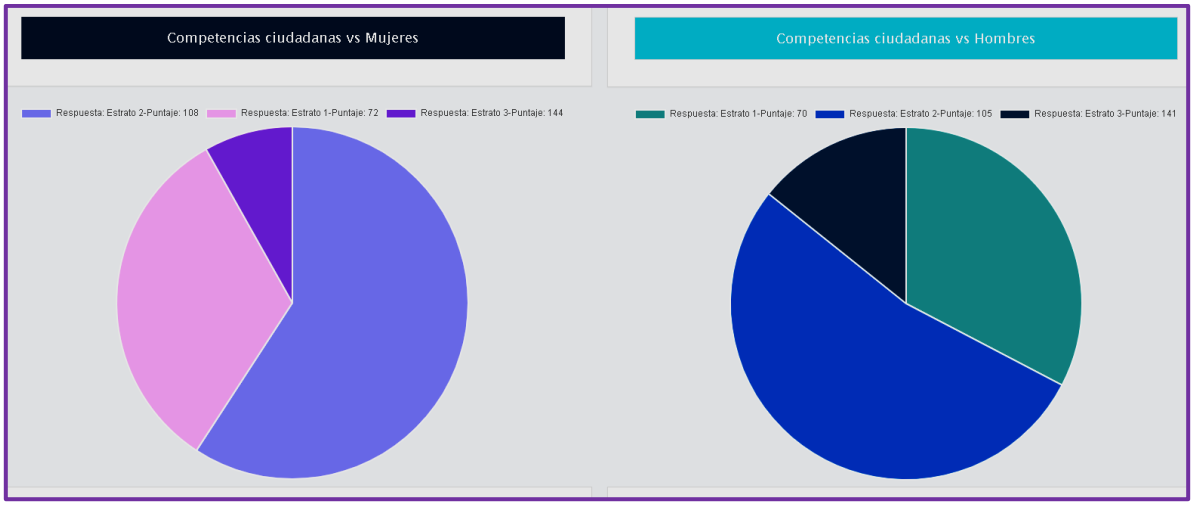


Figura 319. Competencias ciudadanas vs Estrato vivienda, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, para las mujeres se encontraron probabilidades en Estrato 1 (80 ptos), en Estrato 2 (95 ptos) y en Estrato 3 (134 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (74 ptos), en Estrato 2 (99 ptos), en Estrato 3 (130 ptos) y en Estrato 4 (161 ptos). En ambas géneros es más probable le Estrato 2.

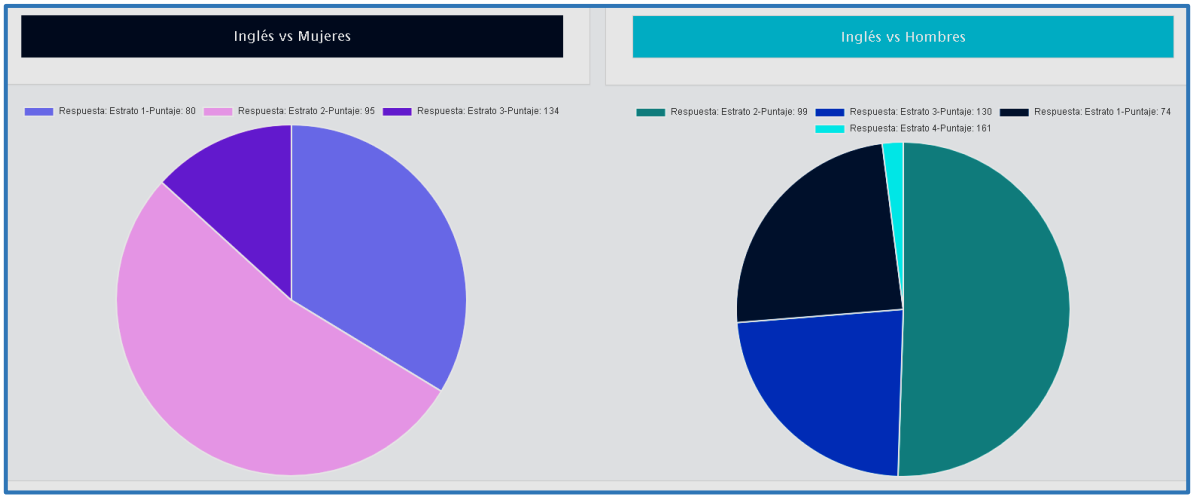


Figura 320. Inglés vs Estrato vivienda, Predicción Saber TyT
Autores: (Manrique, Celeita, Barahona, 2020)

6.3. SABER PRO

Los resultados encontrados a partir del análisis del algoritmo J48 se muestran a continuación representados en gráficas de torta donde se encuentran las principales reglas obtenidas de los nodos de los árboles de decisión, para cada uno de los indicadores evaluados y que son comunes para los diferentes años de presentación de la prueba Saber Pro.

Las reglas obtenidas corresponden al cruce de categorías de cada indicador con los respectivos puntajes de calificación, de esta forma se encontraron los puntajes más probables que pueden llegar a presentarse para cada núcleo temático en cada una de las diferentes categorías.

a) Educación madre

Para el núcleo de Lectura crítica se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (130 ptos), Secundaria bachillerato completa (150 ptos), Educación profesional completa (180 ptos), Técnica y tecnológica completa (179 ptos), Primaria completa (86 ptos) y Postgrado (219 ptos), donde las 4 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (81 ptos), Secundaria bachillerato completa (160 ptos), Educación profesional completa (189 ptos), Técnica y tecnológica completa (79 ptos), Primaria incompleta (101 ptos) y Postgrado (300 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa y Educación profesional completa.

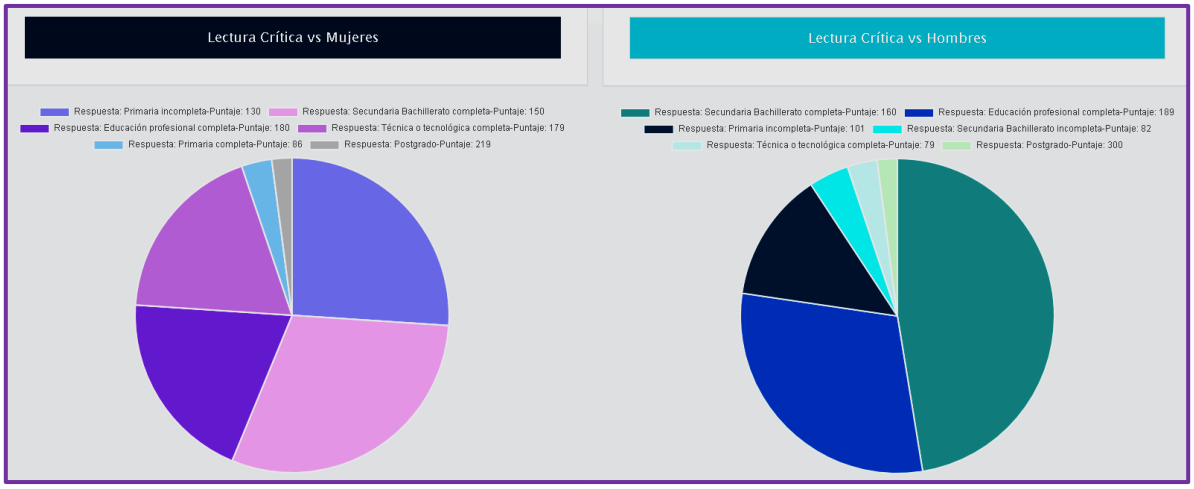


Figura 321. Lectura crítica vs Educación madre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Razonamiento cuantitativo se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (118 ptos), Secundaria bachillerato completa (141 ptos), Educación profesional completa (178 ptos), Técnica y tecnológica completa (184 ptos), Primaria completa (79 ptos) y Postgrado (213 ptos), donde las 4 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato incompleta (89 ptos), Secundaria bachillerato completa (164 ptos), Educación profesional completa (195 ptos), Técnica y tecnológica completa (79 ptos), Primaria incompleta (107 ptos), Postgrado (227 ptos) y Primaria completa (78 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa y Educación profesional completa.

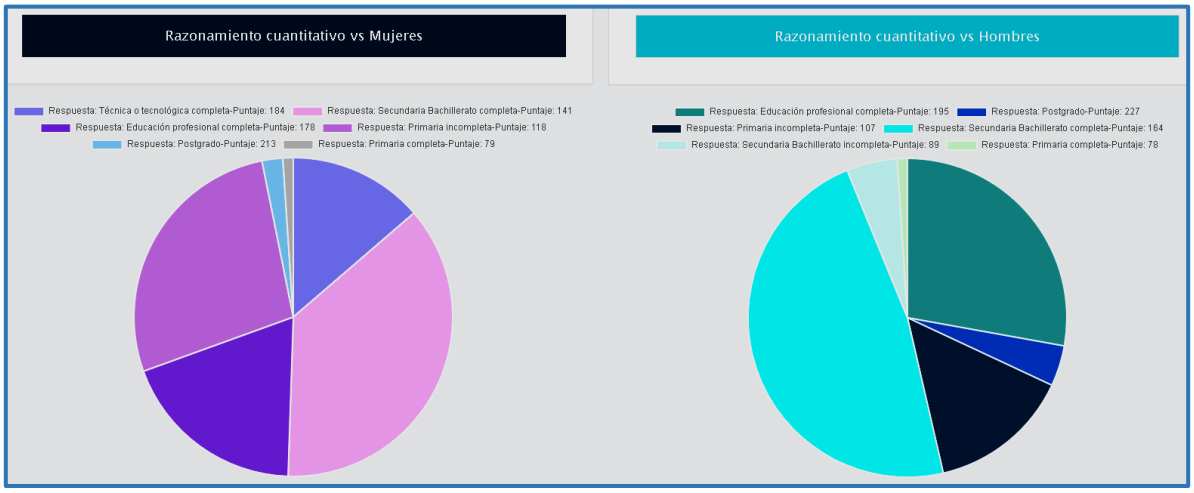


Figura 322. Razonamiento cuantitativo vs Educación madre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

Para el núcleo de Razonamiento cuantitativo se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (116 ptos), Secundaria bachillerato completa (161 ptos), Secundaria bachillerato incompleta (80 ptos), Educación profesional completa (300 ptos), Técnica y tecnológica completa (223 ptos), Primaria completa (79 ptos) y Postgrado (220 ptos), donde las 4 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (160 ptos), Educación profesional completa (192 ptos), Técnica y tecnológica completa (200 ptos), Primaria incompleta (84 ptos), Postgrado (224 ptos) y Primaria completa (87 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.



Figura 323. Comunicación escrita vs Educación madre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (122 ptos), Secundaria bachillerato completa (147 ptos), Educación profesional completa (175 ptos), Secundaria bachillerato incompleta (85 ptos), Técnica y tecnológica completa (208 ptos), Primaria completa (79 ptos) y Postgrado (206 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (145 ptos), Secundaria bachillerato incompleta (0 ptos), Educación profesional completa (178 ptos), Técnica y tecnológica completa (222 ptos), Primaria incompleta (84 ptos), Postgrado (200 ptos) y Primaria completa (70 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.

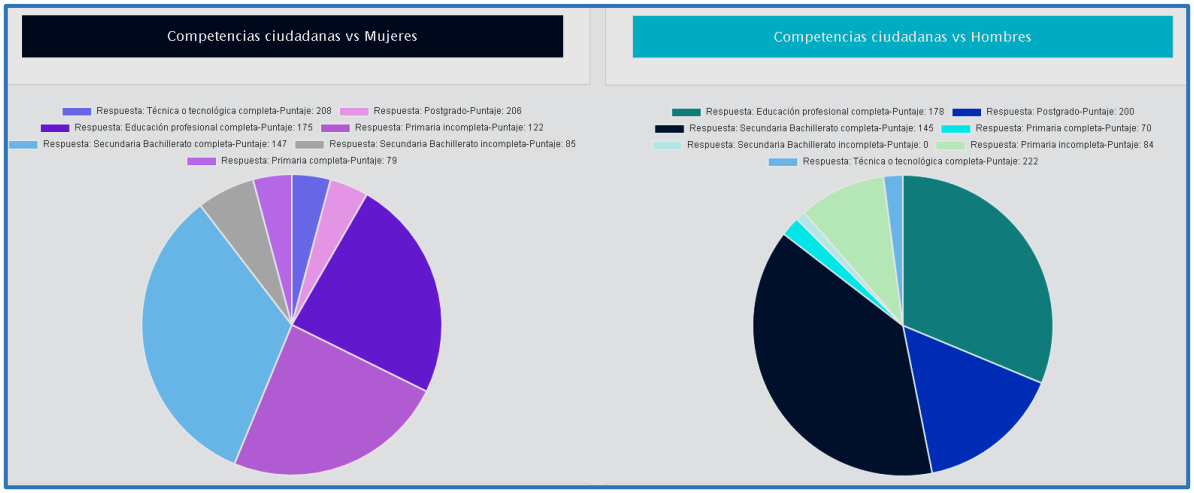


Figura 324. Competencias ciudadanas vs Educación madre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (128 ptos), Secundaria bachillerato completa (135 ptos), Educación profesional completa (184 ptos), y Técnica y tecnológica completa (182 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (138 ptos), Educación profesional completa (188 ptos), Primaria incompleta (114 ptos) y Postgrado (232 ptos), teniendo menos porcentaje de probabilidad la categoría de Postgrado.



Figura 325. Inglés vs Educación madre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

b) Educación padre

Para este indicador en Lectura crítica se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (143 ptos), Secundaria bachillerato completa (150 ptos), Educación profesional completa (177 ptos), Primaria completa (86 ptos) y Postgrado (219 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (160 ptos), Educación profesional completa (184 ptos), Primaria incompleta (111 ptos), Postgrado (216 ptos) y Primaria completa (94 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa y Educación profesional completa.

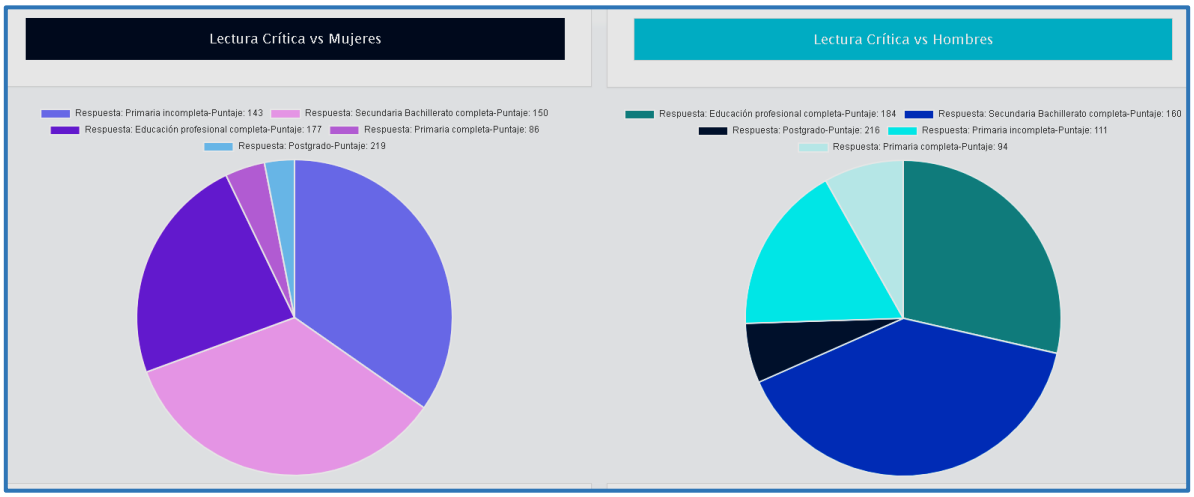


Figura 326. Lectura crítica vs Educación padre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (134 ptos), Secundaria bachillerato completa (141 ptos), Educación profesional completa (174 ptos) y Postgrado (204 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (164 ptos), Secundaria bachillerato incompleta (80 ptos), Educación profesional completa (176 ptos), Primaria incompleta (116 ptos) y Postgrado (227 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa y Educación profesional completa.

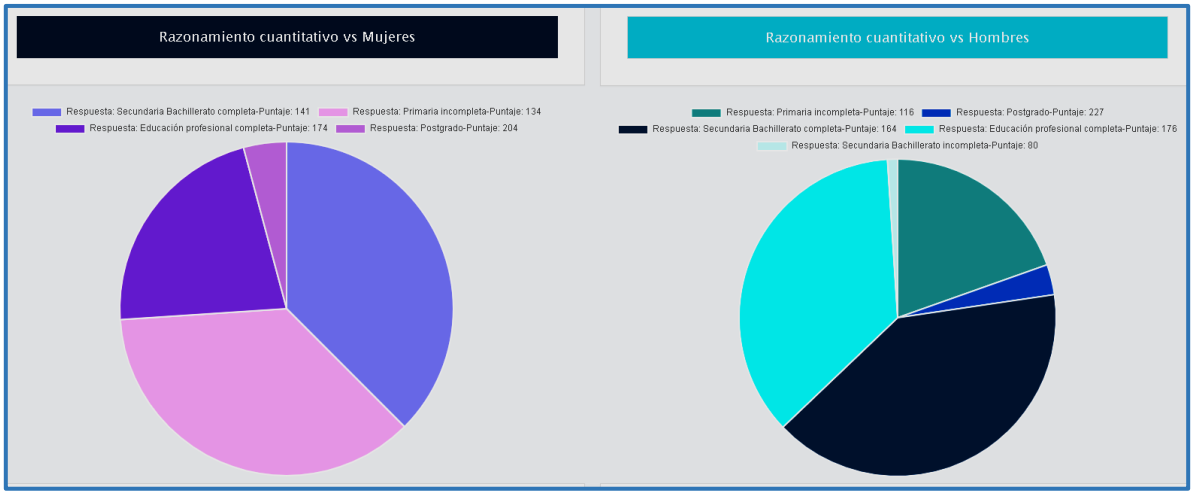


Figura 327. Razonamiento cuantitativo vs Educación padre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (129 ptos), Secundaria bachillerato completa (161 ptos), Educación profesional completa (190 ptos) y Postgrado (215 ptos), donde las 2 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (160 ptos), Secundaria bachillerato incompleta (92 ptos), Educación profesional completa (192 ptos), Primaria incompleta (174 ptos) y Postgrado (218 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa.

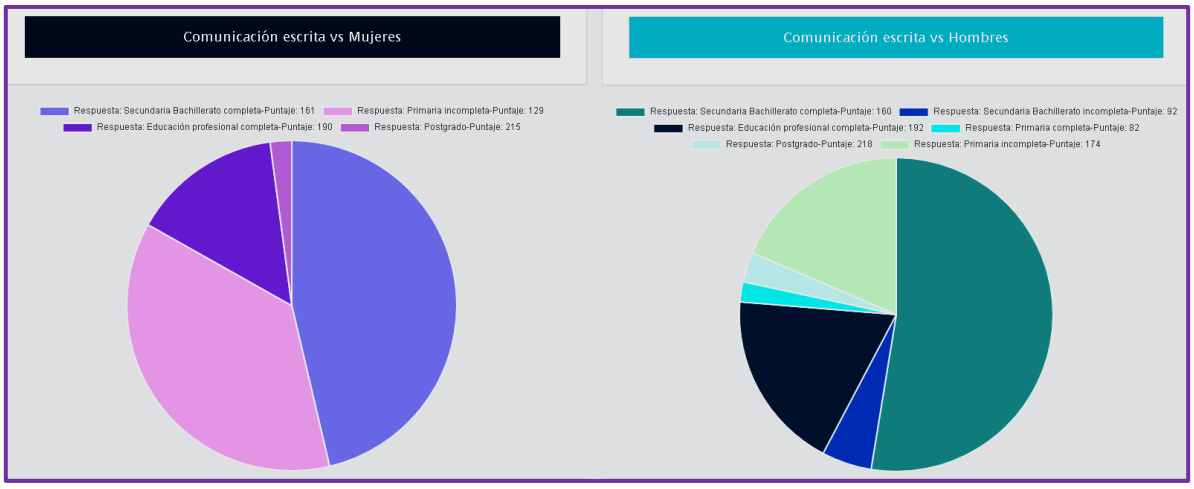


Figura 328. Comunicación escrita vs Educación padre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (150 ptos), Secundaria

bachillerato completa (147 ptos), Educación profesional completa (165 ptos), Técnica y tecnológica completa (75 ptos) y Postgrado (196 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (145 ptos), Secundaria bachillerato incompleta (94 ptos), Educación profesional completa (163 ptos), Primaria incompleta (154 ptos), Primaria completa (86 ptos) y Postgrado (201 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa, Primaria incompleta y Educación profesional completa.

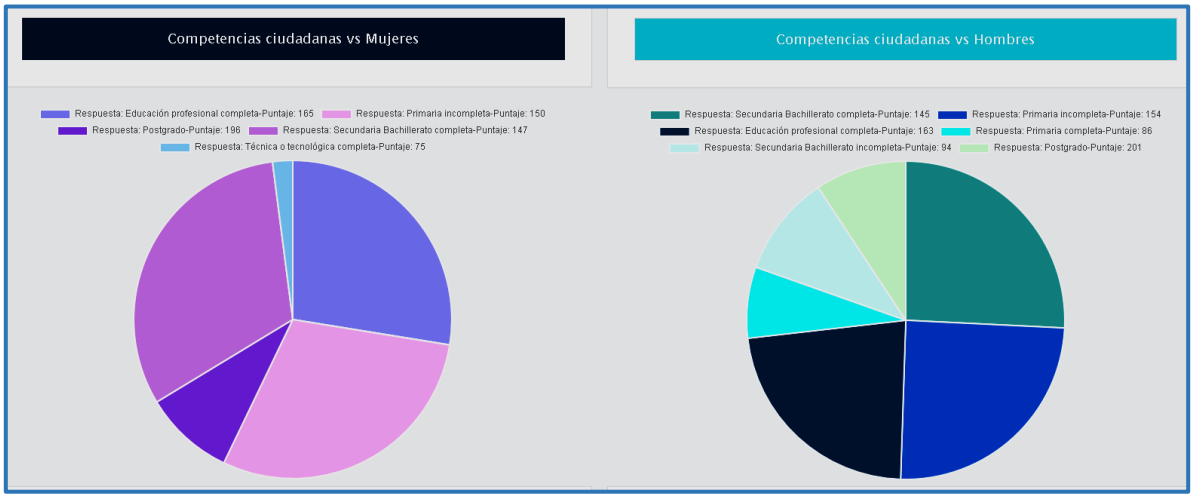


Figura 329. Competencias ciudadanas vs Educación padre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas se encontró que los resultados de las mujeres se presentan en las siguientes categorías: Primaria incompleta (135 ptos), Secundaria bachillerato completa (133 ptos), Educación profesional completa (177 ptos) y Postgrado (217 ptos), donde las 3 primeras presentan una alta probabilidad. Por otro lado, para los hombres los puntajes se tienden a presentar con mayor proporción en las categorías Secundaria bachillerato completa (138 ptos), Educación profesional completa (181 ptos), Primaria incompleta (127 ptos) y Postgrado (215 ptos), teniendo más porcentaje de probabilidad la categoría de Secundaria completa, Primaria incompleta y Educación profesional completa.

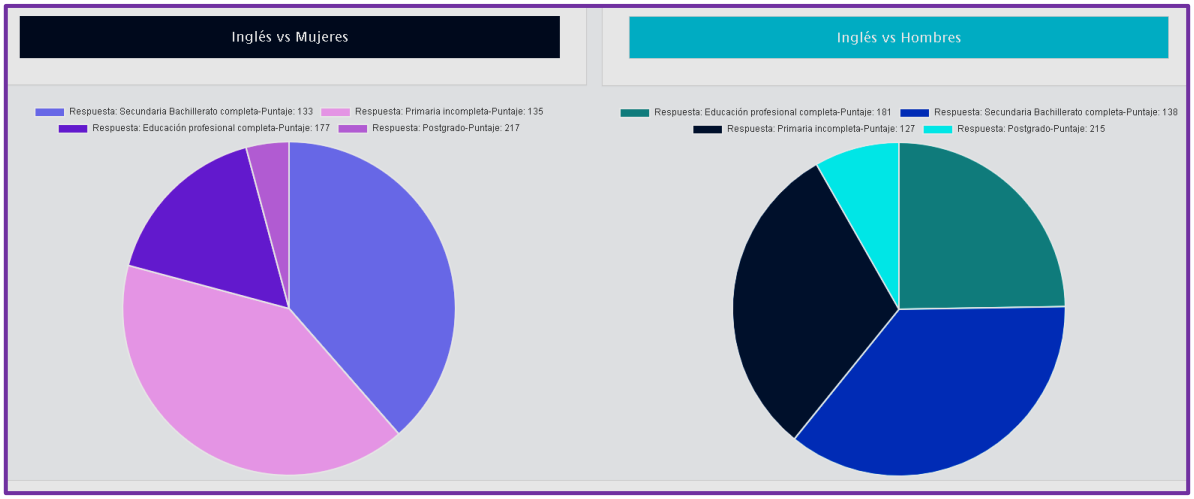


Figura 330. Inglés vs Educación padre, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

c) Estrato de vivienda

En Lectura crítica, para las mujeres se encontraron probabilidades en Estrato 1 (99 ptos), en Estrato 2 (157 ptos), en Estrato 3 (150 ptos) y en Estrato 4 (229 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 4 (216 ptos), en Estrato 2 (155 ptos) y en Estrato 3 (160 ptos). En ambas géneros es más probable el Estrato 2 y 3.

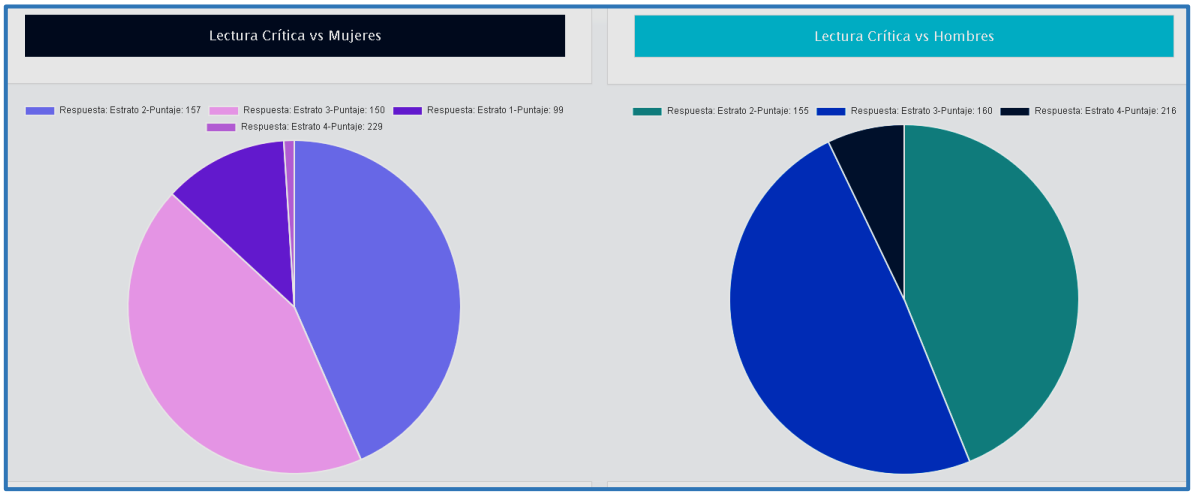


Figura 331. Lectura crítica vs Estrato de vivienda, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, para las mujeres se encontraron probabilidades en Estrato 1 (98 ptos), en Estrato 2 (141 ptos), en Estrato 3 (138 ptos) y en Estrato 4 (213 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (216 ptos), en Estrato 2 (155 ptos), en Estrato

3 (160 ptos) y en Estrato 4 (231 ptos). En ambas géneros es más probable el Estrato 2 y 3.

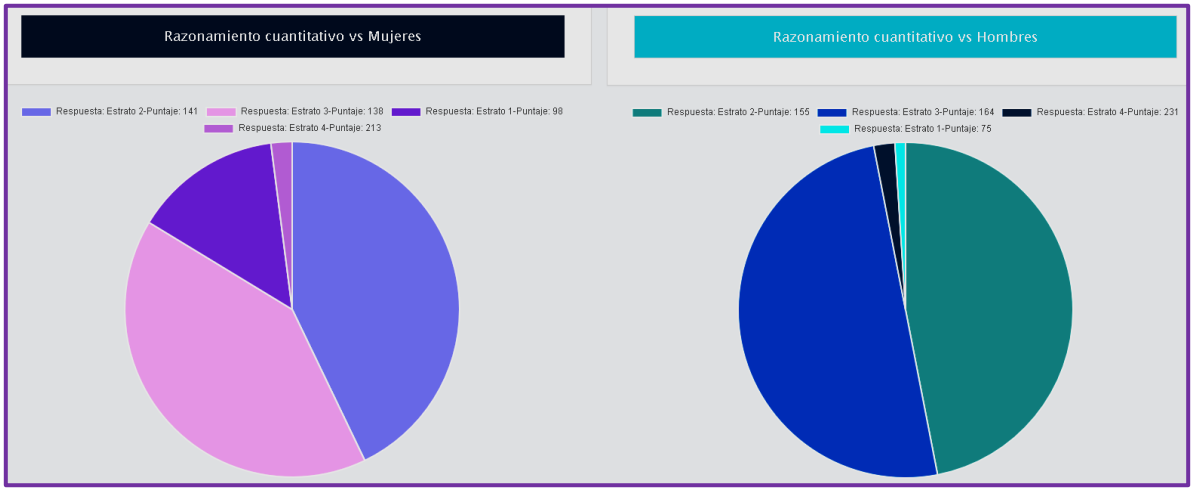


Figura 332. Razonamiento cuantitativo vs Estrato de vivienda, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, para las mujeres se encontraron probabilidades en Estrato 1 (89 ptos), en Estrato 2 (161 ptos) y en Estrato 3 (159 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 2 (129 ptos) y en Estrato 3 (160 ptos). En ambas géneros es más probable el Estrato 2 y 3.

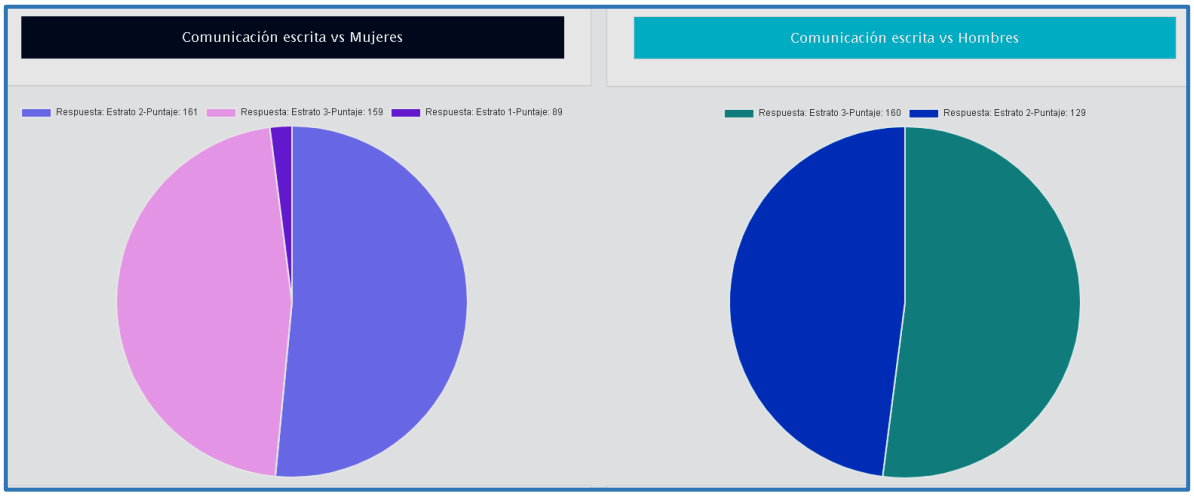


Figura 333. Comunicación escrita vs Estrato de vivienda, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, para las mujeres se encontraron probabilidades en Estrato 1 (76 ptos), en Estrato 2 (154 ptos), en Estrato 3 (153 ptos) y en Estrato 4 (210 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (66 ptos), en Estrato 2 (148 ptos), en Estrato

3 (145 ptos) y en Estrato 4 (220 ptos). En ambas géneros es más probable el Estrato 2 y 3.

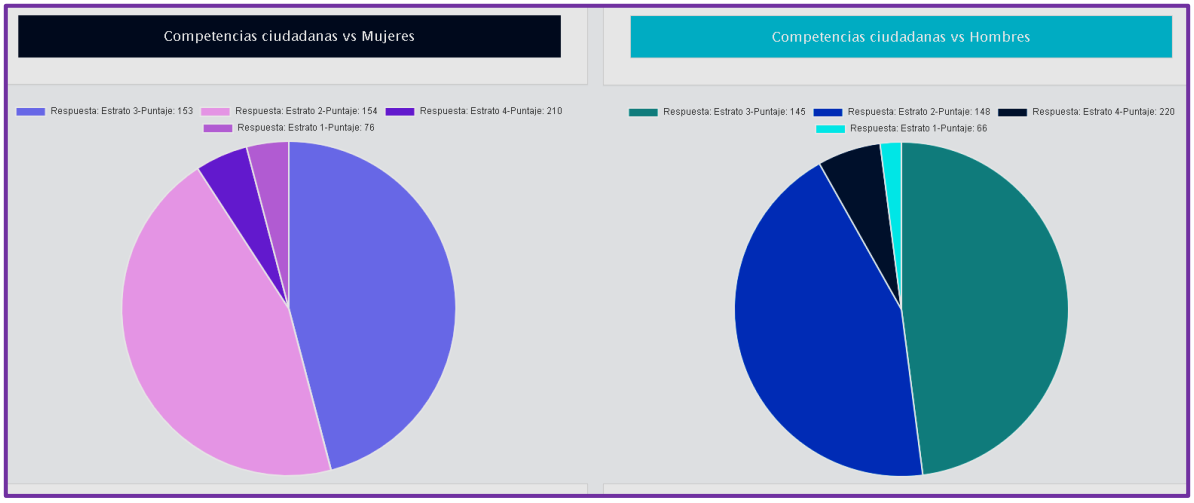


Figura 334. Competencias ciudadanas vs Estrato de vivienda, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, para las mujeres se encontraron probabilidades en Estrato 1 (93 ptos), en Estrato 2 (135 ptos), en Estrato 3 (148 ptos), en Estrato 4 (216 ptos), Estrato 5 (224 ptos) y en Estrato 6 (226 ptos). De igual forma, para los hombres los puntajes se encuentran las siguientes probabilidades en Estrato 1 (79 ptos), en Estrato 2 (138 ptos), en Estrato 3 (145 ptos), en Estrato 4 (212 ptos) y en Estrato 6 (232 ptos). En ambas géneros es más probable le Estrato 2 y 3.

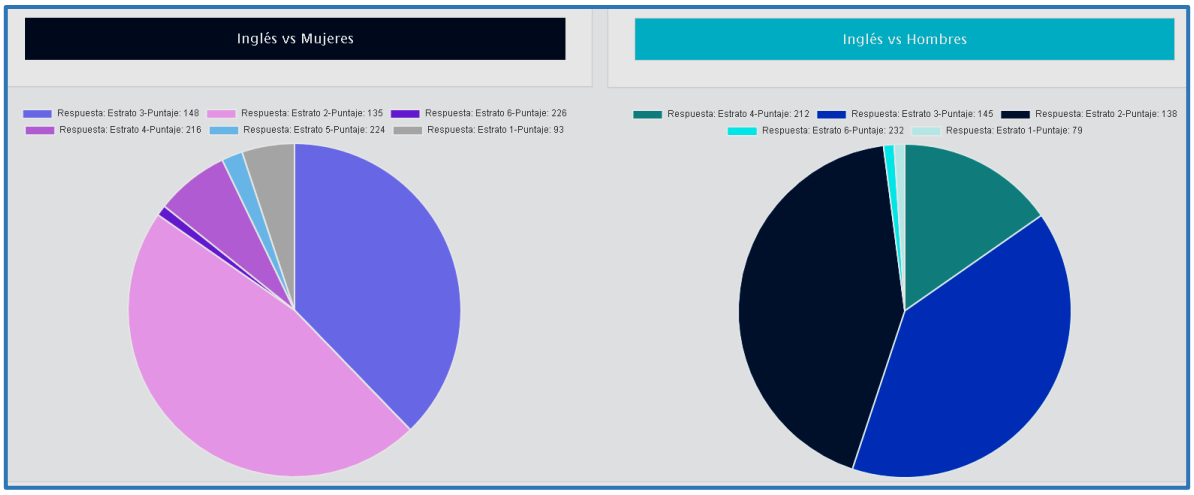


Figura 335. Inglés vs Estrato de vivienda, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

d) Valor matrícula

En este indicador para el núcleo Lectura crítica, se obtiene para el género femenino como probabilidades las siguientes categorías: Entre 1 millón y menos de 2.5 millones (157 ptos), Menos de 500mil (180 ptos), Entre 2.5 millones y menos de 4 millones (156 ptos), Más de 7 millones (220 ptos) y Entre 500 mil y menos de 1 millón (300 ptos). Por otra parte, para el género masculino son Entre 1 millón y menos de 2.5 millones (160 ptos), Menos de 500mil (180 ptos), Entre 2.5 millones y menos de 4 millones (161 ptos), Más de 7 millones (221 ptos) y Entre 500 mil y menos de 1 millón (228 ptos). En ambos géneros los que tienen mayor probabilidad son Entre 1 millón y menos de 2.5 millones, Entre 2.5 millones y menos de 4 millones y Menos de 500mil.



Figura 336. Lectura crítica vs Valor matrícula, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Razonamiento cuantitativo, se obtiene para el género femenino como probabilidades las siguientes categorías: Entre 1 millón y menos de 2.5 millones (141 ptos), Menos de 500mil (181 ptos), Entre 2.5 millones y menos de 4 millones (146 ptos), Más de 7 millones (213 ptos) y Entre 500 mil y menos de 1 millón (210 ptos). Por otra parte, para el género masculino son Entre 1 millón y menos de 2.5 millones (158 ptos), Menos de 500mil (183 ptos), Entre 2.5 millones y menos de 4 millones (168 ptos) y Entre 500 mil y menos de 1 millón (224 ptos). En ambos géneros los que tienen mayor probabilidad son Entre 1 millón y menos de 2.5 millones, Entre 2.5 millones y menos de 4 millones y Menos de 500mil.

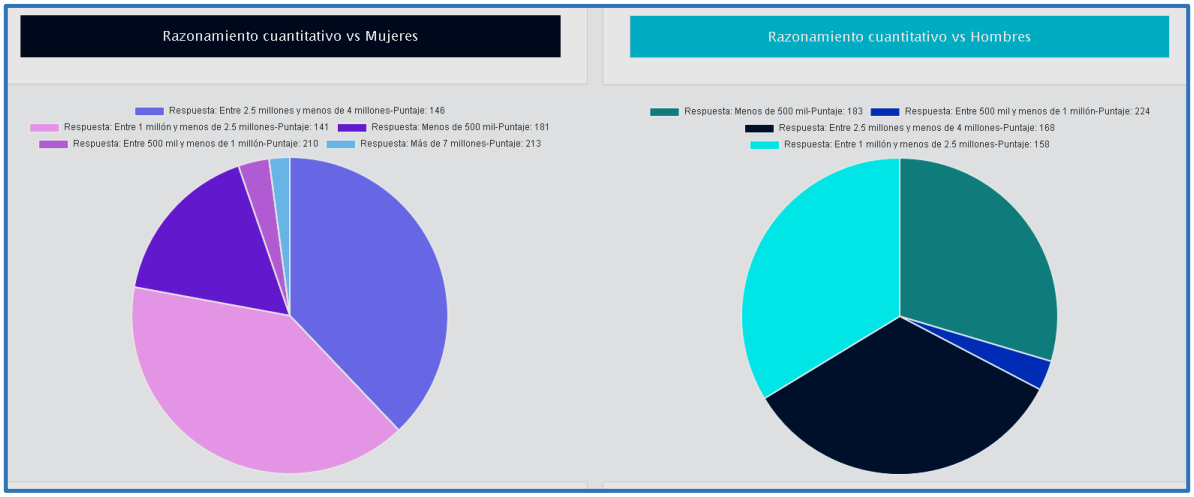


Figura 337. Razonamiento cuantitativo vs Valor matrícula, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Comunicación escrita, se obtiene para el género femenino como probabilidades las siguientes categorías: Entre 1 millón y menos de 2.5 millones (161 ptos), Menos de 500mil (209 ptos), Entre 2.5 millones y menos de 4 millones (134 ptos), Más de 7 millones (227 ptos) y Entre 4 millones y menos 5.5 millones (228 ptos). Por otra parte, para el género masculino son Entre 1 millón y menos de 2.5 millones (139 ptos), Menos de 500mil (196 ptos), Entre 2.5 millones y menos de 4 millones (160 ptos), Más de 7 millones (300 ptos) y Entre 500 mil y menos de 1 millón (218 ptos). En ambos géneros los que tienen mayor probabilidad son Entre 1 millón y menos de 2.5 millones, Entre 2.5 millones y menos de 4 millones y Menos de 500mil.

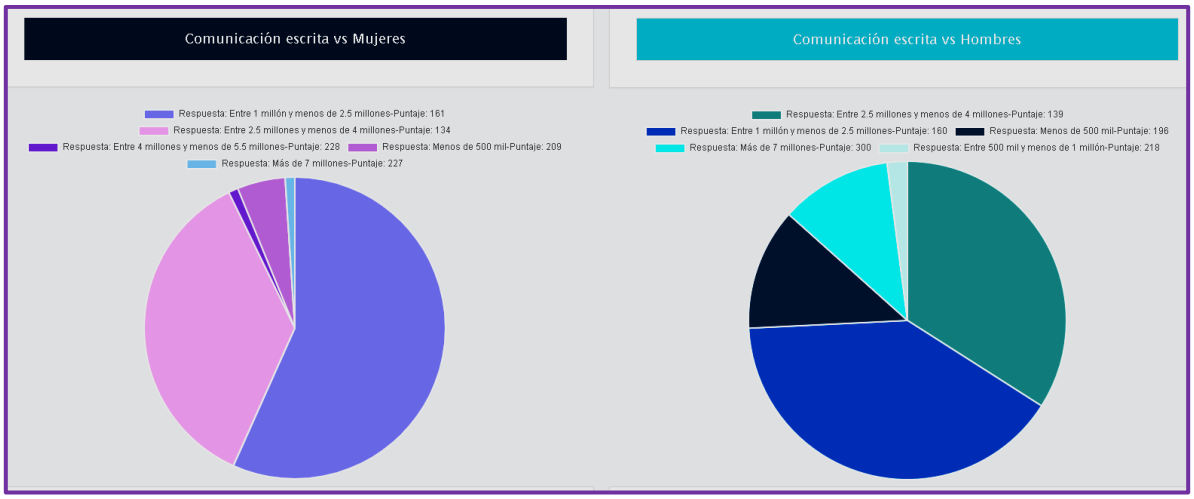


Figura 338. Comunicación escrita vs Valor matrícula, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En Competencias ciudadanas, se obtiene para el género femenino como probabilidades las siguientes categorías: Entre 1 millón y menos de 2.5 millones

(147 ptos), Menos de 500mil (180 ptos), Entre 2.5 millones y menos de 4 millones (158 ptos), Más de 7 millones (195 ptos) y Entre 500 mil y menos de 1 millón (68 ptos). Por otra parte, para el género masculino son Entre 1 millón y menos de 2.5 millones (145 ptos), Menos de 500mil (182 ptos), Entre 2.5 millones y menos de 4 millones (167 ptos), Más de 7 millones (300 ptos), Entre 4 millones y menos de 5.5 millones (226 ptos) y Entre 500 mil y menos de 1 millón (231 ptos). En ambos géneros los que tienen mayor probabilidad son Entre 1 millón y menos de 2.5 millones, Entre 2.5 millones y menos de 4 millones y Menos de 500mil.

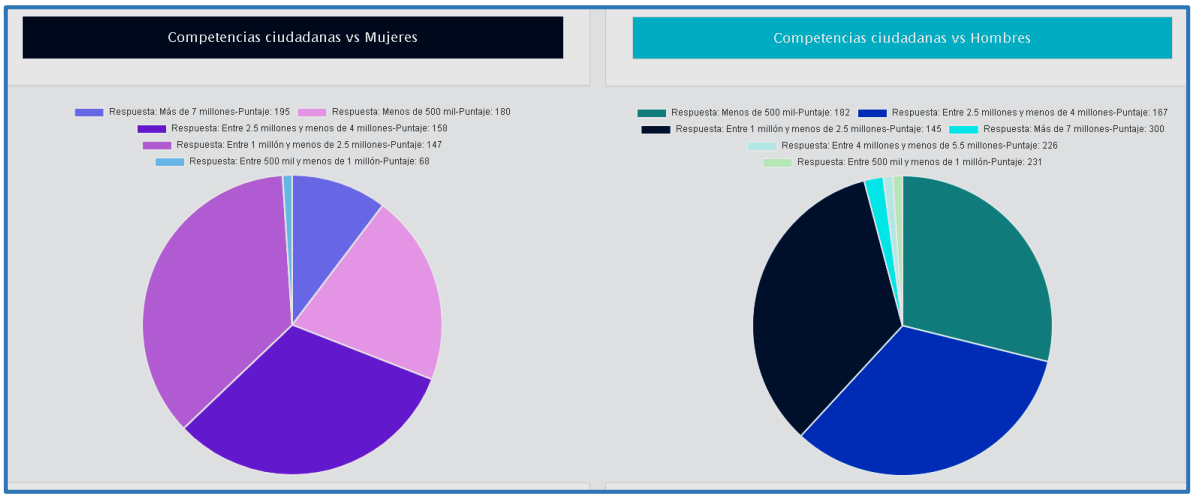


Figura 339. Competencias ciudadanas vs Valor matrícula, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

En inglés, se obtiene para el género femenino como probabilidades las siguientes categorías: Entre 1 millón y menos de 2.5 millones (135 ptos), Menos de 500mil (187 ptos), Entre 2.5 millones y menos de 4 millones (147 ptos), Más de 7 millones (199 ptos), Entre 5.5 millones y menos 7 millones (228 ptos) y Entre 500 mil y menos de 1 millón (89 ptos). Por otra parte, para el género masculino son Entre 1 millón y menos de 2.5 millones (138 ptos), Menos de 500mil (181 ptos), Entre 2.5 millones y menos de 4 millones (139 ptos), Más de 7 millones (202 ptos) y Entre 500 mil y menos de 1 millón (79 ptos). En ambos géneros los que tienen mayor probabilidad son Entre 1 millón y menos de 2.5 millones, Entre 2.5 millones y menos de 4 millones y Menos de 500mil.

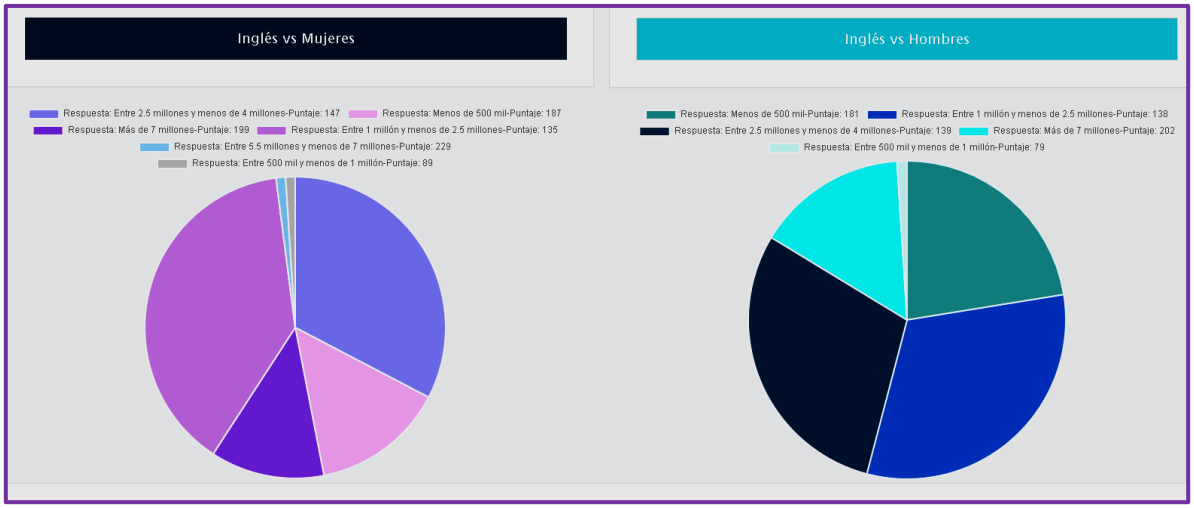


Figura 340. Inglés vs Valor matrícula, Predicción Saber Pro
Autores: (Manrique, Celeita, Barahona, 2020)

CONCLUSIONES

- El uso de minería de datos para el proceso de análisis de datos permite tener acceso a información oculta que con los métodos comúnmente usados de estadística no se puede obtener, además, facilita el procesamiento de grandes volúmenes de datos.

- Gracias a las técnicas de minería de datos, en este caso, de agrupamiento se determinó que es posible generar modelos descriptivos consistentes de la realidad observada, encontrando atributos comunes para a partir de estas asociaciones crear clústeres representativos de la información analizada.

- El análisis predictivo se ocupa principalmente de anticipar resultados futuros basados en minería de datos existentes. Esta información se convierte en importante a la hora de ser analizada porque permite comprender su comportamiento futuro, o en términos coloquiales lo que puede suceder. Esta capacidad de predecir da un poder inmenso de planificación y brinda la posibilidad de solución a diferentes problemáticas ya que entiende cuál es la conducta de una o varias cosas en distintos escenarios de evaluación.

- WEKA facilita la aplicación de los algoritmos de minería de datos siendo una herramienta fiable y de fácil uso para el usuario.

- Con la aplicación del algoritmo K-means se descartaron atributos que no presentaran cambios o diferencias siendo evaluados con respecto al género. Solo quedaron los que de alguna manera influyeron en las brechas.

- Hay áreas de carrera donde se destacan las mujeres, hablando desde el punto de vista de participación como las carreras afines a la administración, ciencias de la salud, entre otras; mientras que hay grupos como las ingenierías y mecánica o afines donde el hombre se destaca.

- Los indicadores de Educación de los padres y el Estrato son indicadores comunes para las 3 pruebas, siendo este último el principal factor de medición económica para los estudiantes, y está bastante relacionado con el nivel de Sisbén, valor de matrícula, ingreso familiar, entre otros.

- En la actualidad, los jóvenes universitarios dependen económicamente de sus familias por lo cual se da a entender que las mujeres, al pedir mayor cantidad de créditos son quienes tienen más apoyo familiar.

- Con respecto al proceso de carga y ETL de los resultados, en la primera entrega de requerimientos se especificó que el proceso de extracción se realizaría de forma automática por el aplicativo. Debido al bloqueo de puertos y a otros limitantes presentados por la compañía prestadora del servicio de internet, esta tarea se dejó para que sea realizada por el administrador, puesto que al ser inestable e intermitente la conexión con el servidor FTP no se podía asegurar la transferencia de archivos.

- En el requerimiento encargado de la actualización de la información para la conexión al servidor FTP fue removido debido a su falta de funcionalidad y en su reemplazo se construyó un formulario de contáctenos donde los usuarios que desean acceder al aplicativo hacen la solicitud de su usuario y contraseña, y los que ya están registrados pueden enviar sugerencias, dudas o cualquier tipo de preguntas respecto al funcionamiento del aplicativo.

- Con el desarrollo de esta investigación se busca contribuir a futuros estudios sobre brecha de género, recomendando así hacer un comparativo de estas pruebas (Saber) con otras realizadas a nivel internacional como las PISA. Y hacer énfasis en los indicadores hallados y su comportamientos respecto a las diferentes áreas temáticas evaluadas.

- Para futuras versiones del aplicativo, se podría optimizar el proceso de carga de los resultados, así como permitirles a los usuarios realizar el cambio de contraseña sin necesidad de contactar con el administrador del sistema.

BIBLIOGRAFÍA

Carmona, C. J., Pulgar-Rubio, F. J., García-Vico, A. M., González, P., & del Jesus, M. J. (2015). Análisis descriptivo mediante aprendizaje supervisado basado en patrones emergentes. *Proc. of the VII Simposio Teoría y Aplicaciones de Minería de Datos*, 685–694.

Caro Acero, B. L., & Casas M., A. F. (2013). *Análisis de las diferencias de género en el desempeño de estudiantes colombianos en matemáticas y lenguaje*.

Cataldi, Z., Salgueiro, F., & Lage, F. J. (2007). *Predicción del rendimiento de los estudiantes y diagnóstico usando redes neuronales*. http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/capred.pdf

Chaves Sánchez, L. J., Pertuz Arroyave, C. J., & Valenzuela Sabogal, G. M. (2016). *INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DE CUNDINAMARCA EN LA EXTENSIÓN FACATATIVÁ UTILIZANDO MODELOS PREDICTIVOS DE MINERÍA DE DATOS*. Universidad de Cundinamarca.

Cobos, C., Zuñiga, J., Guarin, J., León, E., & Mendoza, M. (2010). CMIN - Herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingenieria e Investigacion*, *30*(3), 45–56.

LEY NÚMERO 23 DE 1982, (1982).

Ley 1266 de 2008, (2008). http://www.secretariasenado.gov.co/senado/basedoc/ley_1266_2008.html

Ley 1273 de 2009, 2 (2009).

Corte Constitucional. (1991). *Constitución política de 1991*. http://www.secretariasenado.gov.co/senado/basedoc/constitucion_politica_1991.html

De Moya Amaris, ME, R. R. J. (2003). La contribución de las reglas de asociación a la minería de datos. *Tecnura*, *13*(13), 94–109.

Educación, M. de. (2016). *Revisión de políticas nacionales de educación La educación en Colombia*.

Fernández, A. P., & Díaz, P. (2002). *Investigación cuantitativa y cualitativa*. *Figura 1*, 11–14.

Franco, D. G. (2019). *Proceso ETL en Minería de datos mediante PDI*.

García, S., Ramírez Gallego, S., Luengo, J., & Herrera, F. (2016). *Big Data: Preprocesamiento y calidad de datos*. www.highlycited.com

Gómez Pinzón, D. A., & Valenzuela Sabogal, G. M. (2017). *INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DE CUNDINAMARCA EN LA EXTENSIÓN FACATATIVÁ UTILIZANDO MODELOS PREDICTIVOS DE MINERÍA DE DATOS (FASE 2)*. Universidad de Cundinamarca.

Gutiérrez O., J. A., & Molina, B. (2016). Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales. *Revista Ontare*, *3*(2), 33. https://doi.org/10.21158/23823399.v3.n2.2015.1440

Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., & Big, ". (2015). *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación Big Data: an exploration of research, technologies and application cases*.

Hernández Bonilla, J. M. (2018). *La difícil situación de las escuelas rurales en Colombia - ELESPECTADOR.COM*. https://www.elespectador.com/colombia2020/territorio/la-dificil-situacion-de-las-escuelas-rurales-en-colombia-articulo-856698

Hernández Sampieri, R., Fernandez Collado Carlos, & Baptista Lucio, P. (2006). Los enfoques cuantitativo y cualitativo en la investigación científica. In *Metodología de la instigación*. https://doi.org/10.6018/turismo.36.231041

Hurtado, M. (2019). *Subsisten brechas de género en acceso a educación | El Nuevo Siglo Bogotá*. https://www.elnuevosiglo.com.co/articulos/03-2019-subsisten-brechas-de-genero-en-acceso-educacion

Ipsos. (2019). *Global attitudes towards gender equality*. *February*.

León, C. (2018). *BigMark : análisis y predicción de calificaciones mediante Big Data*. Universidad Complutense Madrid.

López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. https://doi.org/10.18235/0001082

Loury, L. D. (2016). *The Gender Earnings Gap among College-Educated Workers*

Author (s): Linda Datcher Loury Source : Industrial and Labor Relations Review , Vol . 50 , No . 4 (Jul ., 1997), pp . 580-593 Published by : Sage Publications , Inc . Stable URL : http://www.jsto. *50*(4), 580–593.

Marchionni, M., Gasparini, L., & Edo, M. (2018). *Brechas de género en América Latina. Un estado de situación*. www.scioteca.caf.com

Decreto número 1377 de 2013, 11 (2013).

Moine, I. J. M., Gordillo, D. S., & Haedo, D. A. (2011). *Análisis comparativo de metodologías para la gestión de proyectos de minería de datos*.

Molina López, J. M., & García Herrero, J. (2006). *Técnicas de análisis de datos* [Universidad Carlos III]. http://matema.ujaen.es/jnavas/web_recursos/archivos/weka master recursos naturales/apuntesAD.pdf

ONU Mujeres, UNDP, UNFPA, & USAID. (2016). Brechas de género y desigualdad: De los Objetivos de Desarrollo del Milenio a los Objetivos de Desarrollo Sostenible. In *Libre pensamiento* (Issue 88). https://dialnet.unirioja.es/servlet/extart?codigo=5897357

Oracle Corporation. (n.d.). *¿Qué es Java y para qué es necesario?* Retrieved February 26, 2020, from https://www.java.com/es/download/faq/whatis_java.xml

Palacio, J., & Ruata, C. (2011). *Gestión de proyectos Scrum Manager*. http://www.hacienda.go.cr/cifh/sidovih/spaw2/uploads/images/file/Gesti?n de proyectos.pdf

Pérez López, C., & Santín González, D. (2008). *Minería de datos: técnicas y herramientas*. https://books.google.com.co/books?id=wz-D_8uPFCEC&printsec=frontcover&dq=mineria+de+datos&hl=es-419&sa=X&ved=0ahUKEwiSnc_s-rTkAhXG1FkKHUtfCNcQ6AEIKTAA#v=onepage&q=mineria de datos&f=false

Platzi. (2015). *¿Qué es PostgreSQL y cuáles son sus ventajas?* https://platzi.com/blog/que-es-postgresql/

Reyes Nava, A., Flores Fuentes, A., Alejo, R., & Rendón Lara, E. (2017). Minería de datos aplicada para la identificación de factores de riesgo en alumnos.

Research in Computing Science, *139*(ISSN 1870-4069), 177–189.

Rodrigo, J. A. (2018). *Reglas de asociación y algoritmo Apriori con R*. https://github.com/JoaquinAmatRodrigo/Estadistica-con-R/blob/master/PDF_format/43_Reglas_de_asociacion.pdf

Romero, C., & Ventura, S. (2016). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Suárez Soto, L. E. (2014). *TÉCNICAS DE MINERÍA DE DATOS PARA LA DETECCIÓN Y PREVENCIÓN DEL LAVADO DE ACTIVOS Y LA FINANCIACIÓN DEL TERRORISMO (LA/FT)*.

Timaran, R., & Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. *Formacion Universitaria*, 1–19. https://www.oei.es/historico/congreso2014/memoriactei/758.pdf

Universia. (2013). *Estudio demuestra que hombres y mujeres varían su rendimiento en matemáticas*. https://noticias.universia.net.co/en-portada/noticia/2013/09/13/1049251/estudio-demuestra-hombres-mujeres-varian-rendimiento-matematicas.html

University of Waikato. (n.d.). *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. Retrieved February 27, 2020, from https://www.cs.waikato.ac.nz/ml/weka/

Villafañe Blanco, P. V. (2015). *Análisis del Desempeño Académico del Examen de Estado para el Ingreso a la Educación Superior Aplicando Minería de Datos Análisis del Desempeño Académico del Examen de Estado para el Ingreso a la Educación Superior Aplicando Minería de Datos*. http://www.bdigital.unal.edu.co/51414/1/39004913.2015.pdf

Zambrano, C., Rojas, D., Carvajal, K., & Acuña, G. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales. Analysis of students academic performance using data warehouse and neural networks. *Revista Chilena de Ingeniería*, *19*, 369–381.

Zavala Guirado, M. A., Álvarez, M. M., Vázquez, M. A., González, I., & Bazán-Ramírez, A. (2018). Factores internos, externos y bilaterales asociados con la deserción en estudiantes universitarios. *Interacciones. Revista de Avances En Psicología*, *4*(1), 59–69. https://doi.org/10.24016/2018.v4n1.103

7. ANEXOS

7.1. MANUAL DE USUARIO



TABLA DE CONTENIDO

INTRODUCCIÓN ---------- 276

1. **Inicio de Sesión** ---------- 277
2. **Contáctenos** ---------- 278
3. **Página principal usuario** ---------- 279
4. **Análisis descriptivo de pruebas Saber 11** ---------- 281
5. **Análisis descriptivo de pruebas Saber TyT** ---------- 285
6. **Análisis descriptivo de pruebas Saber Pro** ---------- 289
7. **Análisis predictivo de las Pruebas Saber Superior** ---------- 293
8. **Inicio de sesión Administrador** ---------- 298
9. **Insertar usuarios nuevos** ---------- 299
10. **Listar usuarios** ---------- 300
11. **Listado de pruebas** ---------- 301
12. **Cargar archivos pruebas** ---------- 302
13. **Otras funciones Administrador** ---------- 305

INTRODUCCIÓN

El presente manual tiene como objetivo brindar una guía clara y concisa del funcionamiento del aplicativo web Ingapge, enfocado a las pruebas Saber 11, TyT y Pro; y describir paso a paso los diferentes procesos a realizarse en el mismo por parte de los diferentes usuarios del aplicativo.

MANUAL DE USUARIO

1. Inicio de Sesión

En este formulario el usuario debe digitar el nombre de usuario que tiene asignado y su respectiva contraseña, seguido debe darle clic al botón “Acceder” para poder ser redirigido según el rol que tenga asignado.



En caso de que usted no se encuentre registrado, haga clic en “Contáctenos”.

2. Contáctenos

Este formulario tiene como objetivo que los usuarios se contacten con el administrador del software. Puede ser usado, por ejemplo, para la creación de cuentas, cambio de contraseña, dudas, sugerencias, entre otros.

En primer lugar, ingrese sus datos de contacto (nombre, apellido, correo institucional), enseguida escriba el motivo por el cual se está contactando con nosotros, y por último de clic en el botón Enviar.



3. Página principal usuario

En este formulario se puede observar la información referente a la cantidad de mujeres que presentan las diferentes pruebas Saber.



Permite que el usuario realice filtro según los años de presentación y el nombre de la prueba. A continuación, se muestran algunos de ellos.

- Filtro para visualizar cantidad de mujeres en el año 2017:



- Filtro para visualizar cantidad de mujeres en la prueba Saber 11:



- Filtro para visualizar cantidad de mujeres en la prueba Saber Pro:

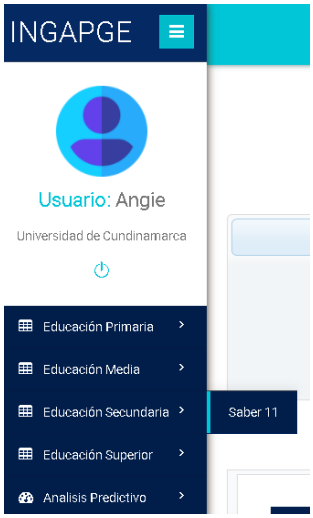


- Filtro para visualizar cantidad de mujeres en la prueba Saber TyT:

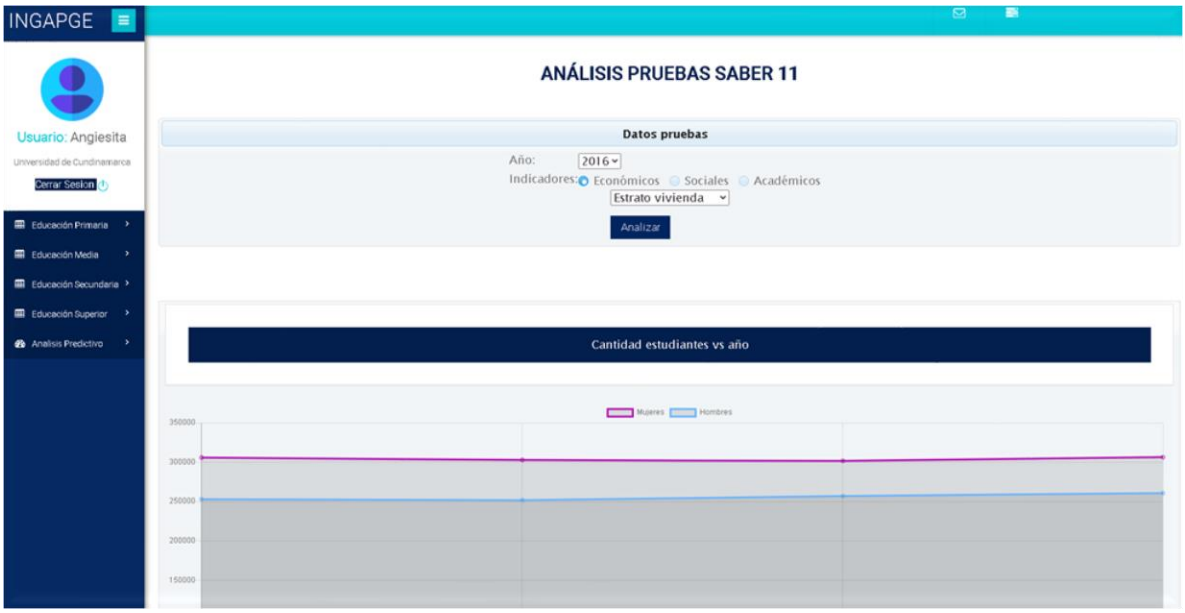


4. Análisis descriptivo de pruebas Saber 11

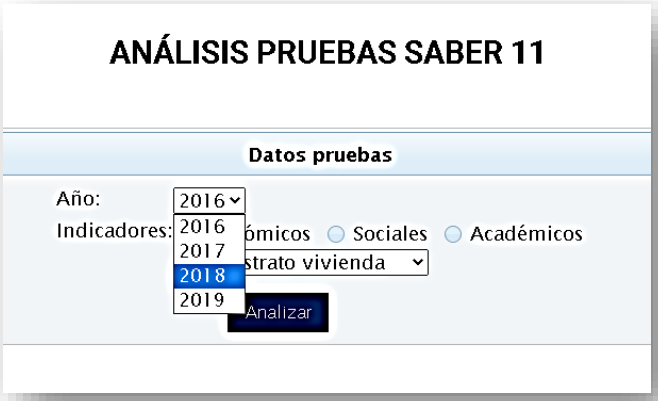
Para la visualización de resultados del análisis descriptivo, el usuario debe elegir la prueba Saber 11 del menú.

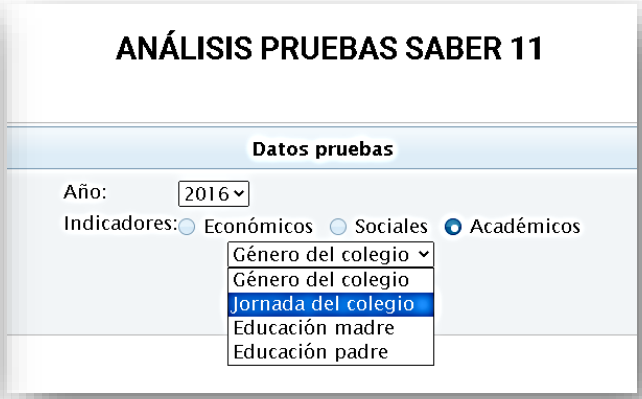


Este formulario permite mostrar los resultados obtenidos del análisis descriptivo de las pruebas Saber 11 realizadas desde el año 2016, en adelante. Tiene visualización por medio de gráficas de línea y barras para mayor entendimiento.

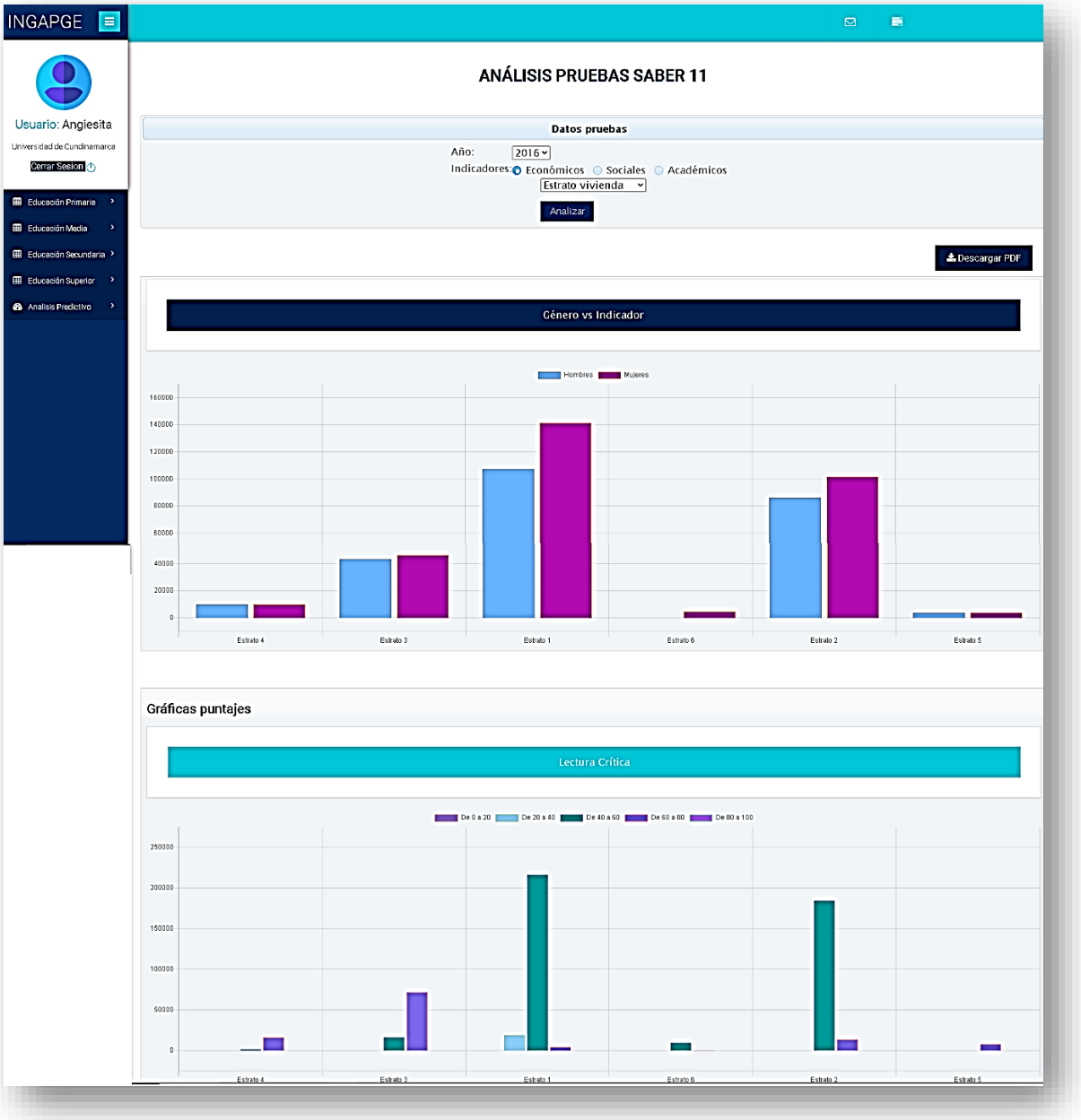


Para realizar el análisis, primero debe seleccionar el año que desee, enseguida el indicador de brecha de género que puede ser de tipo económico, social o académico y finalizar dando clic en el botón Analizar.





El resultado será de la siguiente forma:



El usuario tiene la posibilidad de tener estos resultados en formato PDF, para que haga uso de los mismos en investigaciones futuras, para ello simplemente deberá hacer clic en el botón de Generar PDF, ubicado en la parte superior derecha de las gráficas.

Obteniendo como resultado:

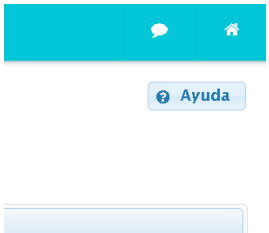


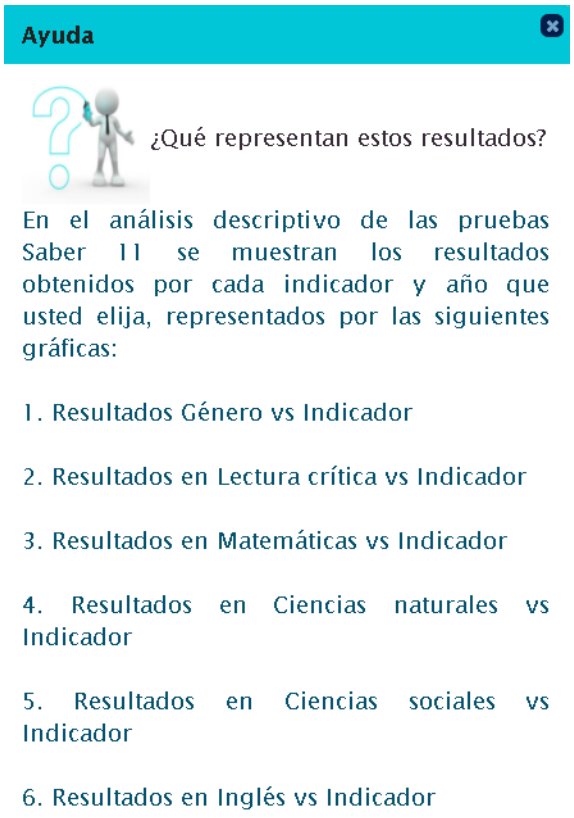






Para entender un poco más lo mostrado, se cuenta con un botón de ayuda para el usuario.



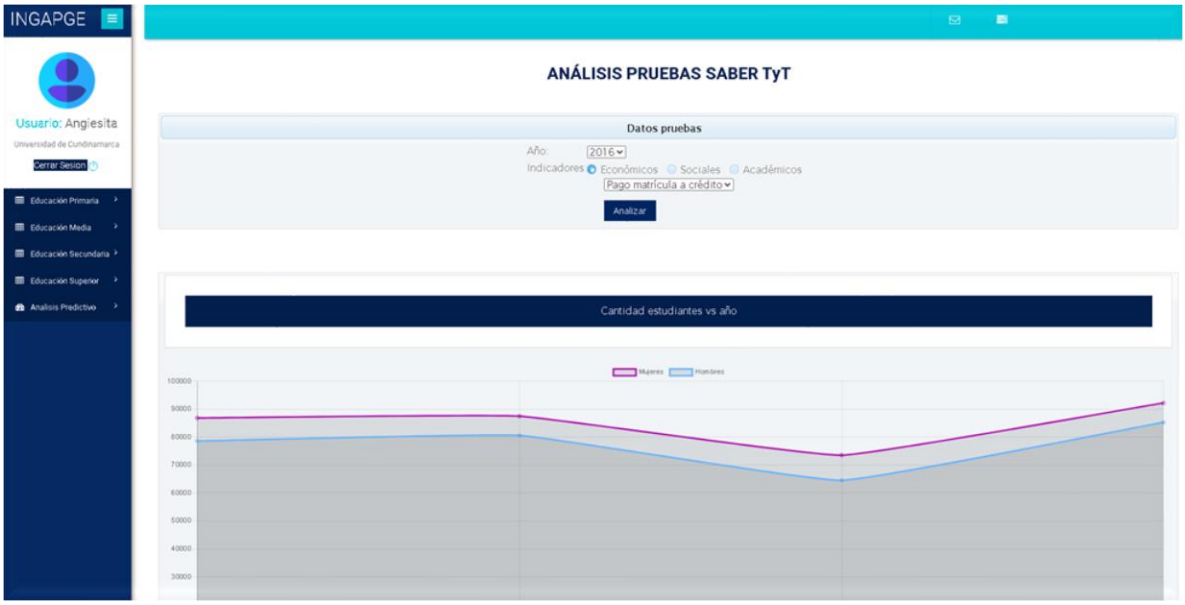


5. Análisis descriptivo de pruebas Saber TyT

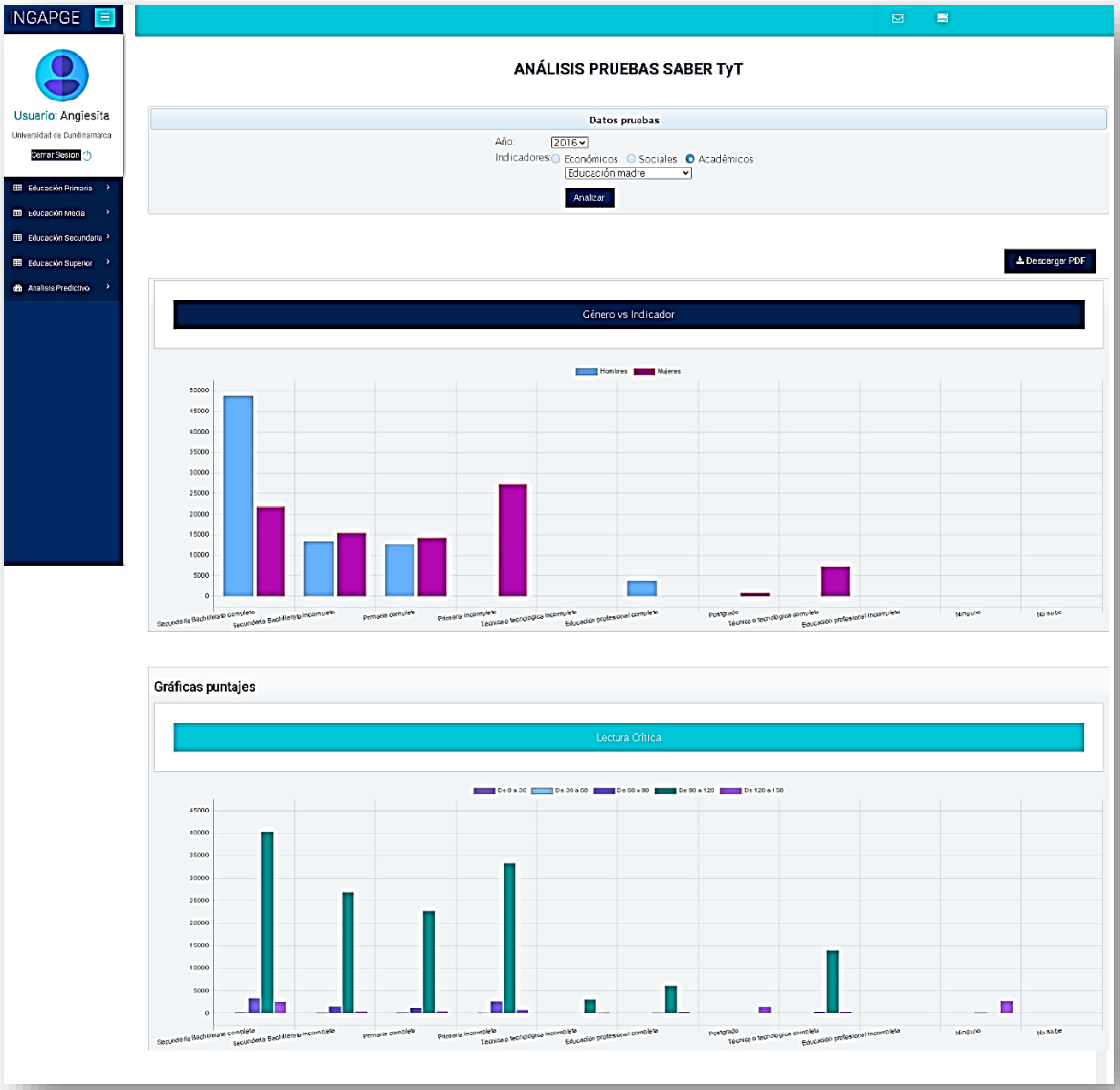
Para la visualización de resultados del análisis descriptivo, el usuario debe elegir la prueba Saber TyT del menú.



Este formulario permite mostrar los resultados obtenidos en el análisis descriptivo de las pruebas Saber TyT realizadas desde el año 2016, en adelante. Tiene visualización por medio de gráficas de línea y barras para mayor entendimiento.



Para generar los resultados del análisis, primero debe seleccionar el año que desee, enseguida el indicador de brecha de género que puede ser de tipo económico, social o académico y finalizar dando clic en el botón Analizar, obteniendo lo siguiente.



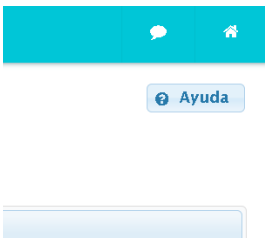
El usuario tiene la posibilidad de obtener estos resultados en formato PDF, para que haga uso de los mismos en investigaciones futuras, para ello simplemente deberá hacer clic en el botón de Generar PDF, ubicado en la parte superior derecha de las gráficas.

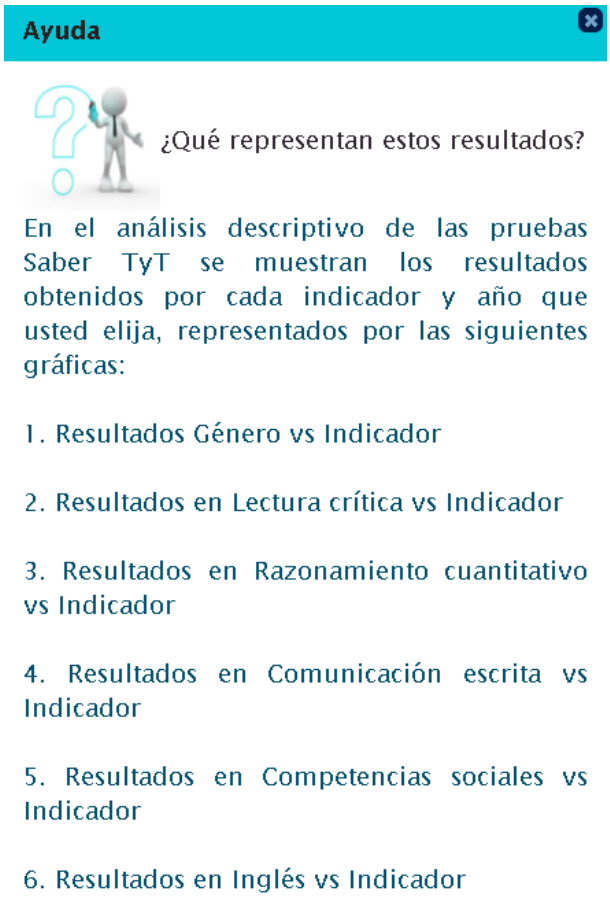


Obteniendo como resultado:



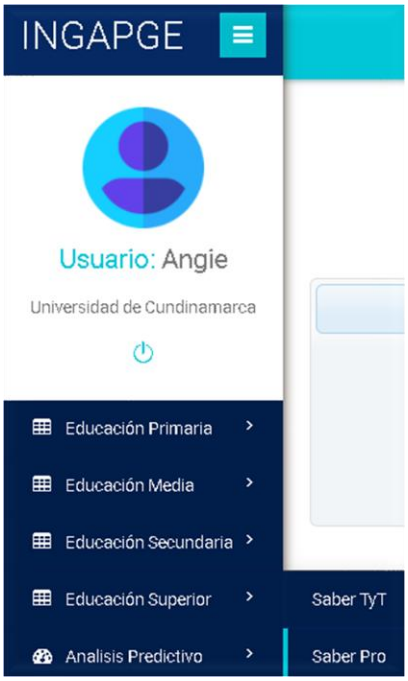
Con el fin de entender un poco más lo mostrado, se cuenta con un botón de ayuda para el usuario.



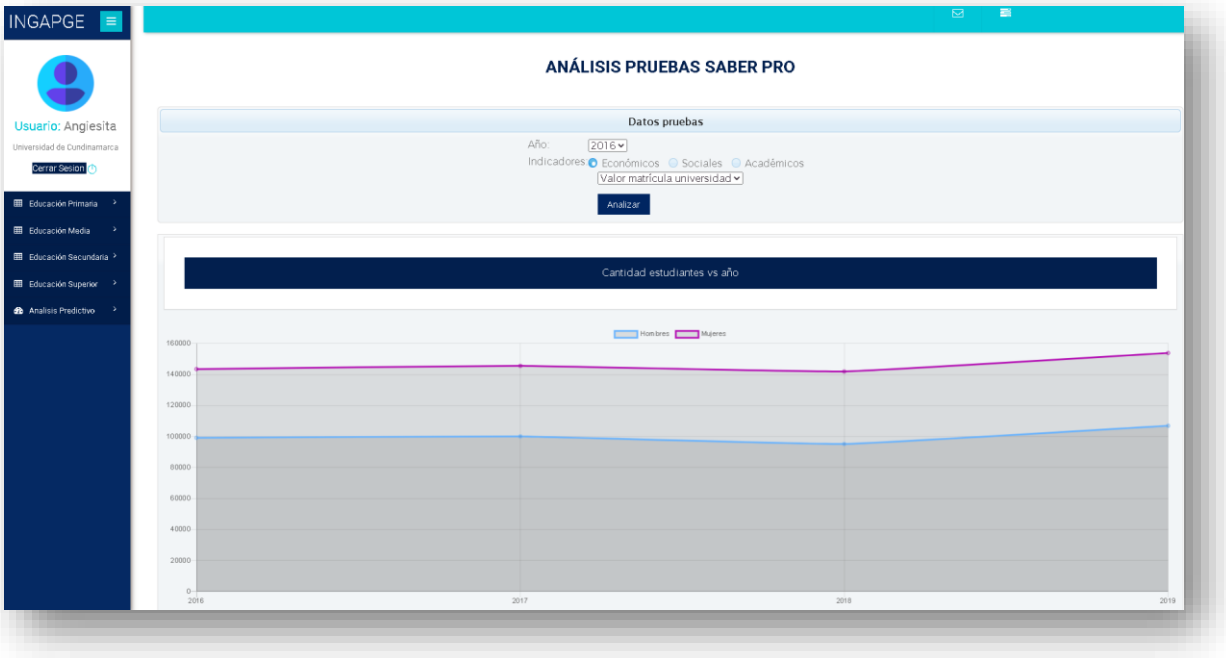


6. Análisis descriptivo de pruebas Saber Pro

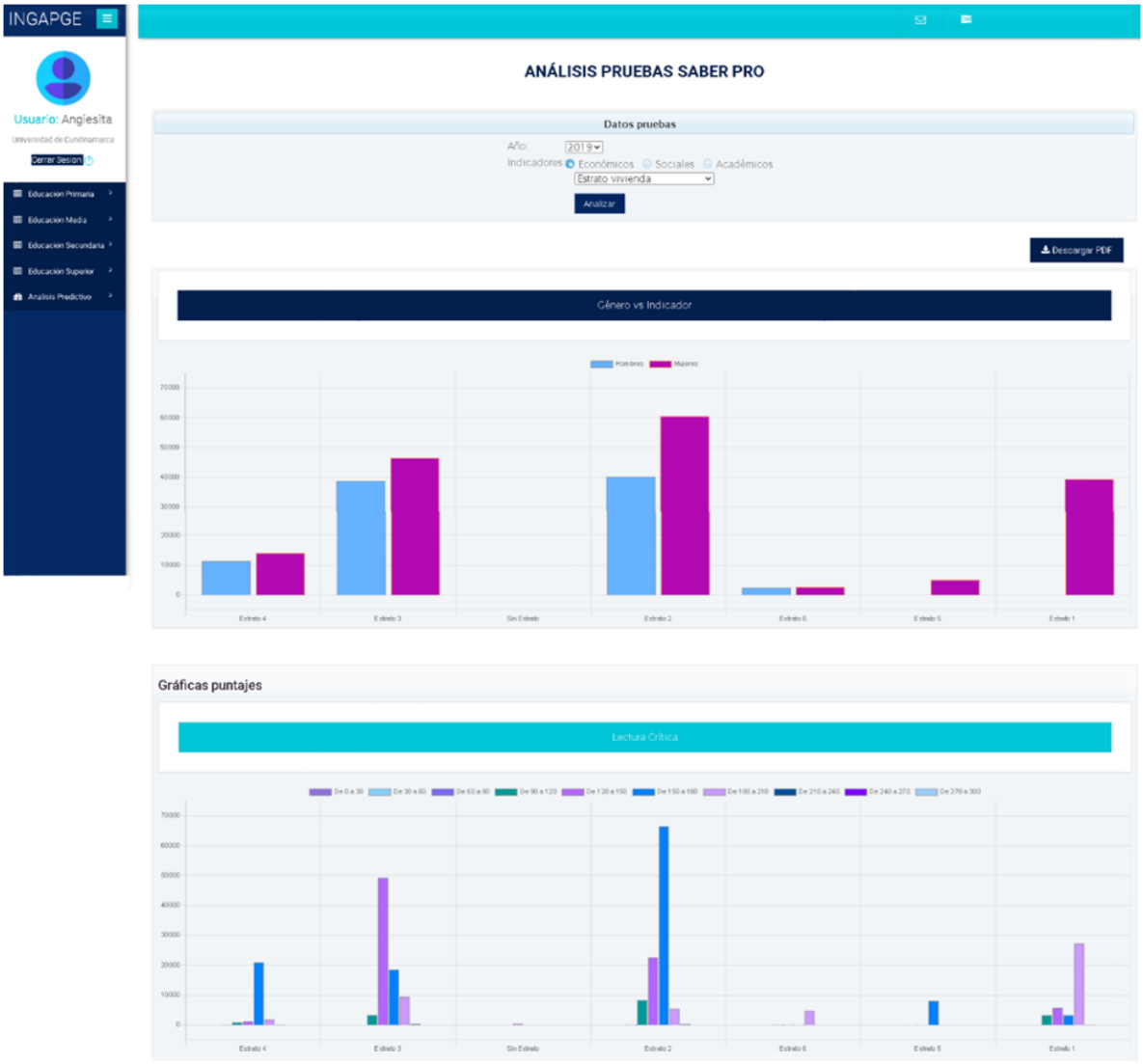
Para la visualización de resultados del análisis descriptivo, el usuario debe elegir la prueba Saber Pro del menú.



Este formulario permite mostrar los resultados obtenidos en el análisis descriptivo de las pruebas Saber Pro realizadas desde el año 2016, en adelante. Tiene visualización por medio de gráficas de línea y barras para mayor entendimiento.



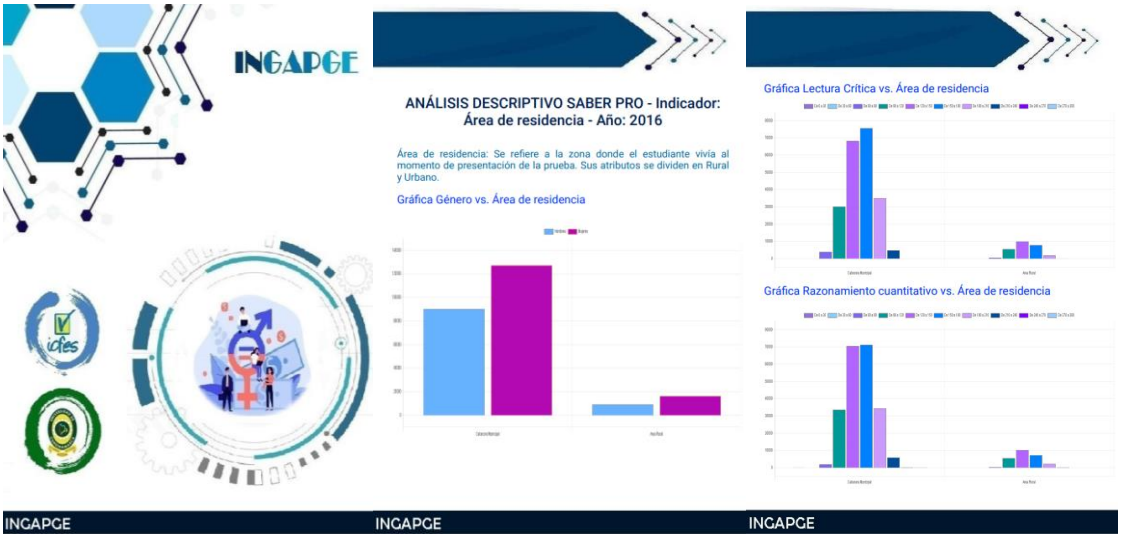
Para generar los resultados del análisis, primero debe seleccionar el año que desee, enseguida el indicador de brecha de género que puede ser de tipo económico, social o académico y finalizar dando clic en el botón Analizar, obteniendo lo siguiente.

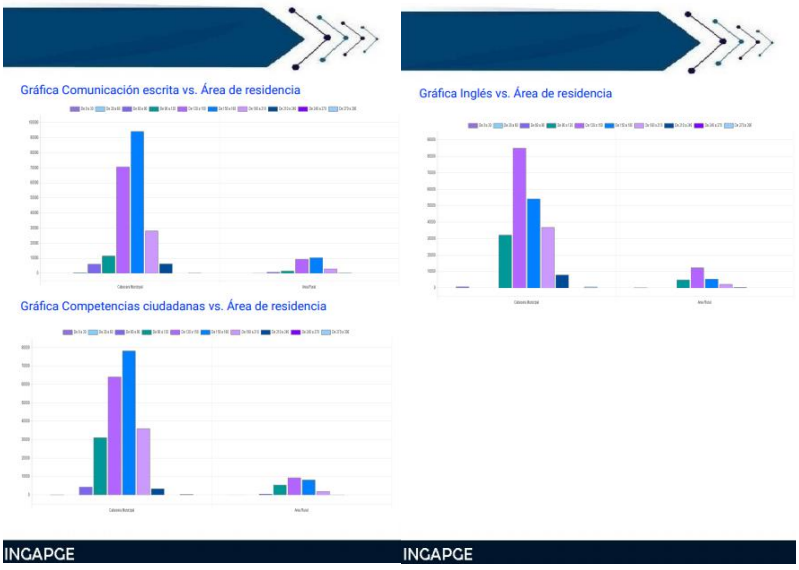


El usuario tiene la posibilidad de obtener estos resultados en formato PDF, para que haga uso de los mismos en investigaciones futuras, para ello simplemente deberá hacer clic en el botón de Generar PDF, ubicado en la parte superior derecha de las gráficas.

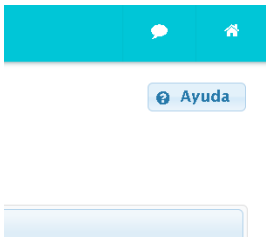


Obteniendo como resultado:



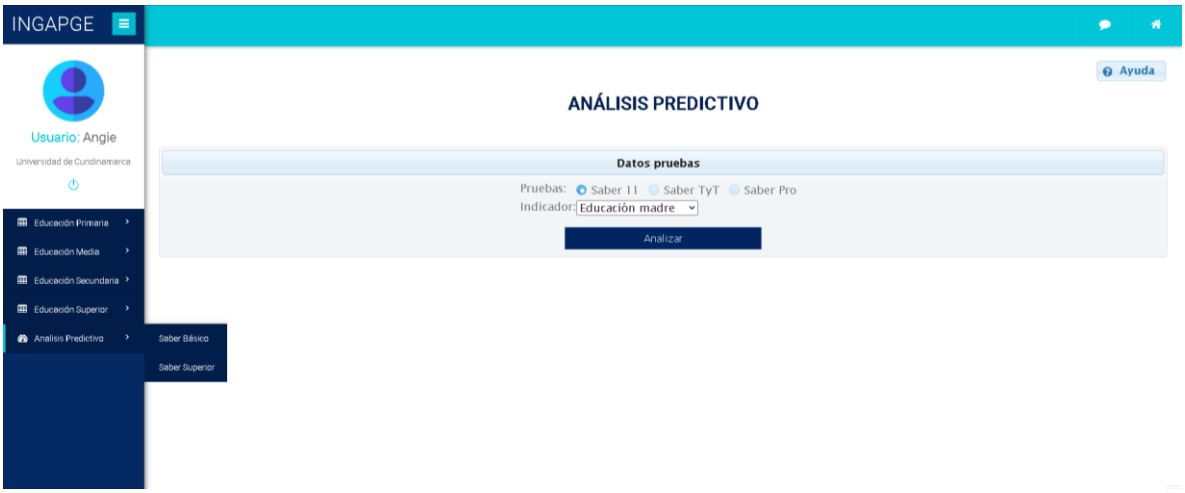


Con el fin de entender un poco más lo mostrado, se cuenta con un botón de ayuda para el usuario.

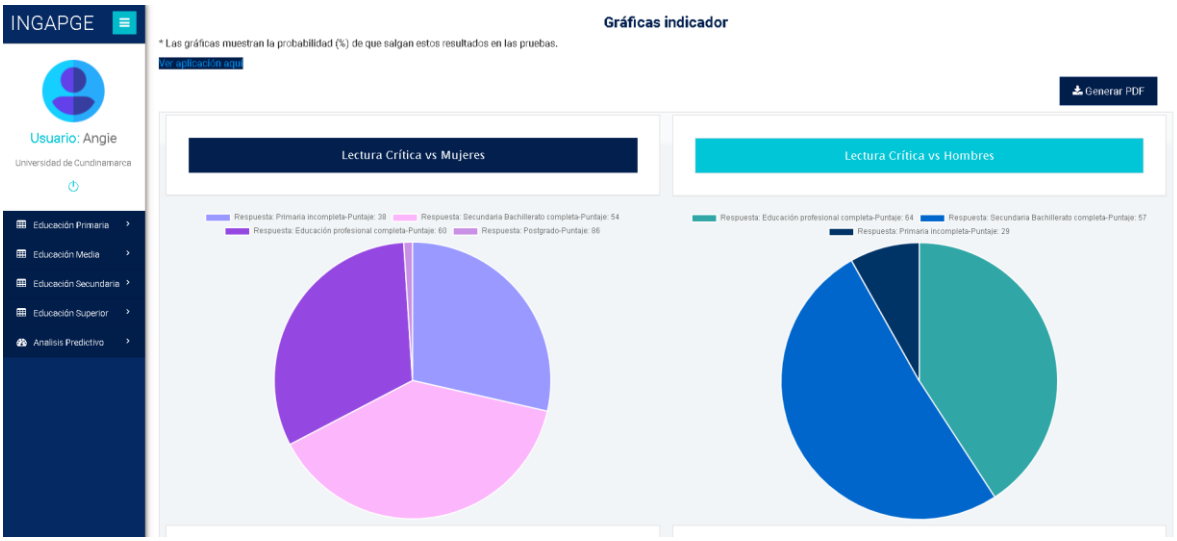


7. Análisis predictivo de las Pruebas Saber Superior

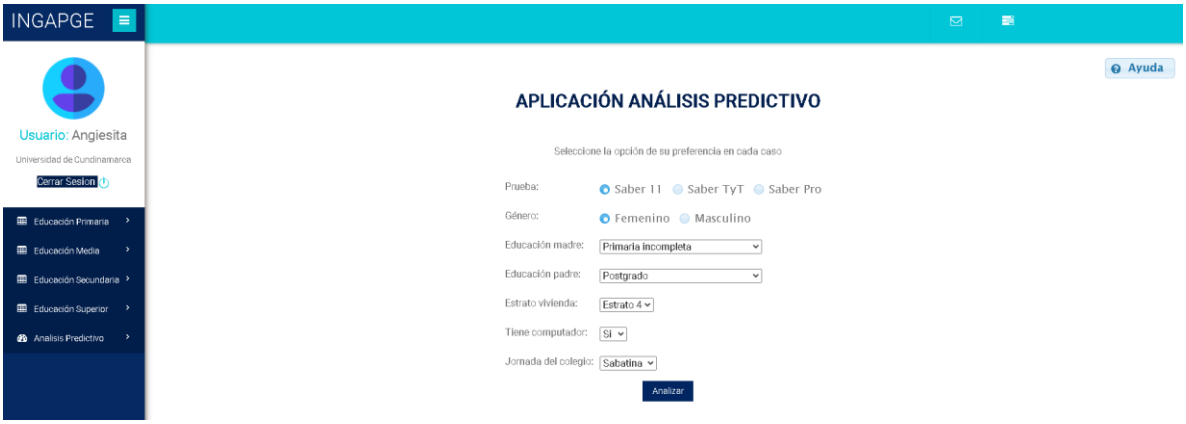
Este formulario permite generar el análisis predictivo de las pruebas Saber 11, Saber TyT y Saber Pro. Se accede por medio de la opción Saber Superior del menú.



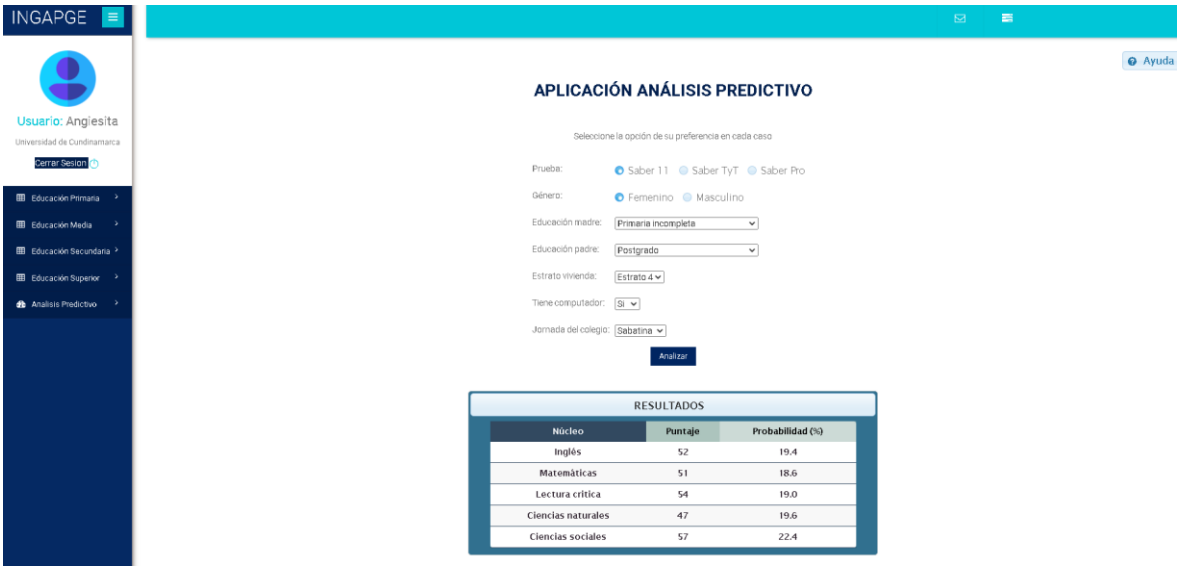
De acuerdo a la selección de prueba e indicador, se muestran gráficas de torta con sus respectivos resultados.



Para la aplicación de los resultados del análisis predictivo se realiza accediendo al link Ver aplicación aquí.



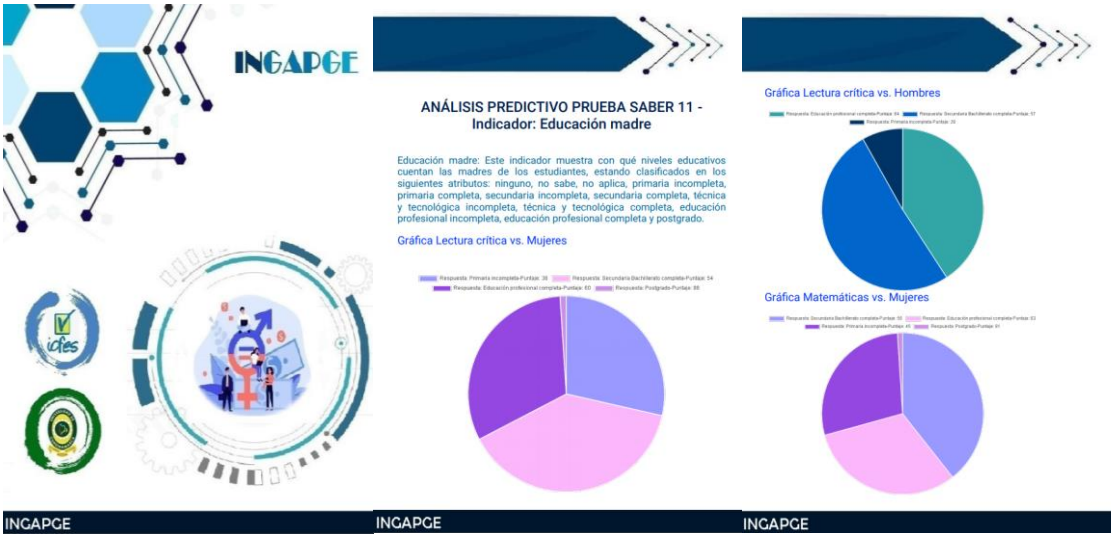
Al seleccionar la prueba, género y una de las categorías en cada indicador, el usuario deberá dar clic en Analizar donde se visualizan los resultados más probables a presentarse en las pruebas de los siguientes años.



El usuario tiene la posibilidad de obtener estos resultados en formato PDF, para que haga uso de los mismos en investigaciones futuras, para ello simplemente deberá hacer clic en el botón de Generar PDF, ubicado en la parte superior derecha de las gráficas.

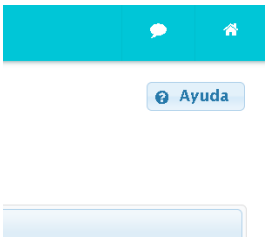


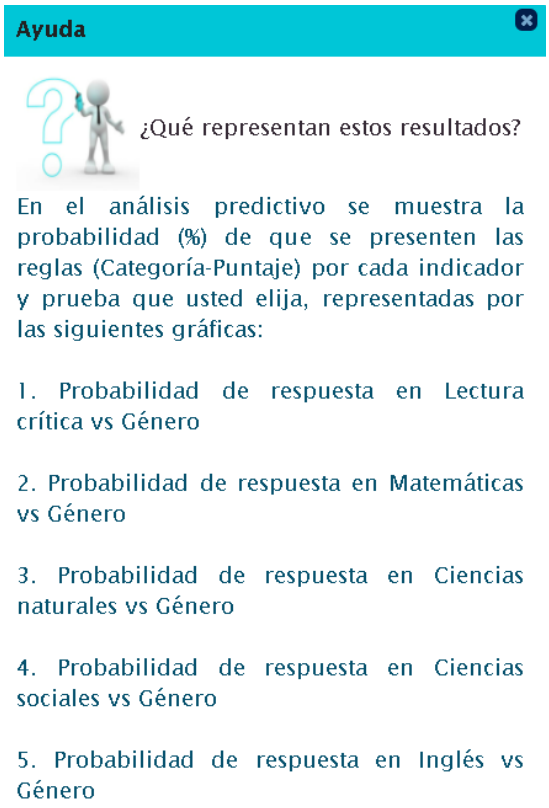
Los reportes muestran los resultados de la siguiente forma:



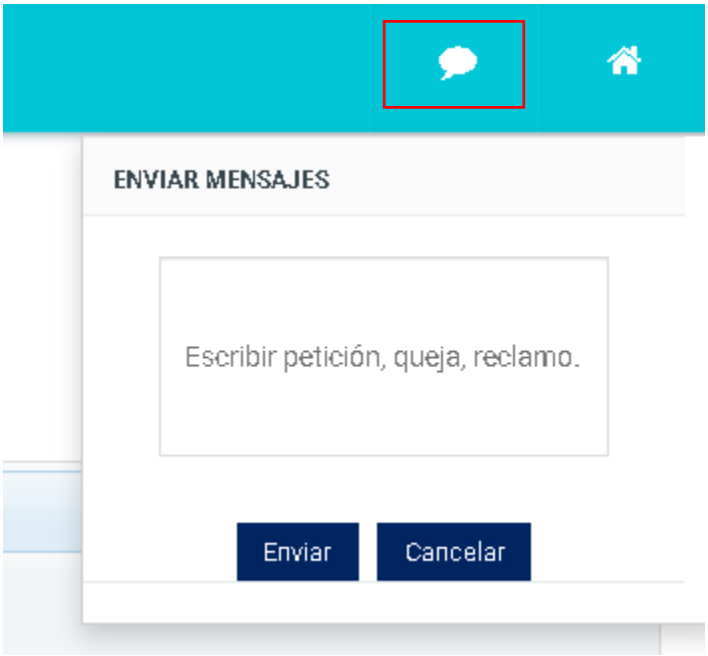


Con el fin de entender un poco más lo mostrado, se cuenta con un botón de ayuda para el usuario.

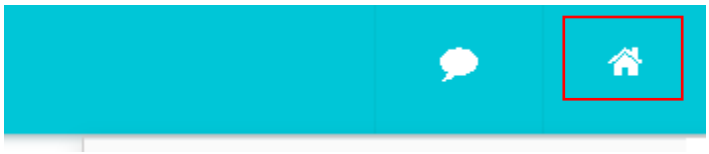




Por otra parte, si el usuario desea contactarse con los desarrolladores del aplicativo puede dar clic al logo de chat y escribir el mensaje respectivo. Y seleccionar la opción Enviar.



Para regresar a la página principal, deberá dar clic en el logo de la casita y será redirigido.



8. Inicio de sesión Administrador

Para el ingreso al aplicativo como Administrador se debe hacer uso del nombre de usuario *admininggapge* y su respectiva contraseña.

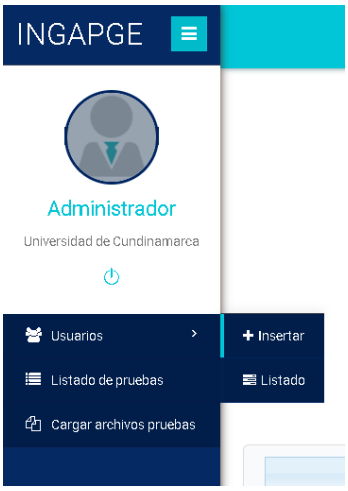


Se tiene como página principal un formulario donde se muestran los registros existentes para cada una de las pruebas Saber existentes en el aplicativo.

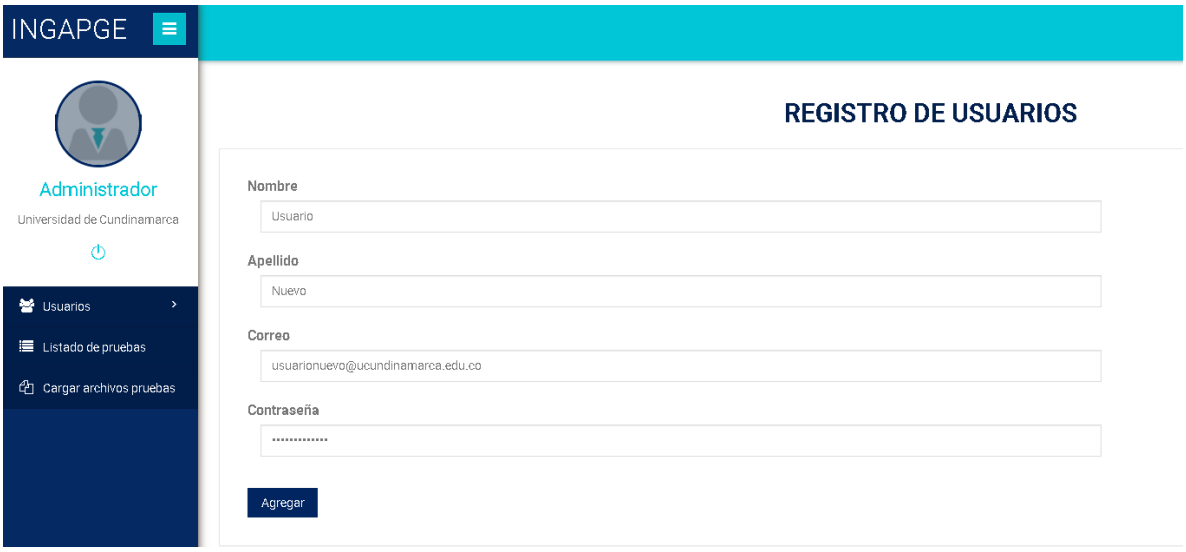


9. Insertar usuarios nuevos

Una de las funciones del administrador es realizar el registro de los usuarios nuevos, de acuerdo a los datos suministrados en el formulario de contáctenos. Para esto debe dar clic en la opción Insertar del menú.



Se llenan cada uno de los campos y se da clic en el botón Agregar.

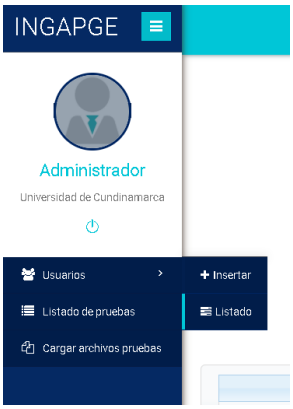


Si está correctamente diligenciado el registro será exitoso, de forma contraria saldrán las validaciones que deben cumplirse.



10. Listar usuarios

Para la visualización de los usuarios registrados se tiene la opción de Listado en el menú.



En este mismo formulario se puede realizar la edición del Nombre, Apellido y Contraseña dando clic sobre el logo de lápiz ubicado al lado derecho de cada usuario.

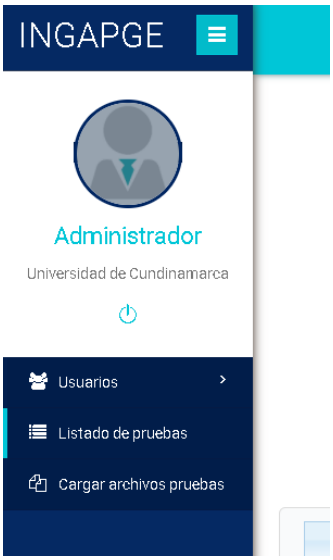


Para la eliminación se debe dar clic sobre el logo de basurita ubicado al lado derecho de cada usuario.



11. Listado de pruebas

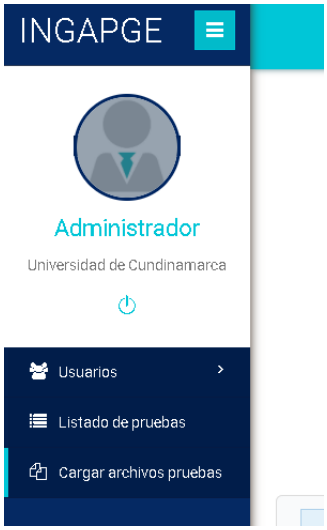
Para visualizar las pruebas cargadas en el aplicativo en sus diferentes años, se tiene la opción Listado de pruebas en el menú.





12. Cargar archivos pruebas

Si desea cargar pruebas nuevas deberá dirigirse a la opción Cargar archivos pruebas del menú.



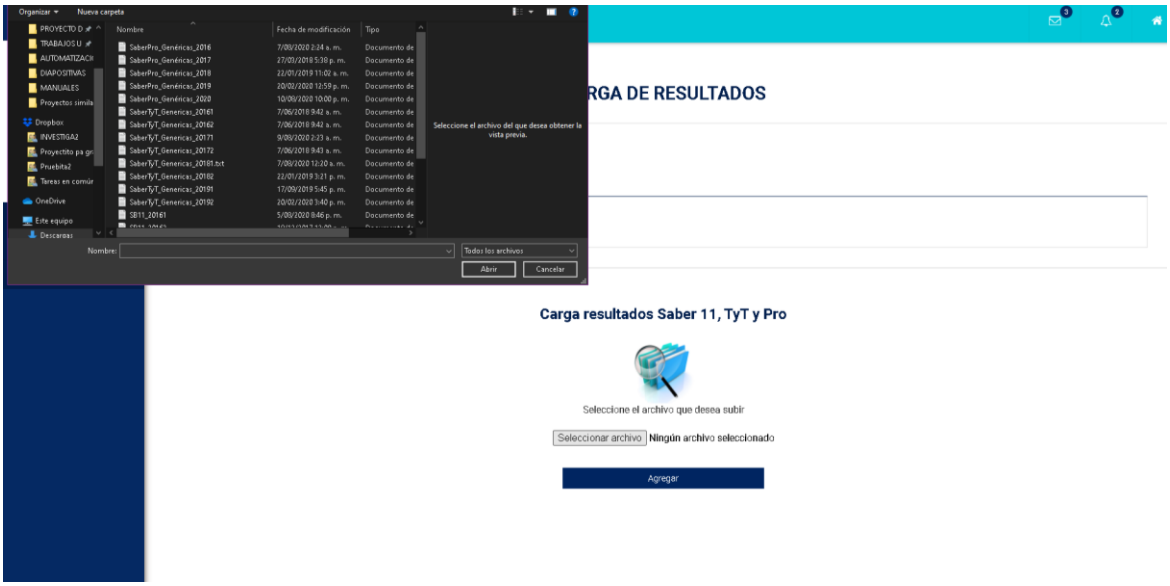
En este formulario deberá elegir la opción Resultados Saber 11, TyT y Pro.



Posteriormente, dar clic en seleccionar archivo.



Allí aparecerá las carpetas del computador, y podrá elegir el archivo de la prueba que sea subir.



Para proceder a la carga dar clic en Agregar.

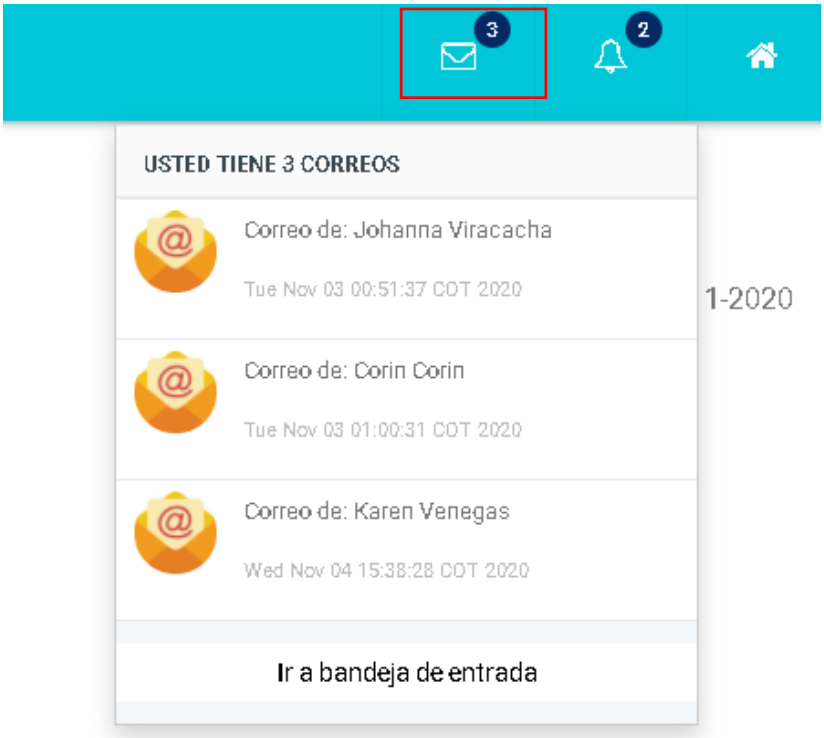


Si el archivo corresponde a un formato .txt y con el nombre e información adecuada se realizará la carga con éxito, de lo contrario saldrá un mensaje de error.



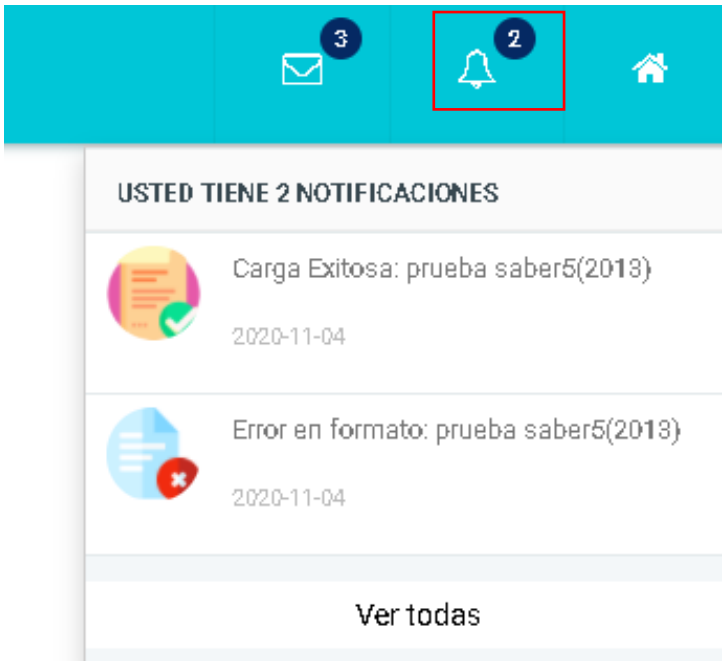
13. Otras funciones Administrador

En la barra superior se cuenta con 3 botones, el primero de ellos es para visualizar cuantos correos sin leer hay, representado por el logo de carta. Al dar clic sobre este se visualiza lo siguiente:



En este desplegable se cuenta con la opción de Ir a la bandeja de entrada, si se da clic aquí lo redireccionará a su correo.

El botón con logo de campana permite ver notificaciones sin leer o pendientes.



Finalmente, para regresar a la página principal, deberá dar clic en el logo de la casita y será redirigido.



7.2. MANUAL TÉCNICO

INGAPGE

MANUAL TÉCNICO



REQUERIMIENTOS DEL SISTEMA

PostgreSQL

- 512 Mb de memoria RAM.
- 1024 Mb máquina virtual.
- 1 GB de espacio de disco duro.
- Sistema operativo: Windows, Linux y Unix.
- Arquitectura del sistema 32/64 bit.
- Protocolo de red TCP/IP.
- Proyector con una resolución de al menos 800 x 600.

NetBeans

Sistema operativo: Microsoft Windows XP Professional SP3/ Vista SP1/ Windows 7 Professional:

- **Procesador:** 800MHz Intel Pentium III o equivalente.
- **Memoria:** 512MB.
- **Disco:** 750 Mb de espacio de disco duro.

Sistema operativo: Ubuntu 9.10:

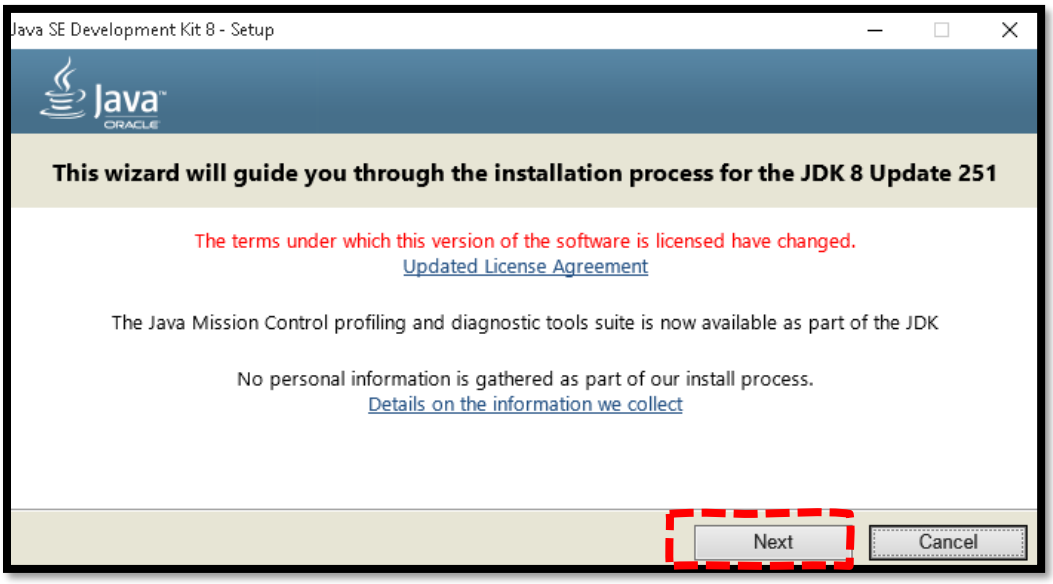
- **Procesador:** 800MHz Intel Pentium III o equivalente.
- **Memoria:** 512MB.
- **Disco:** 650 Mb de espacio de disco duro.

Sistema operativo: Macintosh OS x10.7 Intel:

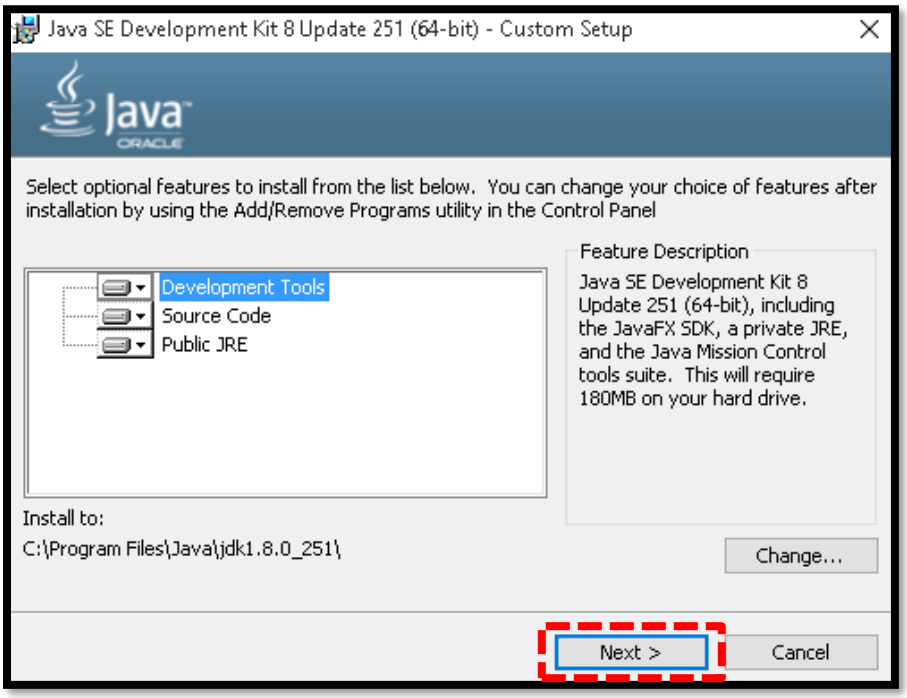
- **Procesador:** Intel Dual - Core.
- **Memoria:** 2 GB.
- **Disco:** 650 Mb de espacio de disco duro.

1. Instalación JDK 8u251.

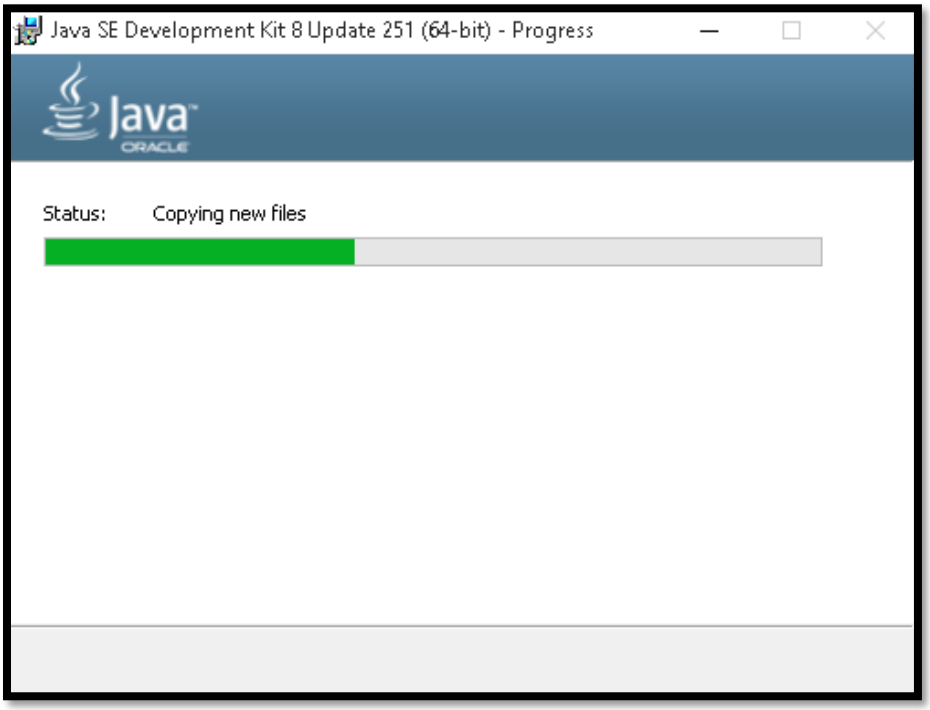
1.1. Ejecutamos el instalador para comenzar con la instalación del JDK 8, seguidamente presionamos el botón “Next”.



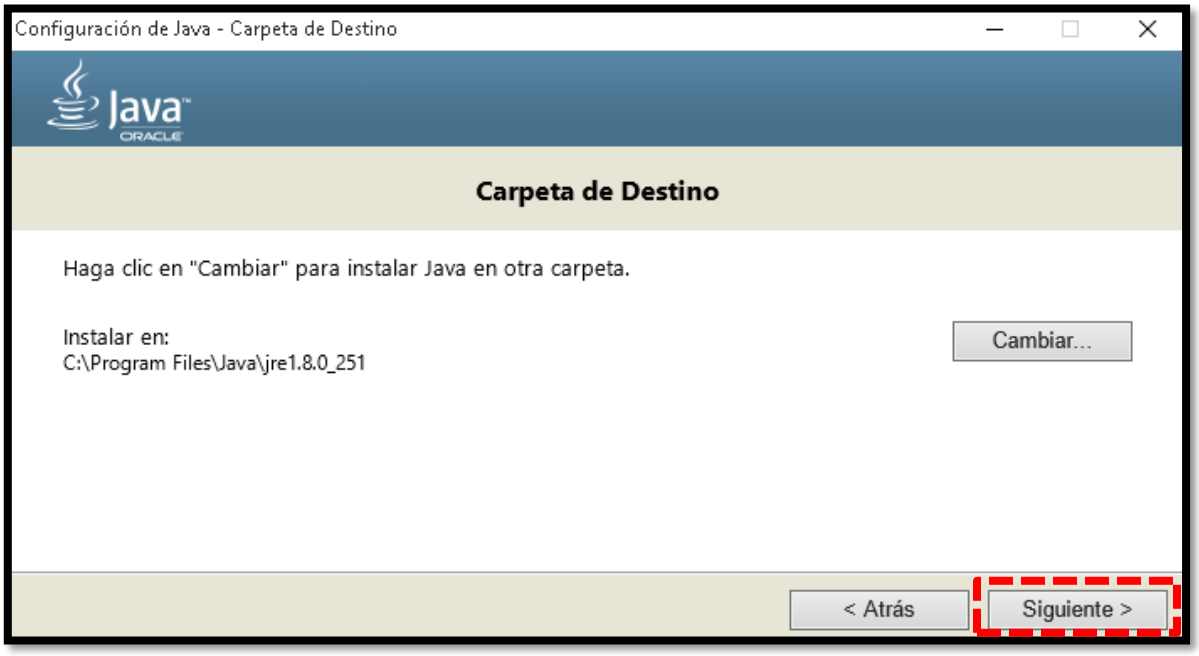
1.2. Aparecerá una pantalla en la cual se pueden escoger las características con las que se desea instalar el programa. Para este caso puntual, se dejan las que se encuentran por defecto y presionamos el botón “Next”.



1.3. En la siguiente ventana se visualiza la barra de instalación del programa, este proceso puede durar varios minutos.



1.4. En la siguiente ventana se especifica la dirección en la cual deseamos que se instale el JDK, luego de ello se da clic en el boton “Siguiente”.



1.5. En este paso se observa el estado de la instalación, este proceso puede demorar algunos minutos.

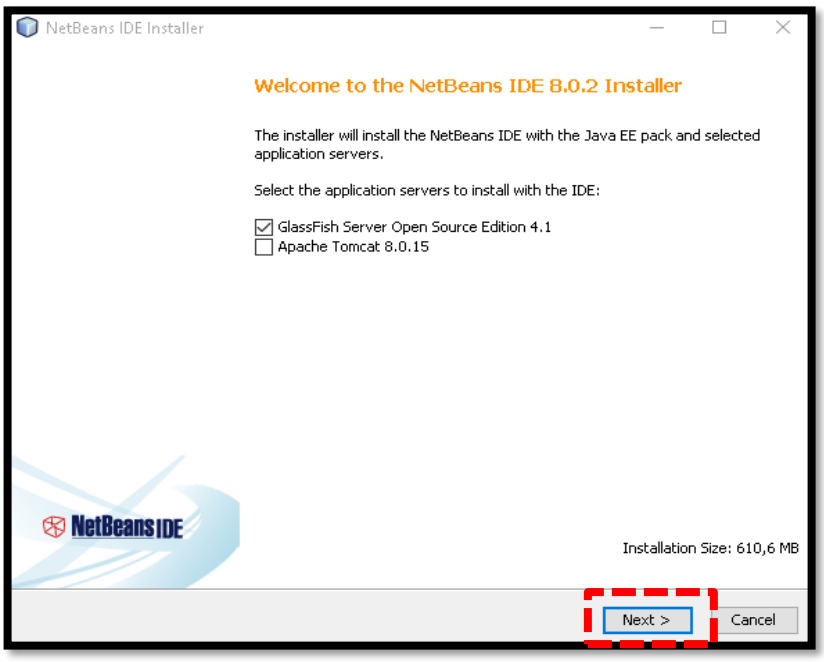


1.6. Después de algunos minutos, se finaliza el proceso de instalación.

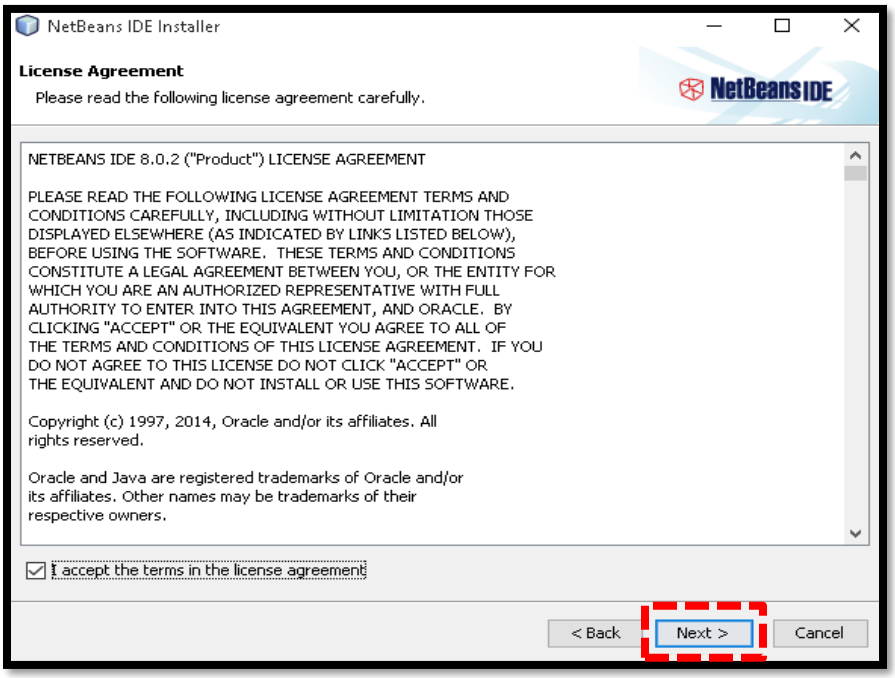


2. Instalación NetBeans 8.0.2.

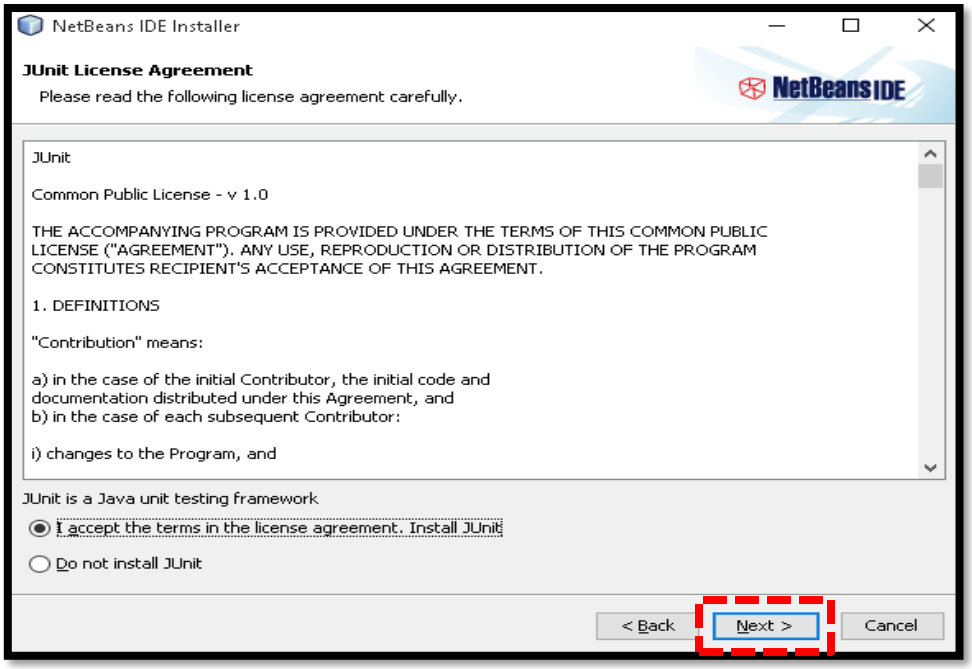
2.1. Ejecutamos el instalador de NetBeans 8.0.2, seleccionamos la instalación del servidor de aplicaciones Glassfish, luego damos clic en el botón “Next”.



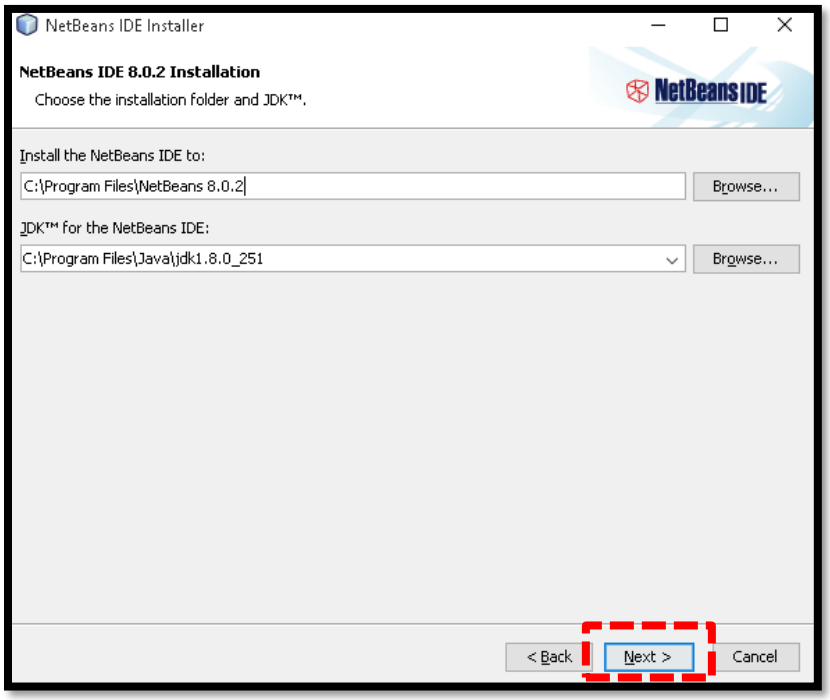
2.2. El siguiente paso es marcar la casilla señalada para aceptar los términos y condiciones de la licencia del software. seguido darle clic al botón “Next”.



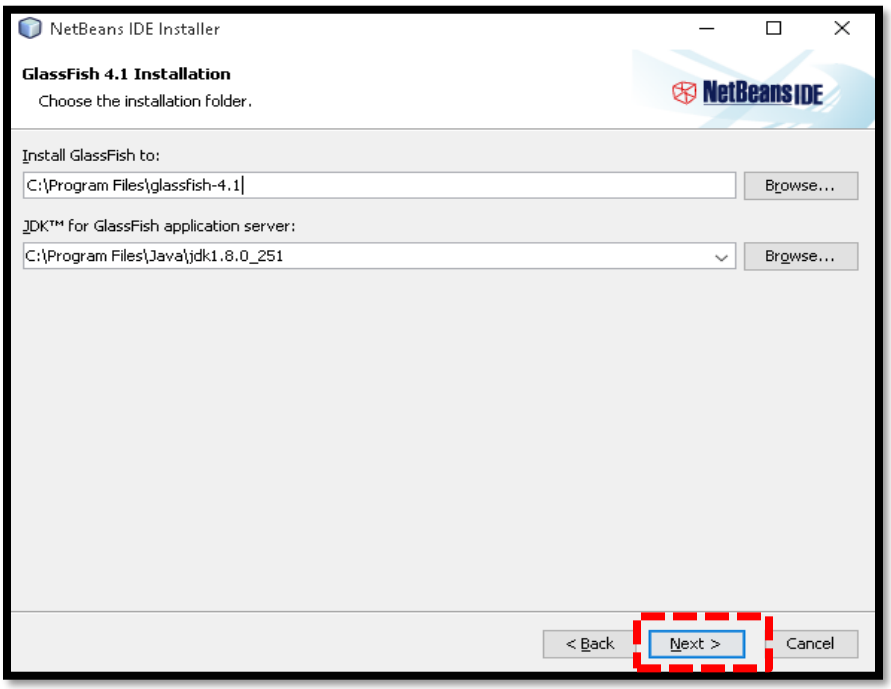
2.3. En la ventana siguiente el instalador da la opción para instalar Junit, un componente para realizar pruebas unitarias. Luego de esto se selecciona el botón “Next”.



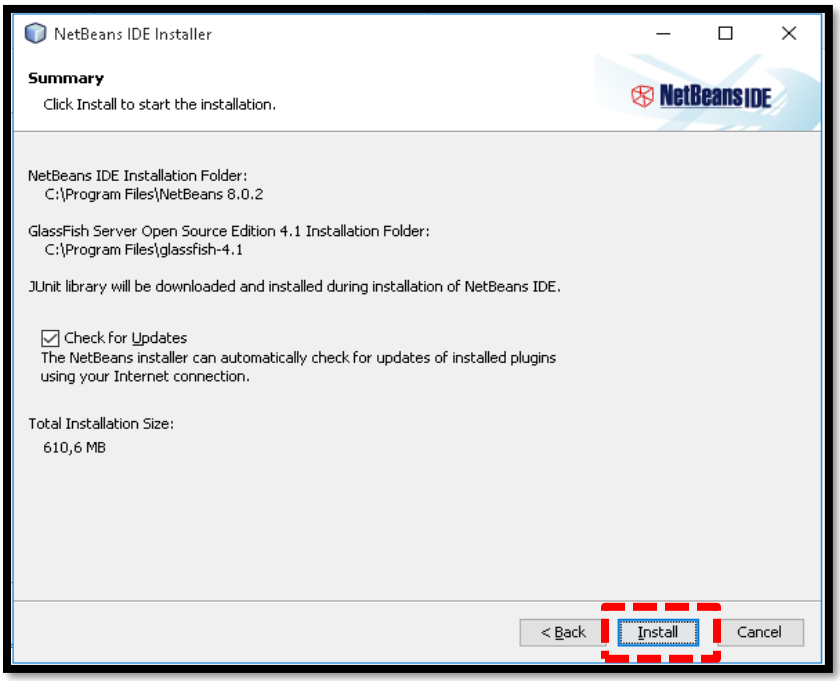
2.4. En el siguiente paso, se escoge la ruta donde se desea instalar el software. La segunda opción es la dirección del JDK instalado en el equipo. Después de realizar los cambios seleccionamos el botón “Next”.



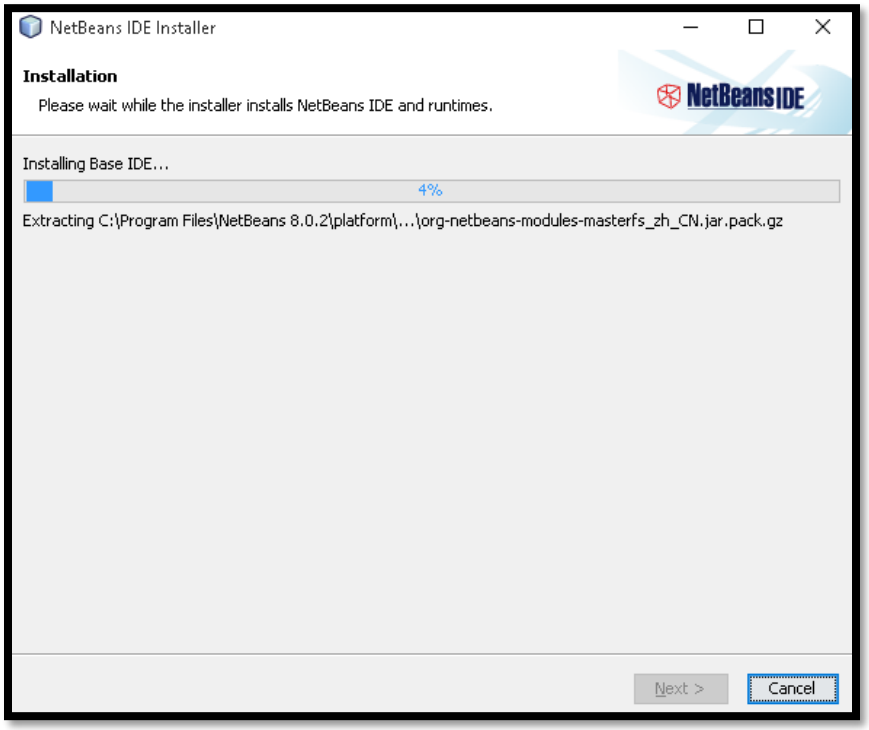
2.5. En la ventana que se visualiza a continuación, se selecciona la carpeta donde se desea instalar el servidor de aplicaciones Glassfish. Damos clic en el botón “Next”.



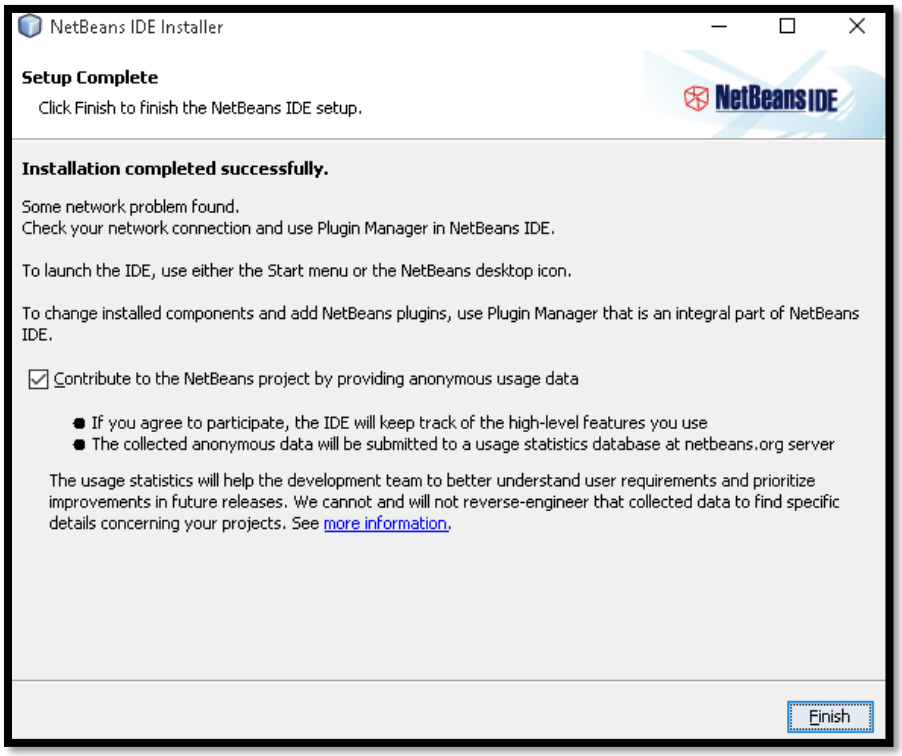
2.6. En el siguiente paso se visualiza cuáles fueron las configuraciones realizadas. Se recomienda dejar la casilla de las actualizaciones activada. Seguidamente seleccionar el botón “Install”.



2.7. En la siguiente ventana se visualiza el proceso de carga, esto puede durar unos cuantos minutos.



2.8. Y ya con esto se finaliza el proceso de instalación de NetBeans 8.0.2.

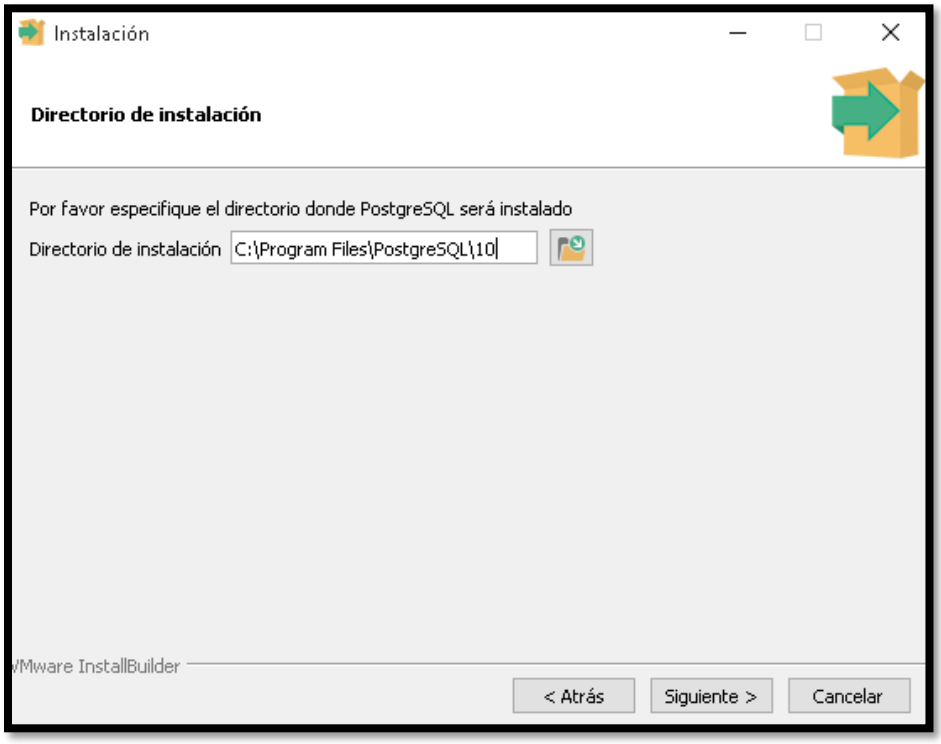


3. Instalación PostgreSQL 10.

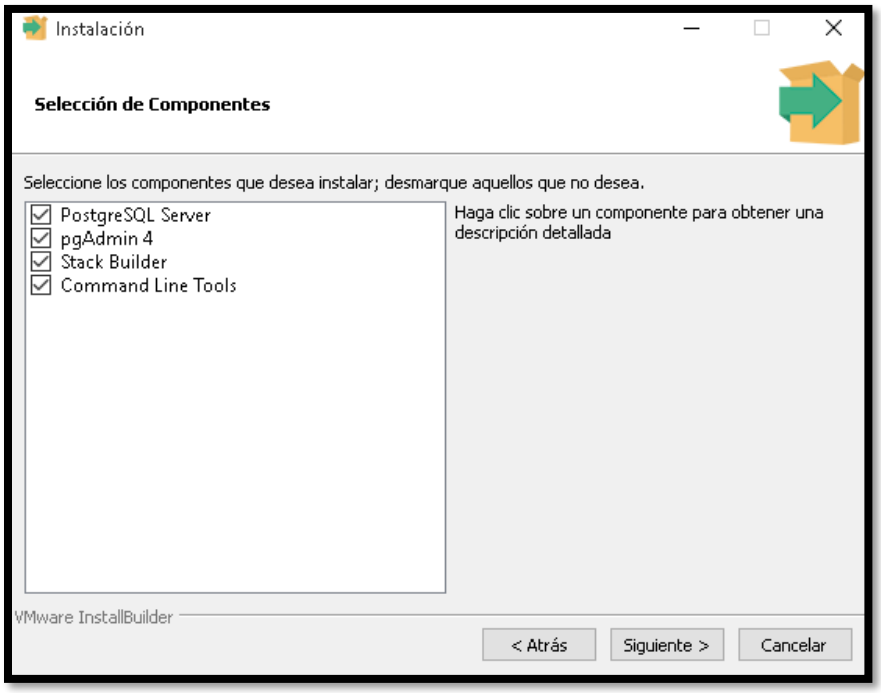
3.1. Ejecutamos el instalador para comenzar el proceso de instalación de PostgreSQL. Aparece la pantalla de Bienvenida y presionamos el botón “Siguiente” para continuar con el proceso de instalación.



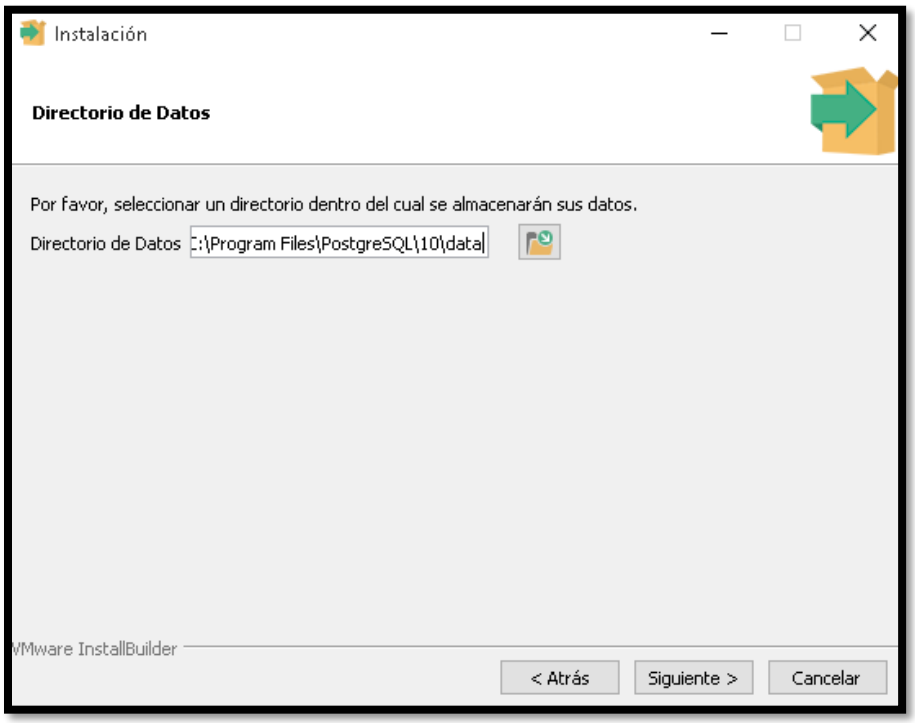
3.2. Especificamos el directorio donde deseamos que se almacenen los archivos y le damos clic en “Siguiente”.



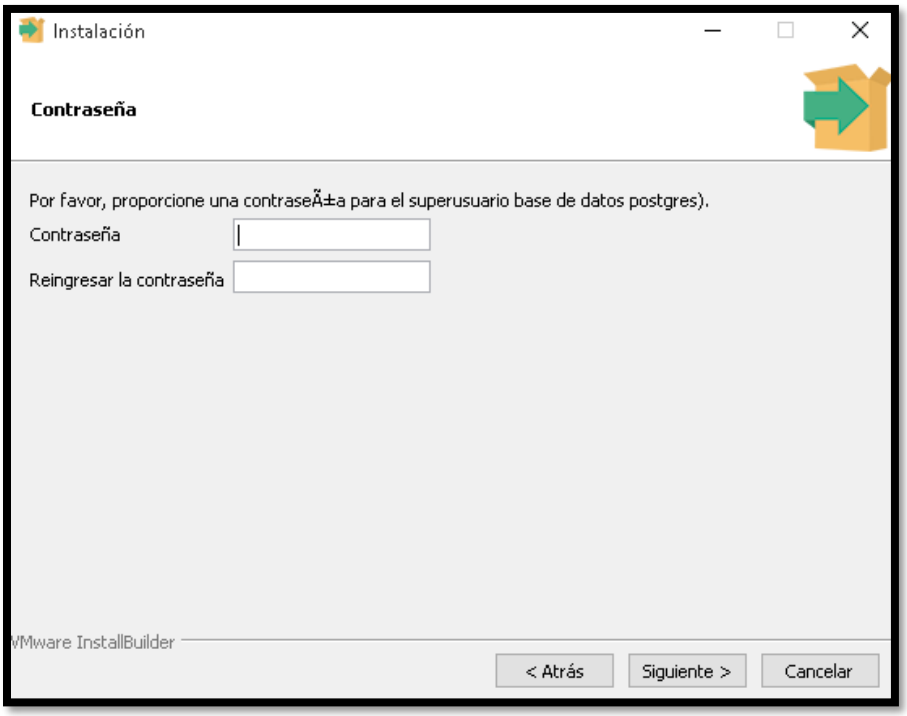
3.3. El instalador nos pide a continuación que seleccionemos cuales son los componentes que deseamos instalar. Para este caso, seleccionamos todas las opciones y damos clic en el botón “Siguiente”.



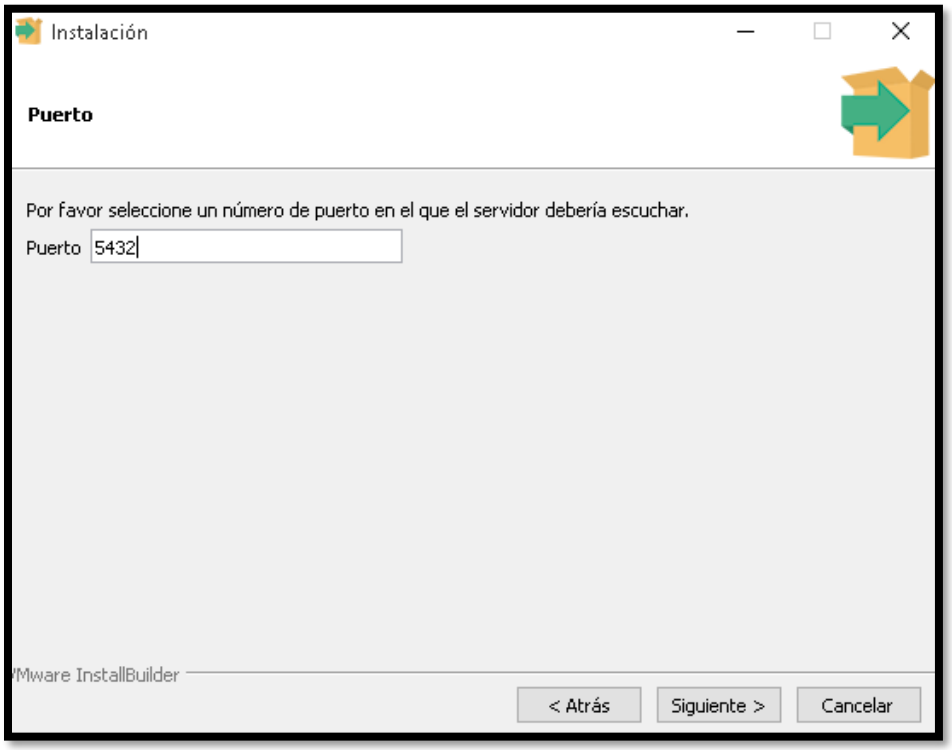
3.4. El instalador nos pide a continuación el directorio para almacenar los datos según la versión del software y presionamos siguiente.



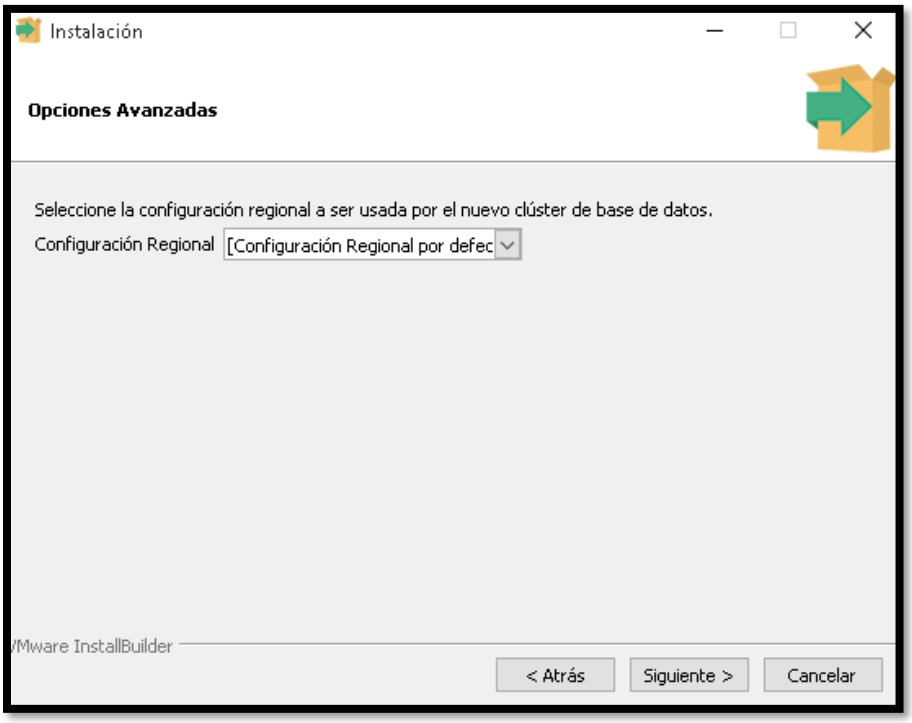
3.5. En este paso debemos dar una contraseña para el manejo de las bases de datos, después de ingresada damos siguientes.



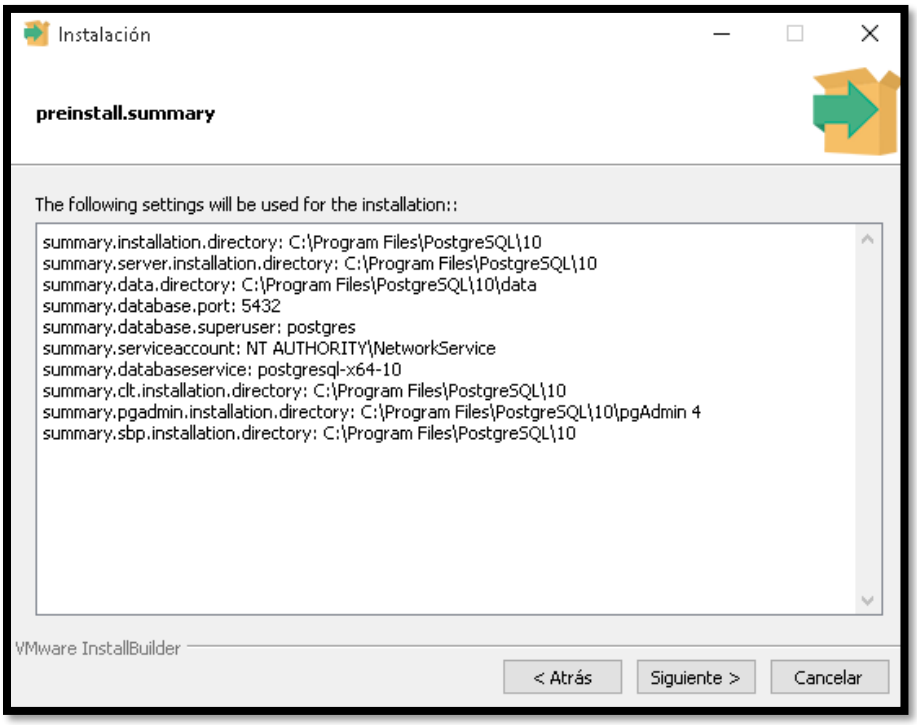
3.6. En este paso diligenciamos el puerto que vamos a usar para el servidor, podemos dejar el puerto que se ve por default, presionamos siguiente.



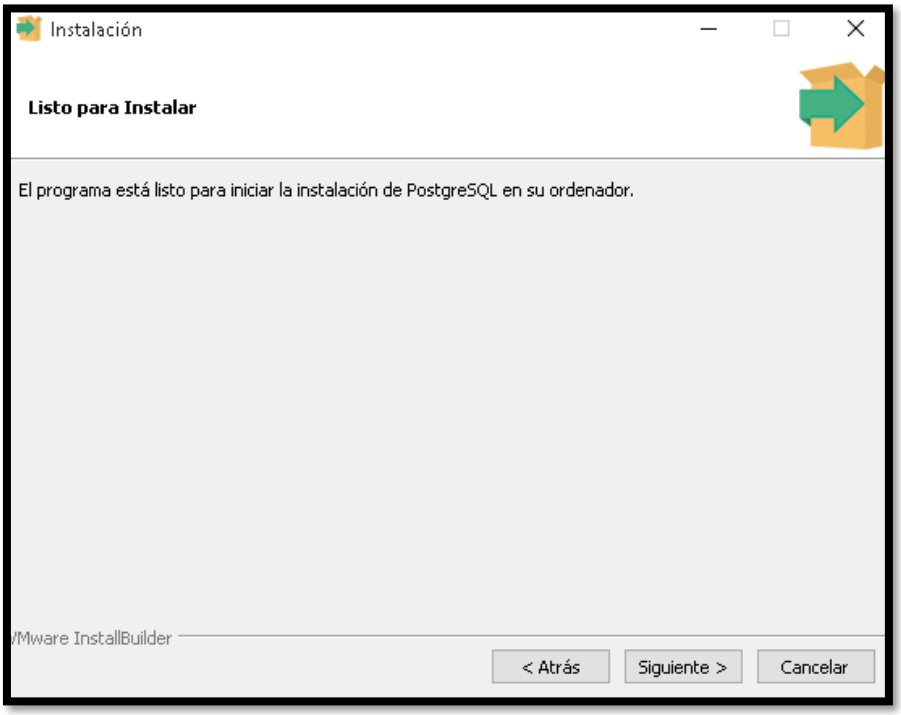
3.7. Seleccionamos la configuración regional usada por el nuevo servidor de base de datos, podemos dejar la que viene por default y presionamos siguiente.



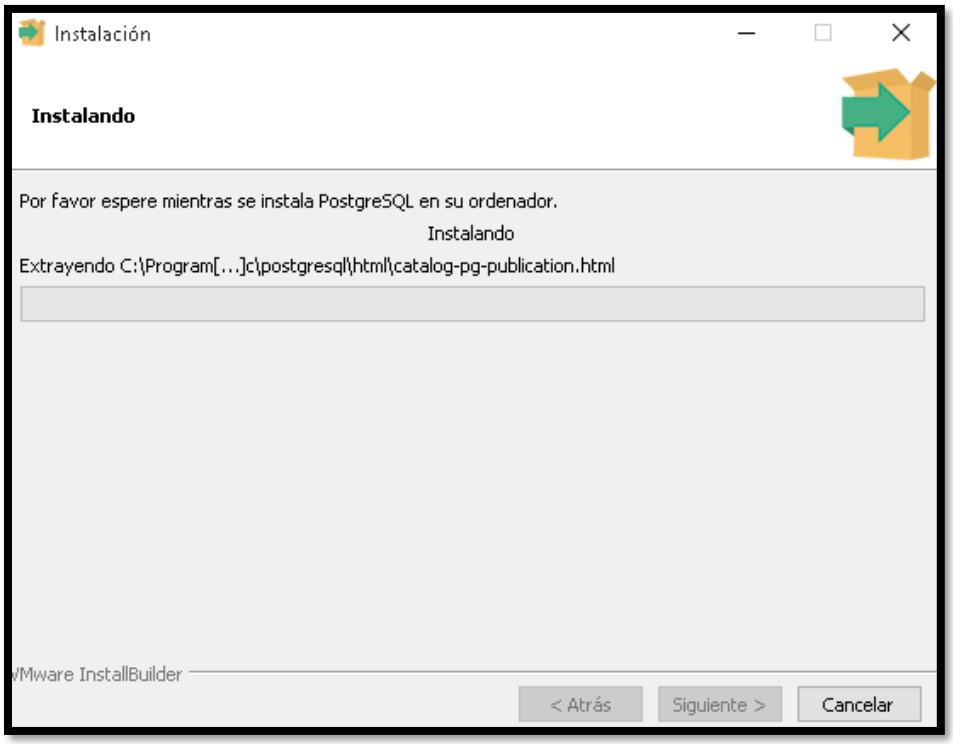
3.8. En la siguiente ventana se muestran aquellas configuraciones que va a usar el software, seleccionamos el botón “Siguiente”.



3.9. Ya establecidos todos los parámetros, comienza la instalación del programa. El instalador nos avisa sobre esto, ahí le damos siguiente.



3.10. Allí comenzara a instalarse el programa, este proceso puede durar unos minutos.



3.11. Con esto ya se finaliza el proceso de instalación de PostgreSQL.



7.3. ARTÍCULO CACIED

Análisis descriptivo y predictivo de brechas de género en las pruebas Saber
Descriptive and predictive analysis of gender gaps in Saber tests

Alisson Catalina Celeita Gutiérrez https://orcid.org/0000-0001-9750-2848
Universidad de Cundinamarca. Facatativá, (Colombia)
acceleita@ucundinamarca.edu.co

Angie Paola Manrique Ravelo https://orcid.org/0000-0003-0302-406X
Universidad de Cundinamarca. Facatativá, (Colombia)
apaolamanrique@ucundinamarca.edu.co

Corín Jazmín Viracacha Viracacha https://orcid.org/0000-0003-4487-938X
Universidad de Cundinamarca. Facatativá, (Colombia)
cviracacha@ucundinamarca.edu.co

Cesar Yesid Barahona Rodríguez https://orcid.org/0000-0001-7673-7381
Universidad de Cundinamarca. Facatativá, (Colombia)
cbarahona@ucundinamarca.edu.co

Resumen

A través del tiempo, en Colombia, se han presentado brechas de género en gran parte de los campos económicos y socioculturales. Asimismo, centrándose en el ámbito educativo, hay estudios que demuestran que allí es donde comienza esta diferencia. Por esta razón, es importante realizar un análisis de los resultados de las pruebas del estado desde el año 2016 y de esta forma hallar los factores que más inciden en esta problemática.

El objetivo principal de este proyecto es analizar los datos brindados por el ICFES de las pruebas Saber para encontrar indicadores de brechas de género y poder relacionar los núcleos temáticos en común desde las pruebas presentadas en la educación básica hasta la educación superior y lograr predecir estos comportamientos. El análisis de la información se realizará por medio de técnicas de minería de datos, tanto descriptivas como predictivas, por su parte, se hará uso de algoritmos matemáticos para calcular la brecha de género. Gracias a esto, se le presentará al usuario una representación detallada de los indicadores hallados, con el fin de que sean usados en futuras investigaciones de entes que se interesen en el estudio y la reducción de esta problemática, las cuales contribuyan en la toma de decisiones que impulsen el desarrollo país. Concluyendo, para brindar la información de una forma más clara se requiere el desarrollo de un aplicativo web.

Abstract

Over time, in Colombia, gender gaps have occurred in much of the economic and socio-cultural fields. Also, focusing on the educational field, there are studies that show that this is where this difference begins. For this reason, it is important to carry out an analysis of the results of the state tests since 2016 and in this way find the factors that most affect this problem.

The main objective of this project is to analyze the data provided by the ICFES of the Saber tests to find indicators of gender gaps and to be able to relate the thematic nuclei in common from the tests presented in basic education to higher education and to predict these behaviors. The analysis of the information will be carried out through data mining techniques, both descriptive and predictive, on the other hand, mathematical algorithms will be used to calculate the gender gap. Thanks to this, the user will be presented with a detailed representation of the indicators found, so that they can be used in future investigations of entities that are interested in the study and the reduction of this problem, which contribute to decision-making. that promote the development of the country. In conclusion, to provide the information in a clearer way, the development of a web application is required.

Palabras clave: Minería de datos; Brechas de género; Educación; Big Data; Análisis descriptivo, Análisis predictivo.

I. INTRODUCCION

Durante mucho tiempo se han presentado brechas de género en gran parte de los campos económicos y socioculturales del mundo. Al observar cada uno de los ámbitos se puede concluir que es una problemática que se encuentra muy lejos

de ser solucionada. En el ámbito laboral se aprecia como las ganancias del sexo femenino son 20.2% menores respecto a las generadas por los hombres, a pesar de que está comprobado que ellas son las que más horas trabajan [1]. La situación no cambia

de dependencia y limitaciones a la hora de conseguir empleo. Esta dificultad como ya se ha mencionado anteriormente no es exclusiva de un área en específico, por ejemplo, de acuerdo con los reportes de la UNESCO en el año 2014, solamente en la educación primaria persistía la disparidad de género en el 37% de los países, así mismo las niñas han conformado el 53% de la población que no recibe algún tipo de educación y se sabe que la mitad de estas nunca asistirán a una escuela [2].

En Colombia según el "Global Gender Gap Report" las brechas de género más significativas se encuentran en el liderazgo político y el fortalecimiento económico. En el caso del consolidamiento político, conforme la resolución sobre la participación de la mujer en la política aprobada por la Asamblea General de las Naciones Unidas en 2011, en todo el mundo, las mujeres siguen siendo marginadas en este campo a causa de diversos factores como lo son los estereotipos, bajos niveles de educación, afectación de manera desproporcionada de la pobreza, etc. Para el año 2014 el puesto del país descendió 12 puestos en la medición en empoderamiento político del Índice de Brechas de Género, ya que a pesar de ser el 52% de la población colombiana, solo el 12% de esta cifra ocupa algún cargo de elección popular [3].

El origen de toda esta problemática por lo general se da desde la primera infancia, puesto que

con los cargos gerenciales, para estos las mujeres ocupan únicamente un 16.2% según el estudio "Report on Women Board Directors of Latin America's 100 Largest Companies" del año 2016, debido a la existencia de estereotipos

la educación recibida de una u otra manera contribuye a que los hombres se destaquen por encima del género femenino, o en otras palabras, "Desde la escuela no se están propiciando los espacios suficientes para que las mujeres sean líderes y lleguen a donde tradicionalmente han estado subrepresentadas" [4]. Este tipo de deficiencias y algunas similares hacen que el alcance de las brechas de género incremente, mostrando así que la calidad educativa empezando por el nivel más básico es indispensable en la disminución de esta inequidad.

En varios países se han realizado estudios que demuestran que al tener un sistema educativo de calidad las diferencias son casi nulas para hombres y mujeres, como es el caso de Estados Unidos que ha venido disminuyendo la inequidad entre los dos sexos. Si en 1970 sólo una de cada 13 personas capaces de obtener 700 puntos en la prueba SAT (Scholarship Aptitude Test) antes de los 13 años era mujer, hoy la cifra es de 1 a 3 y se sigue cerrando camino a la equidad. Otro caso importante para destacar se observa en Marruecos, donde los niños estudian en centros educativos distintos a los de las niñas y con este tipo de enseñanza no se presenta una presión social lo que hace que la brecha de género educativa sea inexistente [5].

Por otra parte, al hacer énfasis en la educación superior en las áreas de la ciencia, tecnología,

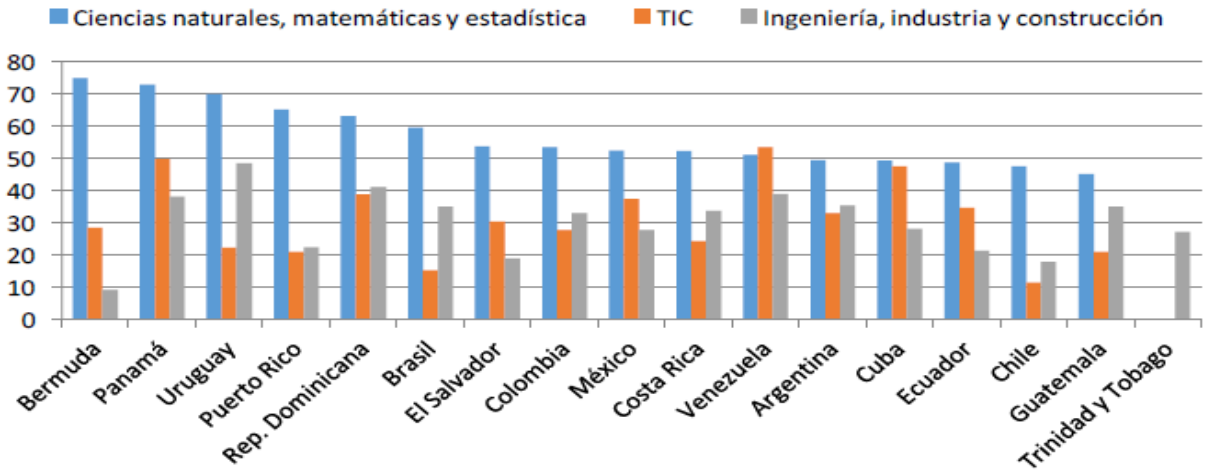


Fig. 1 Porcentaje de mujeres entre los graduados de programas de educación en campos STEM. Fuente: López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. [6]

ingeniería y matemáticas (STEM, siglas en inglés), en los países de América Latina y el Caribe (ALC) se han venido presentando importantes diferencias en el porcentaje de mujeres graduadas de programas educativos pertenecientes a estos núcleos, entre 20% y 50% en el año 2015; mientras que en otras zonas de estudio como lo son las ciencias naturales y afines el porcentaje es mucho mayor encontrándose en un rango del 50% y 80%. A continuación, se muestra en la Fig. 1 que las mujeres profesionales en las tecnologías de la información y la comunicación (TIC) o en ingeniería, industria y construcción es significativamente más bajo que el encontrado en ciencias naturales, matemáticas y estadística. Cabe aclarar que los datos no demuestran los diferentes niveles de estudios, como los son por ejemplo maestrías, especializaciones y/o doctorados, pero sí refleja una conducta generalizada en ALC donde las mujeres no presentan el mismo comportamiento en los diversos campos del conocimiento STEM [6].

En el país para el año 2012, investigadores de la Universidad Tecnológica de Bolívar emplearon técnicas de contraste de medias (t Student) y regresiones por cuantiles en los resultados de las pruebas Saber 11 del año 2009, haciendo énfasis en los núcleos de matemáticas y lenguaje; con los resultados obtenidos se pudo concluir que en los componente de matemáticas los hombres tienen un mejor desempeño que las [7]. Son diversos los factores que pueden afectar el crecimiento o la reducción de la brecha de género, se busca hacer un análisis tanto descriptivo como predictivo en las pruebas estandarizadas nacionales de todos los niveles, desde la educación básica hasta la educación técnica, tecnológica y superior; con el fin de determinar estos indicadores de brecha de género.

En la actualidad, lo más usado para realizar análisis de información es la minería de datos por su capacidad de predicción al ser "un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos [8]. En este proyecto se hará uso de algunos algoritmos de Data Mining como mecanismo de estudio y evaluación de la información de las pruebas, no sin antes hacer un análisis del comportamiento de los mismos para así conocer cuál o cuáles presentan mejores resultados.

El análisis predictivo se define como el proceso de descubrir patrones significativos de datos utilizando técnicas de reconocimiento, estadísticas, aprendizaje automático, inteligencia artificial y minería de datos [9]. Es todo un proceso analítico que se desarrolla de forma automática o sistematizada con el fin de descubrir datos futuros a partir de datos pasados. La selección del modelo más conveniente para el estudio se hace en relación de los datos disponibles, en términos generales cualquier algoritmo de predicción se encarga de encontrar patrones y relaciones ya sea de variables dependientes y/o independientes; la efectividad de los mismos se determina a través de coeficientes de relación de regresión donde los errores deben ser mínimos para que el modelo sea considerado como óptimo.

Existen 2 tipos de modelos predictivos, los de clasificación y de regresión. Los primeros se encargan de predecir la pertenencia a una clase específica, teniendo como resultado binarios, o un sí o un no (en forma de 0 y 1) con su respectivo grado de probabilidad; mientras que los segundos permiten la predicción de un valor, por ejemplo, para el caso de estudio actual son prácticos para hallar porcentajes de relación en cada uno de los indicadores de brecha de género y entre núcleos temáticos de las diferentes pruebas. Por otra parte, las técnicas más usadas son los árboles de decisión, que son modelos de clasificación que dividen los datos en subconjuntos basados en categorías de variables de entrada; las técnicas de regresión lineal y logística siendo uno de los métodos más empleados en estadística por medio de las cuales se estiman las relaciones entre variables, y por último, las redes neuronales que consisten en técnicas sofisticadas capaces de modelar relaciones extremadamente complejas [10].

II. DESARROLLO

La metodología de desarrollo del proyecto consta de 3 grandes etapas, la primera de ellas es el análisis descriptivo de cada una de las pruebas Saber, la segunda es el análisis predictivo de las mismas relacionando los núcleos temáticos comunes entre sí y por último el desarrollo de un aplicativo web que permita mostrar la información hallada en las dos etapas anteriores.

Etapa 1: Para el análisis de los resultados de las pruebas Saber, en primer lugar, se debe hacer uso de los datos abiertos proporcionados por el ICFES (Instituto Colombiano para la Evaluación de la Educación) en su repositorio de acceso abierto, siendo esto un gran volumen de información, a la hora de ser examinada se vuelve en una tarea muy

compleja, por esta razón se hará uso de la minería de datos educativa puesto que facilita el procesamiento de la misma. Además, tiene como ventaja con respecto a los métodos estadísticos, el descubrimiento de información oculta y el manejo óptimo de los datos, convirtiéndose en una herramienta más eficaz al momento de la investigación.

Como primer opción de técnica de análisis se ha optado por usar el algoritmo K-Medias, sin embargo, antes de iniciar este proceso descriptivo se ha decidido realizar un estudio del comportamiento de varios algoritmos de minería de datos para poder escoger el más adecuado. A partir de este momento, el desarrollo del proyecto implementará una de las metodologías más reconocidas en Data Mining, como lo es CRISS-DM. que es considerada la "metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial" [11], la cual incorpora a su vez la gestión de proyectos y permite un movimiento bidireccional entre sus etapas, a continuación, se muestra los pasos a seguir en esta fase:

1) Comprensión del negocio: Se pretende desarrollar un modelo descriptivo de los indicadores de brecha de género, el cual relacione aspectos económicos, socioculturales, familiares y personales. Condicionados por los datos disponibles de cada una de las pruebas se ha decidido empezar el análisis desde el periodo académico 2016 hasta su última actualización.



Fig. 2 Análisis del negocio. Fuente: Autor(es)

2) Comprensión de los datos: Recopilación de la información de los repositorios del ICFES. En la cual se obtiene los puntajes globales y específicos presentados por los estudiantes, factores socioeconómicos tanto familiares como personales, localización del estudiante y los datos relacionados con la institución (en unas pruebas sería el tipo, mientras que en otras el nombre de la misma).

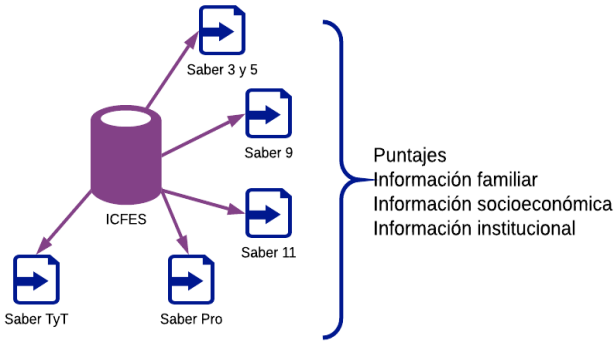


Fig. 3 Recopilación de resultados. Fuente: Autor(es)

3) Preparación de los datos: Se empezará en este punto de la investigación el tratamiento de los datos. Se deberá extraer, transformar y cargar la información de manera que quede seleccionada para su procesamiento.



Fig. 4 ETL (Extracción, transformación y carga).
Fuente: Autor(es)

4) Modelado: En esta fase se hará uso del algoritmo K-Medias como regla de agrupación para caracterizar a los estudiantes según su comportamiento y similitudes encontradas.

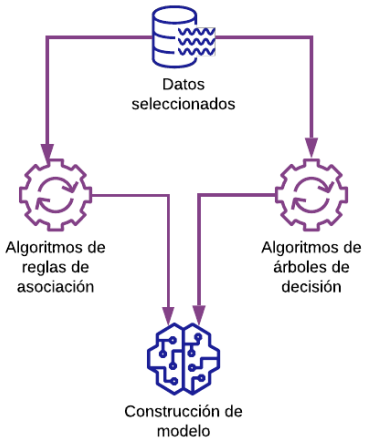


Fig. 5 Creación del modelo con algoritmos de minería de datos. Fuente: Autor(es)

5) Evaluar: Se evaluará la calidad del modelo desarrollado, determinando su margen de error y exactitud.

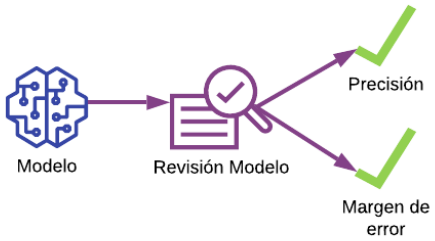


Fig. 6 Verificación del modelo. Fuente: Autor(es)

¿Cuál es el procedimiento para el análisis de conglomerados K-Medias?

Para entrar en contexto, el análisis de conglomerados (Cluster Analysis) es una técnica multivariante que permite agrupar los casos o variables de un archivo de datos en función del parecido o similaridad existente entre ellos [12]. Esta técnica se considera propicia para la investigación porque permite encontrar variables a partir de las características de la información. El uso de la misma inicia con la definición del conjunto de variables que van a ser tratadas, posteriormente se realiza el agrupamiento de los datos. Si la distancia entre los elementos del mismo clúster es pequeña (intra-cluster), indicaría que la similitud para estos casos es alta, de manera que para los elementos externos la similitud es baja. De forma resumida, se puede decir que la fase final es la verificación de convergencia, siendo esta la más importante puesto que describe los resultados obtenidos durante todo el proceso y si alguno de ellos no resulta convergente entonces se deberá volver a iniciar al paso inicial.

Una vez culminado el proceso descriptivo se hará uso de esta información para generar tablas y gráficas en el cual se evidencien los resultados analíticos de las pruebas, las variables halladas y los años de estudio; de manera clara y concisa para el usuario final.

Etapa 2: Posteriormente, se realizará un análisis predictivo para determinar las brechas de género en todos los niveles educativos, para ello se hará uso de la técnica de árboles de decisión, siendo esta una de las más potentes a la hora de analizar decisiones y probabilidades, establecer normas para predecir comportamientos e identificar relación entre las variables halladas. Dentro de esta técnica se busca seleccionar de un conjunto de resultados los que mejor precisión tengan sobre los datos seleccionados.

Con la determinación de las predicciones se quiere mostrar la relación, si es que existe, entre los núcleos temáticos en común, visualizando su comportamiento a través de cada uno de los niveles de la educación haciendo énfasis en las áreas de las matemáticas, ciencias, tecnológicas e ingenierías (STEM).

Para la exploración de los datos y la aplicación de los algoritmos de minería, se hará uso de la herramienta WEKA (entorno para análisis del conocimiento de la Universidad de Waikato) ya que, “proporciona muchos algoritmos diferentes para el análisis de datos y el aprendizaje automático. WEKA es de código abierto y está disponible gratuitamente. También es independiente de la plataforma [13].

Etapa 3: La etapa final del proyecto consiste en la visualización de los indicadores de brecha de género por medio de gráficas y estadísticas en un aplicativo web, de manera que para el usuario final los resultados sean claros y de fácil comprensión. En la construcción de esta herramienta se implementará la metodología de desarrollo ágil SCRUM, ya que es un modelo general de gestión de entornos de producción basados en rutinas, donde el factor más importante es el conocimiento aplicado de las personas que hacen parte del proyecto, aún más primordial que el mismo proceso tecnológico [14]. Teniendo en cuenta que, su importancia radica también en las iteraciones

que permiten el desarrollo incremental del proyecto desde su concepción hasta que se dé por terminado.

Todo el proceso descrito anteriormente no tendría sentido sin que exista una forma de que el usuario final visualice los resultados de ambos análisis, por esta razón se desarrollará un aplicativo web que cumpla con la función de mostrar las gráficas y estadísticas de cada una de las pruebas y en cada uno de los períodos evaluados, permitiendo que esta información pueda ser descargada en forma de reporte tipo PDF.

III. CONCLUSIONES

El uso de minería de datos para el proceso de análisis permite tener acceso a información oculta que con los métodos comúnmente usados de estadística no se puede obtener, además, facilita el procesamiento de grandes volúmenes de datos.

El análisis predictivo se ocupa principalmente de predecir o anticipar resultados futuros basados en minería de datos existentes. Esta información se convierte en importante a la hora de ser analizada porque permite comprender su comportamiento futuro, o en términos coloquiales lo que puede suceder. Esta capacidad de predecir da un poder inmenso de planificación y brinda la posibilidad de solución a diferentes problemáticas ya que entiende cuál es la conducta de una o varias cosas en distintos escenarios de evaluación.

Gracias a las técnicas de minería de datos, en este caso, de agrupamiento y clasificación ad hoc se puede generar modelos descriptivos y predictivos consistentes de la realidad observada.

Con el desarrollo de esta investigación se contribuirá a futuros estudios sobre brecha de género, recomendando así hacer un comparativo de estas pruebas (Saber) con las realizadas a nivel internacional como las PISA.

WEKA facilita la aplicación de los algoritmos de minería de datos siendo una herramienta fiable y de fácil uso para el usuario.

REFERENCIAS

[1] Semana, "¿Por qué las mujeres siguen ganando menos?", 2016.

[2] "La Paridad de Género", Unesco, 2019. [Online]. Available: https://gem-report-2016.unesco.org/es/chapter/la-paridad-de-genero/. [Accessed: 01- Oct- 2019].

[3] ONU Mujeres, UNDP, UNFPA, and USAID, De los Objetivos de Desarrollo del Milenio a los Objetivos de Desarrollo Sostenible, no. 88. 2016.

[4] M. Junguito, "El empoderamiento de las mujeres tiene que empezar desde la infancia", Semana.com, 2019. [Online]. Available: https://www.semana.com/Item/ArticleAsync/615282?nextId=617220. [Accessed: 11- Oct- 2019].

[5] T. Devitt, "Culture, not biology, underpins math gender gap", *University of Wisconsin–Madison*, 2009.

[6] V. López-Bassols, M. Grazzi, C. Guillard, and M. Salazar, "Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe.," *Banco Interam. Desarro.*, p. 52, 2018

[7] C. Cárcamo and J. Mola, "Diferencias por sexo en el desempeño académico en Colombia: Un análisis regional," Econ. Región, vol. 6, no. 1, pp. 133–169, 2012

[8] C. Pérez López and D. Santín González, Minería de datos. [Madrid]: Paraninfo Cengage Learning, 2008.

[9] D. Abbott, Applied predictive analytics. 2014.

[10] C. Espino Timón, "Análisis predictivo técnicas y modelos utilizados y aplicaciones del mismo," p. 65, 2017.

[11] R. Timaran and J. Jiménez, "Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM," Form. Univ., pp. 1–19, 2014.

[12] L. Nel Quezada, *Estadística con SPSS 22*. Perú: Macro E.I.R.L, 2015.

[13] S. Singhal and M. Jena, "A Study on WEKA Tool for Data Preprocessing , Classification and Clustering," Int. J. Innov. Technol. Explor. Eng., vol. 2, no. 6, pp. 250–253, 2013.

[14] J. Palacio and C. Ruata, *Gestión de proyectos Scrum Manager*. 2011.

7.4. ARTÍCULO EAN

APROXIMACIÓN DE ANÁLISIS DESCRIPTIVO Y PREDICTIVO DE BRECHAS DE GÉNERO EN LAS PRUEBAS SABER

Angie Manrique, Alisson Celeita, Corín Viracacha, Cesar Barahona

Universidad de Cundinamarca, Facatativá, Cundinamarca

apaolamanrique@ucundinamarca.edu.co

acceleita@ucundinamarca.edu.co

cviracacha@ucundinamarca.edu.co

cbarahona@ucundinamarca.edu.co

Abstract **- Esta investigación busca analizar los resultados de las pruebas Saber, desde el año 2016, para encontrar indicadores de brechas de género y de esta manera hallar una relación entre los núcleos temáticos que tienen en común los exámenes presentados desde la educación básica hasta la educación superior.**

INTRODUCCIÓN

A través del tiempo, en Colombia, se han presentado brechas de género en gran parte de los campos económicos y socioculturales, tal como lo expresa el "Global Gender Gap Report" las principales brechas de género están en el liderazgo político y el empoderamiento económico. Si nos enfocamos en el ámbito laboral podemos ver que, por ejemplo, las mujeres ganan 20.2% menos que los hombres, aun cuando son las que más horas trabajan (Deutshe Welle, 2016). De igual manera, los cargos gerenciales son ocupados tan solo en un 16.2% por mujeres, según el estudio "Report on Women Board Directors of Latin America's 100 Largest Companies" del año 2016.

Estos indicadores tienen su origen en la educación, más específicamente en la primer infancia, puesto que, "Desde la escuela no se están propiciando los espacios suficientes para que las mujeres sean líderes y lleguen a donde tradicionalmente han estado subrepresentadas" (Junguito M., 2019). En su trabajo de investigación Abadía & Bernal (2016), concluyeron que "hay una brecha académica significativa entre niños y niñas que se amplía a favor de ellos a lo largo de la distribución de los puntajes en matemáticas"; adicionalmente, las investigadoras también encontraron que las características, familiares, escolares y de cada estudiante afectan en una proporción poco significativa con esta diferencia.

En otros países se han realizado estudios que demuestran que al tener un sistema educativo de calidad las diferencias son casi nulas entre hombres y mujeres, como es el caso de Estados Unidos que ha venido disminuyendo la brecha. Si en 1970 sólo una de cada 13 personas capaces de obtener 700 puntos en la prueba SAT (Scholarship Aptitude Test) antes de los 13 años era mujer, hoy la cifra es de 1 a 3 y se sigue cerrando camino a la equidad.

En el año 2012 en Colombia, fueron empleadas técnicas de contraste de medias (t Student) y regresiones por cuantiles en los resultados de las pruebas Saber 11 del año

2009 por parte de investigadores de la Universidad Tecnológica de Bolívar, haciendo énfasis en los núcleos de matemáticas y lenguaje revelando que en el componente de matemáticas los hombres tienen un mejor desempeño que las mujeres (Cárcamo C. & Mola J., 2012). Teniendo en cuenta lo anterior, buscamos hacer un análisis tanto descriptivo como predictivo en las pruebas estandarizadas de todos los niveles en el país, desde la educación básica hasta la educación técnica, tecnológica y superior; con el fin de determinar indicadores de brecha de género.

I. DESARROLLO

Para el análisis de los resultados de las pruebas Saber, en primer lugar, se debe hacer uso de los datos abiertos proporcionados por el ICFES, siendo esto un gran volumen de información que a la hora de ser examinada se vuelve una tarea muy compleja, por esta razón la minería de datos educativa será usada en esta labor puesto que facilita el procesamiento de la misma. Además, tiene como ventaja con respecto a los métodos estadísticos, el descubrimiento de información oculta y el manejo óptimo de los datos, convirtiéndose en una herramienta más eficaz al momento de la investigación.

Metodología

CRISS-DM considerada la "metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial" (Timaran & Jiménez, 2014), incorpora a su vez la gestión de proyectos y permite un movimiento bidireccional entre sus etapas.

Etapas:

6) Comprensión del negocio: Se pretende desarrollar un modelo descriptivo de los indicadores de brecha de género, el cual relacione aspectos económicos, socioculturales, familiares y personales. Condicionados por los datos disponibles de cada una de las pruebas se ha decidido empezar el análisis desde el periodo académico 2016 hasta su última actualización.
7) Comprensión de los datos: Recopilación de la información de los repositorios del ICFES. En la cual se obtiene los puntajes globales y específicos presentados por los estudiantes, factores socioeconómicos tanto familiares como personales, localización del estudiante y los datos relacionados con la institución (en unas pruebas sería el tipo, mientras que en otras el nombre de la misma).
8) Preparación de los datos: Se empezará en este punto de la investigación el tratamiento de los datos. Se deberá extraer, transformar y cargar la información de manera que quede seleccionada para su procesamiento.
9) Modelado: En esta fase se hará uso del algoritmo K-Medias como regla de agrupación para caracterizar a los estudiantes según su comportamiento y similitudes encontradas.
10) Evaluar: Se evaluará la calidad del modelo desarrollado, determinando su margen de error y exactitud.

Una vez culminado el proceso descriptivo se iniciará con el predictivo donde se optará por la aplicación de la técnica de Árboles de Decisión, haciendo uso de la herramienta WEKA puesto que, "proporciona muchos algoritmos diferentes para el análisis de datos y el aprendizaje automático. WEKA es de código abierto y está disponible gratuitamente. También es independiente de la plataforma" (Singhal & Jena, 2013).

Con la determinación de las predicciones se quiere mostrar la relación, si es que existe, entre los núcleos temáticos en común, visualizando su comportamiento a través de cada uno de los niveles de la educación haciendo énfasis en las áreas de las

matemáticas, ciencias, tecnológicas e ingenierías (STEM).

II. CONCLUSIONES

El uso de minería de datos para el proceso de análisis permite tener acceso a información oculta que con los métodos comúnmente usados de estadística no se puede obtener, además, facilita el procesamiento de grandes volúmenes de datos.

Gracias a las técnicas de minería de datos, en este caso, de agrupamiento y clasificación ad hoc se puede generar modelos descriptivos y predictivos consistentes de la realidad observada.

Con el desarrollo de esta investigación se contribuirá a futuros estudios sobre brecha de género, recomendando así hacer un comparativo de estas pruebas (Saber) con las realizadas a nivel internacional como las PISA.

WEKA facilita la aplicación de los algoritmos de minería de datos siendo una herramienta fiable y de fácil uso para el usuario.

III. REFERENCIAS

[1] Camejo, I. M., Medina, G. A., Beltrán, P. M., & Carrillo, K. G. (2015). Desempeño de algoritmos de minería en indicadores académicos: Árbol de Decisión y Regresión Logística Performance of data mining algorithms in academic indicators : Decision Tree and Logistic Regression. 9(4), 104–116

[2] Cárcamo C. & Mola J. (2012). Diferencias por sexo en el desempeño académico en Colombia: un análisis regional.

[3] Deutshe Welle (2016). ¿Por qué las mujeres siguen ganando menos? Retrieved from: https://www.semana.com/nacion/multimedia/brecha-salarial--por-que-las-mujeres-ganan-menos-que-los-hombres/565698

[4] Devitt T. (2009). Culture, not biology, underpins math gender gap. Retrieved from: https://news.wisc.edu/culture-not-biology-underpins-math-gender-gap/

[5] Junguito M. (2019). El empoderamiento de las mujeres tiene que empezar desde la infancia. Retrieved from: https://www.semana.com/Item/ArticleAsync/615282?nextId=617220

[6] Singhal, S., & Jena, M. (2013). A Study on WEKA Tool for Data Preprocessing , Classification and Clustering. International Journal of Innovative Technology and Exploring Engineering, 2(6), 250–253. Retrieved from

[7] Timaran, R., & Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. Formacion Universitaria, 1–19

7.5. ARTÍCULO CIETA

DESCRIPTIVE ANALYSIS OF THE GENDER GAPS IN THE SABER 11, SABER PRO AND SABER TYT TESTS

ÁNALISIS DESCRIPTIVO DE LAS BRECHAS DE GÉNERO EN LAS PRUEBAS SABER 11, SABER PRO Y SABER TYT

Angie Paola Manrique Ravelo, Alisson Catalina Celeita Gutierrez, Cesar Yesid Barahona Rodriguez

Universidad de Cundinamarca

Grupo de investigación GISTFA

Facatativá, Cundinamarca, Colombia

Calle 14 con Avenida 15 | (+57 1) 892 0706

E-mail: {apaolamanrique, acceleita, cbarahona}@ucundinamarca.edu.co

Abstract: For many years there have been gender gaps in different fields of society and nowadays there has been an increase in interest in learning about their behavior, especially in the educational field; For this reason, it has been decided to carry out an investigation that analyzes the data provided by the ICFES from the Saber tests, finding indicators of gender gaps and, in turn, being able to relate the common thematic nuclei, from the tests presented in secondary education to education. superior, in order to predict these behaviors in the future. Thanks to this, the user will be presented, by means of software, a detailed representation of the indicators found in order to be used in future investigations of entities that are interested in the study and reduction of this problem.

Keywords: Data mining, Gender gaps, Education, Descriptive analysis.

Resumen: Desde hace muchos años se han presentado brechas de género en diferentes campos de la sociedad y en la actualidad ha aumentado el interés de conocer el comportamiento de estas, en especial en el ámbito educativo; por esa razón se ha decidido realizar una investigación donde se analicen los datos brindados por el ICFES de las pruebas Saber, encontrando indicadores de brechas de género y a su vez poder relacionar los núcleos temáticos en común, desde las pruebas presentadas en la educación media hasta la educación superior, para lograr predecir estos comportamientos a futuro. Gracias a esto, se le presentará al usuario, por medio de un software, una representación detallada de los indicadores hallados con el fin de que sean usados en futuras investigaciones de entes que se interesen en el estudio y la reducción de esta problemática.

Palabras clave: Minería de datos, Brechas de género, Educación, Análisis descriptivo.

1. INTRODUCCIÓN

Las brechas de género son una problemática que en la actualidad se encuentra en auge. Cada vez son más los grupos y organizaciones que se unen a la causa de disminuir las diferencias existentes entre hombres y mujeres. A través del tiempo, se han presentado injusticias para el género femenino, tanto en los ámbitos educativos, laborales, familiares como en los socioculturales observando ventajas competitivas para los hombres, las cuales se convierten en múltiples indicadores que contribuyen directamente con esta disparidad.

A nivel mundial se presentan diversas diferencias, por ejemplo, en los países de América Latina y el Caribe (ALC) es preocupante ver como los hombres siguen ganando más que las mujeres, la violencia doméstica es elevada, las mujeres indígenas y rurales son quienes menos educación reciben (López-Bassols et al., 2018). Este proyecto centra su interés en la identificación de indicadores de brechas de género en Colombia, más específicamente en el ámbito estudiantil, para poder contrarrestar su crecimiento.

Para mirar el comportamiento de las mismas, se está realizando un análisis descriptivo por medio de técnicas de minería de datos aplicadas a las pruebas Saber realizadas en la educación media y superior, haciendo uso de los resultados entregados por el ICFES de cada una de las pruebas en sus respectivos periodos de presentación. Específicamente hablando, del año 2016 a la actualidad.

2. DESARROLLO

Antes de dar inicio con el análisis de los datos existe una fase de preprocesamiento que tiene como fin seleccionar aquellos datos relevantes para su estudio, más conocida como ETL. Vale la pena aclarar, que no hay un procedimiento determinado a seguir, sino que es realizado en función del objetivo que se va a llevar a cabo. Cuenta con 3 etapas de funcionamiento, mostradas a continuación:

- Etapa 1: Extracción

Consiste en la obtención de la información desde su origen, en este caso el servidor FTP proporcionado por el ICFES, hasta el repositorio local, donde se desea hacer uso de la misma. Para llevar a cabo esto, se realizó una conexión por medio del software libre WinSCP, de forma que

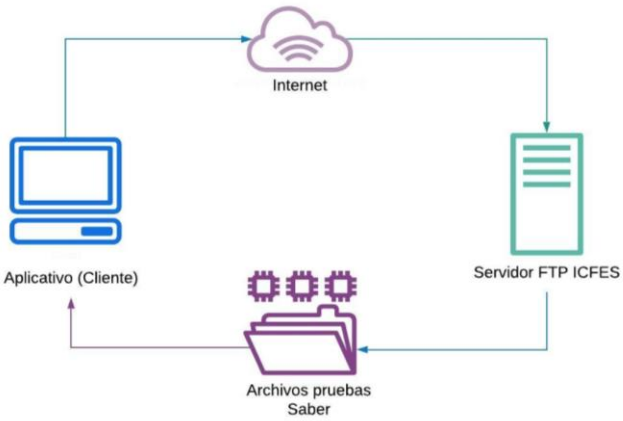
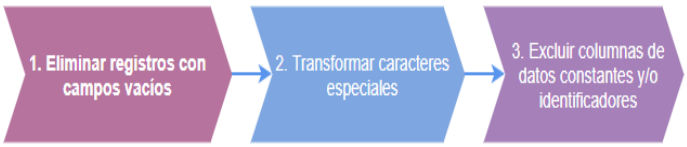


Fig. 1. Extracción de información

facilitara la tarea de descarga de los archivos por parte del administrador del sistema. Con este proceso de almacenamiento se evita que en caso de caída del servidor haya pérdida de los datos y se vea interrumpido el proceso de análisis. WinSCP fue escogido por la practicidad que tiene en la sincronización de elementos, puesto que se encarga de hacer una comparación entre los existentes y faltantes, permitiendo así que no se genere redundancia y/o duplicación en la información.

- Etapa 2: Transformación



Esta etapa es de vital importancia debido a que se

Fig. 2. Transformación de información

encarga de verificar la integridad y consistencia de los datos. Posteriormente, se realiza una selección de los atributos que cumplen con los estándares requeridos. Haciendo énfasis en nuestros requerimientos, en primer lugar, fue necesario realizar la eliminación de todos aquellos registros que no contenían los campos mínimos diligenciados para su evaluación. En segundo lugar, se hizo la transformación de todo aquel carácter que pudiese presentar conflicto en su lectura y su carga. Por último, para evitar una sobrecarga en la base de datos se excluyeron las columnas o atributos que presentaban datos constantes e identificadores que no suman relevancia al estudio.

- Etapa 3: Carga

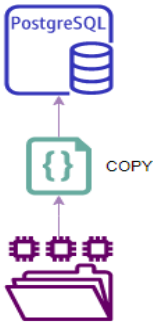


Fig. 3. Carga de archivos a base de datos

Para el almacenamiento se decidió hacer uso del gestor de base de datos relacional PostgreSQL (SGBD), por su flexibilidad en cuanto a lenguajes de programación, su robustez, eficiencia y estabilidad. Además, que es uno de los más recomendados para trabajar con grandes volúmenes de información y ampliamente usado a nivel empresarial. El proceso de carga se realiza por medio del comando "COPY" propio del SGBD, que permite la subida de la información directamente desde el archivo .txt, el cual proviene de descomprimir el archivo rar obtenido del repositorio FTP del ICFES, donde se encuentran alojados los documentos de cada una de las pruebas.

3. RESULTADOS

Ya con las pruebas cargadas a la base de datos, y su respectivo proceso de transformación se procede al análisis descriptivo de las mismas. En primer lugar, las pruebas Saber 11 respecto al género se encontraron diferencias en los puntajes de Lectura crítica, Matemáticas, Sociales y Ciencias Naturales; se puede evidenciar que los puntajes más altos están en mayor proporción para los hombres que para las mujeres. Mientras que, en inglés y Competencias ciudadanas los puntajes se muestran muy parejos para ambos géneros.

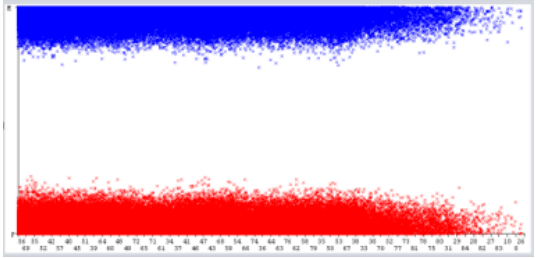


Fig. 4. Puntajes de lectura crítica según género

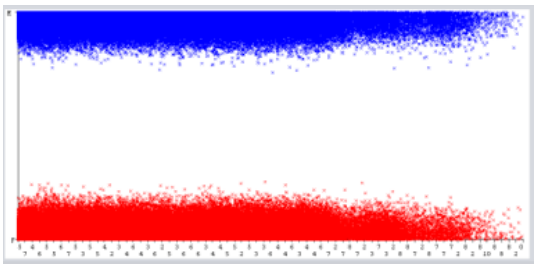


Fig. 5. Puntajes de matemáticas según género

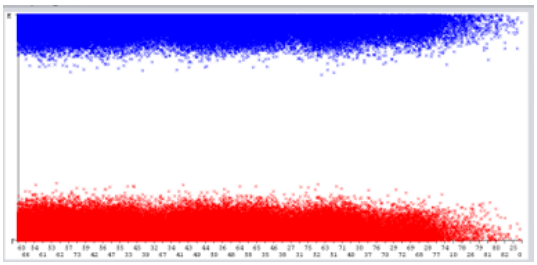


Fig. 6. Puntajes ciencias naturales según género

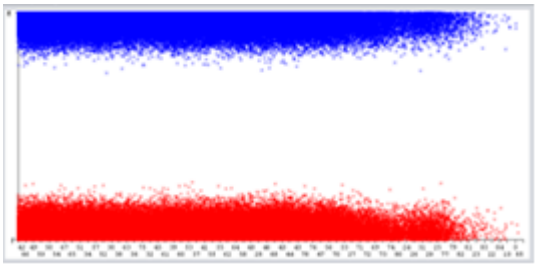


Fig. 7. Puntajes competencias ciudadanas según género

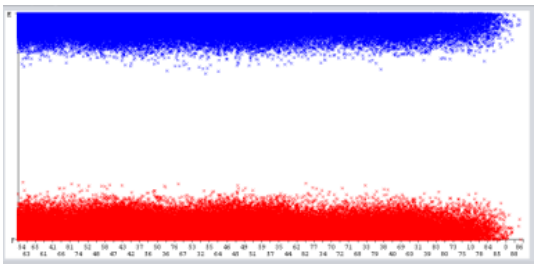


Fig. 8. Puntajes ingles según género

Aparte de los puntajes obtenidos en cada núcleo, se encontraron unos indicadores que afectan un poco el resultado final. Como lo son: el estrato, género del colegio, nivel de Sisbén, ingreso familiar y jornada.

Con **el estrato**, se nota diferencia en los 2 más altos niveles, es decir que en los estratos 5 y 6 se encuentra que hay más cantidad de mujeres que de hombres. Además, la mayoría de la población en ambos géneros está concentrada en el estrato 1 y 2,

lo que nos indica que sus ingresos económicos a nivel familiar son bajos o reducidos.

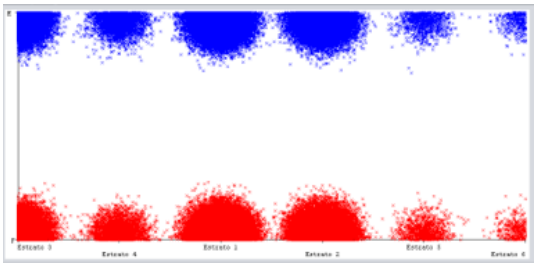


Fig. 9. Estrato según género

Al evaluar los resultados obtenidos en cada uno de los núcleos respecto a los estratos, se puede observar cómo los estratos más bajos logran obtener puntajes más sobresalientes en lectura crítica, matemáticas, ciencias naturales y competencias ciudadanas. Cosa contraria pasa con el núcleo de inglés, en el cual los puntajes más altos fueron obtenidos por los estratos más altos.

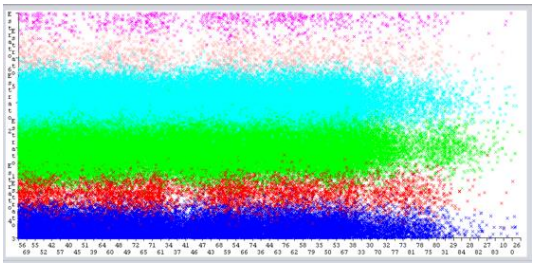


Fig. 10. Puntaje lectura crítica según estrato

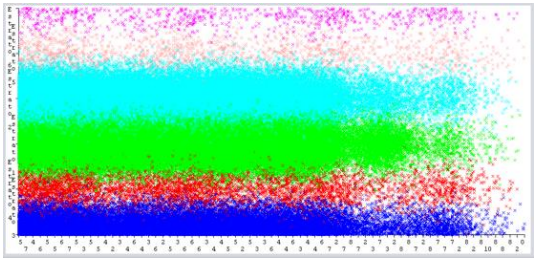


Fig. 11. Puntaje matemáticas según estrato

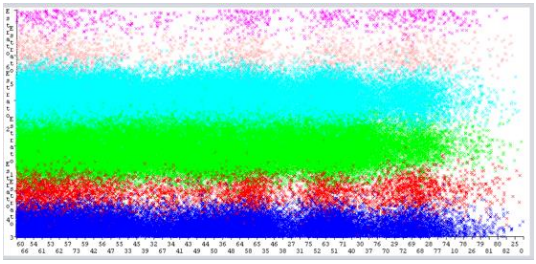


Fig. 12. Puntaje ciencias naturales según estrato

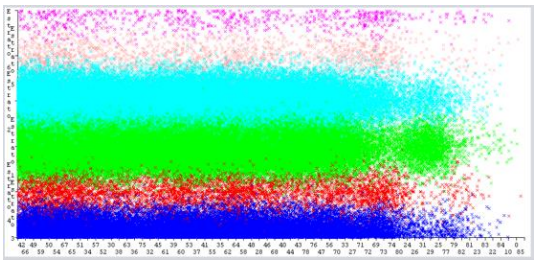


Fig. 13. Puntaje competencias ciudadanas según estrato

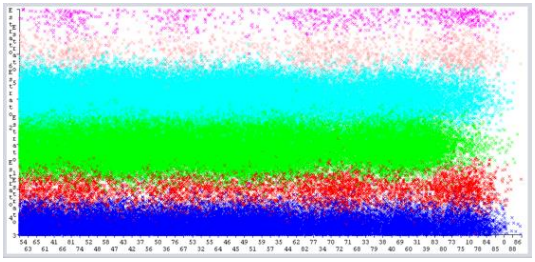


Fig. 14. Puntaje ingles según estrato

La situación que se ve reflejada con el **género del colegio** es la siguiente: Son pocos los colegios de carácter masculino, sin embargo, se destacan mujeres en esta clasificación, mientras que, en los colegios de tipo femenino no se ve una presencia significativa de hombres, y la gran mayoría son de tipo mixto. La cantidad de colegios femeninos a nivel nacional es 5 veces mayor con respecto a la de colegios masculinos.

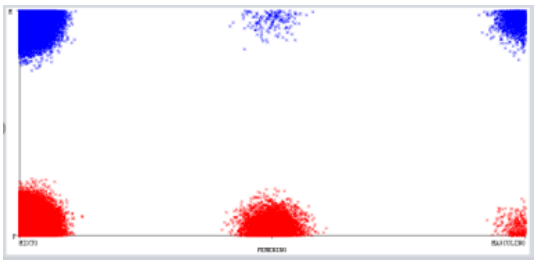


Fig. 15. Género del colegio

Además, dentro de este mismo indicador se encontraron departamentos donde no hay presencia de colegios ni femeninos ni masculinos, solo hay mixtos. Los departamentos en mención son: Guaviare, Vaupés, Vichada, Casanare, Arauca, San Andrés y Guainía. Y donde la presencia de colegios femeninos es mínima son Quindío, Nariño, La Guajira, Sucre, Cesar y Amazonas. Como es mostrado en la **Fig. 16,** a continuación.

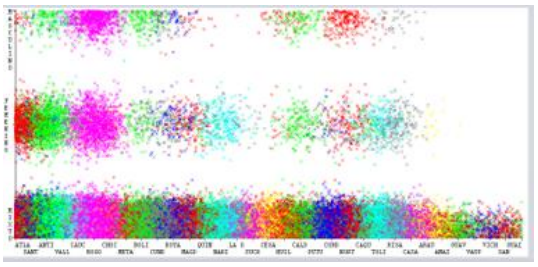


Fig. 16. Género del colegio según departamento

En el análisis de este indicador respecto a los puntajes podemos visualizar como los colegios clasificados por género tienen una gran desventaja en los puntajes obtenidos. En el caso puntual de inglés se ve un poco equilibrada la situación de las calificaciones, pero en los otros núcleos la desventaja aumenta, exceptuando matemáticas, en el cual los colegios masculinos logran emparejar a mixtos dejando atrás a las instituciones femeninas.

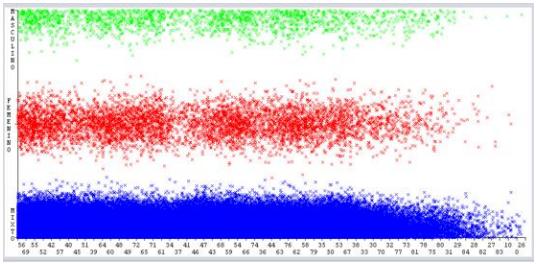


Fig. 17. Puntajes lectura crítica según genero del colegio

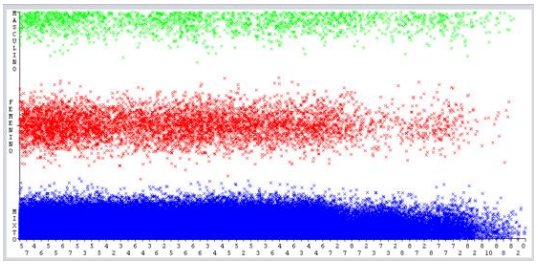


Fig. 18. Puntajes de matemáticas según género del colegio

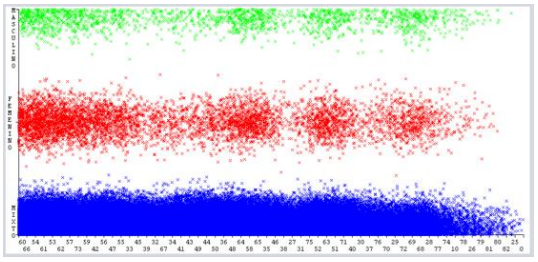


Fig. 19. Puntajes ciencias naturales según género del colegio

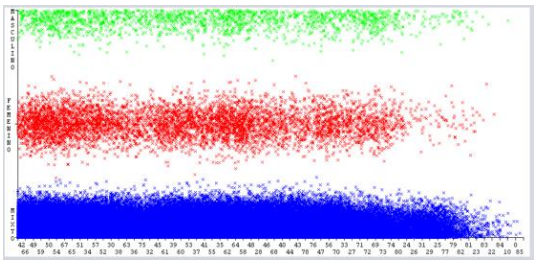


Fig. 20. Puntajes competencias ciudadanas según género del colegio

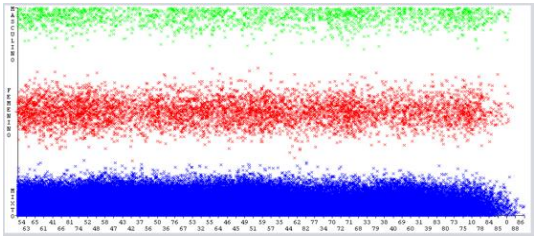


Fig. 21. Puntajes ingles según género del colegio

Otro indicador es **el nivel de Sisbén,** la cantidad de mujeres es mayor en los niveles superiores a 3, o como lo muestra la **Fig. 22** está clasificado en otro nivel, este comportamiento es bastante similar al presentado con el estrato. También en el nivel 2 se puede ver más presencia de mujeres que de hombres.

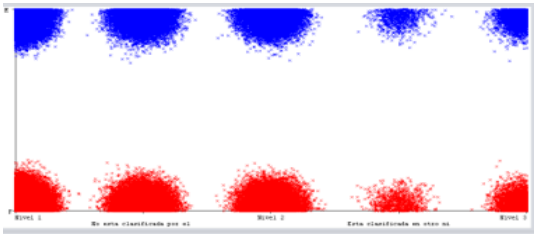


Fig. 22. Nivel de Sisbén según género

En el caso del **ingreso familiar,** quienes reciben menos de 1 SMMV son en mayor proporción los del género femenino, entre 2 y menos de 3 el comportamiento es parejo, entre 3 y menos de 5 hay más hombres que mujeres, entre 5 y menos de 7 también son menos las mujeres, entre 7 y menos de 10 pasa lo mismo y para 10 o más salarios el comportamiento es similar en ambos géneros.

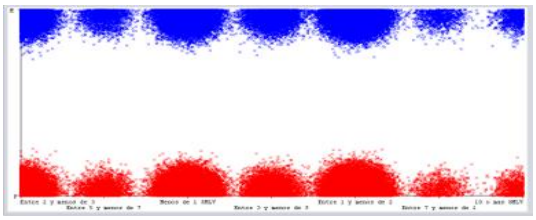


Fig. 23. Ingreso familiar según género

Su comportamiento frente a las valoraciones de las materias demuestra que este indicador solamente afecta los resultados en inglés, donde las personas con mayor ingreso familiar obtienen puntajes más altos.

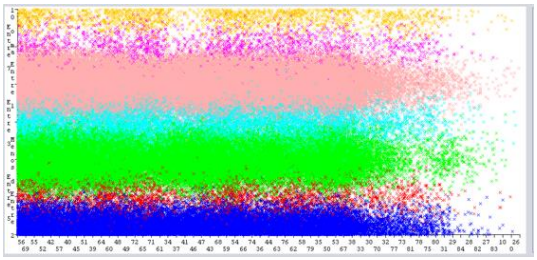


Fig. 24. Puntajes lectura crítica según nivel de Sisbén

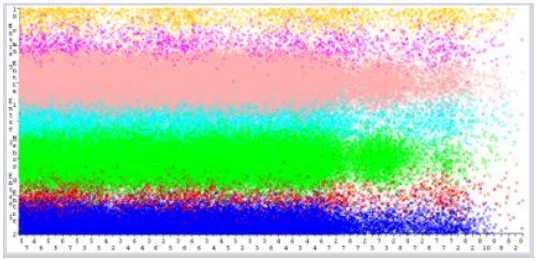


Fig. 25. Puntaje matemáticas según nivel de Sisbén

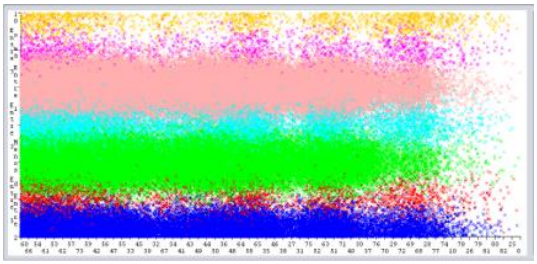


Fig. 26. Puntaje ciencias naturales según nivel de Sisbén

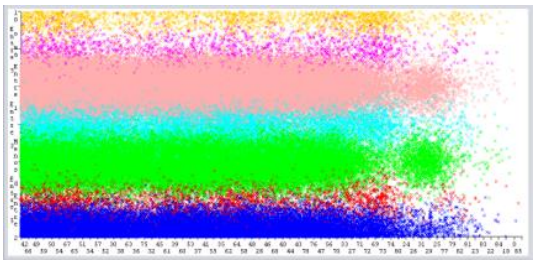


Fig. 27. Puntaje competencia ciudadana según nivel del Sisbén

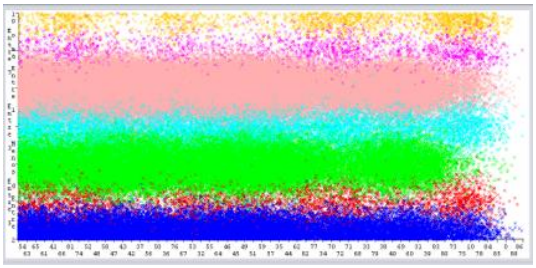


Fig. 28. Puntaje ingles según nivel del Sisbén

El último indicador encontrado en esta prueba es la **jornada,** para las 3 primeras categorías que son de tipo completa, tarde y mañana el comportamiento es parejo en ambos géneros. Las diferencias se encuentran en las jornadas sabatina, noche y única donde hay mayor presencia femenina que masculina.

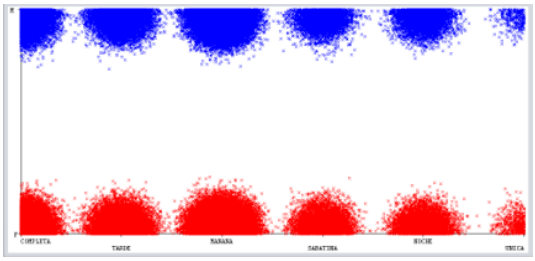


Fig. 29. Jornada académica según género

Se demuestra que este indicador afecta en gran proporción los resultados en cada núcleo. En los casos de matemáticas, lectura crítica, ciencias naturales e inglés las jornadas mañana y completa presentan mejores resultados, dejando en últimos lugares a jornada sabatina y tarde. Competencia ciudadana por su parte tiene puntajes más equilibrados frente a este indicador.

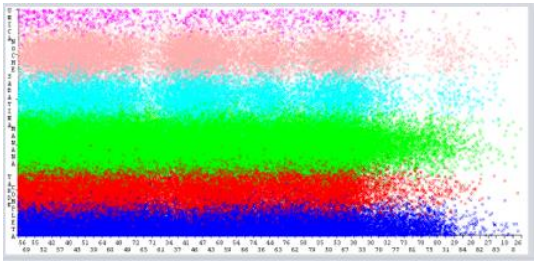


Fig. 30. Puntaje lectura crítica según jornada

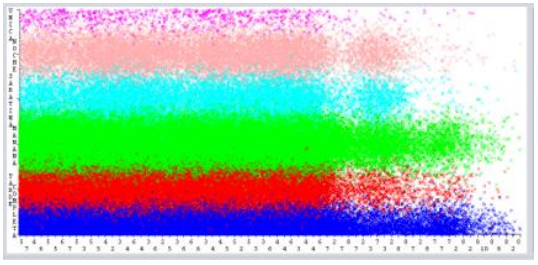


Fig. 31. Puntaje matemáticas según jornada

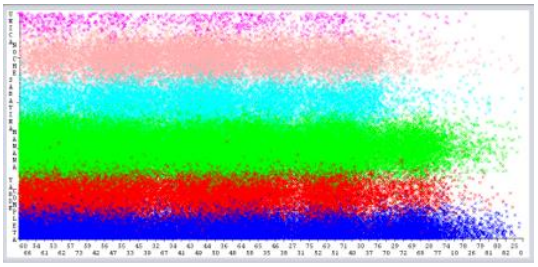


Fig. 32. Puntaje ciencias naturales según jornada

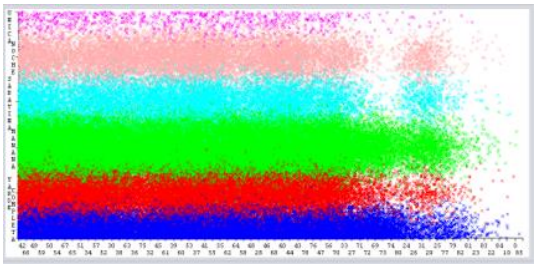


Fig. 33. Puntaje competencia ciudadana según jornada

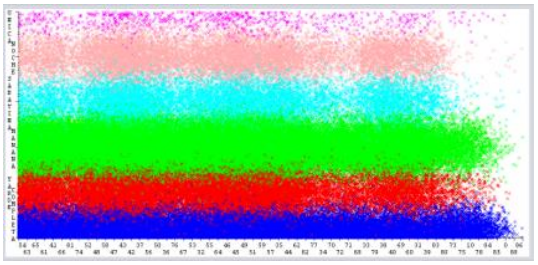


Fig. 34. Puntaje ingles según jornada

Ahora, al analizar la prueba Saber TyT, se encontraron como nuevos indicadores: el estado civil, el valor de matrícula, el tipo de pago matrícula y grupo de referencia de la carrera. Para el **estado civil,** llama la atención que las mujeres con respecto a los hombres son en mayor cantidad casadas, separadas y/o viudas, esto nos indica que ya con un hogar formado bastantes mujeres deciden estudiar carreras técnicas y tecnológicas.

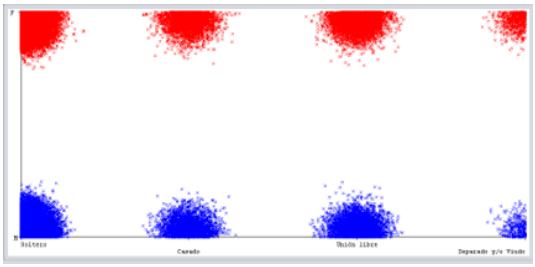


Fig. 35. Estado civil según género

Con el **valor de la matrícula**, podemos observar que el 58% del total de personas que presentaron la prueba no pagaron nada por la matricula. De los que, si pagaron, más hombres que mujeres pagaron un valor menor a 500mil, entre 1 y 4 millones el balance está equilibrado para ambos géneros, entre 4 millones y 5.5 pagaron más hombres que mujeres, entre 5.5 y 7 millones es parejo y más de 7 millones el género femenino tuvo más pagos.

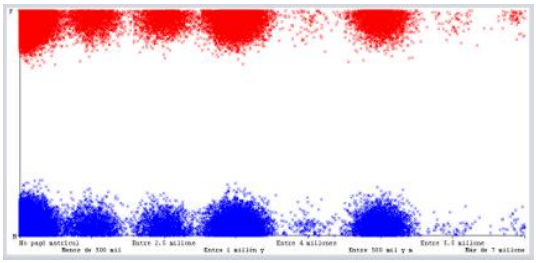


Fig. 36. Valor de la matricula según género

Para el **tipo de pago de matrícula**, en el caso específico de crédito son más las mujeres que afirman pagar de este modo que los hombres que lo hacen.

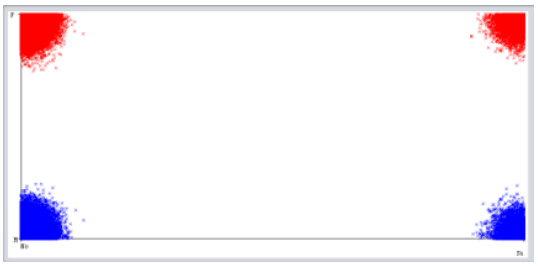


Fig. 37. Pago por crédito según género

Y como último indicador para esta prueba se tiene el **grupo de referencia**, que es el atributo que menciona las carreras de los estudiantes siendo agrupadas entre sus semejantes. Aquí se encontró que las carreras afines a la agronomía, electricidad, mecánica y minas son estudiadas en mayor cantidad por los hombres. Mientras que las mujeres prefieren estudiar en áreas afines a la salud y la administración.

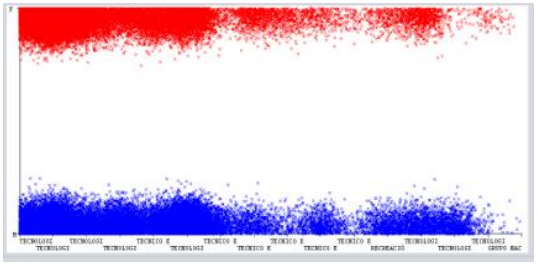


Fig. 38. Grupo de referencia según género

Para finalizar el análisis, fue aplicado el mismo algoritmo para las pruebas Saber Pro, que se presentan en las instituciones de educación superior. Allí los nuevos indicadores hallados son: Cabeza de familia, personas a cargo y lectura diaria. En el primero de ellos, **cabeza de familia**, se puede ver que hay mayor cantidad de mujeres que lo son con respecto a los hombres.

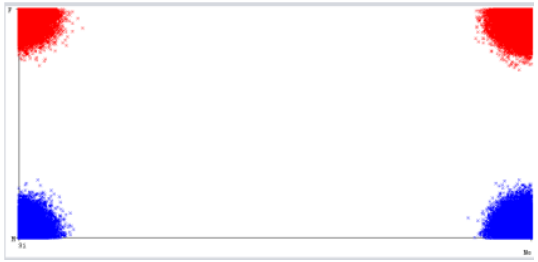


Fig. 39. Cabeza de familia según género

Por el **número de personas a cargo,** se puede apreciar que gran cantidad de los estudiantes no tienen personas a cargo, en especial las mujeres. Entre los que si tienen encontramos que entre 1 y 8 personas el comportamiento entre ambos géneros no presenta diferencias significativas, para 9 personas en los hombres casi no hay registros mientras que en las mujeres sí, para 11 hay mayor cantidad en los hombres que en las mujeres y doce o más el género femenino vuelve a tener mayoría.

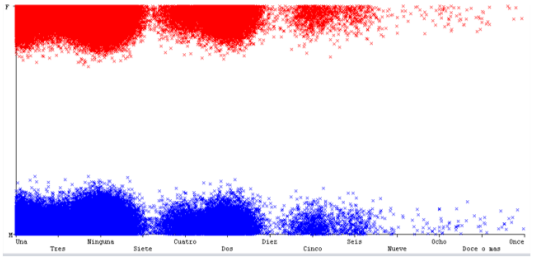


Fig. 40. Número de personas a cargo según género

Por último, con el indicador de **lectura diaria** se logró observar como las mujeres tienen mejores hábitos de lectura diaria, ya que en la mayoría de las opciones se ve un porcentaje significativo del género femenino.

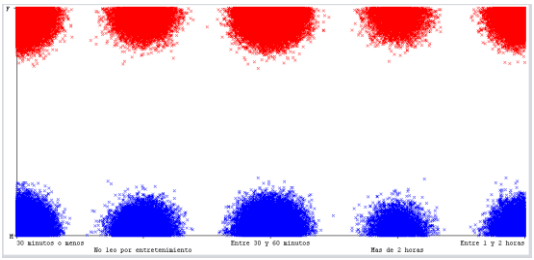


Fig. 41. Lectura diaria según género

4. CONCLUSIONES

Del análisis realizado con las 3 pruebas se puede concluir que los factores que influyen en las brechas de género están relacionados con los campos socioeconómicos y familiares.

Hay áreas donde se destacan las mujeres, hablando desde el punto de vista de participación en carreras afines a la administración, ciencias de la salud, entre otras; mientras que hay grupos como las ingenierías y mecánica o afines donde el hombre se destaca. Esto está directamente relacionado con los

puntajes obtenidos en estos mismos núcleos temáticos

El estrato es un indicador común para las 3 pruebas, es el principal factor de medición económica para los estudiantes, y está bastante relacionado con otros como el nivel de Sisbén, valor de matrícula, ingreso familiar, entre otros.

Los colegios y su departamento de ubicación son 2 de los indicadores que más movimiento presentaron, el caso de los colegios masculinos llama la atención de manera significativa para el estudio, puesto que presenta un comportamiento mucho más sesgado que el de los colegios femeninos. Además, que no tiende a ser muy común hallar este tipo de instituciones en varias partes del país.

Con la aplicación del algoritmo K-means se descartaron atributos que no presentaran cambios o diferencias siendo evaluados con respecto al género. Solo quedaron los que de alguna manera influyeron en las brechas.

En la actualidad, los jóvenes universitarios dependen económicamente de sus familias (Rojas, 2011) por lo cual se da a entender que las mujeres, al pedir mayores créditos es porque tienen más apoyo familiar.

REFERENCIAS

López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. https://doi.org/10.18235/0001082

Rojas, Mauricio (2011) "Autonomía postergada: jóvenes, familia y educación superior". Recuperado el 20 de Junio de 2016. http://www.redalyc.org/pdf/1942/194218961010.pdf.

7.6. ARTÍCULO REVISTA PENSAMIENTO UDECINO

BRECHAS DE GÉNERO EN ADMINISTRACIÓN, CONTADURÍA, INGENIERÍA Y PSICOLOGÍA EN LAS PRUEBAS SABER 2016 EN COLOMBIA

GENDER GAPS IN ADMINISTRATION, ACCOUNTING, ENGINEERING AND PSYCHOLOGY IN THE TESTS SABER 2016 IN COLOMBIA

Autores: Angie Paola Manrique Ravelo, colombiana.[1]

Alisson Catalina Celeita Gutiérrez, colombiana.[2]

Cesar Yesid Barahona Rodríguez, colombiano.[3]

Tipo de artículo: Artículo de investigación científica y tecnológica.

RESUMEN

A través del tiempo, alrededor del mundo se han venido presentando múltiples brechas de género en diferentes campos de la sociedad como los socioeconómicos, familiares, académicos, culturales, entre otros. Así mismo, en los últimos años, ha crecido el interés por conocer cuál es el comportamiento de las mismas, haciendo que sea de gran importancia realizar estudios acerca de la información relacionada con el tema. En este proyecto, se tiene como enfoque el ámbito educativo en Colombia, más específicamente para áreas como la administración, la contaduría, la psicología y la ingeniería en las pruebas Saber realizadas por el ICFES para los futuros técnicos, tecnólogos y profesionales a nivel nacional, puesto que son una herramienta de medición excepcional de la situación del país. Para llevar a cabo este objetivo, se realizó un análisis descriptivo de las pruebas Saber TyT y Saber Pro, por medio de la aplicación de técnicas y algoritmos de minería de datos, obteniendo como resultado final indicadores que inciden en esta problemática y un aplicativo web que permite visualizar de forma gráfica la información encontrada, con el fin de que pueda ser usada en futuras investigaciones.

Palabras claves: Acceso a la educación, Análisis de datos, Educación, Mujer y desarrollo.

ABSTRACT

Throughout time, multiple gender gaps have been presented around the world in different fields of society such as socioeconomic, family, academic, cultural, among others. Likewise, in recent years, interest in knowing their behavior has grown, making it of great importance to carry out studies on information related to the topic. In this project, the focus is on the educational field in Colombia, more specifically for areas such as administration, accounting, psychology and engineering in the Saber tests carried out by ICFES for future technicians, technologists and professionals at the national level, since they are an exceptional measurement tool for the country's situation. To carry out this objective, a descriptive analysis of the Saber TyT and Saber Pro tests was carried out, through the application of data mining techniques and algorithms, obtaining as a final result indicator that affect this problem and a web application that It allows the information found to be viewed graphically, so that it can be used in future research.

Keywords: Access to education, Data analysis, Education, Women and development.

[1] **Estudiante de noveno semestre de Ingeniería de sistemas, Auxiliar de investigación semillero Aplicaciones Móviles, Universidad de Cundinamarca, apaolamanrique@ucundinamarca.edu.co, https://orcid.org/0000-0003-0302-406X, Celular: 3107669063**

[2] **Estudiante de noveno semestre de Ingeniería de sistemas, Auxiliar de investigación semillero Aplicaciones Móviles, Universidad de Cundinamarca, acceleita@ucundinamarca.edu.co, https://orcid.org/0000-0001-9750-2848, Celular: 3115109407**

[3] **Magister en Sistemas Computacionales, Especialista en Gestión de Proyectos, Ingeniero en Telecomunicaciones. Líder semillero Aplicaciones Móviles, Universidad de Cundinamarca, cbarahona@ucundinamarca.edu.co, https://orcid.org/0000-0001-7673-7381, Celular: 3176991109**

1. INTRODUCCIÓN

Las brechas de género, son una problemática y realidad vivida por largo tiempo, las cuales constituyen un caso de discriminación que no es exclusiva de clases sociales, comunidades o etnias, minorías raciales, entre otros, sino que afecta a todo tipo de grupos y poblaciones por una inseparable mitad, que hoy en día ha llegado a convertirse en un campo de estudio que se encuentra en auge. Cada vez son más los grupos y organizaciones que se unen a la causa de disminuir la desigualdad existente entre hombres y mujeres. Sin embargo, a nivel mundial aún se presentan múltiples diferencias, por ejemplo, en los países de América Latina y el Caribe (ALC), según un estudio realizado por el Banco Interamericano de Desarrollo (BID), entre los años 2006 y 2015 las mujeres han venido presentando una mayor asistencia en educación superior, pero a pesar de esto es preocupante observar cómo los hombres son quienes se desempeñan mejor en las pruebas estandarizadas; además siguen ganando más que las mujeres y teniendo mejores ventajas competitivas en el mercado laboral (Arias Ortiz & Bornacelly, 2017). Por otra parte, la violencia doméstica es elevada y desafortunadamente las mujeres indígenas y de áreas rurales son quienes menos educación reciben (López-Bassols et al., 2018).

Continuando con la disparidad presentada a nivel educativo, desde hace años se han reportado brechas en la educación superior siendo evidenciadas en diversos enfoques, la gran mayoría de éstas son indicadas en las tasas de acceso o graduación (Becerra-Gualdrón et al., 2018). Pero no son los únicos aspectos perjudicados, algunos estudios demuestran como el aprendizaje alcanzado por los estudiantes se puede ver realmente afectado por esta problemática (Abadía & Bernal, 2016). Así mismo, "Las desigualdades de género en cuanto a logro académico marcan diferencias en las rutas educativas y los desempeños laborales posteriores" (Caro Acero & Casas M., 2013).

Debido a estos hallazgos, la ONU en un estudio sobre brechas de género y desigualdad, recomienda encontrar indicadores de brecha de género, todo esto con el fin de lograr una medición de forma cuantitativa de las barreras o afectaciones que contribuyen con la permanencia de dichas desigualdades (ONU Mujeres, 2018). Debido a esto se determinó hacer un análisis descriptivo y cuantitativo (Hernández Sampieri et al., 2006) de los resultados en las pruebas Saber Pro, siendo este examen "un instrumento estandarizado para la evaluación externa de la calidad de la Educación Superior" (Decreto 3963 de 2009, 2009) y Saber TyT que evalúa las competencias específicas de acuerdo con el grupo de referencia (Universidad de los Llanos, 2018); brindados por el ICFES como datos de acceso abierto.

Todo esto haciendo uso de la minería de datos educativa, puesto que es una disciplina emergente que está interesada en el desarrollo y aplicación de métodos para entender mejor a los estudiantes y su entorno educativo (Villafañe Blanco, 2015).

El siguiente documento se enfoca en el estudio de los puntajes entre mujeres y hombres con respecto a indicadores hallados con la aplicación del algoritmo k-means (Espino Timón, 2017), específicamente en las puntuaciones obtenidas en cada uno de los núcleos temáticos según los programas de pregrado de los grupos de referencia: Administración y afines, Contaduría y afines, Psicología e Ingeniería en el examen Saber Pro 2016; y estudiantes de Tecnólogo en Administración y Turismo, Tecnólogo en Ingeniería, Industria y Minas, Técnico en Administración y Turismo y Técnico en Ingeniería, Industria y Minas que presentaron el examen Saber TyT en el mismo año. Son precisamente estos conceptos: *puntajes por núcleo temático* e *indicadores de brechas de género*, las variables graficadas según la relación que están presentando en conjunto con el número de estudiantes clasificados dentro de las categorías resultantes (Orjuela, 2014).

Con el fin de contar con una correcta visualización y de lograr una mayor comprensión de los resultados obtenidos en el análisis descriptivo, se desarrolló también un aplicativo web como herramienta práctica para el usuario donde se expone la brecha de género presente en los cinco módulos temáticos de las pruebas y su relación con cada indicador.

2. METODOLOGÍA

Antes de empezar con el cálculo de los indicadores de brecha de género, hay una fase de extracción y posterior preprocesamiento de la información para que pueda ser correctamente analizada, y haya un margen de error mucho menor en los resultados (García et al., 2016).

- Fase 1: Extracción

Esta fase se encarga de obtener los resultados de las pruebas suministrados por el ICFES, los cuales están alojados en un servidor FTP; estos servidores en términos generales, se dedican al envío y descarga de información de acuerdo con los protocolos existentes para la transferencia de datos por internet (González Castañeda, 2012). Lo anterior permite que se tenga una copia local de

los archivos y puedan ser usados para la investigación, así como lo muestra la figura 1.

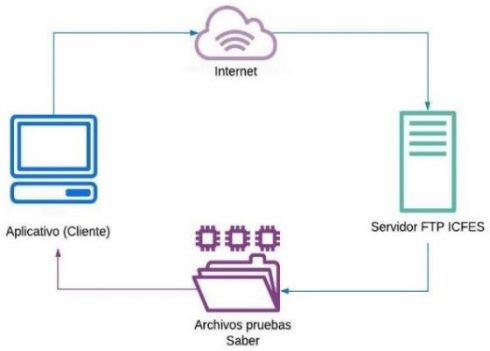


Figura 1. Extracción de los resultados de las pruebas
Fuente: Autores

- Fase 2: Preprocesamiento

En esta etapa el aplicativo tiene como objetivo hacer un filtrado de la información dejando solo la que sea relevante y depurando toda la que no sea de utilidad, por ejemplo, aquella que presente campos no diligenciados o que sean campos que no representan información de interés de las pruebas deberán ser eliminados, como lo son los números de identificación de las personas o departamentos. Así mismo, también se realiza una transformación en donde se ubican caracteres especiales que no puedan ser leídos y deban ser cambiados. Al finalizar este proceso, se procede a subir la información ya tratada en una base de datos. La importancia de esta fase radica en que permite realizar procesos de transformación y análisis de datos (Sierra et al., 2015).

- Fase 3: Cálculo de los indicadores de brecha de género

También conocido como análisis descriptivo, esta es la fase más extensa del proyecto donde a partir de la información, se hace un cálculo de cuáles son los factores que influyen en las brechas de género reflejadas en los resultados de las pruebas, en términos generales tiene como función principal extraer conocimiento interesante de los datos sin la necesidad de trabajar con estructuras (Carmona et al., 2015). De acuerdo con lo anterior, en primer lugar, lo que se hace es aplicar un algoritmo de minería de datos que permita mirar la incidencia de cada campo obtenido en las pruebas. ¿Cómo se hace esto? K-means es un algoritmo que realiza agrupaciones, conocidas en términos técnicos como “clústeres” (Romero & Ventura, 2016), que en forma práctica son asociaciones según ciertas características en común de los datos, que entrega la cantidad de registros, en este caso, estudiantes tanto hombres como mujeres, que se encuentran en cada una de estas asociaciones.

Los datos estadísticos que arroja este proceso y permiten realizar con el software gráficas para la comprensión de los usuarios son: *Centroide:* Es un registro que contiene la respuesta promedio de los estudiantes en cada una de las asociaciones generadas, es decir, por cada agrupación hay un centroide (Palacio et al., 2015). *Atributos:* Son aquellas opciones de respuesta o de clasificación de cada indicador (Rodríguez, 2012). *Tamaño del clúster:* Es el número correspondiente a la cantidad de estudiantes de cada asociación encontrada (Rodrigo, 2017).

Como ya se ha mencionado, a partir del análisis se generan unas gráficas que expresan de forma clara estos datos, y el aplicativo web le permite al usuario elegir qué año y qué indicador quiere ver. Con esta selección, son generadas 6 gráficas de barras. La primera de ellas relaciona el género con los atributos presentados en ese indicador. Las siguientes son 5 gráficas de puntajes con respecto de nuevo al indicador; cada una de éstas lo que expresa es la afectación del indicador en los diferentes atributos dentro los rangos de puntajes obtenidos en los exámenes, repitiendo este proceso para los 5 distintos núcleos temáticos.

3. RESULTADOS

La educación cumple un papel muy importante en el desarrollo del país; en aspectos como el social, económico y político; independiente de las diferentes etapas de la educación (Peréz et al., 2018). Los indicadores de brecha de género que se establecieron en las pruebas Saber Pro y Saber TyT gracias a los datos suministrados por el ICFES en el año 2016, fueron clasificados en 3 dimensiones que son las siguientes: económicos, sociales y académicos (Orozco, 2010).

Tabla *33.*
Indicadores de género Saber Pro

INDICADORES SABER PRO		
ECONÓMICOS	**SOCIALES**	**ACADÉMICOS**
• Valor matrícula universidad. • Tipo de pago matrícula. • Estrato de vivienda. • Tiene internet.	• Estado civil. • Cabeza de familia. • Número de personas a cargo. • Área de residencia.	• Educación madre. • Educación padre. • Dedicación lectura diaria. • Área de la carrera. • Método de programa.

Tabla 34.
Indicadores de género Saber TyT

INDICADORES SABER TyT		
ECONÓMICOS	**SOCIALES**	**ACADÉMICOS**
• Tipo de pago matrícula. • Estrato de vivienda. • Tiene internet.	• Estado civil. • Cabeza de familia. • Número de personas a cargo. • Área de residencia.	• Educación madre. • Educación padre. • Dedicación lectura diaria. • Área de la carrera.

En las tablas 1 y 2 se mencionan cuáles son los indicadores finales de cada prueba Saber. A continuación, se explica en qué consiste cada uno de ellos.

- Valor matrícula universidad: Este indicador representa el rango de valores pagados por los estudiantes en la inscripción del programa que están cursando en la universidad.
- Tipo de pago matrícula: En esta prueba se tienen varias opciones de respuesta para los tipos de pago entre los cuales se encuentra si es pagado con ayuda de padres, si es por medio de crédito, si fue pago propio o si el estudiante se encuentra becado. Entre estos métodos, el pago por crédito fue aquel indicador que arrojó una afectación significativa. Los atributos con los que cuenta el indicador, son sí y no.
- Estrato de vivienda: Indica la clasificación socioeconómica jerarquizada del país, sus categorías comprenden desde el estrato 0 hasta el estrato 6 (DANE, 2016).
- Tiene internet: Este indicador señala qué estudiantes cuentan con el servicio de internet en sus hogares y quiénes no.
- Estado civil: Representa la condición al momento de la realización de la prueba de los estudiantes determinada por sus relaciones de pareja, se dividen entre 4 opciones: si está casado, soltero, en unión libre y separado y/o viudo.
- Cabeza de familia: Determina si la persona es quien tiene a cargo, o no, toda la responsabilidad económica del hogar.
- Número de personas a cargo: Este indicador está directamente relacionado con el de cabeza de familia, y pretende hacer la misma medición, pero desde un punto de vista cuantitativo estando en diferentes rangos comprendidos entre ninguna y11 personas.
- Área de residencia: Se refiere a la zona donde el estudiante vivía al momento de presentación de la prueba. Sus atributos se dividen en Rural y Urbano.
- Educación madre y educación padre: Estos dos indicadores muestran qué niveles educativos cuentan los padres de los estudiantes, estando clasificados en los siguientes atributos: ninguno, no sabe, no aplica, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica y tecnológica incompleta, técnica y tecnológica completa, educación profesional incompleta, educación profesional completa y postgrado.
- Dedicación lectura diaria: Muestra el tiempo diario que le dedican los estudiantes al ejercicio de la lectura, siendo clasificada por rangos de tiempo de la siguiente manera: 30 minutos o menos, no leo por entretenimiento, entre 30 y 60 minutos, más de 2 horas, y entre 1 y 2 horas.
- Área de la carrera: Este indicador en las pruebas se conoce como grupo de referencia, y especifica en qué área se encuentra clasificada la carrera que se está cursando. Existen múltiples categorías, pero para este estudio se escogieron 4, las cuales son: Administración y afines, Contaduría y afines, Psicología e Ingeniería, en el caso de Saber Pro. Y para Saber TyT, son Tecnólogo en Administración y Turismo, Tecnólogo en Ingeniería, Industria y Minas, Técnico en Administración y Turismo y Técnico en Ingeniería, Industria y Minas.
- Método del programa: Hace referencia al modo de acceso al programa educativo. Los 3 tipos de clasificación son presenciales, a distancia y virtual.

Con el contexto claro de los indicadores, se procede a mostrar las gráficas que determinan de forma visual la afectación de cada uno de ellos. Puesto que, por cada indicador se generan 6 gráficas, en este texto se hará enfoque solo en uno; en este caso, fue escogida el Área de residencia. El primer gráfico de barras, muestra en la Figura 2 la cantidad de estudiantes que están ubicados en la cabecera municipal y en el área rural del país. La población en su mayoría vive en cabecera municipal, y en ambas áreas, son más las mujeres que los hombres (Ruiz, 2003).

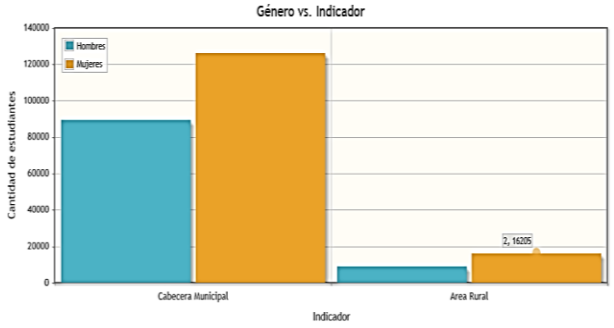


Figura 2. Género vs. Área de residencia
Fuente: Autores

Posteriormente, se presentan las gráficas de puntaje por núcleo temático contra el indicador escogido. En el caso de lectura crítica, representada en la gráfica 3, se observa que las mujeres de la cabecera municipal obtuvieron en su mayoría puntajes que están entre los rangos de 100 a 150 y de 150 a 200. De igual manera, para el área rural solo se generaron clústeres en estos 2 rangos, pero los otros 4 clústeres no presentaron resultados significativos, evidenciando puntajes más bajos en las mujeres del campo. (Vergara & Ávila, 2012).

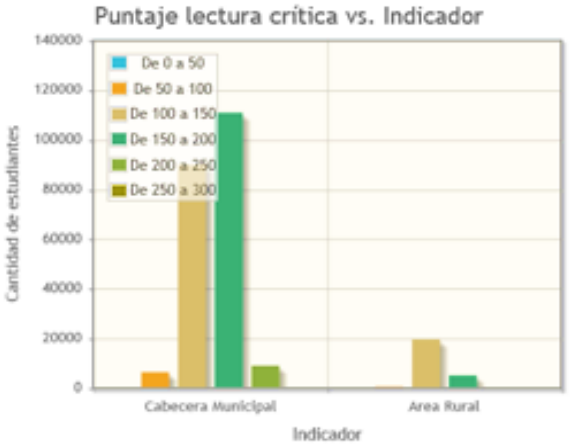


Figura 3. Puntaje lectura crítica vs. Área de residencia
Fuente: Autores

El núcleo de razonamiento cuantitativo mostrado en la Figura 3 presenta un comportamiento bastante similar al de lectura crítica. De hecho, si se compara con comunicación escrita (Figura 4) y competencias ciudadanas (Figura 5) sucede lo mismo, de lo cual se puede decir que las mujeres tienen un puntaje promedio para estos núcleos temáticos entre 100 y 200. Y teniendo en cuenta que el máximo es 300, la población obtiene resultados aceptables pero que pueden ser mejorados; además, se demuestra que las mujeres tienen mayor acceso a la educación en las ciudades que en el campo (Universia, 2013).

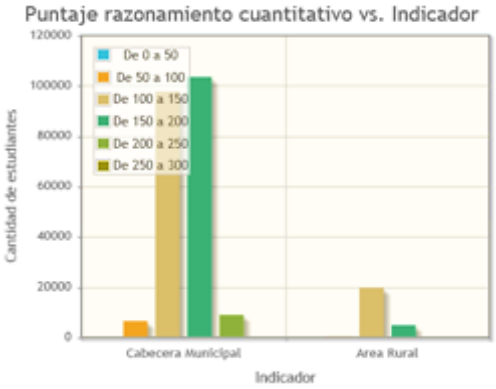


Figura 4. Puntaje razonamiento cuantitativo vs. Área de residencia Fuente: Autores

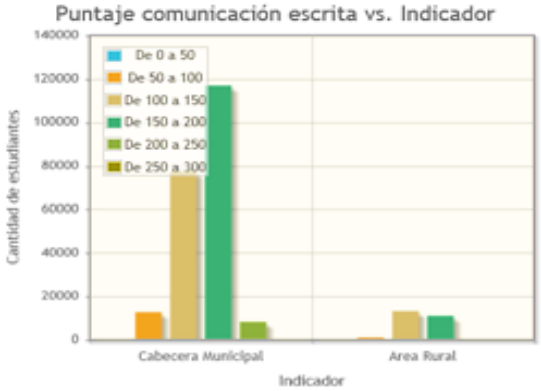


Figura 5. Puntaje comunicación escrita vs. Área de residencia Fuente: Autores



Figura 6. Puntaje competencias ciudadanas vs. Área de residencia Fuente: Autores

En inglés, el indicador solo presenta resultados significativos en tres rangos para la cabecera municipal, que están comprendidos entre 100 y 250. Desafortunadamente, este es el núcleo temático donde los estudiantes muestran los puntajes más bajos a nivel nacional, como lo señala gráficamente la Figura 7. En el área rural, es incluso más preocupante la situación, puesto que solo se encuentran clústeres con registros importantes entre los rangos de 100 y 200, bajando el promedio con respecto a la cabecera municipal. Toda esta situación

también indica que los hombres de la mayoría de ciudades del país presentan mejor nivel de inglés que las mujeres.

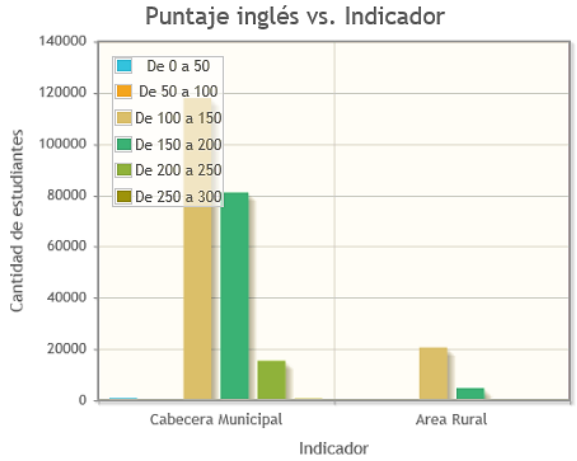
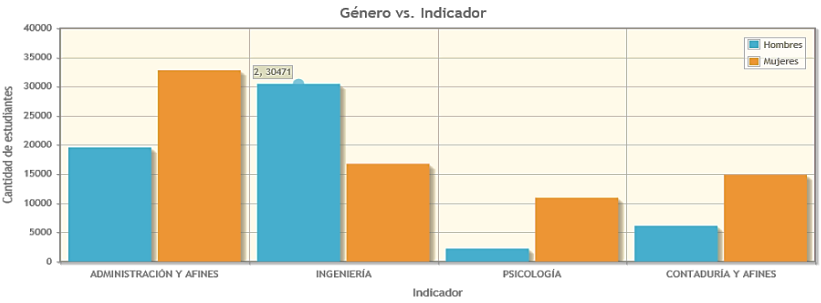


Figura 7. Puntaje inglés vs. Área de residencia
Fuente: Autores

Analizando los grupos de referencia, ya específicamente de Saber Pro, se puede observar que las mujeres tienden a estudiar carreras relacionadas con administración. Aunque hay una cantidad importante en las otras áreas representadas que son en orden descendente Ingeniería, Contaduría y Psicología; mientras que los hombres prefieren estudiar Ingeniería, luego Administración, Contaduría y Psicología. Vale la pena aclarar que, por ejemplo, para el área de Ingeniería esta gran cantidad de mujeres estudiando tiene que ver con que hay en el país una gran variedad de ingenierías, totalmente opuesto a psicología que solo es una carrera y en comparación a estas

Figura 8. Género vs. Área de la carrera
Fuente: Autores



4 áreas tiene pocos estudiantes como es mostrado en la Figura 8 (López, 2015).

Con respecto a los puntajes de los núcleos temáticos, en lectura crítica se puede ver en la Figura 9 que en promedio los estudiantes registran puntajes entre los rangos de 90 a 210. Ocurre un comportamiento para destacar negativamente en ingeniería, los puntajes son más bajos que para las otras 3 áreas que tienen como clúster más significativo el comprendido entre 150 y 180 puntos. En cambio, los futuros ingenieros del país tienen una mayor concentración en el rango de 120 a 150, encontrándose una brecha de desempeño académico en este núcleo.

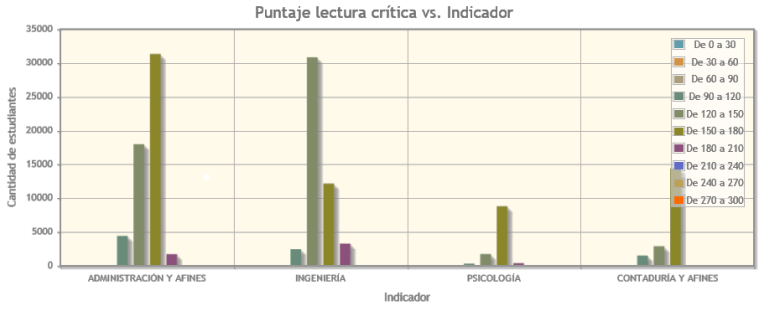


Figura 9. Puntaje lectura crítica vs. Área de la carrera
Fuente: Autores

En razonamiento cuantitativo, quienes tienen mejores puntajes son los ingenieros, seguidos de los administradores, representado así en la Figura 10. Mientras que, para el área de contaduría y psicología el promedio de puntajes obtenidos disminuye significativamente, siendo los psicólogos los de desempeño más bajo, no muy lejos de los contadores. Es decir que los puntajes más altos fueron de los hombres ingenieros y las mujeres administradoras.



Figura 10. Puntaje razonamiento cuantitativo vs. Área de la carrera Fuente: Autores

Para el núcleo de comunicación escrita mostrado en la Figura 11, las administradoras obtuvieron un mejor desempeño que las mujeres en las otras carreras (Rincón-baez & Ciencias-física, n.d.); están seguidas de las contadoras e ingenieras, con un comportamiento similar entre ellos, mientras que las psicólogas son quienes obtuvieron un desempeño más bajo.

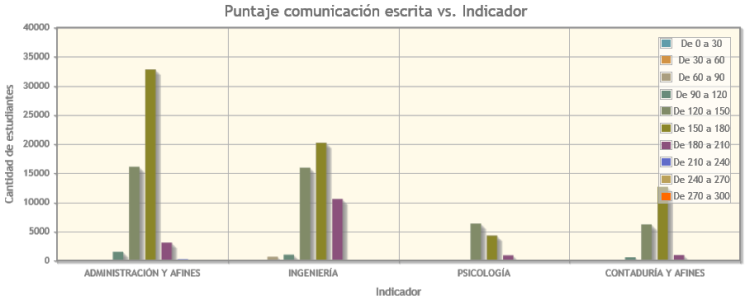


Figura 11. Puntaje comunicación escrita vs. Área de la carrera
Fuente: Autores

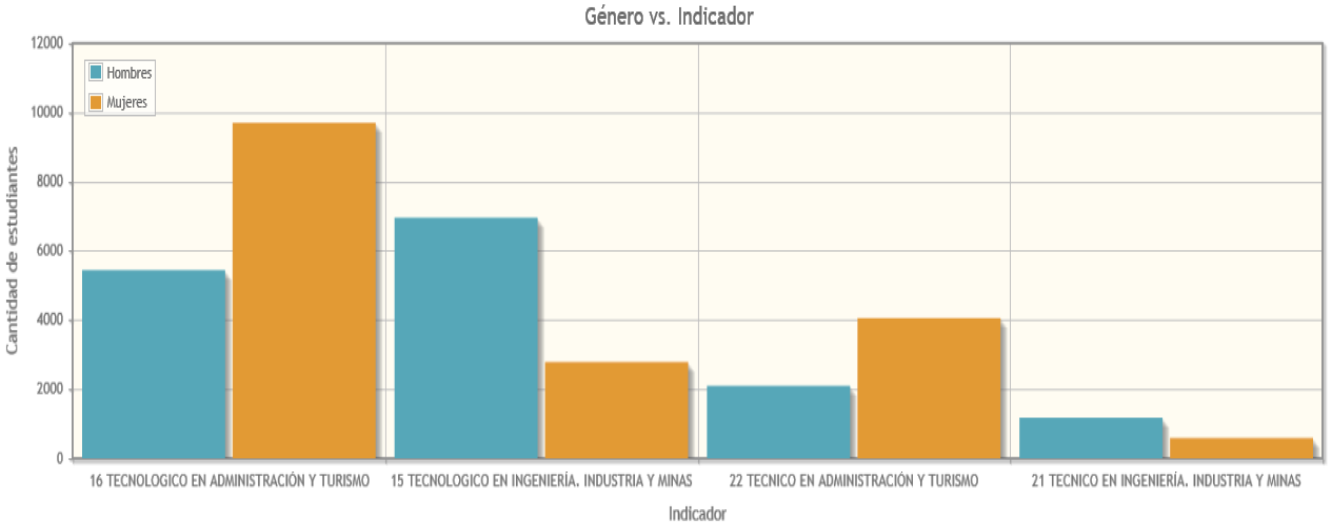


Figura 12. Género vs. Área de TyT
Fuente: Autores

En competencias ciudadanas, los psicólogos obtuvieron buenos puntajes, por su parte los ingenieros también presentan un comportamiento similar en los 2 rangos principales, pero se observa en la Figura 13 que tienen también una cantidad importante en un rango inferior. En cambio, los contadores presentan los puntajes más bajos y muy de cerca los acompañan los administradores, la diferencia radica en que los últimos presentan una cantidad considerable en un rango superior. Este núcleo es el que menos brecha de género presenta en las pruebas Saber.

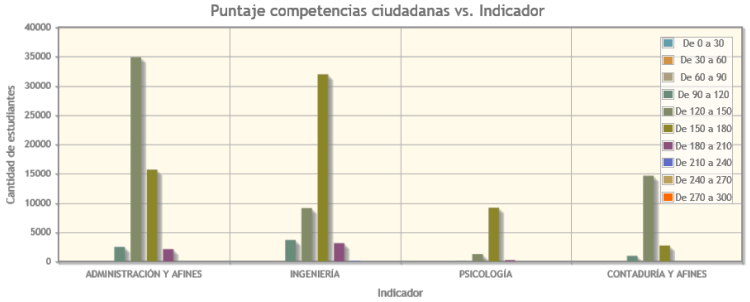


Figura 13. Puntaje competencias ciudadanas vs. Área de la carrera Fuente: Autores

Finalizando con esta prueba, en el núcleo de inglés representado en la Figura 14 no se notan diferencias considerables entre las 4 áreas de estudio. Lo que si se puede resaltar es que los estudiantes registran un promedio bajo que debe ser mejorado para su subir el nivel educativo del país.

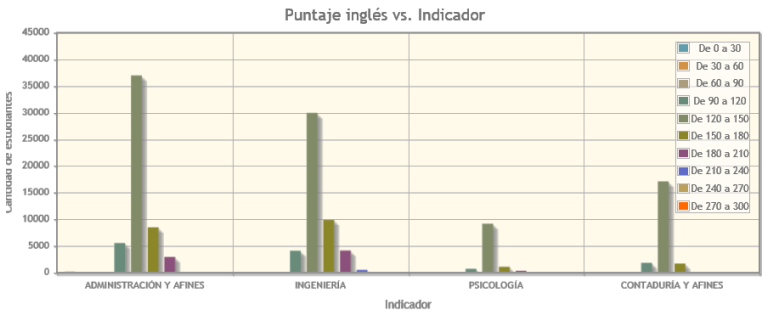


Figura 14. Puntaje inglés vs. Área de la carrera
Fuente: Autores

Ahora, para la prueba Saber TyT se realizó un análisis descriptivo de igual manera para 4 grupos de referencia, como se aprecia en la Figura 12. Los resultados se expresan de forma visual en las siguientes gráficas. En primer lugar, la relación que hay entre las 4 áreas de carrera y la cantidad de hombres y mujeres, permite observar que tanto el técnico como el tecnológico en Administración y Turismo, tienen en su mayoría al género femenino, caso contrario para el técnico y tecnológico de Ingeniería, Industria y Minas.

El núcleo temático de lectura crítica para esta prueba, tiene brechas académicas en 2 sentidos; la primera de ellas entre técnicos y tecnólogos, puesto que los tecnólogos presentan mejores puntajes en ambas áreas con respecto a los técnicos. La otra diferencia es que en el campo es decir que las mujeres obtuvieron mejores resultados, mostrado así en la Figura 15.



Figura 15. Puntajes lectura crítica vs. Área de TyT
Fuente: Autores

Para razonamiento cuantitativo, en esta materia no se notan diferencias significativas, la disparidad de género no es representativa. Sólo se continúa con el mismo comportamiento donde a los tecnólogos se desempeñan mejor que los técnicos, como lo expresa gráficamente la Figura 16.

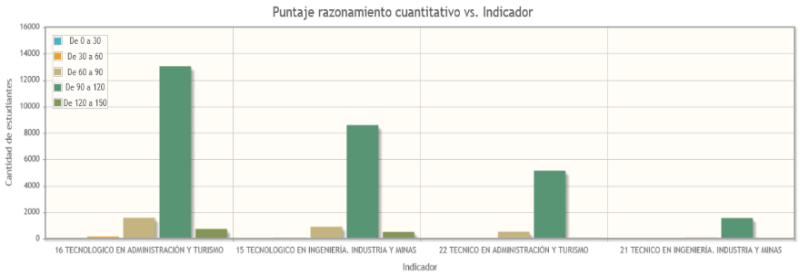


Figura 16. Puntajes razonamiento cuantitativo vs. Área de TyT Fuente: Autores

En la evaluación del núcleo de Comunicación escrita se presenta un comportamiento contrario como lo muestra la Figura 17, esto quiere decir que los técnicos tuvieron mejor desempeño que los tecnólogos. Analizando por áreas, las mujeres de tecnólogos de administración presentan mejores puntajes que los de ingeniería.

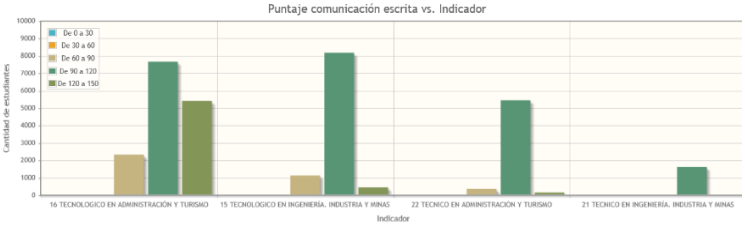


Figura 17. Puntajes comunicación escrita vs. Área de TyT Fuente: Autores

En este nivel educativo, los estudiantes obtuvieron puntajes similares en el núcleo de competencias ciudadanas con el de lectura crítica. Tanto el campo administrativo como de ingenieros alcanzaron un promedio de puntajes medio. Vale la pena aclarar que en esta prueba los puntajes ya no son de 0 a 300 como el Saber Pro, sino de 0 a 150, tal como se refleja en la Figura 18.



Figura 18. Puntaje competencias ciudadanas vs. Área de TyT Fuente: Autores

Por último, los estudiantes de los distintos grupos de referencia en inglés tienen un comportamiento bastante similar, mostrado así en la Figura 19, donde su clúster más representativo es el comprendido entre 90 y 120 puntos, esto indica que el nivel de inglés es aceptable.

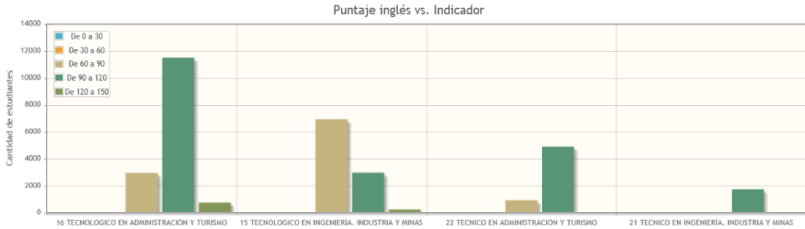


Figura 19. Puntaje inglés vs. Área de TyT Fuente: Autores

4. CONCLUSIONES

En el país, hoy en día aún persisten brechas de género en la educación, tales como: acceso a la universidad o institución educativa, en los resultados de las pruebas estandarizadas, en la elección de qué carreras estudiar. A pesar de que las mujeres son el género que más ingresa a la educación técnica y superior, son también las que presentan desempeños más bajos en los exámenes. A través del tiempo, se ha venido tomando una tendencia para las mujeres en optar por carreras administrativas, contables y/o afines. Mientras que, el género masculino tiende a estudiar carreras ingenieriles, de industria o relacionados (Reyes Nava et al., 2017).

A partir del análisis de los indicadores con respecto a cada núcleo temático, se puede concluir que en términos generales los estudiantes no tienen un alto desempeño en las pruebas Saber 2016. Además, existen algunas diferencias marcadas en los puntajes finales obtenidos entre las áreas de carrera, por ejemplo, los ingenieros tienen mejores resultados en razonamiento cuantitativo y por su parte los administradores se destacan en lectura crítica. Lo anterior señala que en Matemáticas la brecha es a favor de los hombres; mientras que, en Comprensión lectora está inclinada hacia las mujeres.

Se puede concluir también que el género femenino tiende a tener una menor la probabilidad de adquirir un nivel alto de bilingüismo, cursar una carrera técnica si se compara con una tecnóloga y así mismo, estudiar en grupos de referencia tales como ingeniería, industria y minas (Núñez et al., n.d.).

Por último, como fue mostrado el comportamiento del indicador de Área de residencia, de éste se deduce que la mayoría de población está concentrada en el área urbana o cabeceras municipales del país. Tanto en el área rural como urbana, las mujeres constituyen una mayoría, lo que está relacionado con el gran porcentaje de mujeres que

acceden a la educación. Con este indicador, también se logró identificar que existe una brecha negativa en el desempeño de ciertos núcleos temáticos para los estudiantes del área rural, siendo más notable en inglés. Aunque la constitución política de Colombia señala que "la educación es gratuita y obligatoria para todos los ciudadanos" (Giraldo & Fajardo, 2018), se puede observar que población como campesinos, desplazados y personas de muy bajos recursos económicos no tienen las mismas posibilidades de acceso, factor que se recomienda estudiar en profundidad en una próxima investigación.

BIBLIOGRAFÍA

Abadía, L. K., & Bernal, G. (2016). Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: evidencia pruebas Saber 11. *Vniversitas Económica*, *16*(9), 1–34.

Arias Ortiz, E., & Bornacelly, I. (2017). *¿Les va mejor a las niñas en educación?*

Becerra-Gualdrón, C. J., Gallardo-Sánchez, C. F., & Becerra-Gualdrón, S. B. (2018). Índice de brecha de género en Colombia a nivel departamental 2008 – 2014. *Revista Logos, Ciencia & Tecnología*, *10*(3). https://doi.org/10.22335/rlct.v10i3.555

Carmona, C. J., Pulgar-Rubio, F. J., García-Vico, A. M., González, P., & del Jesus, M. J. (2015). Análisis descriptivo mediante aprendizaje supervisado basado en patrones emergentes. *Proc. of the VII Simposio Teoría y Aplicaciones de Minería de Datos*, 685–694.

Caro Acero, B. L., & Casas M., A. F. (2013). *Análisis de las diferencias de género en el desempeño de estudiantes colombianos en matemáticas y lenguaje*.

DANE, D. A. N. de E. (2016). Estratificación. *Información Vía Web*, 7. https://www.dane.gov.co/files/geoestadistica/Preguntas_frecuentes_estratificacion.pdf

Decreto 3963 de 2009, Diario Oficial 4 (2009).

Espino Timón, C. (2017). *Análisis predictivo técnicas y modelos utilizados y aplicaciones del mismo*. 65. http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117memòria.pdf

García, S., Ramírez Gallego, S., Luengo, J., & Herrera, F. (2016). *Big Data: Preprocesamiento y calidad de datos*. www.highlycited.com

Giraldo, C., & Fajardo, R. (2018). LA EDUCACIÓN PÚBLICA FRENTE A LA EDUCACIÓN PRIVADA: UNA COMPARACIÓN DE BENEFICIOS. *Pensamiento Udecino REVISTA DE PROYECCIÓN ACADÉMICA, October 2018*, 6–13. https://doi.org/10.1017/CBO9781107415324.004

González Castañeda, R. (2012). *Implementación y ejecución de un protocolo de transferencia de archivos (FTP)*.

Hernández Sampieri, R., Fernandez Collado Carlos, & Baptista Lucio, P. (2006). Los enfoques cuantitativo y cualitativo en la investigación científica. In *Metodología de la instigación*. https://doi.org/10.6018/turismo.36.231041

López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. https://doi.org/10.18235/0001082

López, D. (2015, March 11). *Hay poca presencia de las mujeres en carreras relacionadas con ciencias exactas y tecnología*. Universidad Nacional de Colombia. https://minas.medellin.unal.edu.co/noticias/facultad/143-mujeres

Núñez, J., Olga, T., & Juan Manuel, M. (n.d.). *Diagnóstico del estado del bilingüismo en Bogotá Informe final*.

ONU Mujeres. (2018). *El progreso de las mujeres en Colombia 2018: Transformar la economía para realizar los derechos*. http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های نوین&option=com_dbook&task=readonline&book_id=13650&page=73&chkhashk=ED9C9491B4&Itemid=218&lang=fa&tmpl=component

Orjuela, J. (2014). Análisis del Desempeño Estudiantil en las Pruebas de Estado para Educación Media en Colombia mediante Modelos Jerárquicos Lineales. *Ingeniería*, *18*(2), 54–67. https://doi.org/10.14483/udistrital.jour.reving.2013.2.a04

Orozco, L. E. (2010). Calidad académica y relevancia social de la educación superior en América Latina. *Revista Iberoamericana de Educación Superior*, *1*(1), 24–36. http://www.scielo.org.mx/scielo.php?pid=S2007-28722010000100003&script=sci_arttext

Palacio, A., Bosque, J., & García, D. (2015). *TÉCNICAS DE MINERÍA DE DATOS APLICADO A LA MONITORIZACIÓN DE SISTEMAS*.

Pérez, E., Vásquez, J., & González, M. (2018). TASAS ESTADÍSTICAS Y CONCEPTOS ASOCIADAS A LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA – CASO ESPECÍFICO PROGRAMAS DE CONTADURÍA PÚBLICA ENTRE LOS AÑOS 2008 AL 2013. Pensamiento Udecino REVISTA DE PROYECCIÓN ACADÉMICA, (October 2018), 59–70. https://doi.org/10.1017/CBO9781107415324.004

Reyes Nava, A., Flores Fuentes, A., Alejo, R., & Rendón

Lara, E. (2017). Minería de datos aplicada para la identificación de factores de riesgo en alumnos. *Research in Computing Science*, *139*(ISSN 1870-4069), 177–189.

Rincón-baez, W. U., & Ciencias-física, M. (n.d.). *GÉNERO EN SABER PRO EN PROGRAMAS DE*.

Rodrigo, J. A. (2017). Clustering y heatmaps: aprendizaje no supervisado. *RPubs*. https://rpubs.com/Joaquin_AR/310338

Rodríguez, J. E. (2012). *Desarrollo de herramientas para mineria de datos " UDMiner ."* 21–40.

Romero, C., & Ventura, S. (2016). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Ruiz, M. E. (2003). *Desafíos y oportunidades para la equidad de género en América Latina y el Caribe*. 20. http://www.oei.es/genero/documentos/internacionales/desafios_y_oport_genero_AL_2003.pdf

Sierra, R., Núñez, E., Martín-Asín, G., Sevilla, C., & Delgado, J. (2015). *Transformación y explotación de datos hidrográficos del IGN con herramientas ETL*. *1*.

Universia. (2013). *Estudio demuestra que hombres y mujeres varían su rendimiento en matemáticas*. https://noticias.universia.net.co/en-portada/noticia/2013/09/13/1049251/estudio-demuestra-hombres-mujeres-varian-rendimiento-matematicas.html

Universidad de los Llanos. (2018). *INFORME DE RESULTADOS PRUEBAS SABER PRO y SABER TyT*.

Vergara, C. C., & Ávila, J. A. M. (2012). Diferencias por sexo en el desempeño académico en Colombia: Un análisis regional. *Economía & Región*, *6*(1), 133–169. http://publicaciones.unitecnologica.edu.co/index.php/economia-y-region/article/view/109/108

Villafañe Blanco, P. V. (2015). *Análisis del Desempeño Académico del Examen de Estado para el Ingreso a la Educación Superior Aplicando Minería de Datos Análisis del Desempeño Académico del Examen de Estado para el Ingreso a la Educación Superior Aplicando Minería de Datos*. http://www.bdigital.unal.edu.co/51414/1/39004913.2015.pdf

7.7. ARTÍCULO CICI

DETERMINACIÓN DE INDICADORES DE BRECHAS DE GÉNERO Y ANÁLISIS DESCRIPTIVO EN LAS PRUEBAS SABER

Alisson Catalina Celeita Gutierrez [1[0000-0001-9750-2848]], Angie Paola Manrique Ravelo [2[0000-0003-0302-406X]], Cesar Yesid Barahona Rodriguez [3[0000-0001-7673-7381]] y Corín Jazmín Viracacha Viracacha [4[0000-0003-4487-938X]]

[1] Universidad de Cundinamarca
[2] Facatativá, Cundinamarca, Colombia
acceleita@ucundinamarca.edu.co apaolamanrique@ucundinamarca.edu.co
cbarahona@ucundinamarca.edu.co
cviracacha@ucundinamarca.edu.co

Resumen. Las brechas de género son una problemática que perjudica varios aspectos de la sociedad, por esta razón se tiene como objetivo la identificación de aquellos factores que inciden en la educación del país más exactamente en los resultados de las pruebas realizadas por el ICFES, para ello se implementaron técnicas de minería de datos y algoritmos de análisis de agrupamiento (clustering), generando como parte final del proyecto un aplicativo web donde el usuario podrá visualizar de forma gráfica la información resultante para que sea objeto de estudio de futuras investigaciones.

Palabras clave: Brechas de género, minería de datos, clustering, educación, ICFES

INTRODUCCIÓN

Con el paso de los años han existido brechas de género en los ámbitos sociales, económicos y políticos afectando significativamente el desarrollo humano de todos los países en especial a los que conforman a América Latina y el Caribe. Tal es el caso de la gravedad de la problemática en algunos sectores, que por ejemplo en la participación política de las mujeres en Colombia, tan solo el 55.2% ejercen el derecho al sufragio. En cuanto al ámbito económico [1], para los años comprendidos entre el 2014 y 2017 la participación laboral de las mujeres se ha estancado en un 54% dejando al país dos puntos por debajo del promedio de la región [2].

El panorama en el área educativa tampoco es alentador, puesto que, por ejemplo, las mujeres indígenas y rurales son quienes menos educación reciben [3]. Así mismo, la participación femenina para el 2018 en las áreas del STEM (Ciencias, tecnología, ingeniería y matemáticas) disminuyo a un 34.9% con respecto al 2001 la cual se ubica en un 38.4%, demostrando que la brecha de género en la educación superior va aumentando [4].

La manera más sencilla para comprender el comportamiento de las brechas de género es guiarse de factores en específico o en áreas puntuales, por esta razón, haciendo uso de los resultados de las pruebas estandarizadas del ICFES se realizó por medio de técnicas de minería de datos un análisis descriptivo donde se evidencie a través de un aplicativo web qué indicadores han incidido en los últimos años de presentación de las pruebas.

1. METODOLOGIA

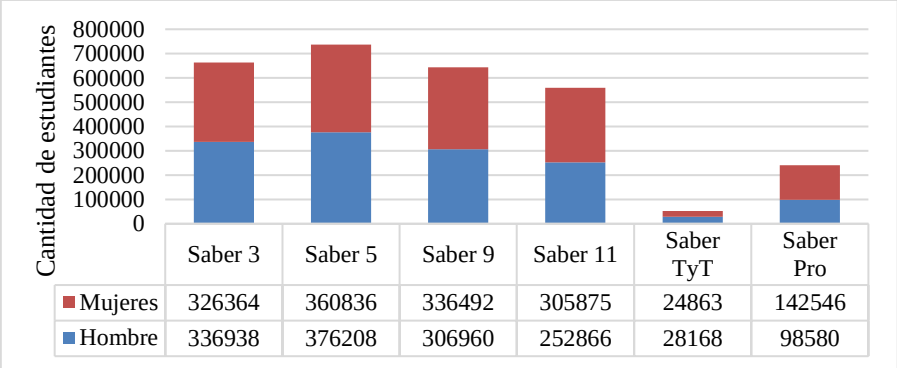
Para el análisis de las pruebas se hizo uso de la metodología de minería de datos CRISP-DM, la cual tiene varias etapas a seguir. En primera instancia se identificó el problema a tratar, luego se realizó una fase de preprocesamiento de los datos, la cual inició con la extracción de los resultados del FTP; una vez obtenidos los archivos ".CSV" se procedió a la eliminación de columnas innecesarias y datos nulos, para posteriormente realizar la carga de la información a la base de datos.

Para encontrar los indicadores se hizo uso del algoritmo K-Means, el cual consiste en clasificar los datos y asignarlos a un clúster de acuerdo al centroide más cercano, esto para poder determinar en donde existe mayor

concentración y qué relación hay entre las variables seleccionadas para finalmente mostrar dentro del aplicativo una sección con gráficos estadísticos de los resultados hallados.

3. RESULTADOS

Como parte de la información recolectada en la siguiente figura se evidencia la cantidad de datos procesados y analizados por prueba. Es decir, la cantidad de estudiantes por género en cada uno de los niveles evaluados.



	Saber 3	Saber 5	Saber 9	Saber 11	Saber TyT	Saber Pro
Mujeres	326364	360836	336492	305875	24863	142546
Hombre	336938	376208	306960	252866	28168	98580

Fig. 1. Cantidad de estudiantes prueba SABER.

De acuerdo a diversos autores para llevar un análisis más organizado se debe categorizar la información en dimensiones académicas, económicas y sociales; por lo anterior, los resultados obtenidos de la etapa descriptiva fueron divididos de esta forma. Para Gálvez los aspectos sociales ejercen una gran influencia en la autodeterminación, seguridad y habilidad para controlar el dinero, el trabajo propio y las relaciones sociales [5]. Mientras que desde el punto de vista académico existe una gran variación en las instituciones educativas, en cuanto a la clasificación profesional y la transmisión académica, que contribuye al aumento de las brechas de género en los países de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) [6].

De todas las variables analizadas y procesadas, finalmente se estableció cuáles son los indicadores con mayor influencia en la brecha de género por prueba.

Tabla 1. Indicadores por dimensiones

Económicos	Académicos	Sociales
Valor pensión	Educación Madre	Estado civil
Estrato de vivienda	Educación Padre	Cabeza de familia
Ingreso familiar	Género del colegio	Número de personas a cargo
Nivel de Sisbén	Jornada del colegio	Área de residencia
Servicio de TV	Dedicación lectura diaria	
Trabaja Actualmente	Área de la carrera	
Tiene Computador	Método de programa	
Tiene Consola de video-juegos		
Tiene internet		
Valor matrícula universidad		
Tipo de pago matrícula.		

- Educación madre y educación padre: Estos indicadores representan los niveles de escolaridad alcanzados por los padres, el cual se clasifica en los siguientes atributos: ninguno, no sabe, no aplica, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica y tecnológica incompleta, técnica y tecnológica completa, educación profesional incompleta, educación profesional completa y postgrado.

- Género y jornada del colegio: Estos indicadores definen las características principales del colegio; en primer lugar, el género, se clasifica en femenino, masculino y mixto, mientras que la jornada hace referencia al horario de presentación de las clases, clasificado en los siguientes atributos: completa, tarde, mañana, sabatina, noche y única.

- Dedicación lectura diaria: Representa el tiempo de dedicación a la lectura por parte del estudiante, clasificada en rangos de tiempo, de la siguiente manera: 30 minutos o menos, no leo por entretenimiento, entre 30 y 60 minutos, más de 2 horas y entre 1 y 2 horas.

- Área de la carrera: Hace referencia al grupo de clasificación de la carrera que se está cursando. Existen diversas categorías, pero para el análisis se tuvieron en cuenta: Administración y afines, Contaduría y afines, psicología e ingeniería, en el caso de Saber Pro. Y para Saber TyT son: tecnológico en administración y turismo, tecnológico en ingeniería, industria y minas, técnico en administración y turismo, y técnico en ingeniería, industria y minas.

- Método del programa: Hace referencia a la modalidad de estudio del programa, el cual se clasifica en presencial, distancia y virtual.

- Valor pensión: Este indicador representa el rango de valores de la pensión y matricula, que se paga por el acceso a la educación, está clasificado por rangos de precio de la siguiente manera: entre 87.000 y menos de 120.000, no paga pensión, 250.000 o más, entre 150.000 y menos de 250.000, menos de 87.000, entre 120.000 y menos de 150.000.

- Ingreso familiar: Representa la cantidad de dinero que percibe la familia mensualmente, clasificado en: entre 2 y menos de 3 SMLV, entre 5 y menos de 7 SMLV, menos de 1 SMLV, entre 3 y menos de 5 SMLV, entre 1 y menos de 2 SMLV, entre 7 y menos de 10 SMLV, 10 o más SMLV.

- Nivel del Sisbén y estrado de vivienda: Estos indicadores representan la clasificación de la población de acuerdo a sus condiciones socioeconómicas, para el caso del Sisbén se clasifica en: nivel 1, nivel 2, nivel 3, no está clasificada por el SISBÉN y está clasificada en otro nivel del SISBÉN, en cuanto al estrato sus categorías comprenden desde el estrato 0 hasta el estrato 6.

- Trabaja actualmente: Nos indica si el estudiante se encuentra laborando, clasificada de la siguiente manera: si menos de 20 horas a la semana, si 20 horas o más a la semana y no se encuentra trabajando.

- Servicio de internet y televisión: Nos indica si el estudiante cuenta o no con el servicio de televisión e internet en su hogar.

- Tiene computador y consola video-juegos: Indica si el estudiante cuenta o no con computador y consola de video-juegos en su hogar.

- Valor matricula universidad: Este indicador representa el costo de la matrícula del programa que está cursando el estudiante, clasificado en rangos de pago, de la siguiente manera: entre 5.5 millones y menos de 7 millones, entre 2.5 millones y menos de 4 millones, menos de 500 mil, entre 4 millones y menos de 5.5 millones, entre 500 mil y menos de 1 millón, entre 1 millón y menos de 2.5 millones, Mas de 7 millones y no pago matricula.

- Tipo de pago matricula: Este indicador cuenta con varias opciones de respuesta, pero debido a su afectación nos centramos en el tipo de pago crédito, el cual cuenta con dos opciones de respuesta sí y no.

- Estado civil: Representa la situación de la persona de acuerdo a sus relaciones de pareja, se dividen en 4 opciones: si está casado, soltero, unión libre y separado y/o viudo.

- Cabeza de familia y número de personas a cargo: Determina si el estudiante es o no responsable de la economía de su hogar, en caso de ser responsable se relaciona la cantidad de personas a cargos.

- Área de residencia: Indica la zona de residencia del estudiante, en dicho momento. Sus atributos se clasifican en rural y urbano.

CONCLUSIONES

Gracias al uso de técnicas de minería de datos se pudo descartar variables innecesarias e inconsistentes para el objeto de estudio, determinando así rasgos característicos en común entre los indicadores, y de esta manera clasificarlos en ámbitos académicos, sociales y económicos. A partir de esto se encontró en qué grado afecta cada una de las subcategorías permitiendo establecer una relación entre núcleos temáticos y dichos resultados.

Para futuras investigaciones se recomienda realizar un comparativo entre las pruebas PISA, TIMSS y Saber, el cual permita estimar que otros factores inciden en los resultados académicos, también es recomendable hacer un análisis con las pruebas saber 4°, 6° y 8°, las cuáles serán implementadas próximamente por el ICFES.

BIBLIOGRAFIA

A. R. L. Francisco, *Colombia: La hora de la paridad política*, vol. 53, no. 9. 2013.

ONU Mujeres, UNDP, UNFPA, and USAID, *De los Objetivos de Desarrollo del Milenio a los Objetivos de Desarrollo Sostenible*, no. 88. 2016

Lopez, V., Grazzi, M., Guillard, C. and Salazar, M., 2018. *Las Brechas De Género En Ciencia, Tecnología E Innovación En América Latina Y El Caribe: Resultados De Una Recolección Piloto Y Propuesta Metodológica Para La Medición*. Banco Interamericano de Desarrollo, accessed 13 June 2020.

Patiño, L., 2020. *16 años después, hay más mujeres, pero la brecha no cede. El Tiempo*, Available at: https://www.eltiempo.com/tecnosfera/novedades-tecnologia/brecha-de-genero-mujeres-estudian-ciencia-y-tecnologia-en-colombia-41213, accessed 9 June 2020.

Gálvez, T., 2002. *Aspectos económicos de la Equidad de Género, Publicaciones de las Naciones Unidas*, accessed 11 June 2020.

Schrøter, J. and Skyt, H., 2013. *Math And Gender: Is Math A Route To A High-Powered Career?*. IZA, accessed 10 June 2020.

7.8. ARTÍCULO ENCUENTRO DE SEMILLEROS

APLICATIVO WEB PARA EL ANÁLISIS DESCRIPTIVO Y PREDICTIVO DE BRECHAS DE GÉNERO EN LAS PRUEBAS SABER

Angie Paola Manrique Ravelo[4]
Alisson Catalina Celeita Gutiérrez[5]

RESUMEN

En este proyecto, se tiene como objetivo encontrar brechas de género en el ámbito educativo en Colombia, más específicamente en los resultados obtenidos por los estudiantes que presentan las pruebas Saber 11, Saber TyT y Saber Pro realizadas por el ICFES, puesto que son una herramienta de medición excepcional de la situación del país. Por medio de la aplicación de técnicas y algoritmos de minería de datos se realiza un análisis descriptivo de las pruebas, obteniendo como resultado final indicadores que inciden en esta problemática y un aplicativo web que permite visualizar de forma gráfica la información encontrada, con el fin de que pueda ser usada en futuras investigaciones. Además, se construyó también un modelo predictivo donde los resultados son las probabilidades más altas en cada una de las pruebas mencionadas anteriormente.

Palabras claves:

Análisis descriptivo, Análisis predictivo, Brechas de género, Educación, Minería de datos.

ABSTRACT

In this project, the objective of this project is to find gender gaps in the educational field in Colombia, more specifically in the results obtained by the students who take the Saber 11, Saber TyT and Saber Pro tests carried out by ICFES, since they are a tool exceptional measurement of the country's situation. Through the application of data mining techniques and algorithms, a descriptive analysis of the tests is carried out, obtaining as a final result indicators that influence this problem and a web application that allows the information found to be displayed graphically, in order to that can be used in future research. In addition, a predictive model was also built where the results are the highest probabilities in each of the aforementioned tests.

Keywords:

Descriptive analysis, Predictive analysis, Gender gaps, Education, Data mining.

[4] **Estudiante de noveno semestre de Ingeniería de sistemas, Auxiliar de investigación semillero Aplicaciones Móviles, Universidad de Cundinamarca, apaolamanrique@ucundinamarca.edu.co, https://orcid.org/0000-0003-0302-406X**
[5] **Estudiante de noveno semestre de Ingeniería de sistemas, Auxiliar de investigación semillero Aplicaciones Móviles, Universidad de Cundinamarca, acceleita@ucundinamarca.edu.co, https://orcid.org/0000-0001-9750-2848**

INTRODUCCIÓN

A través del tiempo, en Colombia, se han presentado brechas de género en gran parte de los campos económicos y socioculturales. En consecuencia, de estas diferencias presentadas es importante realizar un análisis de los resultados de las pruebas del estado desde el año 2016 y de esta forma hallar los factores que más inciden en esta problemática en el ámbito educativo del país. Este enfoque se justifica, por múltiples cifras o estadísticas que muestran el desequilibrio que se ha venido presentando; por ejemplo, en América Latina y el Caribe (ALC) para el año 2015 el índice de mujeres graduadas en estos núcleos estuvo comprendido entre 20% y 50%; mientras que en otros campos de estudio como lo son las ciencias naturales y afines, el porcentaje es mucho mayor encontrándose en un rango del 50% y 80% (López-Bassols et al., 2018).

Por consiguiente, a través de la minería de datos, se encontrarán indicadores de brecha de género en las pruebas Saber realizadas en el grado 11, la educación superior, técnica y tecnóloga; con el fin de evaluar su comportamiento y mirar que factores contribuyen en su crecimiento. A partir de este análisis, se procederá a encontrar una relación con los núcleos temáticos en común para predecir cuál será su tendencia en las próximas pruebas.

DESARROLLO

En primer lugar, se realizó el desarrollo de un modelo descriptivo que muestre si existe relación entre ciertos factores recolectados en las Pruebas Saber y la brecha de género haciendo uso de técnicas de minería de datos, específicamente técnicas de asociación, las cuales “establecen las posibles relaciones entre diferentes acciones aparentemente independientes entre sí, reconociendo como pueden ocurrir los eventos o acciones a partir de la aparición de otros” (Reyes Nava et al., 2017). Además, se implementó la metodología CRISP-DM, que existe para orientar el proceso de minería de datos, facilita la realización de proyectos optimizando su planificación, reduciendo su complejidad y permitiendo un mejor seguimiento de los mismos. Tiene como ventaja la centralización de los objetivos realizando un análisis del problema para transformarlo en un problema técnico de minería de datos, para este proyecto ha sido escogida por su capacidad de definir una estructura y orientar su ejecución (Cobos et al., 2010).

Durante este proceso de construcción del modelo se evaluó cual algoritmo tiene mejor porcentaje de acierto, con el fin de obtener resultados lo más precisos posibles. Asimismo, en la construcción del aplicativo web se implementó la metodología de desarrollo ágil SCRUM, ya que es un modelo general de gestión de entornos de producción basados en rutinas, donde el factor más importante es el conocimiento aplicado de las personas que hacen parte del proyecto, aún más primordial que el mismo proceso tecnológico (Palacio & Ruata, 2011). Por otra parte,

su importancia radica también en las iteraciones que permiten el desarrollo incremental del proyecto desde su concepción hasta que se dé por terminado. Finalmente, se visualizan por medio de graficas los resultados obtenidos del análisis descriptivo y predictivo de las Pruebas Saber en Colombia, haciendo énfasis en los indicadores de brecha de género encontrados.

CONCLUSIONES

- El uso de minería de datos para el proceso de análisis de datos permite tener acceso a información oculta que con los métodos comúnmente usados de estadística no se puede obtener, además, facilita el procesamiento de grandes volúmenes de datos.
- Gracias a las técnicas de minería de datos, en este caso, de agrupamiento y clasificación ad hoc se puede generar modelos descriptivos y predictivos consistentes de la realidad observada. Con el desarrollo de esta investigación se contribuirá a futuros estudios sobre brecha de género, recomendando así hacer un comparativo de estas pruebas (Saber) con las realizadas a nivel internacional como las PISA.
- El análisis predictivo se ocupa principalmente de anticipar resultados futuros basados en minería de datos existentes. Esta información se convierte en importante a la hora de ser analizada porque permite comprender su comportamiento futuro, o en términos coloquiales lo que puede suceder. Esta capacidad de predecir da un poder inmenso de planificación y brinda la posibilidad de solución a diferentes problemáticas ya que entiende cuál es la conducta de una o varias cosas en distintos escenarios de evaluación.
- WEKA facilita la aplicación de los algoritmos de minería de datos siendo una herramienta fiable y de fácil uso para el usuario.

REFERENCIAS BIBLIOGRÁFICAS

Cobos, C., Zuñiga, J., Guarin, J., León, E., & Mendoza, M. (2010). CMIN - Herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingenieria e Investigacion*, *30*(3), 45–56.

López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. https://doi.org/10.18235/0001082

Palacio, J., & Ruata, C. (2011). *Gestión de proyectos Scrum Manager*. http://www.hacienda.go.cr/cifh/sidovih/spaw2/uploads/images/file/Gesti?n de proyectos.pdf

Reyes Nava, A., Flores Fuentes, A., Alejo, R., & Rendón Lara, E. (2017). Minería de datos aplicada para la identificación de factores de riesgo en alumnos. *Research in Computing Science*, *139*(ISSN 1870-4069), 177–189.

7.9 FORMATOS DE SEGUIMIENTO

A continuación, se adjunta los formatos de seguimiento llenados de forma presencial y pantallazos de las reuniones virtuales que se tuvieron con nuestro director de proyecto. Dichas reuniones fueron cada 8 o 15 días según los avances programados.

UNIVERSIDAD DE CUNDINAMARCA

Programa de Ingeniería de Sistemas

UDEC
UNIVERSIDAD DE CUNDINAMARCA

CONTROL Y SEGUIMIENTO PROYECTOS DE GRADO

FECHA: 13 de Febrero 2020

NOMBRE DEL PROYECTO:
Analisis descriptivo y predictivo de brechas de genero en las pruebas Saber 11, saber Pro y Saber TyT

CODIGO: 46121620 ESTUDIANTE: Alisson Catalina Celeita Gutierrez

CODIGO: 46121642 ESTUDIANTE: Angie Paola Manrique Ravelo

DIRECTOR DE PROYECTO: Cesar Yesid Barahona

TEMA TRATADO:
Avance del proyecto en vacaciones

TEMA SIGUIENTE AVANCE:
Finalizar diagramas y avance del articulo

FECHA SIGUIENTE AVANCE: 27 de Febrero

OBSERVACIONES:

FIRMAS

ESTUDIANTE — Angie Manrique — ESTUDIANTE — DIRECTOR DEL PROYECTO

Figura 341. Formato de seguimiento 1
Autores: (Manrique, Celeita, Barahona, 2020)

UNIVERSIDAD DE CUNDINAMARCA

Programa de Ingeniería de Sistemas

UDEC
UNIVERSIDAD DE CUNDINAMARCA

CONTROL Y SEGUIMIENTO PROYECTOS DE GRADO

FECHA: 27 Febrero 2020

NOMBRE DEL PROYECTO:
Analisis descriptivo y predictivo de brechas de genero en las pruebas Saber 11, Saber Pro, Saber TyT

CODIGO: 461216120 ESTUDIANTE: Alisson Catalina Celeita Gutierrez

CODIGO: 461216142 ESTUDIANTE: Angie Paola Manrique Rincon

DIRECTOR DE PROYECTO: Cesar Yesid Barahona

TEMA TRATADO:
Finalizar diagramas y avance del articulo

TEMA SIGUIENTE AVANCE:
Entregar ETL y FTP.

FECHA SIGUIENTE AVANCE: 12 Marzo 2020

OBSERVACIONES:

FIRMAS

ESTUDIANTE

Angie Manrique
ESTUDIANTE

DIRECTOR DEL PROYECTO

Figura 342. Formato de seguimiento 2
Autores: (Manrique, Celeita, Barahona, 2020)

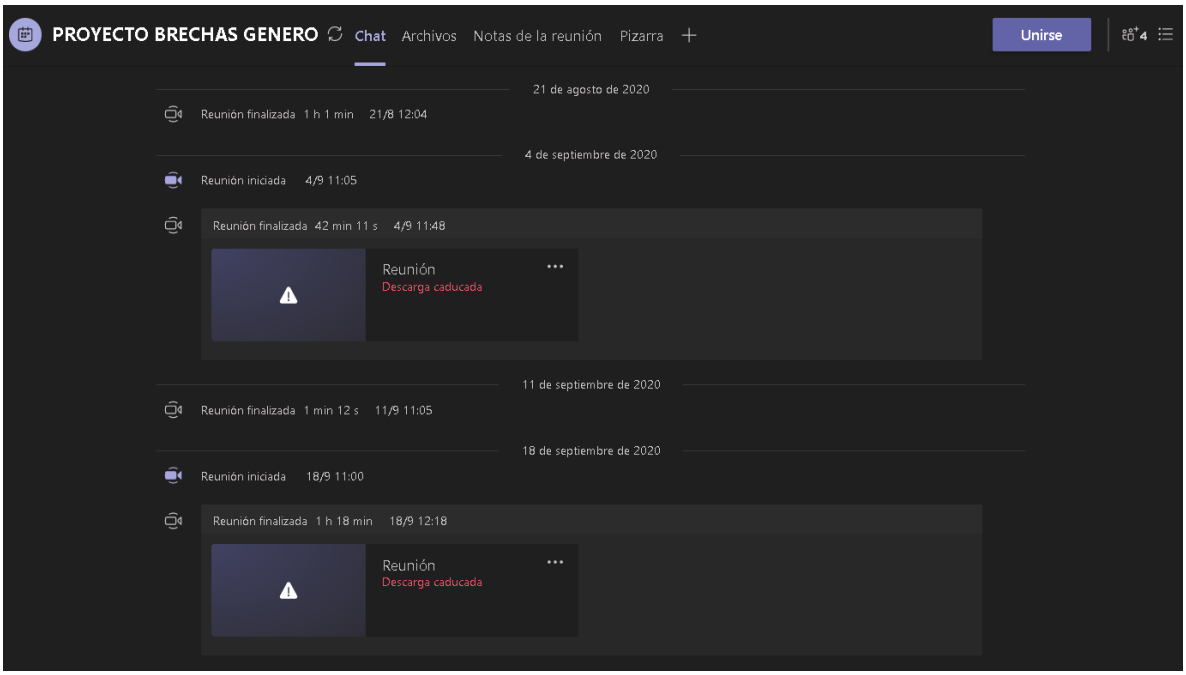


Figura 343. Reuniones virtuales 1
Autores: (Manrique, Celeita, Barahona, 2020)

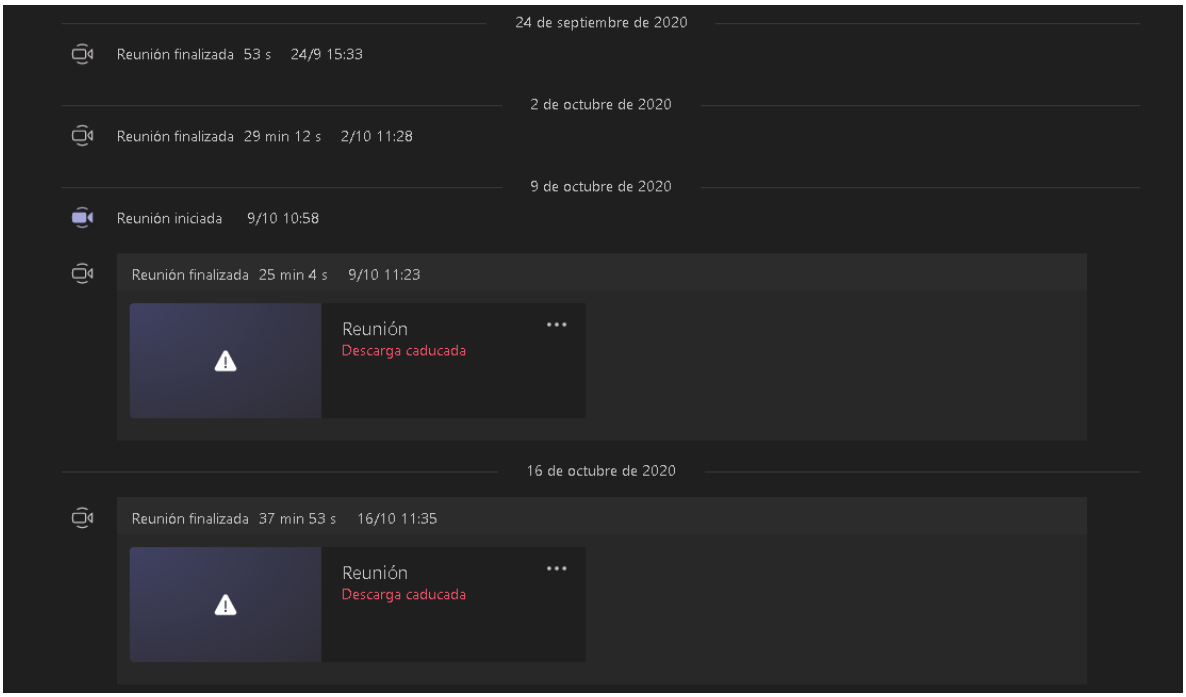


Figura 344. Reuniones virtuales 2
Autores: (Manrique, Celeita, Barahona, 2020)

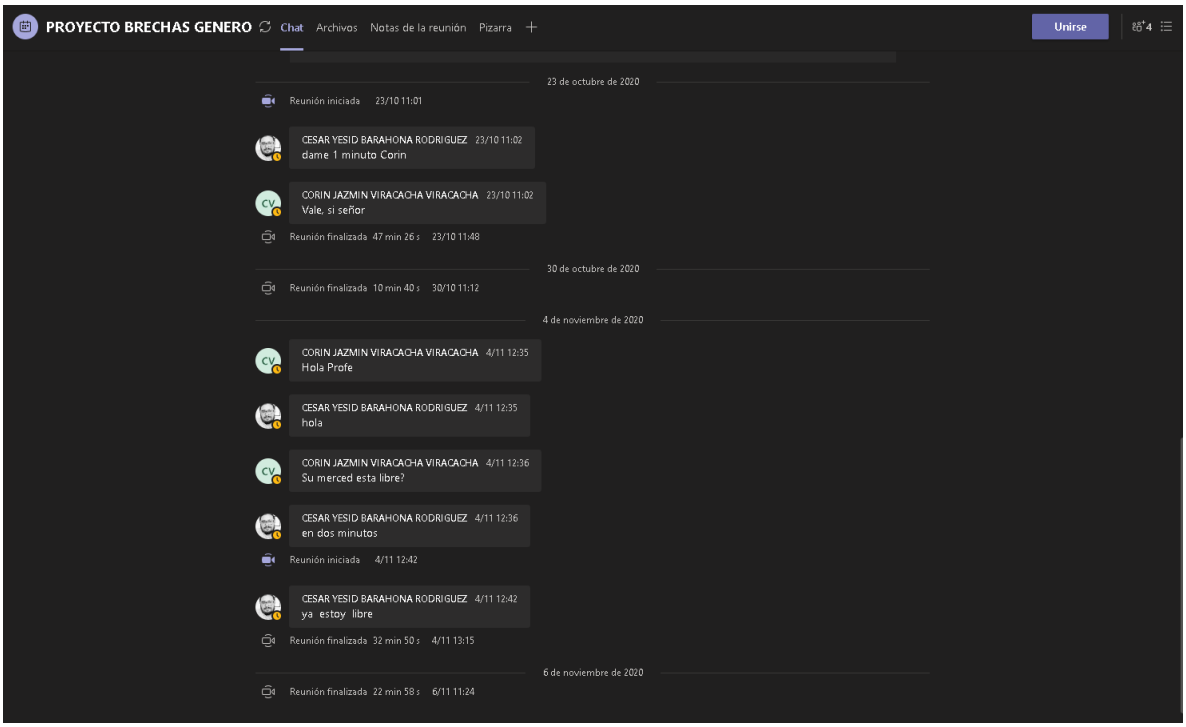


Figura 345. Reuniones virtuales 3
Autores: (Manrique, Celeita, Barahona, 2020)