

**ANÁLISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD DE  
CUNDINAMARCA SEDE FUSAGASUGÁ UTILIZANDO HERRAMIENTAS DE  
INTELIGENCIA DE NEGOCIOS  
(Caso de Uso-Muestra de Datos suministrado: Programa de Ingeniería de Sistemas  
Pensum (Ingeniería de Sistemas 2013) Períodos ( IPA 2013- IPA 2018)).**

ELIZABETH OJEDA MARIN  
SANTIAGO GONZALEZ GOMEZ



UNIVERSIDAD DE CUNDINAMARCA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS  
FUSAGASUGÁ  
2019

**ANÁLISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD DE  
CUNDINAMARCA SEDE FUSAGASUGÁ UTILIZANDO HERRAMIENTAS DE  
INTELIGENCIA DE NEGOCIOS.**

**(Caso de Uso-Muestra de Datos suministrado: Programa de Ingeniería de Sistemas  
Pensum (Ingeniería de Sistemas 2013) Períodos ( IPA 2013- IPA 2018)).**

**ELIZABETH OJEDA MARIN  
SANTIAGO GONZALEZ GOMEZ**

**DOCUMENTO PROYECTO DE GRADO PARA OPTAR EL TÍTULO DE INGENIERO  
DE SISTEMAS**

**DIRECTOR(A):  
ANGELA PATRICIA ARENAS AMADO**

**UNIVERSIDAD DE CUNDINAMARCA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS  
FUSAGASUGÁ**

**2019**

Copyright © 2019 por Elizabeth Ojeda Marín & Santiago González Gómez. Todos los derechos reservados.

## **Agradecimientos**

*Principalmente agradezco a Dios por permitirme terminar este proyecto de la mejor manera porque me dio la inteligencia y la sabiduría suficiente para cada uno de los pasos que se llevaron a cabo en el transcurso, él siempre tuvo el control de todo.*

*Agradezco a mi familia por estar presente en todo el proceso, desde traspasadas en adelante, por sus palabras de aliento, por su amor y por todo el esfuerzo que han hecho para apoyarme.*

*Agradezco a mi compañero Santiago por su apoyo en la ejecución del proyecto.*

*Agradezco a nuestra directora de proyecto, la profesora Ángela que fue una excelente guía y por todo lo que me enseñó, su paciencia y dedicación con su labor de maestra.*

*También agradezco a mis amigos que de una u otra manera siempre fueron un apoyo de conocimiento y moral.*

*Elizabeth Ojeda Marín*

*Le agradezco por sobre todo a mis padres, por brindarme su apoyo incondicional, por sus consejos y enseñanzas, por su amor y su paciencia, por los valores inculcados, por permitirme ser la persona que soy.*

*Agradezco a mi compañera Elizabeth, por ser una persona tan comprometida, tan constante y tan paciente.*

*Agradezco a la profesora Ángela, por su apoyo y sus enseñanzas, por su confianza y fe en este proyecto y sus integrantes.*

*Santiago González Gómez*

## Tabla de Contenido

Abreviaturas	12
Resumen	13
Abstract	15
INTRODUCCIÓN	17
PLANTEAMIENTO DEL PROBLEMA	18
Descripción del problema	18
JUSTIFICACIÓN	19
OBJETIVOS	20
Objetivo General	20
Objetivos Específicos:	20
MARCO REFERENCIAL	21
Universidad de Cundinamarca	21
ANTECEDENTES	24
ESTADO DEL ARTE	27
Deserción Estudiantil Universitaria en Colombia	27
Ministerio de Educación Nacional (MEN)	28
Instituciones de Educación Superior (IES)	28
Sistema Nacional de Información de la Educación Superior (SNIES)	28
Universidad de Cundinamarca Respecto a la Deserción Estudiantil	30
Procesos de seguimiento a la Deserción Estudiantil en la Universidad de Cundinamarca	31
MARCO TEÓRICO	33
Inteligencia de Negocios (Business Intelligence)	33
Arquitectura Inteligencia de Negocios	33
Data Warehouse	35
ETL (Extracción, Transformación y Carga)	35
Bases de Datos OLTP y OLAP	35
OLTP (On-Line Transactional Processing)	35
OLAP (On-Line Analytical Processing)	36
Diferencias entre OLTP y OLAP	36
Análisis OLAP	37
Modelos de Datos Multidimensionales	38
Esquema Estrella	39
Esquema de Copo de Nieve	39

Tablas de Dimensiones y Hechos	40
Tabla de Hechos	40
Tabla de Dimensiones	40
Sistemas ROLAP, MOLAP y HOLAP	41
Sistema MOLAP	41
Sistema ROLAP	41
Sistema HOLAP	41
Data Mart	42
Gestión de la calidad de datos (DQM)	42
MARCO TÉCNICO	43
Cuadrante Mágico de Gartner	43
Cuadro comparativo de Herramientas Tecnológicas	44
Power BI	48
Excel	49
Microsoft Access	50
Cliente Oracle (ODAC 12.2c versión 1 (12.1.0.0.2))	50
Sophos VPN Clients	51
METODOLOGÍA CRISP-DM	52
FASE 1: COMPRENSIÓN DEL NEGOCIO	55
Determinar los Objetivos del Negocio	55
Compilación de la Información de la Universidad	55
Valoración de la Situación	57
Determinar los Objetivos de los resultados	59
Plan del Proyecto	59
FASE 2: COMPRENSIÓN DE LOS DATOS	61
Recolección de los Datos Iniciales	61
Descripción de los datos	64
Exploración de los Datos	64
Verificación de la Calidad de Datos	66
FASE 3: PREPARACIÓN DE LOS DATOS	69
Seleccionar los datos	69
Limpiar los datos	70
Construcción de Nuevos Datos	72
Integración de datos	72
FASE 4 MODELADO	74
Construcción del Data Warehouse	74

Proceso de ETL	74
Construcción Modelo Multidimensional Principal en Power BI	76
Evaluación de los Modelos	78
FASE 5 EVALUACIÓN	84
Evaluación de Resultados	84
Proceso de Revisión	94
Determinar los próximos pasos	94
FASE 6 DESPLIEGUE	95
Cartillas Digitales	95
Capacitación en la Universidad de Cundinamarca	95
Requerimientos para la implementación	95
GLOSARIO	97
ANEXO 1	99
ANEXO 2	100
CONCLUSIONES	105
RECOMENDACIONES	106
REFERENCIAS	107

## **LISTA DE TABLAS**

Tabla 1 Cuadro Comparativo entre OLTP y OLAP	37
Tabla 2 Tabla Comparativa de Herramientas de Inteligencia de Negocios	45
Tabla 3 Plan del Proyecto	60
Tabla 4 Tabla de Hechos y Dimensiones	78



## LISTA DE FIGURAS

Figura 1 Estructura Organizacional. Copyright por Universidad de Cundinamarca	23
Figura 2 Consolidación de la Información de la Educación Superior en Colombia	30
Figura 3 Consulta Universidad de Cundinamarca en Sistema SPADIES.	31
Figura 4 Estadística deserción en la Universidad de Cundinamarca por Periodo Académico	32
Figura 5 Proceso del Estudiante Desertor	33
Figura 6 Arquitectura de una solución mediante la Inteligencia de Negocios	35
Figura 7 ETL	36
Figura 8 Ejemplo Cubo OLAP	39
Figura 9 Cuadrante mágico de Gartner	44
Figura 10 Power BI	49
Figura 11 Excel	50
Figura 12 Instalación del Cliente de Oracle	51
Figura 13 Sophos	52
Figura 14 Vista software Sophos en escritorio de Windows	52
Figura 15 Ciclo de la metodología Crisp- DM	53
Figura 16 . Base de datos Bienestar Universitario	62
Figura 17 Base de datos Académica	63
Figura 18 Informe Admisiones y Registro	63
Figura 19 Acta Facultad de Ingeniería	64
Figura 20 Base de Datos Facultad de Ingeniería	64
Figura 21 Total de la Población	65
Figura 22 Materias Perdidas por Género Masculino	66
Figura 23 Materias Perdidas por Género Femenino	66
Figura 24 Calidad de Datos BD Bienestar	67

Figura 25 Calidad de Datos BD Bienestar	68
Figura 26 Dimensiones de DQ Reflejadas	68
Figura 27 Inconsistencia en BD Académica	69
Figura 28 Estandarización de Datos	71
Figura 29 Limpieza de Datos	72
Figura 30 Inconsistencia de Datos	72
Figura 31 Fusión de Bases de Datos	74
Figura 32 Modelo	75
Figura 33 Extracción de Base de Datos Académica	76
Figura 34 Transformación de los Datos	76
Figura 35 Carga de la Base de Datos Académica	77
Figura 36 Esquema Estrella de la BD principal	77
Figura 37 Esquema Copo de Nieve Base de Datos Académica	78
Figura 38 Seguimiento del Proceso de la UCundinamarca con respecto a la Deserción	80
Figura 39 Total de Estudiantes Desertores	81
Figura 40 Clasificación de Deserciones	81
Figura 41 Deserción Temprana	82
Figura 42 Deserción Tardía	82
Figura 43 Calidad de Datos Suministrados	83
Figura 44 Ubicación Geográfica	83
Figura 45 Posible Deserción del Año 2019	84
Figura 46 Cantidad de Estudiantes Desertores	85
Figura 47 Edad Promedio de los Estudiantes Desertores	86
Figura 48 Deserción Precoz	87
Figura 49 Deserción Temprana	88
Figura 50 Deserción Tardía	89
Figura 51 Deserción Temprana	90
Figura 52 Deserción Tardía	91

Figura 53 Ubicación Geográfica	92
Figura 54 Calidad de Datos Suministrados	92
Figura 55 Seguimiento del Proceso de la UCundinamarca con respecto a la Deserción	93
Figura 56 Posible deserción para el año 2019	94

## **Abreviaturas**

**BI:** Business Intelligence- Inteligencia de Negocios.

**ETL:** Extract, transform and load- Extracción, Transformación y Carga.

**OLAP:** On-Line Analytical Processing- Procesamiento Analítico en Línea.

**OLTP:** OnLine Transaction Processing -Procesamiento de Transacciones En Línea.

**BD:** Base de Datos.

**CRISP-DM:** Cross Industry Standard Process for Data Mining -Proceso Estándar de la Industria para la Minería de Datos.

## Resumen

En el proyecto Análisis de la Deserción Estudiantil en la Universidad de Cundinamarca Sede Fusagasugá utilizando herramientas de Inteligencia de Negocios, se brinda una solución informática por medio de una arquitectura de Inteligencia de Negocios (Business Intelligence). Entre otros aspectos, el análisis considera como variables a estudiar: la tasa de deserción total en el período evaluado, el tipo de deserción, la edad de los desertores, el género de los desertores, el período académico, el lugar de procedencia de los desertores, las materias reprobadas por los desertores, principalmente. En la etapa denominada procesamiento de datos, seleccionamos, limpiamos, transformamos y construimos los datos que nos fueron facilitados por las diferentes áreas/dependencias de la Universidad (Sistemas y Tecnología, Facultad de Ingeniería, Bienestar Universitario, Admisiones y Registro).

La solución brinda apoyo para la toma de decisiones y la generación de estrategias que aporten a disminuir el índice de abandono estudiantil en la institución por medio del análisis de la información.

La metodología utilizada fue CRISP-DM, siendo ésta una de las mejores prácticas reconocidas en la gestión de volumen de datos (IBM, 2012). Dentro de las fases de la metodología partimos de la Comprensión del Negocio en donde se determinaron los objetivos del proyecto identificando el problema principal como la carencia de un componente de análisis dentro de la Institución, que permitiera generar las estrategias necesarias para mitigar el índice de Deserción Estudiantil. Continuamos con la fase de Comprensión de los Datos en la cual se comprobó la calidad de los datos, una vez comprobada la calidad de los datos se procedió con la fase 3. La Preparación de los Datos fase en la cual, teniendo ya la evaluación de calidad a los datos, se procedió a hacer la calidad de datos necesaria para que estos cumplieran con los principios básicos de una base de datos. Se tomaron los datos correspondientes a la muestra de

5 años y finalmente se crearon nuevos parámetros para cumplir con el alcance del análisis.

Entramos a la fase de Modelado en donde se creó un Data Warehouse, o en otras palabras una base de datos compuesta de varias bases de datos (Las que fueron suministradas por las diferentes áreas). Por otra parte, se tomó la base de datos Académica a la que se le realizó un análisis OLAP (On-Line Analytical Processing). Posteriormente entramos en la fase de Evaluación en donde para darle sentido al análisis se plantearon las preguntas que podrían resolver el problema planteado; dichas preguntas son:

- ¿Cuál es la población Desertora en el programa de Ingeniería de Sistemas entre los períodos ( IPA 2013- IPA 2018)?
- ¿De qué manera la Edad es un factor que está involucrado en la población desertora?
- ¿Por qué La Deserción Temprana es el tipo de deserción con más población de estudiantes desertores?
- ¿Por qué la Deserción Tardía es el tipo de deserción con más población de estudiantes desertores?

Una vez dadas las respuestas, avanzamos hacia la sexta y última fase de la metodología. Distribución o Despliegue fase en la cual teniendo el modelo que ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio. Teniendo los resultados del análisis, se creó y se hizo entrega al área de Sistemas y Tecnología (Cliente) un Manual Técnico y un Manual de Usuario, así mismo también se realizó una capacitación al equipo de trabajo del área y se dejó un equipo adaptado para dar continuidad al proyecto.

## **Abstract**

In the Student Dismissal Analysis project at the University of Cundinamarca, Fusagasugá Headquarters, using Business Intelligence tools, a computer solution is provided through a Business Intelligence (Business Intelligence) architecture. Among other aspects, the analysis considers as variables to study: the total dropout rate in the evaluated period, the type of desertion, the age of the deserters, the gender of the deserters, the academic period, the place of origin of the deserters, the subjects disapproved by the deserters, mainly. In the stage called data processing, we select, clean, transform and construct the data that was provided to us by the different areas / dependencies of the University (Systems and Technology, Faculty of Engineering, University Welfare, Admissions and Registration).

The solution provides support for decision making and the generation of strategies that contribute to reducing the student dropout rate in the institution through the analysis of information.

The methodology used was CRISP-DM, which is one of the best recognized practices in data volume management (IBM, 2012). Within the phases of the methodology, we started from the Business Understanding where the project's objectives were identified, identifying the main problem as the lack of an analysis component within the Institution, which would generate the necessary strategies to mitigate the Dropout rate Student. We continue with the Data Comprehension phase in which the quality of the data was checked, once the quality of the data was verified, we proceeded with phase 3. The Preparation of the Data, phase in which, having already the evaluation of data quality, we proceeded to make the data quality necessary for them to comply with the basic principles of a database. The data corresponding to the 5-year sample were taken and finally new parameters were created to comply with the scope of the analysis. We enter the Modeling phase where a Data Warehouse was created, or in other words

a database composed of several databases (Those that were supplied by the different areas). On the other hand, the Academic database was taken and an OLAP analysis (On-Line Analytical Processing) was carried out. Later we entered the Evaluation phase where, to make sense of the analysis, the questions that could solve the problem were raised; These questions are:

- What is the Deserto population in the Systems Engineering program between the periods (IPA 2013- IPA 2018)?
- How is Age a factor that is involved in the deserting population?
- Why is Early Dropout the type of dropout with the highest student dropout population?
- Why is Late Desertion the type of desertion with the most population of drop-out students?

Once the answers are given, we move towards the sixth and final phase of the methodology. Distribution or Deployment phase in which having the model that has been constructed and validated, the knowledge obtained is transformed into actions within the business process. Taking the results of the analysis, a Technical Manual and a User Manual were created and delivered to the Systems and Technology area (Client), as well as training was given to the work team of the area and an adapted computer was left for give continuity to the project.



## INTRODUCCIÓN

La efectividad de un Sistema Educativo se evidencia en la tasa de permanencia de los estudiantes. Aun teniendo en cuenta la complejidad que abarca este índice, consideramos esencial su estudio. Ahora bien, no se trata de únicamente saber las posibles variables que lo causan, sino también de planear y generar estrategias que las mitiguen.

La Deserción Estudiantil, es un factor que ha aumentado a través de los años, por el abandono de los estudiantes en su carrera universitaria, lo cual es causado por distintas razones ya sean económicas, académicas, entre otras. La Educación Superior en Colombia, lleva un control de esta cifra por medio de estadísticas del SPADIES, que es el sistema de información que se encarga de la Deserción en Colombia de las Instituciones de Educación Superior.

La Universidad de Cundinamarca conoce bien la importancia de la Deserción Estudiantil como problemática, y siempre ha tenido la intención de darle un seguimiento y control por medio de planes de permanencia y procesos. Lo que buscamos es apoyar y mejorar dichos procesos, como también proponer nuevas estrategias que aún no estuvieran contempladas.

La Inteligencia de Negocios, un área de conocimiento de la Gestión de Datos, cuenta con las funciones necesarias para realizar un análisis sobre la Deserción Estudiantil, y como muestra para dicho análisis, la población escogida fue el programa de Ingeniería de Sistemas sede Fusagasugá en su plan de estudios 2013 dentro de los periodos (IPA 2013- IPA 2018).

## **PLANTEAMIENTO DEL PROBLEMA**

### **Descripción del problema**

La Deserción Estudiantil en las Instituciones de Educación Superior (IES) se ha presentado desde hace varios años, según estadísticas del Banco Mundial sobre la situación estudiantil en la Educación Superior en Latinoamérica, de los estudiantes que inician su Educación Superior, solo un 50% se gradúan (Dinero, Educación, Deserción 29/06/2017). Lo anterior permite evidenciar la ausencia de seguimiento al tema y la Universidad de Cundinamarca no es la excepción, la carencia de un análisis claro y conciso de la Deserción Estudiantil, las variables de las causas que lo originan son desconocidas y esto impide hacer el respectivo acompañamiento como también tomar decisiones acertadas para controlar este índice. En el área de Sistemas y Tecnología de la Universidad de Cundinamarca se presenta la necesidad de llevar a cabo un análisis basado en datos históricos como un requisito primordial para hacer seguimiento a la Deserción Estudiantil, que permita la toma de decisiones para mitigar esta problemática.

## **JUSTIFICACIÓN**

El presente proyecto de investigación se enfoca en el análisis de los datos existente y disponibles sobre Deserción Estudiantil suministrados por el área de Sistemas y Tecnología de la Universidad de Cundinamarca, Dado que en la actualidad la Universidad no cuenta con un seguimiento a la Deserción Estudiantil se evidencia la necesidad de realizar un análisis basado en Inteligencia de Negocios (BI) para encontrar las posibles variables que lo ocasionan y de esta manera permitir a las áreas/procesos interesadas(os)

-Bienestar Universitario, Sistemas y Tecnología, Autoevaluación y Acreditación, Planeación Institucional, entre otros, tomar decisiones basadas en estrategias que permitan disminuir este índice.

## **OBJETIVOS**

### **Objetivo General**

Implementar una solución para el análisis de la Deserción Estudiantil en la Universidad de Cundinamarca a través de herramientas de la Inteligencia de Negocios, como apoyo para la toma de decisiones y la generación de estrategias que aporten a disminuir el índice de abandono estudiantil en la institución.

### **Objetivos Específicos:**

- Identificar y analizar la situación actual de la Deserción Estudiantil en la Universidad de Cundinamarca.
- Proveer información útil al usuario final, que permita caracterizar los estudiantes en riesgo de deserción a través de una herramienta tecnológica orientada a la Inteligencia de Negocios.

## **MARCO REFERENCIAL**

### **Universidad de Cundinamarca**

#### **Misión**

“La Universidad de Cundinamarca es una institución pública local del Siglo XXI, caracterizada por ser una organización social de conocimiento, democrática, autónoma, formadora, agente de la transmodernidad que incorpora los consensos mundiales de la humanidad y las buenas prácticas de gobernanza universitaria, cuya calidad se genera desde los procesos de enseñanza - aprendizaje, investigación e innovación, e interacción universitaria.” (Universidad de Cundinamarca, 2016).

#### **Visión**

“La Universidad de Cundinamarca será reconocida por la sociedad, en el ámbito local, regional, nacional e internacional, como generadora de conocimiento relevante y pertinente, centrada en el cuidado de la vida, la naturaleza, el ambiente, la humanidad y la convivencia.” (Universidad de Cundinamarca, 2016).

#### **Objetivos de la Universidad de Cundinamarca**

1. Formar profesionales integrales, que requiere el Departamento de Cundinamarca y el país, con capacidades académicas, científicas y humanas, formados en los aspectos fundamentales del conocimiento posmoderno y coherentes con el contexto socioeconómico nacional y global.
2. Promover el conocimiento y la reafirmación de los valores humanos, la cultura, la incorporación integral de los colombianos a los beneficios que de ella se deriven.
3. Contribuir al desarrollo de las tecnologías para la protección y el aprovechamiento de los recursos naturales, la preservación de la vida en el planeta, la sostenibilidad con la naturaleza y el medio ambiente.

4. Fomentar la formación profesional, en el marco de la cultura, la ciencia, tecnología e innovación, en el ámbito superior y al servicio de la comunidad a través de la interacción social orientada al desarrollo que requiere el departamento y el país.
5. Adelantar programas que respondan a los intereses de las poblaciones vulnerables a nivel urbano y rural del Departamento de Cundinamarca.
6. Fomentar la formación integral sobre las bases científicas, éticas y humanísticas, para la ocupación del trabajo, y el libre desarrollo de la personalidad, para que se integren con eficiencia, y responsabilidad a las dimensiones axiológicas, profesionales, científicas. Así mismo, propender por una educación para la vida que favorezca los valores democráticos, la civilidad, y la libertad.
7. Evaluar de manera continua el quehacer institucional, tomando como referencia lo consagrado en el Proyecto Educativo Institucional y su Plan de Desarrollo.
8. Consolidar la presencia institucional en el contexto científico y académico nacional e internacional, a través de resultados producto de competencias y habilidades de la comunidad universitaria UdeC.
9. Propiciar la integración de la Universidad con los diferentes sectores sociales del orden local, departamental y nacional, para ser factor de desarrollo social, económico, y político.
10. Contribuir al desarrollo de los niveles educativos que le preceden, para facilitar su articulación y el logro de sus correspondientes fines.

(Universidad de Cundinamarca, 2016).

## Estructura Organizacional

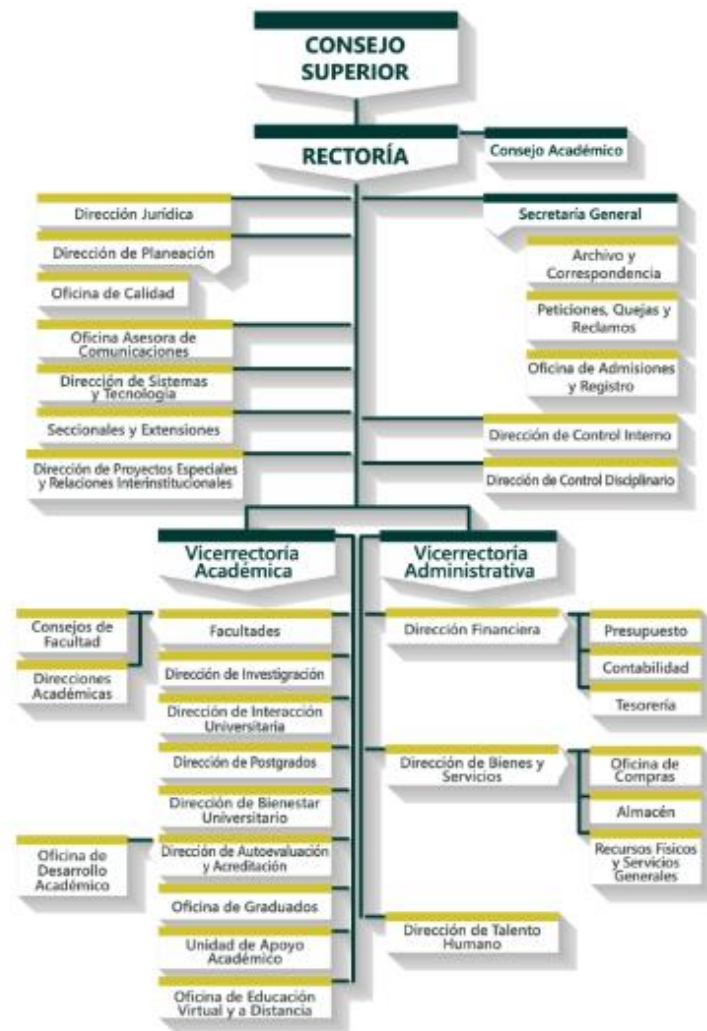


Figura 1 Estructura Organizacional. Copyright por Universidad de Cundinamarca

## ANTECEDENTES

Una de las principales problemáticas que presenta la Educación Superior en Colombia y que se ha caracterizado en los últimos años es la Deserción Estudiantil. Según estadísticas del Ministerio de Educación Nacional, de cada cien estudiantes que ingresan a una institución de educación superior cerca de la mitad no logra culminar su ciclo académico y obtener la graduación (Ministerio de Educación Nacional, 2009).

Teniendo en cuenta los altos índices de deserción que se presentan en la Educación Superior, los entes de control educativo no solo en Colombia sino en la mayoría de países, han adoptado por soluciones basadas en las tecnologías de la información con el objetivo de identificar los factores que afectan la deserción, obteniendo una proyección eficiente, adecuada y coherente de los desertores, así como también un análisis del rendimiento académico, permitiendo con esto mejorar la planificación y definición de estrategias que logren mitigar las altas tasas de Deserción Estudiantil. (Ministerio de Educación Nacional de la República de Colombia, 2009).

La Inteligencia de negocios juega un papel fundamental para la solución de los problemas mencionados anteriormente, debido a que permite orientar los procesos, tecnologías y herramientas para convertir los datos en información y estos en conocimiento con el objetivo de mejorar la toma de decisiones en pro de mejorar la eficiencia y calidad de los procesos de negocio, en este caso de la Universidad de Cundinamarca como Institución de Educación Superior.

A continuación, algunos proyectos que ya se han realizado acerca del tema:

**-Análisis de la Deserción Estudiantil de CECAR utilizando Herramientas de Inteligencia de Negocios con Licencia Libre**



Este proyecto desarrolla un prototipo funcional de la inteligencia de negocios utilizando herramientas de software libre para el análisis y proyección de la deserción de estudiantes en la Corporación Universitaria del Caribe -CECAR.

La solución planteada en este proyecto incluye:

- Desarrollo de un modelo de Data Mart análisis OLAP.
- Integración de datos mediante herramienta ETL.

(Bohorquez, 2016)

### **-Modelo para la Automatización del Proceso de Determinación de Riesgo de Deserción en Estudiantes Universitarios.**

El objetivo de este proyecto consiste en investigar y proponer una metodología que permita identificar en forma automática a los estudiantes con mayor riesgo de deserción de las carreras de Ingeniería de la Universidad de Las Américas.

La solución planteada para este proyecto incluye:

- Analítica de datos.
- Minería de datos mediante metodología CRISP-DM.

(Angulo, 2012)

### **La utilización de Business Intelligence como propuesta para mejorar los indicadores de deserción de los estudiantes de la Escuela de Ingeniería Informática.**

El presente trabajo tiene como finalidad proponer herramientas y una metodología de Business Intelligence (BI), y los diversos beneficios y ventajas de su implementación dentro de la educación universitaria, tomando como caso de estudio el problema de la deserción de los estudiantes de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso (PUCV).

La solución planteada para este proyecto incluye:

- Desarrollo de un modelo de data mart análisis OLAP.

- Integración de datos mediante herramienta ETL.
- Analítica de datos.

(Cerde, 2017)

## **ESTADO DEL ARTE**

Consideramos importante mencionar cuales son los Sistemas existentes que llevan el control del índice de Deserción Estudiantil en Colombia, como también cual es la relación entre ellos y como dependen el uno del otro.

Si bien es cierto que estos Sistemas recolectan la estadística de Deserción Estudiantil por cohorte, no se encuentra información detallada de las razones que ocasionan dicha estadística, no es más que una cifra que les permite predecir la cantidad de desertores para las siguientes cohortes. Por lo tanto, resaltamos una vez más la importancia del estudio detallado de este índice, no solo para corroborar que las estadísticas de estos sistemas coinciden, sino para proponer soluciones a este problema.

### **Deserción Estudiantil Universitaria en Colombia**

La Deserción Estudiantil en la Educación Superior cubre cifras alarmantes, según estadísticas del MEN (Ministerio de Educación Nacional, 2009): “de cada cien estudiantes que ingresan a una institución de educación superior, cerca de la mitad no logra culminar su ciclo académico y obtener la graduación”. La tasa semestral, en promedio, está entre el 12% y 13%.

Dentro de las causas de la elevada tasa de abandono se encuentran, entre otros:

- La falta de preparación académica (en parte, debido a la educación de baja calidad que reciben en la escuela secundaria).
- La inconformidad por la educación de baja calidad (docentes sin la adecuada preparación).
- La falta de medios económicos.
- La duración de los programas y la inflexibilidad para cambiar de carrera.

(Universidad de Cundinamarca, 2018)

## **Ministerio de Educación Nacional (MEN)**

El MEN tiene como objetivo, “la operación del Sistema de Aseguramiento de la Calidad de la Educación Superior, la pertinencia de los programas, la eficiencia y transparencia. Se encarga de consolidar y armonizar las etapas y procesos que generan cultura para así dinamizar e interrelacionar los resultados que emiten esos organismos.

Se conforma por las instituciones de Educación Superior, los docentes, los estudiantes, personal administrativo que laboran en ellas, facultades, pares académicos, científicos que hacen aportes al área educativa, entre otros” (Ministerio de Educación, 2019)

El MEN cuenta con distintos sistemas de información para analizar cada variable que se presenta en los procesos de la Educación Superior.

### **Instituciones de Educación Superior (IES)**

Las Instituciones de Educación Superior, son las entidades que cuentan, con arreglo a las normas legales, con el reconocimiento oficial como prestadoras del servicio público de la educación superior en el territorio colombiano (Ministerio de Educación Nacional, 2010).

Se encarga de clasificar según su modalidad de formación y carácter académico. La IES es la base de datos que alimenta los sistemas de información del MEN.

### **Sistema Nacional de Información de la Educación Superior (SNIES)**

El Sistema Nacional de Información de la Educación Superior (SNIES), es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia.

En este sistema se recopila y organiza la información relevante sobre la Educación Superior que permite hacer planeación, monitoreo, evaluación, asesoría, inspección y vigilancia del sector.

Este sistema como fuente de información, en relación con las instituciones y programas académicos aprobados por el Ministerio de Educación Nacional, consolida y suministra datos, estadísticas e indicadores relevantes, confiables y oportunos.

El SNIES pone a disposición de las Instituciones de Educación Superior, agencias del sector, entes de gobierno y la comunidad en general información que:

- Facilita la gestión, la planeación y toma de decisiones.
- Orienta a las IES en los procesos de mejoramiento a partir de la identificación de mejores prácticas.
- Ayuda a la autorregulación del sector.
- Simplifica el proceso de reporte de información.
- Sirve como marco de referencia.

(Ministerio de Educación, 2019)

A continuación el proceso de la consolidación de la información de la Educación Superior en Colombia



Figura 2 Consolidación de la Información de la Educación Superior en Colombia Copyright IES

## Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES)

Este sistema es la herramienta para hacer seguimiento sobre las cifras de deserción de estudiantes de la educación superior. Con los datos suministrados por las Instituciones de Educación Superior a SPADIES, se identifican y se ponderan los comportamientos, las causas, variables y riesgos determinantes para desertar. Además, con esta información se agrupan los estudiantes de acuerdo con su riesgo de deserción.

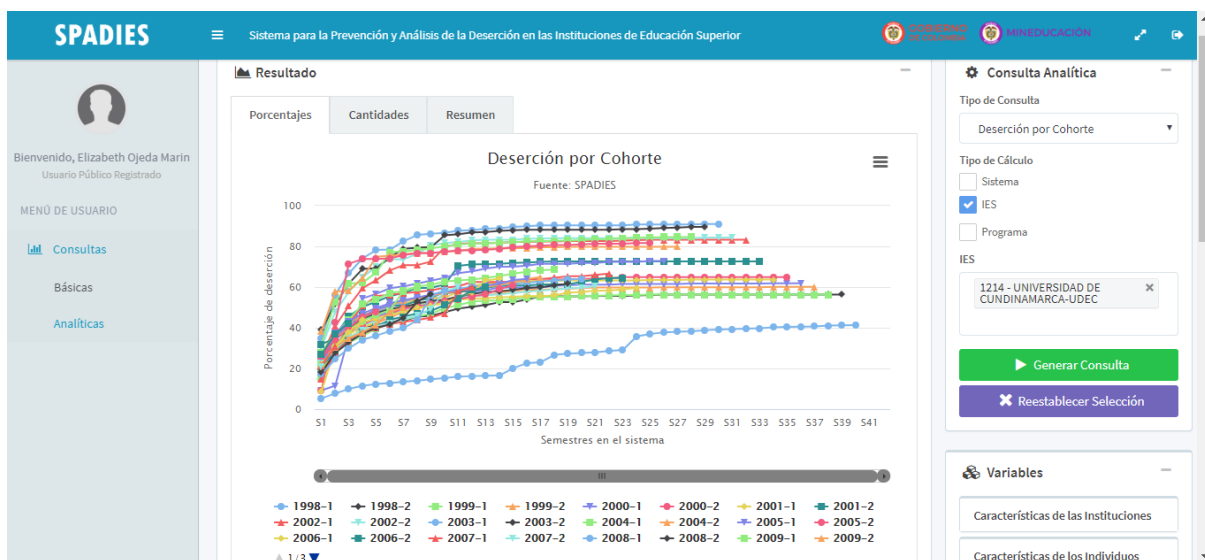


Figura 3 Consulta Universidad de Cundinamarca en Sistema SPADIES.

SPADIES permite acceder a las estadísticas de una manera restringida con un usuario público, en la figura anterior se observa el análisis por cohorte a grandes rasgos a nivel de todos los programas académicos de la Universidad de Cundinamarca.

## Universidad de Cundinamarca Respecto a la Deserción Estudiantil

La Universidad de Cundinamarca no es la excepción en cuanto a Deserción Estudiantil, sufre este problema como el resto de las universidades, sin embargo, según cifras del MEN la deserción que se presenta en la universidad es “baja”, pues en los últimos años se ha logrado

mantener el índice de deserción alrededor del 10%, el cual, frente a todo el sistema universitario del país, es bajo.



Figura 4 Estadística deserción en la Universidad de Cundinamarca por Periodo Académico Copyright UCundinamarca

“En términos generales, los datos reflejan un promedio bajo de deserción en la Universidad de Cundinamarca como resultado de las políticas y estrategias educativas misionales de permanencia que incentiva la Institución como los programas de exoneraciones (becas) por cultura, deporte, medallería, restaurante, y hogar universitario, así como el valor de la matrícula ajustado a un salario mínimo legal vigente en todos los programas académicos de pregrado que oferta la Universidad.” (Universidad de Cundinamarca, 2016)

La Universidad de Cundinamarca establece la deserción estudiantil dividida en 3 tipos:

- Deserción Precoz: La población que conforma esta deserción son las personas que son admitidos en la universidad, pero no cancelan su matrícula.
- Deserción Temprana: La población que conforma esta deserción son los estudiantes que se retiran de sus estudios en los primeros 5 semestres.
- Deserción Tardía: Esta población la conforman los estudiantes que deciden dejar sus estudios después de haber transcurrido la mitad de su carrera profesional.

## Procesos de seguimiento a la Deserción Estudiantil en la Universidad de Cundinamarca

En la actualidad, la Universidad de Cundinamarca cuenta con un proceso para que los estudiantes tengan la opción de retirarse, aplazar, o cancelar el semestre. Dicho proceso consiste en hacer la solicitud por medio de la plataforma institucional, una vez hecha dicha solicitud, el estudiante debe llenar un formato dirigiéndose en primera instancia a Bienestar Universitario donde es atendido por un psicólogo(a), quien entrevista al estudiante y analiza las razones que el estudiante argumenta y por las cuales realizó la solicitud. Una vez terminada la entrevista si no se logra que el estudiante “cambie de opinión”, el estudiante debe dirigirse a su respectiva Facultad para formalizar la solicitud, la cual será evaluada por el Consejo de Facultad, estamento que determina si la solicitud es válida para ser aprobada (de acuerdo a los estándares de la universidad).

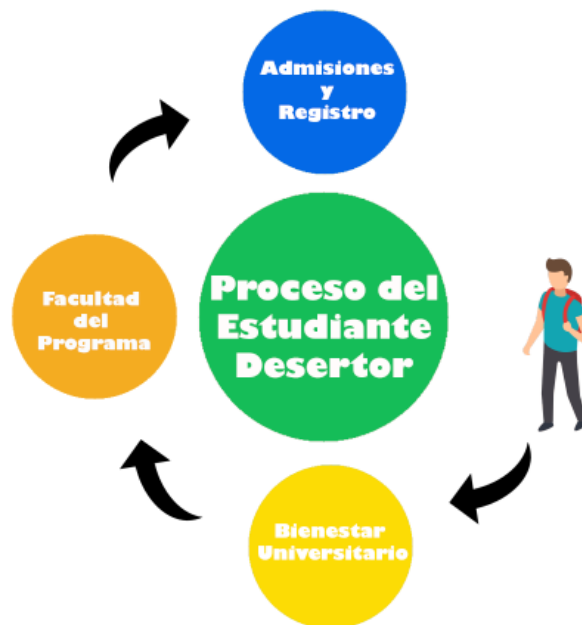


Figura 5 Proceso del Estudiante Desertor Creación Propia.



## **MARCO TEÓRICO**

### **Inteligencia de Negocios (Business Intelligence)**

El contexto de la sociedad de la información ha propiciado la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuida a lo largo de la cadena de valor.

La inteligencia de negocios (Business Intelligence) responde a dicha necesidad, y podemos entender, en una primera aproximación, que es una evolución de los Sistemas de Soporte a las Decisiones (DSS, Decissions Support Systems).

Se entiende por Inteligencia de Negocios al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización (J. Díaz, Introducción al Business Intelligence, 2010).

### **Arquitectura Inteligencia de Negocios**

Una solución de Business Intelligence comienza con los sistemas origen de la organización (bases de datos, ERPs, ficheros de texto, ...), a los cuales es necesario realizar una transformación estructural para optimizar su proceso analítico.

Para ello se realiza una fase de extracción, transformación y carga (ETL) de datos.

La información resultante, ya unificada, depurada y consolidada, se almacena en un Datawarehouse corporativo, que puede servir como base para la construcción de distintos Datamarts departamentales. Estos datamarts se caracterizan por poseer la estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante bases de datos transaccionales (OLTP) o mediante bases de datos analíticas (OLAP).

Los datos albergados en el datawarehouse o en cada datamart se explotan utilizando herramientas comerciales de análisis, reporting, alertas... etc. (S.L, Sinergia e Inteligencia de Negocio, 2007).

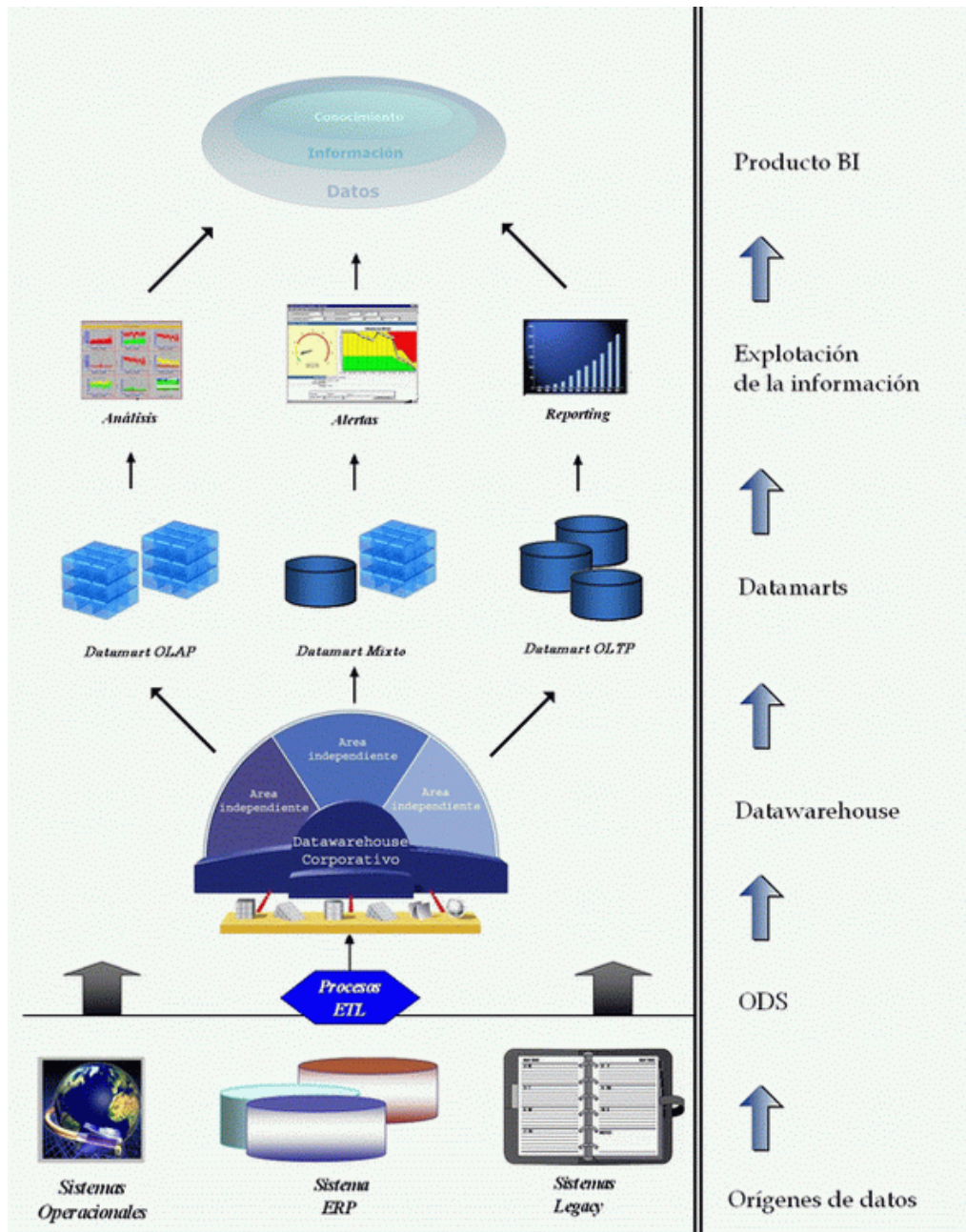


Figura 6 Arquitectura de una solución mediante la Inteligencia de Negocios Copyright Sinergia

## Data Warehouse

Un Data Warehouse es un almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información. Los datos de un data warehouse deben almacenarse de forma segura, fiable, fácil de recuperar y fácil de administrar. (PowerData Especialista en Gestión de Datos, 2013).

## ETL (Extracción, Transformación y Carga)

ETL es un tipo de integración de datos que hace referencia a los tres pasos (extraer, transformar, cargar) que se utilizan para mezclar datos de múltiples fuentes. Se utiliza a menudo para construir un almacén de datos. Durante este proceso, los datos se toman (extraen) de un sistema origen, se convierten (transforman) en un formato que se puede almacenar y finalmente se cargan en un data warehouse u otro sistema.

(SAS, Statistical Analysis System, 2019)

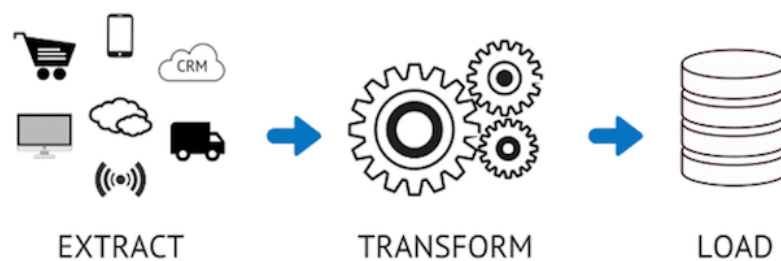


Figura 7 ETL Copyright SAS

## Bases de Datos OLTP y OLAP

### OLTP (On-Line Transactional Processing)

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico, que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es

típico de las bases de datos operacionales. (S.L, Sinergia e Inteligencia de Negocio, 2016)

### **OLAP (On-Line Analytical Processing)**

OLAP (online analytical processing) es un método computacional que permite a los usuarios de manera fácil y selectiva, consultar y extraer datos para analizarlos desde diferentes puntos de vista. Las consultas de Inteligencia de Negocios de OLAP a menudo ayudan en el análisis de tendencias, reportes financieros, previsión de ventas, presupuesto y otros propósitos de planeación.

### **Diferencias entre OLTP y OLAP**

Tabla 1 Cuadro Comparativo entre OLTP y OLAP Creación Propia

	<b>OLTP</b>	<b>OLAP</b>
<b>Objetivo</b>	<b>Soportar actividades transaccionales diarias.</b>	<b>Consultar y analizar información estratégica y táctica.</b>
<b>Tipo de datos</b>	<b>Operacionales.</b>	<b>Para la toma de decisiones.</b>
<b>Datos</b>	<b>Actuales.</b>	<b>Actuales e históricos.</b>

<b>consultados</b>		
<b>Consultas</b>	<b>Consultas simples.</b>	<b>Consultas complejas.</b>
<b>Normalización</b>	<b>Las tablas en una base de datos OLTP están normalizadas (3FN)</b>	<b>Las tablas en una base de datos OLAP no están normalizadas.</b>
<b>Estructura</b>	<b>Generalmente estable.</b>	<b>Puede variar de acuerdo a su propia evolución y utilización.</b>

## **Análisis OLAP**

Un análisis OLAP consiste en la recopilación de datos, los cuales posteriormente se limpian y organizan en cubos de datos. Cada cubo OLAP contiene datos categorizados por dimensiones (como clientes, región geográfica de ventas y periodo de tiempo) derivados de tablas dimensionales en los almacenes de datos. Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Por medio de este proceso se pueden realizar 5 tipos de análisis OLAP:

- Enrollar. También conocida como consolidación o profundización, esta operación resume los datos a lo largo de la dimensión.
- Profundizar. Esto permite que los analistas naveguen más profundamente entre las dimensiones de los datos, por ejemplo, profundizando de "período de tiempo" a "años" y "meses" para registrar el crecimiento de las ventas de un producto.

- Rebanada. Esto permite que un analista tome un nivel de información para mostrar, por ejemplo "ventas en 2017".
- Dado. Esto le permite a un analista seleccionar datos de múltiples dimensiones para analizar, ejemplo: "ventas de pelotas de playa azul en Iowa en 2017".
- Pivote. Los analistas pueden obtener una nueva vista de los datos girando los ejes de datos del cubo.

(Rouse Margaret, TechTarget, 2007).

## Modelos de Datos Multidimensionales

Una base de datos multidimensional (MDB) es un tipo de base de datos que se ha optimizado para data warehouse y aplicaciones de procesamiento analítico en línea (OLAP). Las bases de datos multidimensionales se crean con frecuencia usando entradas de las bases de datos relacionales existentes. A una aplicación OLAP que accede a los datos desde una base de datos multidimensional se le conoce como una aplicación MOLAP (OLAP multidimensional).

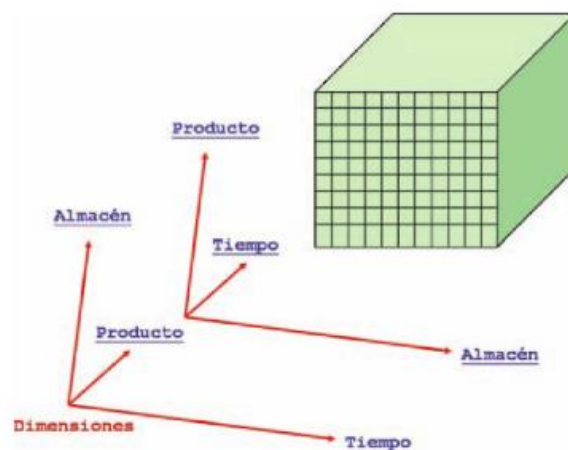


Figura 8 Ejemplo Cubo OLAP

## **Esquema Estrella**

Un esquema estrella consiste en una tabla central (tabla de hechos) y un conjunto de tablas de dimensión. Las características principales de este tipo de esquema son:

- Cada tabla de dimensión contiene una clave primaria no compuesta.
- Cada una de las columnas de la tabla de hechos que forma parte de la clave primaria compuesta es una clave ajena que referencia a la clave primaria de cada dimensión. Es decir, la clave primaria de la tabla de hechos está compuesta por claves ajenas a las tablas de dimensiones. Por lo tanto, la tabla de hechos tiene una relación muchos a muchos con las tablas de dimensiones.

## **Esquema de Copo de Nieve**

El esquema de copo de nieve se diferencia del esquema estrella en que alguna de las dimensiones se normaliza según los niveles de jerarquía. En la tabla dimensión se encuentran los valores del mínimo nivel de jerarquía.

Las consultas de atributo simple son más rápidas en este tipo de esquema ya que se recorren las tablas con menor número de instancias. Las características principales de este esquema son:

- Posee una mayor complejidad de estructura.
- Debido a la normalización de las tablas implica que se requiere menos esfuerzo de diseño.
- Permite realizar clases de jerarquías en donde se podrá analizar los datos más generales a los más detallados.

## **Tablas de Dimensiones y Hechos**

### **Tabla de Hechos**

En el corazón de un esquema en estrella hay una tabla de hechos, que contiene entradas de datos que comprenden un conjunto de hechos relacionados con las operaciones comerciales de una compañía. Cada fila en una tabla de hechos representa una transacción o evento individual; las columnas documentan los diferentes elementos de datos que entran en juego al procesar los capturados en la tabla. Salvador Ramos (2012) los describe así:

“Los Hechos están compuestos por los detalles del proceso de negocio a analizar, contienen datos numéricos y medidas (métricas) de Negocio a analizar. Contienen también elementos (claves externas) para contextualizar dichas medidas, como por ejemplo el producto, la fecha, el cliente, la cuenta contable, etc.”

### **Tabla de Dimensiones**

En una tabla de hechos, las entradas en los campos de datos de la entidad no son los datos reales; en su lugar, son claves externas que apuntan a las claves primarias para entradas relacionadas en tablas de dimensiones, que capturan una variedad de información sobre cada entidad a la que se hace referencia en la tabla de hechos. Una tabla de dimensiones nos permite contextualizar los hechos, agregando diferentes perspectivas de análisis a ellos, (Ramos Salvador, 2012). Una tabla de dimensiones proporciona una forma uniforme de mantener una versión actualizada de los datos asociados con esas entidades.

(Loshi David, TechTarget, 2018).



## **Sistemas ROLAP, MOLAP y HOLAP**

### **Sistema MOLAP**

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

El sistema MOLAP utiliza una arquitectura de dos niveles: las bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato. El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP.

### **Sistema ROLAP**

La arquitectura ROLAP, accede a los datos almacenados en un Datawarehouse para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, el motor ROLAP proporciona la funcionalidad analítica, y el nivel de aplicación que permite la interacción del usuario con el sistema ROLAP. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

### **Sistema HOLAP**

Este sistema es una solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle

(los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

(S.L., Sinergia e Inteligencia de Negocio, 2018)

### **Data Mart**

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. (S.L., Sinergia e Inteligencia de Negocio, 2016)

### **Gestión de la calidad de datos (DQM)**

La gestión de la calidad de datos (DQM) es un proceso de apoyo crítico en la gestión del cambio organizacional. Es sinónimo de calidad de la información ya que los malos resultados de calidad de datos conllevan a información inexacta y bajo rendimiento del negocio.

DQM es también un proceso continuo para definir los parámetros para especificar los niveles aceptables de calidad de datos para satisfacer las necesidades del negocio y de garantizar que la calidad de datos cumple con estos niveles. DQM implica analizar la calidad de los datos, la identificación de anomalías en los datos y la definición de los requerimientos del negocio y las correspondientes reglas de negocio para reglamentar la calidad de los datos requeridos.

(DAMA, Guía de fundamentos para la gestión de datos, 2009)

# MARCO TÉCNICO

## Cuadrante Mágico de Gartner

El Cuadrante Mágico de Gartner es una representación gráfica y sencilla de la situación actual en el mercado de las herramientas de Inteligencia de Negocios, se divide en 4 cuadrantes (Líderes, Retadores o Aspirantes, Visionarios y Jugadores de Nicho).

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2019)

Figura 9 Cuadrante mágico de Gartner

Durante 12 años consecutivos, Gartner ha reconocido a Microsoft como líder en analítica e inteligencia empresarial. Investigaciones en ventas y experiencias de usuario contribuyeron a posicionar a la herramienta en un alto rango.

Gartner ofrece un informe sobre cada una de las herramientas expuestas en el cuadrante mágico, con intención de mostrar el mercado de Inteligencia de Negocios destacando sus fortalezas y debilidades. (Gartner, 2019)

### Cuadro comparativo de Herramientas Tecnológicas

Tabla 2 Tabla Comparativa de Herramientas de Inteligencia de Negocios Creación Propia

Herramienta	Descripción	Ventajas	Costos
Power BI	Herramienta de Microsoft con servicio en la nube con fácil implementación; que permite subir, compartir y tener acceso a informes desde cualquier dispositivo, ya sea un ordenador, una tablet o un smartphone. La integración de Power	*Incremento de la eficiencia en las compañías para extraer informes de manera autónoma cuando lo requieran sin tener un conocimiento específico de la misma. *Power BI presenta una herramienta Quick Insights que determina las	*Power BI Desktop Free *Power BI Pro \$9.99/mes/usuario *Power BI Premium \$ 4,995 / nodo / mes. Aprox. \$ 24,975 / mes para 5000 usuarios con 3 nodos (el costo de Power BI Premium será de \$ 14,895 + \$ 9,900 para 100 usuarios de BI pro). Las licencias de Power BI

	<p>BI con Microsoft Office 365 también permite el acceso a orígenes de datos locales, orígenes de bases de datos y servicios en la nube.</p>	<p>correlaciones y patrones dentro de sus datos, produciendo gráficos y gráficos personalizados.</p> <p>*Integración del análisis avanzado a través de scripts y objetos visuales de R, Microsoft Azure Machine Learning y Azure Stream Analytics.</p>	<p>pro se deben comprar para la suscripción premium de Power BI para contenido autoría.</p>
Qlik	<p>Qlik es una plataforma enfocada al análisis visual de datos y aplicaciones interactivas que tiene por objetivo mejorar el proceso de acceso a los datos de cara al usuario. Como, por</p>	<p>*Los conocimientos de datos se pueden generar rápidamente a partir de un usuario competente, ya que las capacidades de análisis de datos están limitadas sólo por sus capacidades</p>	<p>*QlikView Client License</p> <p>Edición Personal: Gratis</p> <ul style="list-style-type: none"> <li>• Licencia de usuario nombrado: \$ 1,350 / usuario nombrado</li> <li>• Licencia de documento: \$ 350 / usuario nombrado / documento</li> </ul>

	<p>ejemplo, acceder a ciertas visualizaciones ‘limpias’ y fáciles de comprender, diseños de gráficos llamativos, entre otros.</p>	<p>de creación de scripts. *El motor asociativo subyacente realiza uniones naturales en tiempo real en función de las selecciones del usuario, destaca las relaciones entre las entidades para el usuario.</p>	<ul style="list-style-type: none"> <li>• Licencia concurrente: \$ 15,000 / usuario concurrente</li> <li>*QlikView Server License Pricing</li> <li>• Servidor Enterprise Edition: \$ 35,000 / servidor</li> <li>• Small Business Edition Server (hasta 25 nombres usuarios): \$ 8,400 / servidor</li> <li>• Servidores de prueba: 50% del costo de la licencia del servidor.</li> <li>• QlikView Publisher: \$ 21,000 / servidor</li> <li>• Servicio de distribución de informes en PDF: \$ 21,000 / servidor</li> <li>• Conector a SAP NetWeaver®: \$ 22,500 / servidor</li> </ul>
--	---	--	--

			<ul style="list-style-type: none"> <li>• Conector a Salesforce.com: Gratis</li> <li>• Conector a Informatica: Gratis</li> </ul>
Microstrateg y	MicroStrategy es una compañía que ofrece software OLAP, una solución utilizada en el campo de la llamada Inteligencia de Negocio (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes volúmenes de datos.	*Maneja cubos virtuales conectándose directamente a la base de datos *Es realmente una plataforma BI y no conglomerado de productos.	*Licencia local (arquitectura): \$ 5000 / nombrada uso / año • Basado en la web: \$ 600 / usuario nombrado / año, o \$ 300 k por Núcleo de la CPU • Basado en dispositivos móviles: \$ 600 / usuario nombrado / año, o \$ 300k por núcleo de CPU • Servidor: \$ 1200 / usuario nombrado, o \$ 600k por CPU núcleo

(Worldwide, 2019)

## **Power BI**

Power BI es una solución de análisis empresarial que permite visualizar los datos y compartir información con toda la organización, o insertarla en una aplicación o sitio web. Conectarse a cientos de orígenes de datos y dar vida a los datos con paneles e informes dinámicos.

Existen 3 tipos :

- Creadores de contenido

Los creadores de contenido crean visualizaciones de datos e informes de autores. Usan la herramienta gratuita Power BI Desktop para preparar, modelar y crear visualizaciones de datos.

- Editores de contenido

Los editores de contenido requieren colaboración, modelado de datos, creación de contenido, uso compartido de paneles, análisis ad hoc y publicación de informes. Estos usuarios requieren una licencia de Power BI Pro.

- Consumidores de contenido

Los consumidores de contenido utilizan paneles e informes publicados previamente. Hay dos opciones para dar a los consumidores acceso al contenido: licenciar a todos en Power BI Pro o licencia Power BI Premium para usuarios a gran escala.

(Microsoft, 2019)



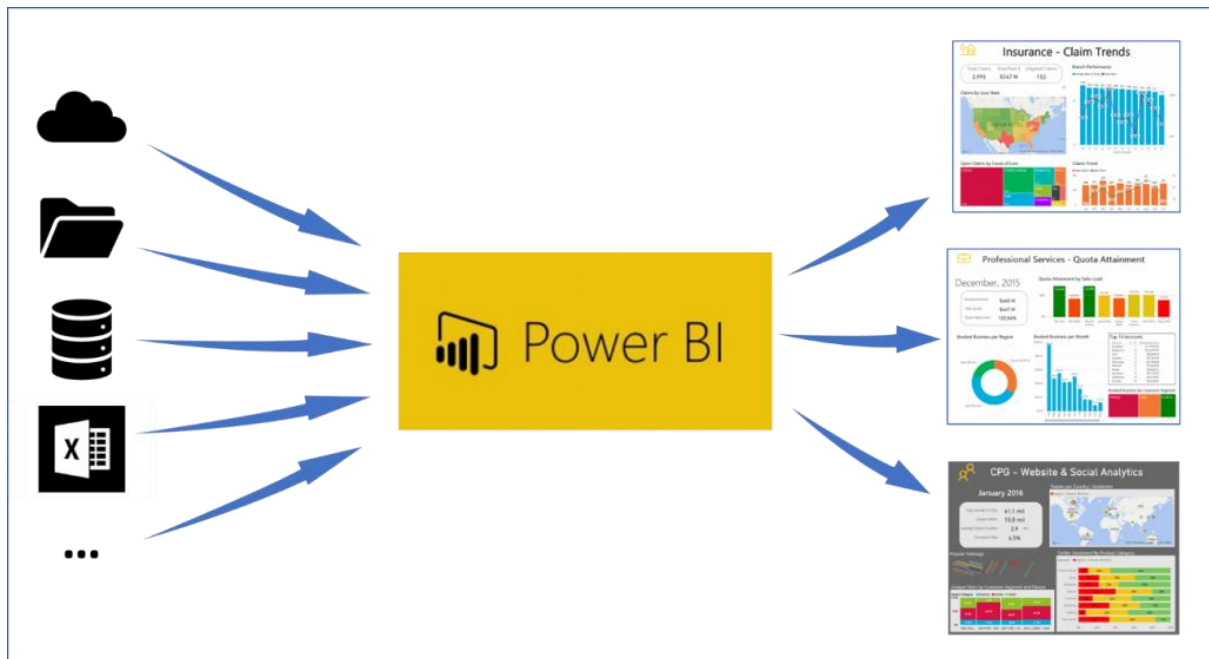


Figura 10 Power BI Copyright Microsoft Power BI

## Excel

Microsoft Excel, MS Excel o simplemente Excel es un software de aplicación publicado por la empresa Microsoft, que brinda soporte digital a las labores contables, financieras, organizativas y de programación, mediante hojas de cálculo. Forma parte del paquete Microsoft Office, que contiene diversos programas de oficina como Microsoft Word, Microsoft Powerpoint, etc.

Excel es posiblemente el programa más popular de su área en el mundo entero, publicado en diferentes versiones y actualizaciones desde su primerísima aparición en 1985. Entre su oferta de capacidades está la gestión de tablas, formatos y fórmulas matemáticas, así como funciones pre programadas, y su correspondiente conversión a gráficos de diversa naturaleza. (Concepto Excel, 2018)



Figura 11 Excel Copyright Microsoft

### **Microsoft Access**

Microsoft Access es un sistema de gestión de base de datos desarrollado por Microsoft y componente de la suite Microsoft Office. Es útil para la creación de pequeñas bases de datos para el hogar y pequeñas empresas. No es multiplataforma, pues sólo funciona en Windows.

Microsoft Access almacena datos en su propio formato basado en el Motor de base de datos de Access Jet. También puede importar o vincular directamente a datos almacenados en otras aplicaciones y bases de datos. Los desarrolladores de software, los arquitectos de datos y los usuarios avanzados pueden usar Microsoft Access para desarrollar software de aplicaciones. (Office, 2019)

### **Cliente Oracle (ODAC 12.2c versión 1 (12.1.0.0.2))**

Componente de acceso a la base de datos de Oracle como cliente.

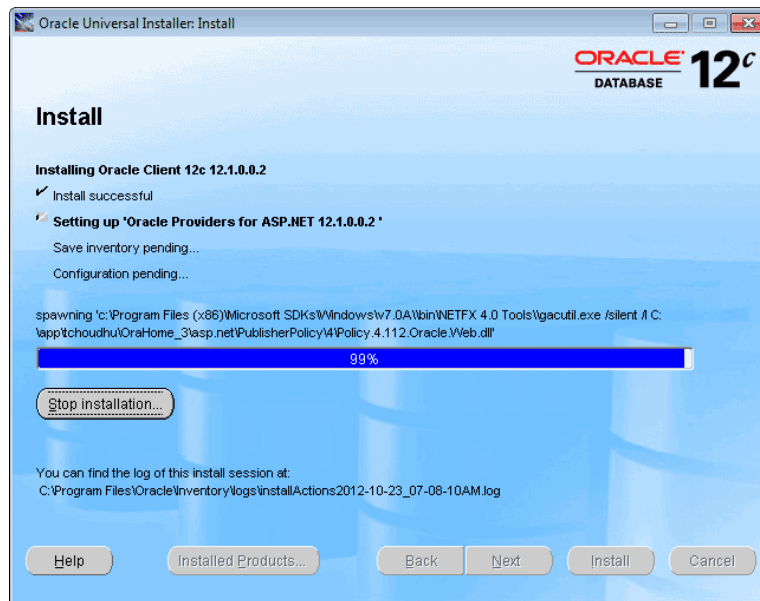


Figura 12 Instalación del Cliente de Oracle Creación Propia

## Sophos VPN Clients

Proporciona acceso remoto transparente y fácil de usar para todos en los procesos de la empresa. Esto se proporciona a través de SSL seguro o VPN IPsec de Windows, Linux, Mac OS y sistemas basados en UNIX.

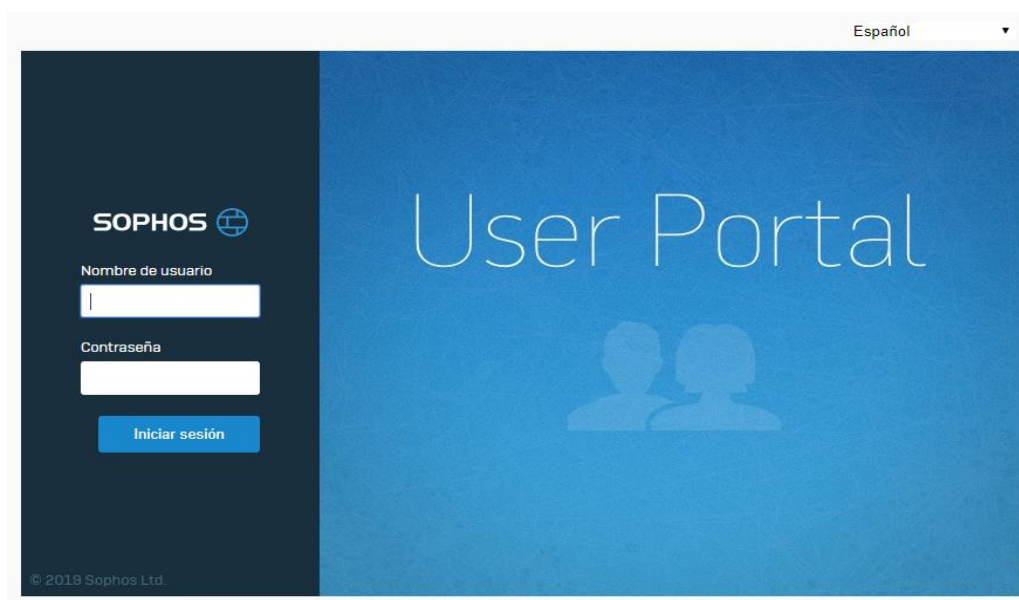


Figura 13 Sophos Creación Propia

Software que permite acceso a la VPN para acceder a la base de datos del proyecto. Después de instalarse un semáforo se asigna en la barra de herramientas como aviso de (Conectado/Desconectado) dependiendo su color como se observa en la imagen.



**Desconectado**

**Conectando**

**Conectado**

Figura 14 Vista software Sophos en escritorio de Windows Creación Propia



## **2. Comprensión de los Datos**

Esta segunda fase comprende la recolección inicial de los datos con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, implica estudiar más de cerca los datos disponibles de minería e identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis. Este paso es esencial para evitar problemas inesperados durante la siguiente fase (preparación de datos) que suele ser la fase más larga de un proyecto.

## **3. Preparación de los Datos**

En esta fase y una vez efectuada la recolección inicial de los datos, se procede a su preparación para adaptarlos a las técnicas de minería de datos que se van a utilizar posteriormente, éstas pueden ser técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para explotación de los datos.

## **4. Modelado**

En esta fase de CRISP-DM se seleccionan las técnicas de modelado más apropiadas para el proyecto de minería de datos específico.

El modelado se suele ejecutar en múltiples iteraciones. Normalmente, los analistas de datos ejecutan varios modelos utilizando los parámetros por defecto y ajustan los parámetros o vuelven a la fase de preparación de datos para las manipulaciones necesarias por su modelo.

## **5. Evaluación**

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis.

## **6. Distribución**

**o**

## **Implantación**

En esta fase, y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio.

## **FASE 1: COMPRENSIÓN DEL NEGOCIO**

### **Determinar los Objetivos del Negocio**

La Universidad de Cundinamarca, no cuenta con un proceso establecido para llevar a cabo el análisis de una variable que ha aumentado su importancia a través de los años en la Educación Superior, la Deserción Estudiantil, un dato que no es exacto por la falta de seguimiento a este índice. El área de Sistemas y Tecnología de la Universidad, quiere contar con un análisis que ayude a mitigar y hallar las causas que la generan.

#### Objetivos Comerciales

- Comprender por medio del análisis, la Deserción Estudiantil actual del programa de Ingeniería de Sistemas en la Universidad de Cundinamarca.
- Identificar los factores que han afectado y ocasionado la Deserción Estudiantil.
- Comprender el proceso que lleva a cabo el estudiante Desertor en la Universidad de Cundinamarca.
- Clasificar la Deserción Estudiantil en 3 tipos como está establecido en el reglamento de la Universidad y evaluar (cantidad, causas).

#### Criterios de Éxito

- Determinar cuál de los 3 tipos de Deserción posee una mayor cantidad de Desertores.
- Identificar las variables que efectúa la Deserción.
- El análisis servirá como estructura para implementar este proceso en todas los Programas Académicos de la Universidad.

### **Compilación de la Información de la Universidad**

#### **Inventario de Recursos**



El área de Sistemas y Tecnología posee el registro académico de los estudiantes durante todo su proceso de permanencia en la Universidad, la cual será base de la investigación.

La Facultad de Ingeniería, en cada Comité de Facultad registra actas, donde se encuentran las solicitudes de estudiantes para Cancelar y/o Aplazar su semestre.

Bienestar Universitario, cuenta con un registro de los estudiantes que hacen su solicitud 080 y se acercan a comentar el motivo de su solicitud.

### **Descripción de la Problemática**

En la actualidad la universidad cuenta con un proceso para que los estudiantes cancelen o aplacen su semestre, y por el cual también se lleva un control a la deserción. Dicho proceso consiste en realizar la la petición de cancelación o aplazamiento de semestre por la plataforma institucional, en donde se puede descargar un formato que deben diligenciar con las firmas de Bienestar Universitario y su respectiva Facultad.

Tanto en Bienestar Universitario como en la Facultad evalúan éstas solicitudes, en Bienestar Universitario un profesional entrevista al estudiante y evalúa si la situación por la cual el estudiante decide cancelar o aplazar su semestre, puede aplicar a un plan de permanencia (beneficios socio-económicos), o se firma por el profesional y es enviado a Facultad para terminar su proceso. La Facultad revisa que la solicitud se haya hecho en el periodo permitido para cancelación o aplazamiento para aprobar la solicitud.

Desafortunadamente no todos los estudiantes que desertan siguen este proceso, en cambio es una pequeña parte los que lo hacen, impidiendo de ésta manera llevar un correcto seguimiento al proceso de deserción e imposibilitando detectar las causas y así mismo maneras de evitar la deserción estudiantil.

### **Descripción de la Solución**

De acuerdo a la situación planteada, la solución está basada en reunir toda la información que poseen cada una de las Dependencias anteriormente nombradas y con ellas

analizar la situación que se está presentando actualmente en la Universidad de Cundinamarca. Los datos principales en la investigación, será la información académica entregada por la Dirección de Sistemas y Tecnología, del registro académico de los estudiantes de los últimos 6 años (IPA 2013- IIPA 2018) .

El análisis se realizará a través de herramientas de Inteligencia de Negocios, como apoyo para la toma de decisiones, se formularán estrategias que ayuden a disminuir índice de Deserción Estudiantil en la Universidad de Cundinamarca.

## **Valoración de la Situación**

### **Datos Disponibles para el Análisis**

- Base de datos académica
- Actas de Facultad de Ingeniería
- Base de datos de Bienestar
- Base de datos de Admisiones y Registro

### **Inventario de Recursos**

#### **Software para implementación de la Solución**

- Microsoft Power BI
- Microsoft Excel
- Microsoft Access
- Sophos SSL VPN Client
- Cliente Oracle (ODAC 12.2c versión 1 (12.1.0.0.2))

### **Origen de los Datos**

Los Datos son suministrados por las diferentes dependencias de la Universidad de Cundinamarca que hacen parte del proceso de Deserción Estudiantil las cuales son: Sistemas y Tecnología, Bienestar Universitario, Facultad de Ingeniería, Admisiones y Registro.

- **Base de Datos Académica**

El área de Sistemas y Tecnología facilita el acceso a la Base de Datos por medio de una VPN conectada al servidor de la Universidad, habilitando una vista construida por consultas a la base principal, extrayendo los datos necesarios para el posterior análisis.

La VPN fue asignada con un usuario y una contraseña, junto con otro usuario y contraseña para acceder a la vista de la base de datos Oracle.

- **Actas de Facultad de Ingeniería**

La facultad de Ingeniería, realiza comités mensuales donde se generan actas especificando los temas tratados y las decisiones tomadas, de esta información se extraen las solicitudes de cancelación, aplazamiento y reingreso de los estudiantes las cuales se agregan a una base de datos en Access.

- **Base de datos de Bienestar**

Bienestar Universitario cuenta con una base de datos en Excel de las solicitudes de cancelación y/o aplazamiento de los estudiantes.

- **Base de datos de Admisiones y Registro**

Admisiones y Registro, se encarga de registrar la cantidad de estudiantes admitidos y matriculados en cada semestre del año, las cuales están en una base de datos de Excel.

## **Requisitos, Restricciones y Supuestos**

### **Requisitos y Restricciones**

La seguridad de los datos, es un factor importante para tratar grandes cantidades de información. Para acceder a la vista de las Bases de Datos anteriormente nombradas se firmó un acuerdo de confidencialidad, donde los integrantes del proyecto nos comprometimos a no compartir, ni usar este contenido de una forma indebida.

## Supuestos

El cliente del presente proyecto es el área de Sistemas y Tecnología junto con Bienestar Universitario.

El resultado esperado por el cliente es un análisis de la cantidad de los estudiantes desertores, por años, motivos, edades, entre otras variables.

El análisis será presentado por medio de un dashboard generado por la herramienta de Inteligencia de Negocios (Power BI). La conexión de la base de datos será simultánea, para actualizaciones futuras y posteriores análisis que desee realizar la Universidad de Cundinamarca.

## Determinar los Objetivos de los resultados

- Incentivar a la universidad a implementar las estrategias generadas a partir del análisis.
- Crear conciencia acerca de la deserción estudiantil, sus causas y sus impactos dentro de la institución.
- Apoyar el proceso existente y dar alternativas para su continua mejora.

## Plan del Proyecto

Plan de estudio del proyecto en Fases, Tiempo, Recursos y Riesgos

Tabla 3 Plan del Proyecto Creación Propia

Fase	Tiempo	Recursos	Riesgos
Comprensión del Negocio	3 Semanas	Todos los analistas y cliente	Tiempo de reunión analista-cliente
Comprensión de los Datos	3 Semanas	Todos los analistas	Problemas de datos, problemas tecnológicos
Preparación de los Datos	10 Semanas	Asesor de minería de datos, tiempo de análisis de base de datos	Problemas de datos, problemas tecnológicos

<b>Modelado</b>	<b>3 Semanas</b>	<b>Asesor de minería de datos, tiempo de análisis de base de datos</b>	<b>Problemas de tecnología, incapacidad para encontrar un modelo adecuado</b>
<b>Evaluación</b>	<b>2 Semanas</b>	<b>Todos los analistas</b>	<b>Cambio económico, incapacidad para implementar resultados</b>
<b>Implementación</b>	<b>1 Semana</b>	<b>Asesor de minería de datos, tiempo de análisis de base de datos</b>	<b>Cambio económico, incapacidad para implementar resultados</b>

## FASE 2: COMPRESIÓN DE LOS DATOS

### Recolección de los Datos Iniciales

Los datos fueron suministrados por 3 dependencias de la Universidad de Cundinamarca, Sistemas y Tecnología dieron acceso a una vista de los todos los datos académicos de los estudiantes de Ingeniería de Sistemas, Bienestar Universitario suministró el formato 080 que manejan dentro de ésta dependencia y por con el cual llevan control de los estudiantes que desean cancelar, retirar o aplazar su semestre, Admisiones y Registro suministró los datos que conciernen a las cifras de los estudiantes que aplican a un programa de pregrado, los que son admitidos y los que finalmente se matriculan.

A continuación, se describen los datos suministrados por cada una de las dependencias:

- **Formato 080 suministrado por Bienestar Universitario**

Base de datos que contiene las solicitudes de los estudiantes al cancelar y/o aplazar semestre.

	A	B	C	D	E	F	G	H	I	J	K	L
	FECHA	NOMBRE DE ESTUDIANTE	PROGRAMA ACADEMICO	SEMESTRE	TELEFONO	E - MAIL	RESERVA DE CUPO	CANCELACION	FECHA POSIBLE DE REINGRESO	RETENCION	RETIRO	MOTIVO
1	05 de junio de 2015	Yesid Darío Villalobos Barrera	ING. De Sistemas	1	318185223	yesid.11@hotmail.com		x	IPA 2016			Personales y A
2	05 de junio de 2015	Samir Santiago Salcedo Sanabria	ING. De Sistemas	1	3132822974	samirsantiago8@hotmail.com		x	IPA 2016			Personales y A
3												
4												
5	febrero 10 de 2015	Alexs Danilo Zipa Guevara	Ingeniería de sistema	Segundo	3213229121	Splinter94@hotmail.com		x	Segundo periodo 2015			Dificultad ec
6												
7												
8	feb-20	Elibiana Rojas Herrera	Ingeniería de sistemas	Octavo	3209403324	elibianarohve@gmail.com		x	Segundo periodo 2015			Problemas
9												
10	14 de mayo de 2015	carlos eduardo mendez acosta	Ingeniería de Sistemas	primero	3209408309	carlozozaco_1881@hotmail.com		x	IPA 2016			por su trabajo asignado no pue
11	01/02/2016	MARIA ALEJANDRA SANABRIA MONCADA	ING SISTEMAS	I	3202333211	ale_031539@hotmail.com					x	RETIRO VOL
12	12/04/2016	NIJEL FERNANDO ESCOBAR CONTRERAS	ING SISTEMAS	I	312491948	escobar556@hotmail.es	x		IPA 2017			MOTIVOS ACA
13	28/04/2016	CARLOS JAVIER DUEÑAS DIAZ	ING SISTEMAS	IV	3173788886	javierduaj@gmail.com					x	RETIRO VOL
14	13/05/2016	ANDRES CAMILO MORALES SANCHEZ	ING SISTEMAS	I	3204453773	andcam22@gmail.com					x	MOTIVOS PEF
15												

Figura 16 . Base de datos Bienestar Universitario Creación Propia.

- **Vista de datos académicos suministrada por Sistemas y Tecnología**

Base de Datos Académica de los estudiantes de Ingeniería de Sistemas de la Universidad de Cundinamarca sede Fusagasugá.

Figura 17 Base de datos Académica Creación Propia.

- **Informe de registrados, admitidos y matriculados suministrado por Admisiones y Registro**

PROGRAMAS ACADEMICOS	SEDE FUSAGASUGA			PERIODO ACADEMICO
	INSCR.	ADMIT.	MATR.	
INGENIERÍA DE SISTEMAS	136	79		2019-1
	94	71	65	2018-2
	123	76	65	2018-1
	99	78	69	2017-2
	147	85	76	2017-1
	100	82	76	2016-2
	137	55	48	2016-1
	56	47	47	2015-2
	90	80	64	2015-1
	37	33	25	2014-2
	83	63	58	2014-1

Figura 18 Informe Admisiones y Registro Creación Propia.

- **Actas del Consejo de Facultad de Ingeniería**

El Consejo de Facultad en cada una de sus reuniones registra los temas tratados entre ellos las solicitudes de los estudiantes en las Actas.

ACTA No. 004 del 2017 – 03 – 07	
<b>CLASE DE REUNIÓN</b>	<b>REUNIÓN CONSEJO DE FACULTAD - FACULTAD DE INGENIERÍA -</b>
<b>FECHA</b>	<b>07 DE MARZO DE 2017</b>
<b>HORA</b>	<b>10:00 A.M.</b>
<b>LUGAR</b>	<b>UNIVERSIDAD DE CUNDINAMARCA</b>
<b>PARTICIPANTES</b>	<b>Ing. WILSON JOVEN SARRIA Decano de Facultad</b>
	<b>Ing. HUMBERTO NUMPAQUE LOPEZ Representante de Directores de Programa Coordinador Programa Ingeniería Electrónica Sede Fusagasugá</b>
	<b>Ing. LEIDER LEXANDRA VASQUEZ OCHOA</b>

Figura 19 Acta Facultad de Ingeniería Creación Propia.

○ **Base de Datos en Access**

Con la información suministrada por las actas del Consejo de Facultad se construyó una base de datos en Access con los datos de cada estudiante.

Id	Nombre	Genero	Documento	Tipo de Aplazamiento	Periodo de Solicitud	Reingreso	Motivos	Aprobado	Tiempo de J	Haga clic para agregar
1	J			Aplazamiento	2015-1	2016-1	Personal	<input checked="" type="checkbox"/>	2 Semestres	
2	E			Cancelacion-Reingresc	2015-1	2015-2	Laboral	<input checked="" type="checkbox"/>	1 Semestre	
3	A			Cancelacion	2015-1		Economico	<input checked="" type="checkbox"/>		
4	B			Cancelacion-Reingresc	2015-1	2015-2	Salud	<input checked="" type="checkbox"/>	1 Semestre	
5	C			Cancelacion	2015-1		Salud	<input checked="" type="checkbox"/>		
6	J			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
7	J			Cancelacion-Reingresc	2015-1	2015-2	Salud	<input type="checkbox"/>	1 Semestre	
8	S			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
9	J			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
10	C			Cancelacion-Reingresc	2015-1	2015-2	Academicos; E	<input type="checkbox"/>	1 Semestre	
11	C			Cancelacion-Reingresc	2015-1	2015-2	Economico	<input type="checkbox"/>	1 Semestre	
12	F			Reingreso	2015-1	2015-2	Suspension	<input checked="" type="checkbox"/>	2 Semestres	
13	H			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
14	C			Cancelacion-Reingresc	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
15	C			Cancelacion	2015-1		Fuerza Mayor	<input checked="" type="checkbox"/>		
16	J			Reingreso	2015-1	2015-2	Traslado	<input checked="" type="checkbox"/>	1 Semestre	
17	L			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	
18	S			Retiro Voluntario	2015-1		Laboral	<input type="checkbox"/>		
19	Y			Retiro Voluntario	2015-1		No tiene	<input type="checkbox"/>		
20	C			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	2 Semestres	
21	J			Reingreso	2015-1			<input type="checkbox"/>		
22	A			Aplazamiento	2015-1	2015-2	Economico	<input checked="" type="checkbox"/>	1 Semestre	
23	Y			Cancelacion	2015-1		Salud	<input checked="" type="checkbox"/>		
24	R			Reingreso	2015-1	2015-2		<input checked="" type="checkbox"/>	1 Semestre	

Figura 20 Base de Datos Facultad de Ingeniería Creación Propia.



## Descripción de los datos

- La muestra de la base de datos principal a analizar consta de 110.346 Filas y 20 Columnas, donde se encuentra el historial académico de 596 Estudiantes.  
Formato: Vista Servidor Oracle.
- La base de datos de la Facultad de Ingeniería se encuentra en un archivo (.accdb) Access, consta de 123 Filas y 9 Columnas.
- La base de datos de Bienestar Universitario se encuentra en un archivo (.xls) Excel, consta de 28 Filas y 12 Columnas.
- El informe de Admisiones y Registro se encuentra en un archivo (.xls) Excel, consta de 10 Filas y 5 Columnas.

## Exploración de los Datos

El propósito principal es analizar de la cantidad total de estudiantes en los últimos 5 años (IPA2013-IPA2018), destacar por periodos cuales de ellos han sido desertores durante el transcurso de la carrera , analizando variables en conjunto como el Género y la Edad.

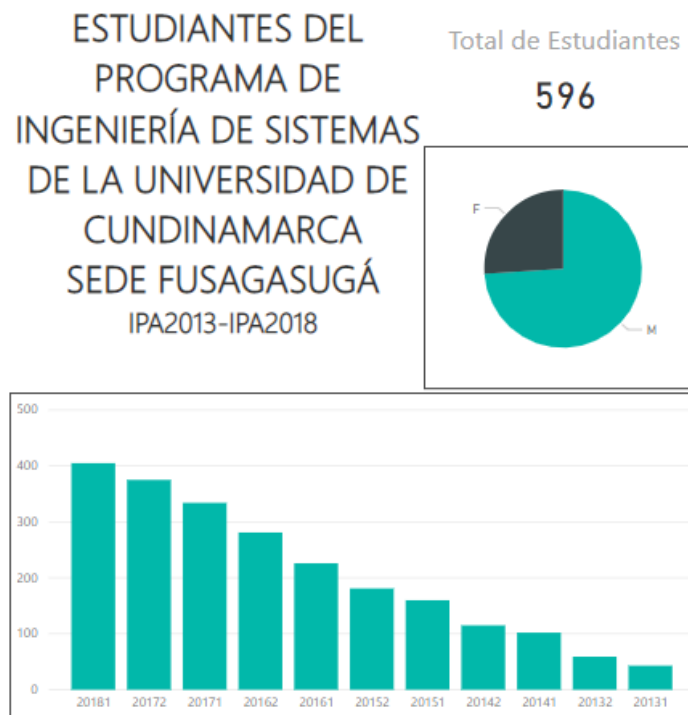


Figura 21 Total de la Población Creación Propia.

Una de las principales causas en la deserción de un estudiante, parte de las materias que se cursan y la dificultad que cada una de ellas puede llegar a tener. De la muestra anterior de 596 estudiantes, a continuación, las materias más perdidas por género.

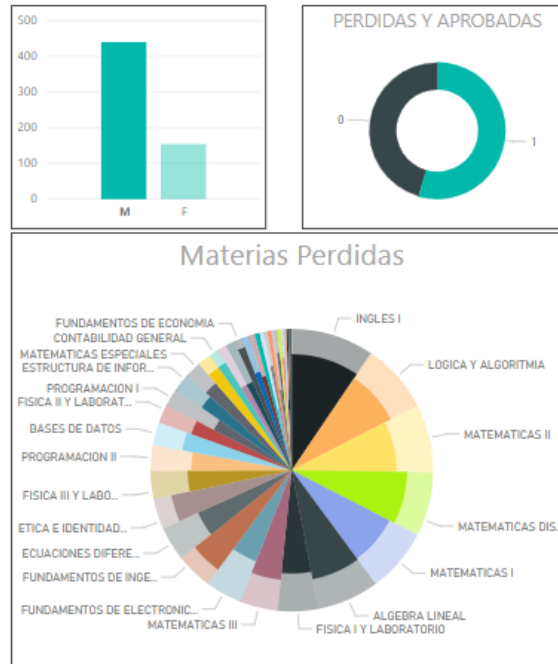


Figura 22 Materias Perdidas por Género Masculino Creación Propia.

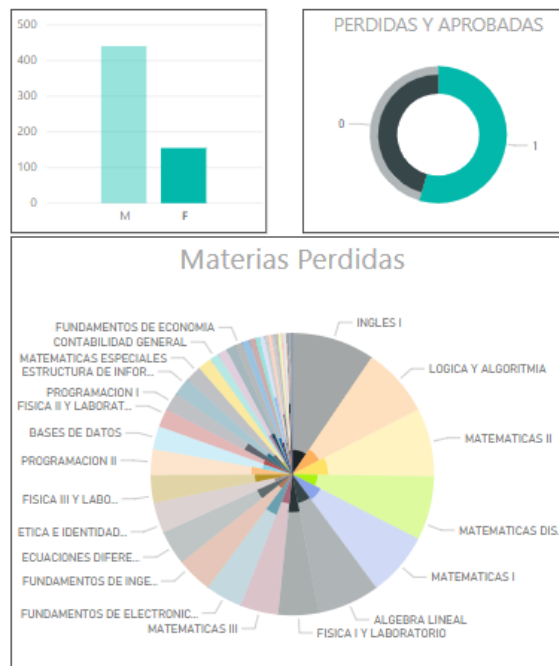


Figura 23 Materias Perdidas por Género Femenino Creación Propia.

## Verificación de la Calidad de Datos

Las dimensiones de la calidad de datos se ven reflejadas a la hora de verificar y nos permiten identificar los tipos de errores que se presentan en las bases de datos. A Continuación se identificarán los tipos de errores en cada una de las bases de datos.

- **Base de Datos de Bienestar Universitario**

En la base de datos se ven reflejadas dos dimensiones de la calidad de datos (Exactitud y Completitud). La exactitud es sintáctica ya que no existe una estandarización de los datos.

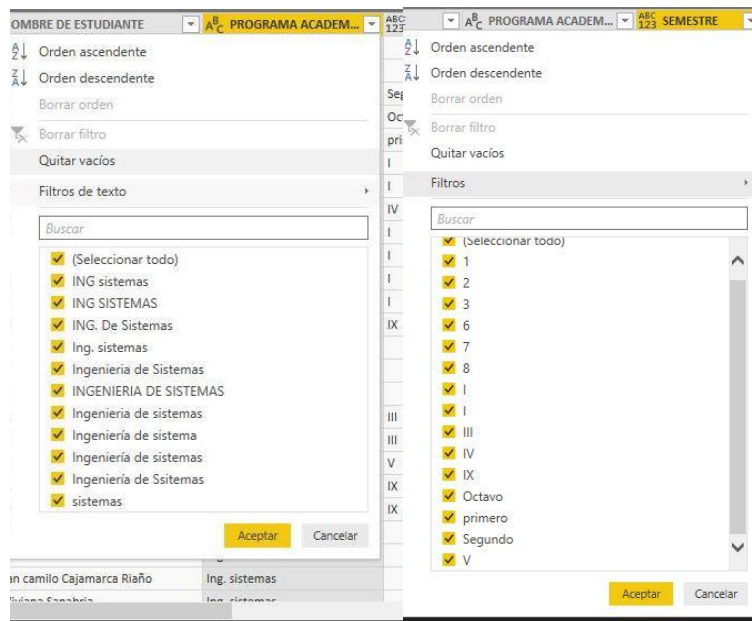


Figura 24 Calidad de Datos BD Bienestar Creación Propia.

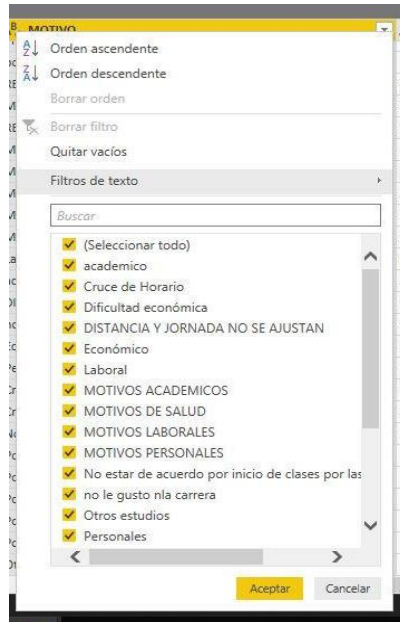


Figura 25 Calidad de Datos BD Bienestar Creación Propia.

También se ve reflejada la dimensión de completitud porque existen valores nulos dentro de la base de datos.

	FECHA	NOMBRE DE ESTUDIANTE	PROGRAMA ACADEM...	SEMESTRE	TELEFONO	E - MAIL
1	05/06/2015	Yesid Dario Villalobos Barrera	ING. De Sistemas		3118185223	yedis.11@hotmail.com
2	05/06/2015	Samir Santiago Salcedo Sanabria	ING. De Sistemas		3132822974	samirsantiagoss@hotmail.cc
3	10/02/2015	Aleks Danilo Zipa Guevara	Ingeniería de sistema	Segundo	3213229121	Splinter94@hotmail.com
4	null	null	null	null	null	null
5	null	null	null	null	null	null
6	null	null	null	null	null	null
7	null	null	null	null	null	null
8	01/02/2020	Bibiana Rojas Herrera	Ingeniería de sistemas	Octavo	3209403824	bibianarohe@gmail.com
9	null	null	null	null	null	null
10	14/05/2015	carlos eduardo mendez acosta	Ingenieria de Sistemas	primero	3209408309	carlosnacio_1861@hotmail.

Figura 26 Dimensiones de DQ Reflejadas Creación Propia.

- **Base de Datos Académica**

La consistencia de los datos se ve afectada bajo las reglas de integridad en la columna código, ya que no todas las filas cumplen con la regla de ser número, como se observa en la siguiente imagen.

DESERCIÓN ESTUDIANTIL - Excel

Herramientas de tabla | Herr... | Elizabeth Ojeda Marin

Archivo | Inicio | Insertar | Disposición de página | Fórmulas | Datos | Revisar | Vista | Kutools™ | Kutools Plus | Ayuda | Power BI | Diseño | Consulta

Calibri 11 | Fuente | Alineación | Número | Estilos | Celdas | Edición

	A	B	C	D	E	F
	CODIGO	DOCUMENTO	NOMBRES	FECHANACIMIENTO	PENS_DESCRIPCION	EPS
2795	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2796	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2797	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2798	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2799	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2800	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
2801	2017161 - UNICO 2013640	...	...	16-11-1996	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3751	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3752	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3753	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3754	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3755	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3756	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3757	2017161 - UNICO 2013662	...	...	11-11-1997	INGENIERIA DE SISTEMAS 2013	SALUDCOOP
3563	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3564	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3565	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3566	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3567	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3568	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3569	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS
3570	2016261 - UNICO 2013550	...	...	18-11-1993	INGENIERIA DE SISTEMAS 2013	NUEVA EPS

Se encontraron 162 de 110345 registros

Figura 27 Inconsistencia en BD Académica Creación Propia.

## **FASE 3: PREPARACIÓN DE LOS DATOS**

### **Seleccionar los datos**

#### **Selección de Elementos**

- **Base de Datos Académica**

La base de datos, consta de 110.345 registros de los cuales se analizaron 13.827, información académica de 5 años (IPA2013-IPA2018) del programa de Ingeniería de Sistemas de Fggá.

- **Base de Datos de Bienestar Universitario**

La información de Bienestar Universitario cuenta con 28 registros, desde el año 2015 hasta el año 2018.

- **Facultad de Ingeniería**

La base de datos consta de 123 registros de las solicitudes de los estudiantes de Ingeniería de Sistemas sede Fusagasugá al consejo de facultad desde el 2013 al 2018.

- **Informe de Admisiones y Registro**

El informe cuenta con registros de Inscritos, Matriculados y Admitidos del 2014 hasta el 2018.

#### **Selección de Atributos**

En las bases de datos se encuentra información confidencial como el código, documento y nombre de los estudiantes, los cuales se usarán para el análisis al relacionar los registros con otras tablas de datos.

En cada base de datos anteriormente nombrada, se encuentran atributos como género, edad, periodo cursado, materias aprobadas y motivos de retiro, se utilizarán para generar un análisis por cada una de ellas.

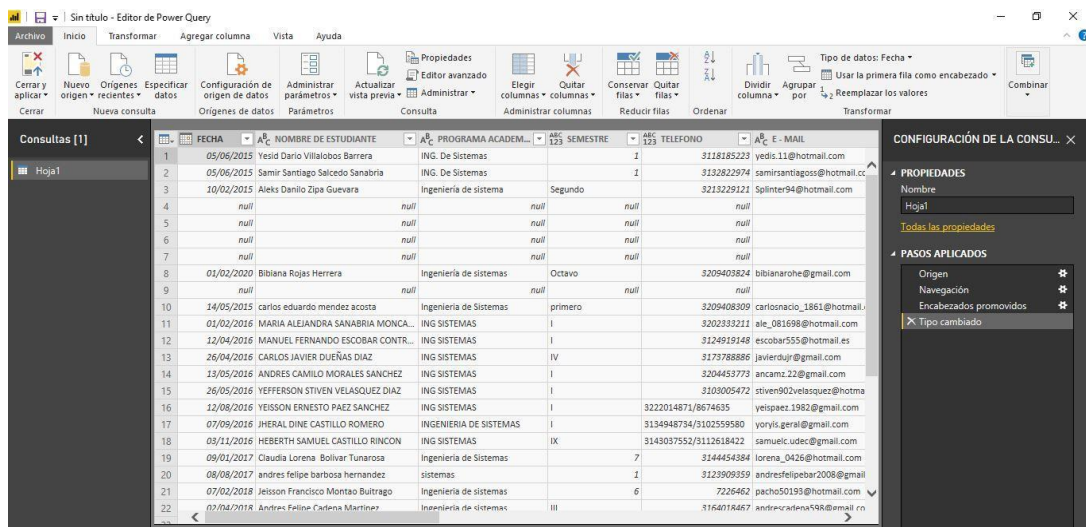
## Limpiar los datos

Basándonos en las dimensiones de la calidad de datos y los errores encontrados en el informe se realizó la siguiente limpieza de datos:

- **Base de Datos de Bienestar Universitario**

Para las inconsistencias encontradas en la base de datos suministrada por Bienestar Universitario se realizaron las correcciones correspondientes; se estandarizaron los datos, se anulaban los datos con campos vacíos para que no crearan errores al ser analizados y se hizo una revisión final.

Antes de la limpieza de datos:



	FECHA	NOMBRE DE ESTUDIANTE	PROGRAMA ACADÉMICO	SEMESTRE	TELEFONO	E-MAIL
1	05/06/2015	Yesid Dario Villalobos Barrera	ING. De Sistemas	I	318185223	yedid.11@hotmail.com
2	05/06/2015	Samir Santiago Salcedo Sanabria	ING. De Sistemas	I	3132822974	samirsantagos@hotmail.com
3	10/02/2015	Aleks Danilo Zipa Guevara	Ingeniería de sistema	Segundo	3213229121	Splinter94@hotmail.com
4	null	null	null	null	null	null
5	null	null	null	null	null	null
6	null	null	null	null	null	null
7	null	null	null	null	null	null
8	01/02/2020	Bibiana Rojas Herrera	Ingeniería de sistemas	Octavo	3209403824	bibianaroh@gmail.com
9	null	null	null	null	null	null
10	14/05/2015	carlos eduardo mendez acosta	Ingeniería de Sistemas	primero	3209408309	carlosnacio_1861@hotmail.com
11	01/02/2016	MARIA ALEJANDRA SANABRIA MONCA...	ING SISTEMAS	I	320233211	ale_081698@hotmail.com
12	12/04/2016	MANUEL FERNANDO ESCOBAR CONTR...	ING SISTEMAS	I	3124919148	escobar555@hotmail.es
13	26/04/2016	CARLOS JAVIER DUEÑAS DIAZ	ING SISTEMAS	IV	3173788886	javierduj@gmail.com
14	13/05/2016	ANDRES CAMILO MORALES SANCHEZ	ING SISTEMAS	I	3204453773	ancamz.22@gmail.com
15	26/05/2016	YEFFERSON STIVEN VELASQUEZ DIAZ	ING SISTEMAS	I	3103005472	stiven902velasquez@hotmail.com
16	12/08/2016	YEISSON ERNESTO PAEZ SANCHEZ	ING SISTEMAS	I	3222014871/8674655	yeispaez.1982@gmail.com
17	07/09/2016	HERALDINE CASTILLO ROMERO	INGENIERIA DE SISTEMAS	I	3134948734/3102559580	yoris.geral@gmail.com
18	03/11/2016	HEBERTH SAMUEL CASTILLO RINCON	ING SISTEMAS	IX	3143037552/3112618422	samuelc.udac@gmail.com
19	09/01/2017	Claudia Lorena Bolivar Tunarosa	Ingeniería de Sistemas	7	3144454584	lorena_0426@hotmail.com
20	08/08/2017	andres felipe barbosa hermandez	sistemas	I	3123909359	andrestfelipebar2008@gmail.com
21	07/02/2018	Jeisson Francisco Montao Bultrago	Ingeniería de sistemas	6	7226462	pachos0193@hotmail.com
22	07/04/2018	Andrés Felipe Cardena Martínez	Ingeniería de sistemas	III	3154017467	andracardena98@gmail.com

Figura 28 Estandarización de Datos Creación Propia.

En la columna de programa académico, semestre y motivos, se realizó una estandarización de datos, ya que no existía una exactitud y completitud en la consistencia de datos, se eliminan filas nulas sin afectar la calidad de la información.

Después de la limpieza de datos:

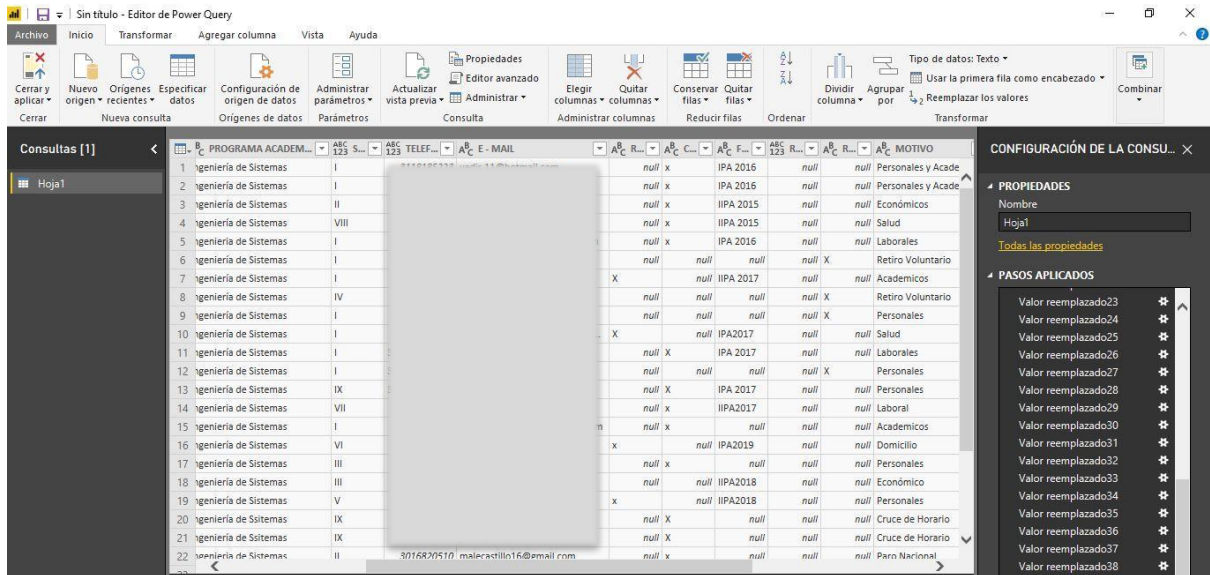


Figura 29 Limpieza de Datos Creación Propia.

- **Base de Datos Académica**

La integridad de los datos se ve afectada por la inconsistencia encontrada en la columna código, ya que esta columna está definida como número y dentro de la casilla existen letras y caracteres, entonces se realizó la siguiente corrección de estos registros de manera manual.

The screenshot shows an Excel spreadsheet with the following columns: 'CODIGO', 'DOCUMENTO', 'NOM', and 'CODIGO'. The data rows show various codes and names. A red box highlights the 'CODIGO' column, and a red arrow points to a cell containing '2017161 - UNICO 2013662'. The spreadsheet also shows sheet tabs at the bottom: 'Hoja2', 'DATA', 'Hoja3', 'Hoja1', and 'DATA ACADEMICA'.

	A	B	A
	CODIGO	DOCUMENTO	CODIGO
1			
2795	2017161 - UNICO 2013640		2013640 101
2796	2017161 - UNICO 2013640		2013640 101
2797	2017161 - UNICO 2013640		2013640 101
2798	2017161 - UNICO 2013640		2013640 101
2799	2017161 - UNICO 2013640		2013640 101
2800	2017161 - UNICO 2013640		2013640 101
2801	2017161 - UNICO 2013640		2013640 101
3751	2017161 - UNICO 2013662		2013662 101
3752	2017161 - UNICO 2013662		2013662 101
3753	2017161 - UNICO 2013662		2013662 101
3754	2017161 - UNICO 2013662		2013662 101
3755	2017161 - UNICO 2013662		2013662 101
3756	2017161 - UNICO 2013662		2013662 101
3757	2017161 - UNICO 2013662		2013662 101
35638	2016261 - UNICO 2013550		2013550 101
35639	2016261 - UNICO 2013550		2013550 101
35640	2016261 - UNICO 2013550		2013550 101
35641	2016261 - UNICO 2013550		2013550 101
35642	2016261 - UNICO 2013550		2013550 101
35643	2016261 - UNICO 2013550		2013550 101
35644	2016261 - UNICO 2013550		2013550 101
35645	2016261 - UNICO 2013550		2013550 101

Figura 30 Inconsistencia de Datos Creación Propia.



Se dejó únicamente con el número que precedía a el texto que generaba la inconsistencia, para así, posteriormente cargar la base de datos a la herramienta y poder generar el análisis.

## **Construcción de Nuevos Datos**

### **Derivación de Atributos**

Con el fin de maximizar la utilidad de los datos y realizar un análisis de calidad se crearon nuevas columnas para identificar en la base de datos académica:

- Estudiantes que desertaron.

Para hallar la deserción en los últimos 5 años se agregó una columna titulada Deserción, la cual muestra la continuidad del estudiante.

- Periodo en que el estudiante desertó.

Para destacar el periodo en que el estudiante decidió desertar, se agrego una columna titulada Fecha de posible deserción.

- Edad en la que el estudiante desertó.

Para hallar la edad exacta en la que desertó el estudiante, se agregaron las siguientes columnas: ( Último dígito periodo, Año retiró, Mes semestre) para crear la columna de la fecha exacta del retiro del estudiante y esa columna se tituló (Fecha retiro), junto con las columnas ya existentes se creó la columna de Edad.

- Cantidad de materias aprobadas y pérdidas por cada estudiante.

Para encontrar la cantidad de materias perdidas y aprobadas, se agregaron 2 columnas llamadas auxiliares y 2 columnas de sumatoria.

## **Integración de datos**

Para la integración de los datos se usó el método de fusión en Power BI, su editor de consultas Power Query, donde se tomaron los datos suministrados por Bienestar Universitario y por la Facultad de Ingeniería.

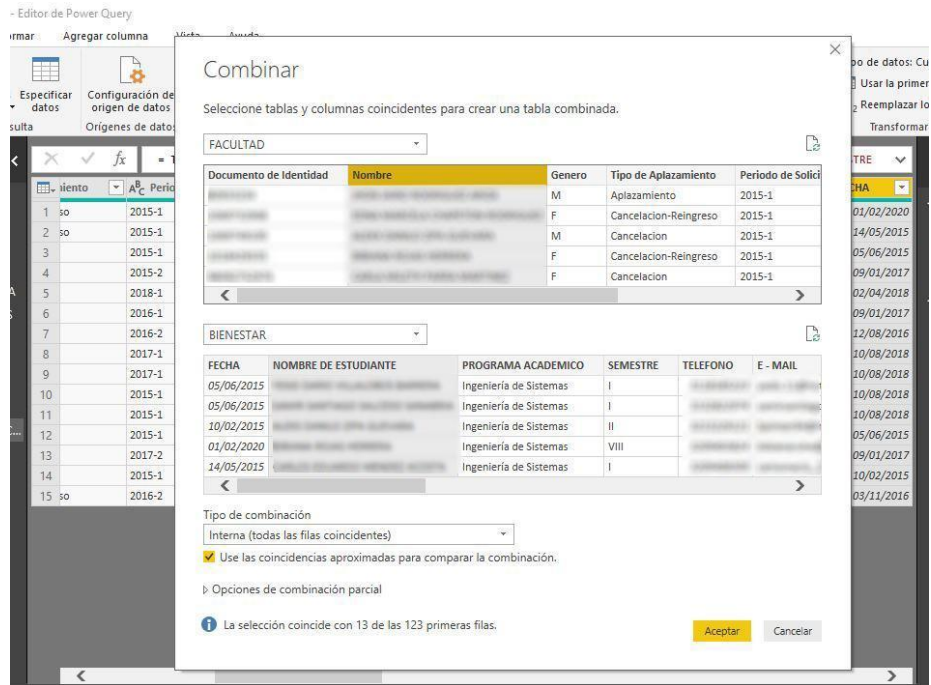


Figura 31 Fusión de Bases de Datos Creación Propia.

La integración permite apreciar que estudiantes siguieron el conducto regular a la hora de aplazar, retirar o cancelar su semestre.

## FASE 4 MODELADO

### Construcción del Data Warehouse

El proceso de construcción de un Data Warehouse, consiste en relacionar todas las bases de datos a analizar.

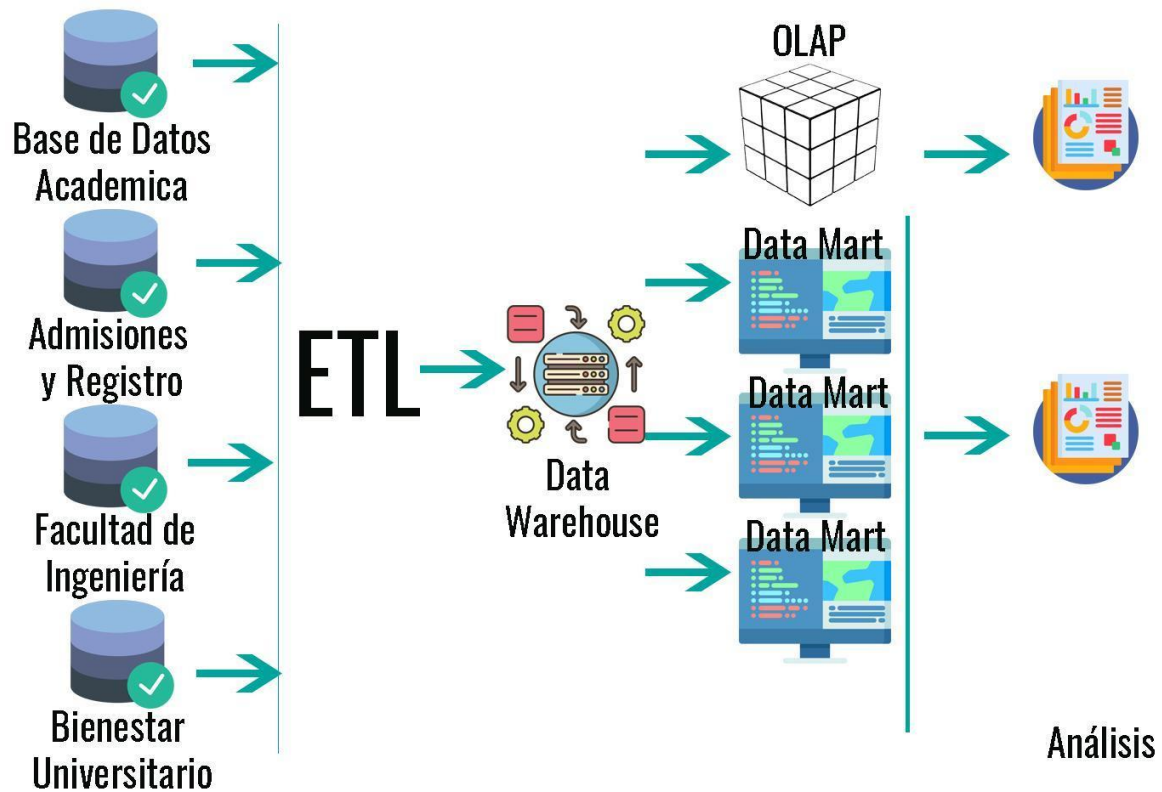


Figura 32 Modelo Creación Propia.

Para la construcción del Data Warehouse, Power BI permite añadir todas las bases de datos y relacionarlas, para que de esta manera se construya el modelo apropiado para el análisis.

### Proceso de ETL

#### Extracción de los Datos

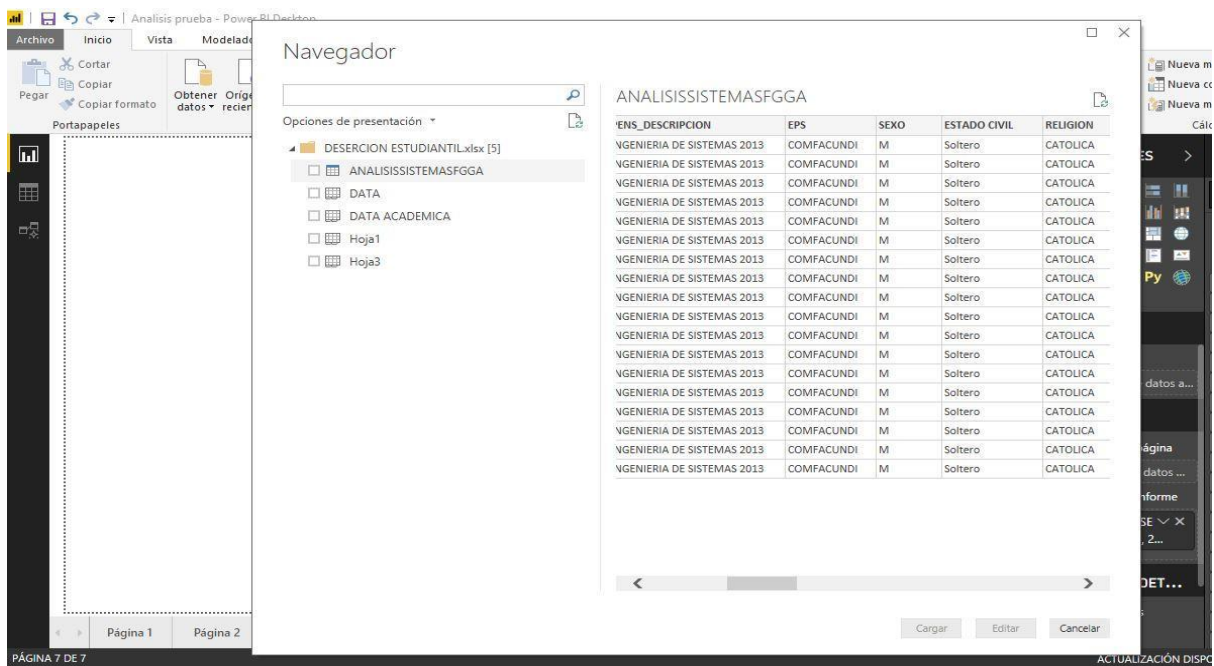


Figura 33 Extracción de Base de Datos Académica Creación Propia.

Datos extraídos de las bases y cargadas en Power BI

### Transformación de los Datos

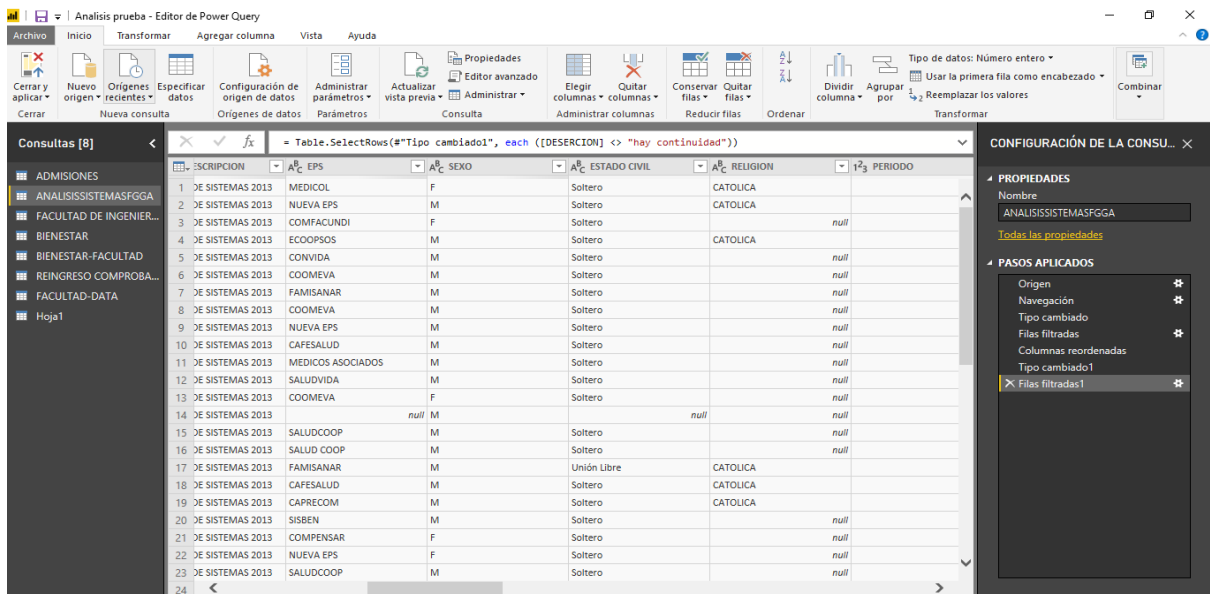


Figura 34 Transformación de los Datos Creación Propia.

Power BI cuenta con Power Query para transformar las bases de datos y analizarlas.

## Carga de los Datos

The screenshot shows the Power BI Desktop interface with a data table loaded. The table has the following columns: EPS, SEXO, ESTADO CIVIL, RELIGION, PERIODO, MATE\_CODIGOMATERIA, MATE\_NOMBRE, CREDITOS, APROBADO, and SUMA APROBADOS. The data rows represent individual student records with their respective academic details.

EPS	SEXO	ESTADO CIVIL	RELIGION	PERIODO	MATE_CODIGOMATERIA	MATE_NOMBRE	CREDITOS	APROBADO	SUMA APROBADOS
S 2013	COOMEVA	M	Soltero		20172	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	NUEVA EPS	M	Soltero		20181	612013101	MATEMATICAS I	4	0
S 2013	MEDICOS ASOCIADOS	M	Soltero		20181	612007103	FUNDAMENTOS DE INGENIERIA	2	0
S 2013	SALUDCOOP	M	Soltero		20171	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	SALUD COOP	M	Soltero		20132	TR1002013106	ETICA E IDENTIDAD INSTITUCIONAL	2	0
S 2013	SALUDCOOP	M	Soltero		20171	TR1002007104	INGLES I	2	0
S 2013	CAFESALUD	M	Soltero		20132	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	CRUZ BLANCA	M	Soltero		20172	612013209	FISICA I Y LABORATORIO	4	0
S 2013	PONAL	M	Soltero		20141	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	CONFACUNDI	M	Soltero		20171	TR1002007104	INGLES I	2	0
S 2013	MEDICOL	M	Soltero		20151	612013420	FISICA III Y LABORATORIO	4	0
S 2013		M	Soltero		20132	TR1002007104	INGLES I	2	0
S 2013	CONVIDA	M	Soltero		20171	612007425	MATEMATICAS DISCRETAS	2	0
S 2013	CAFESALUD	M	Soltero		20142	TR1002013318	ELECTIVA FORMAC. GRAL INTEGRAL I	2	0
S 2013	FAMISANAR	M	Soltero		20162	TR1002007104	INGLES I	2	0
S 2013	EJERCITO	M	Soltero		20181	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	FAMISANAR	M	Soltero		20181	612007320	PROGRAMACION I	3	0
S 2013	NUEVA EPS	M	Soltero		20181	612013208	MATEMATICAS II	4	0
S 2013	SISBEN	M	Soltero		20162	TR1002007104	INGLES I	2	0
S 2013	NUEVA EPS	M	Soltero		20161	TR1002007104	INGLES I	2	0
S 2013	CONVIDA	M	Soltero		20172	612013102	ALGEBRA LINEAL	3	0
S 2013	COMPENSAR	M	Soltero		20171	612013101	MATEMATICAS I	4	0
S 2013	SISBEN	M	Soltero		20171	612013105	LOGICA Y ALGORITMIA	3	0
S 2013	EXAMISANAR	M	Soltero		20181	TR1002013106	ETICA E IDENTIDAD INSTITUCIONAL	2	0

Figura 35 Carga de la Base de Datos Académica Creación Propia.

Bases de datos listas para ser analizadas.

## Construcción Modelo Multidimensional Principal en Power BI

### Esquema Estrella

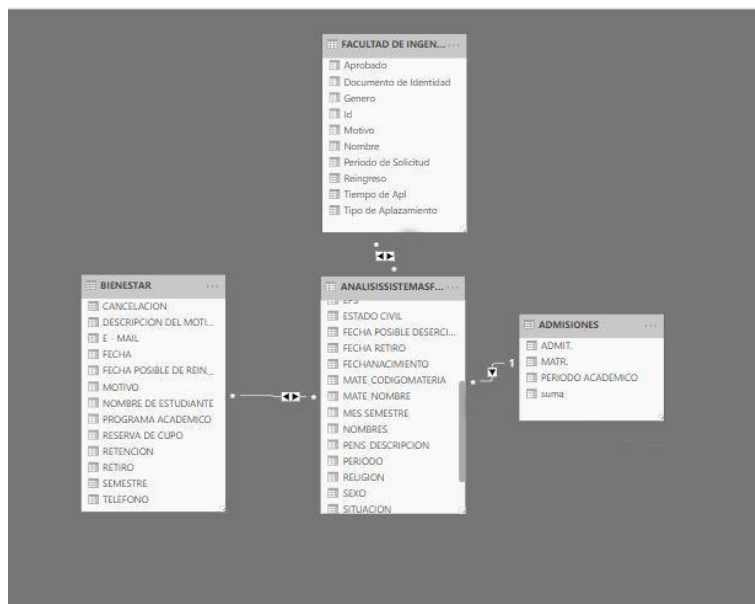


Figura 36 Esquema Estrella de la BD principal Creación Propia.

La tabla de Hechos es ANALISISISTEMASFUSAGASUGA, y las tablas de Dimensiones son: BIENESTAR, FACULTAD DE INGENIERÍA y ADMISIONES.

### 9.4.3 10.4.4 Construcción Modelo Multidimensional (Base de Datos Académica)

Un Data Warehouse está conformado por Data Marts y uno de ellos se analizó mediante un cubo OLAP, ya que su procedencia es de una base de datos extensa.

#### Esquema Copo de Nieve

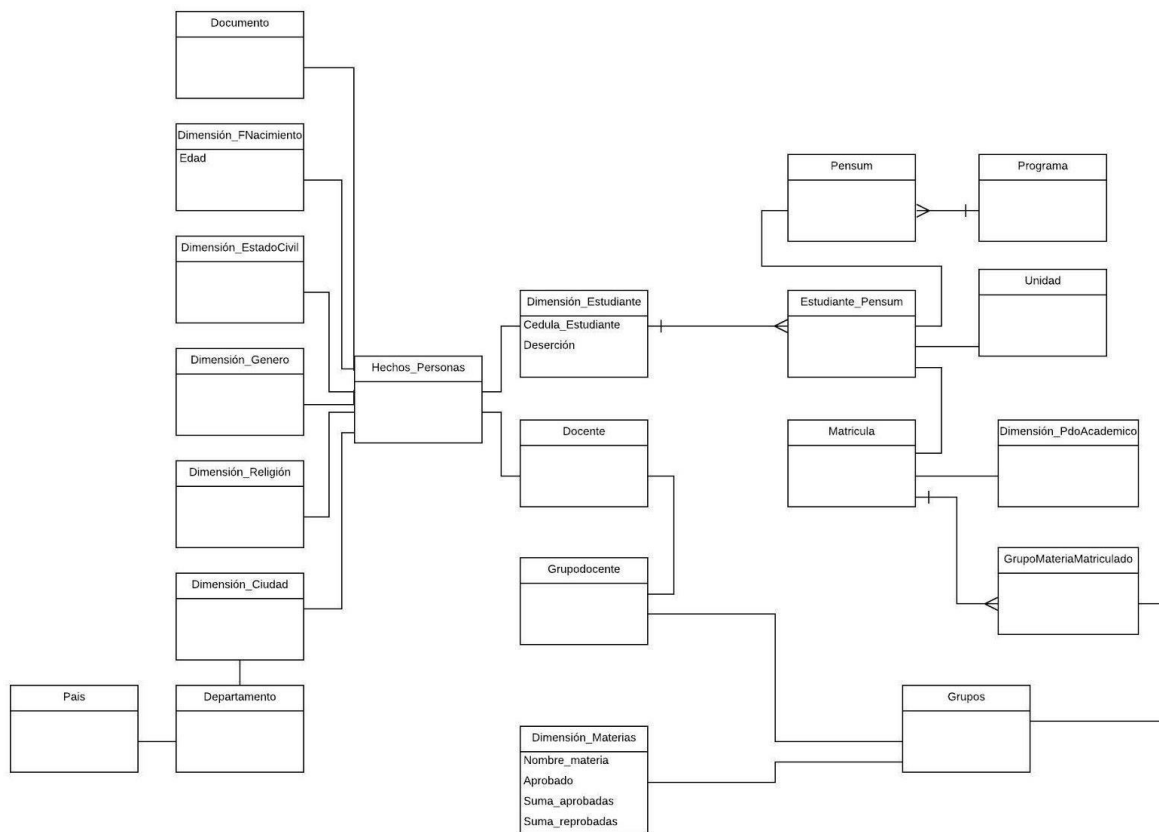


Figura 37 Esquema Copo de Nieve Base de Datos Académica Creación Propia.

Tabla 4 Tabla de Hechos y Dimensiones

#### Tabla de Hechos

Hecho	Fuente
-------	--------

Hechos_Personas	Base de Datos Académica
-----------------	-------------------------

### Tabla de Dimensiones

Dimensiones	Fuente	Campos
Dimensión_Estudiente	Base de Datos Académica	Cedula_Estudiente, Deserción
Dimensión_FechaNacimiento	Base de Datos Académica	Edad
Dimension_Periodo	Base de Datos Académica	Periodo_Académico
Dimensión_Género	Base de Datos Académica	Género_Estudiente
Dimensión _ Estado Civil	Base de Datos Académica	Estado Civil_Estudiente
Dimensión_Religión	Base de Datos Académica	Religión_Estudiente
Dimension_Ciudad	Base de Datos Académica	Ciudad Origen Estudiante
Dimensión_Materia	Base de Datos Académica	Nombre Materia,Aprobado, Suma de Aprobados,Suma de Pérdidas

### Evaluación de los Modelos

El proyecto cuenta con dos modelos multidimensionales, uno de ellos es el principal que contiene el Data Warehouse el cual es un esquema de estrella, que mediante la relación de las tablas se pudo analizar el proceso del estudiante y analizarlo desde diferentes puntos de vista con el cubo de OLAP que nos permite evaluar distintas variables al mismo tiempo.

El segundo modelo es un esquema copo de nieve, se implementó el otro modelo en el proceso porque la base de datos es de una fuente bastante amplia y cuenta con más tablas en su proceso interno. La base de datos académica es la principal del análisis realizado ya que cuenta con las variables principales y mediante el cubo OLAP se realizó el proceso adecuado para un excelente resultado.

La herramienta Power BI es de fácil uso y permite unir bases de datos de distintos orígenes, además de que la actualización de los datos será de manera automática y el Dashboard se podrá ajustar al tiempo actual.

A continuación una vista sobre el Dashboard con su breve explicación:

Los tableros son dinámicos y son entregados en un link que permite conocer los datos de una manera más amigable al seleccionar la edad o el género en los gráficos.

### Gráfico del Proceso del Estudiante Desertor

## PROCESO DE DESERCIÓN DEL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE CUNDINAMARCA SEDE FUSAGASUGÁ

De 130 estudiantes, solo el 11% realizó el proceso establecido en la UCundinamarca

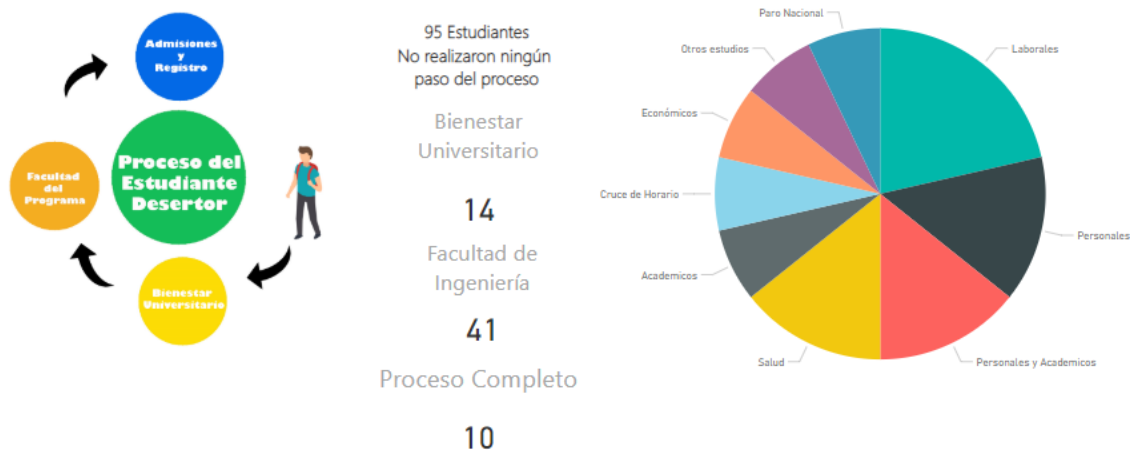


Figura 38 Seguimiento del Proceso de la UCundinamarca con respecto a la Deserción Estudiantil Creación Propia.



En el tablero se puede observar la cantidad de estudiantes que realizaron el proceso correspondiente en cada una de las dependencias y los motivos de retiro en esa pequeña muestra de estudiantes.

### Cantidad de Estudiantes Desertores

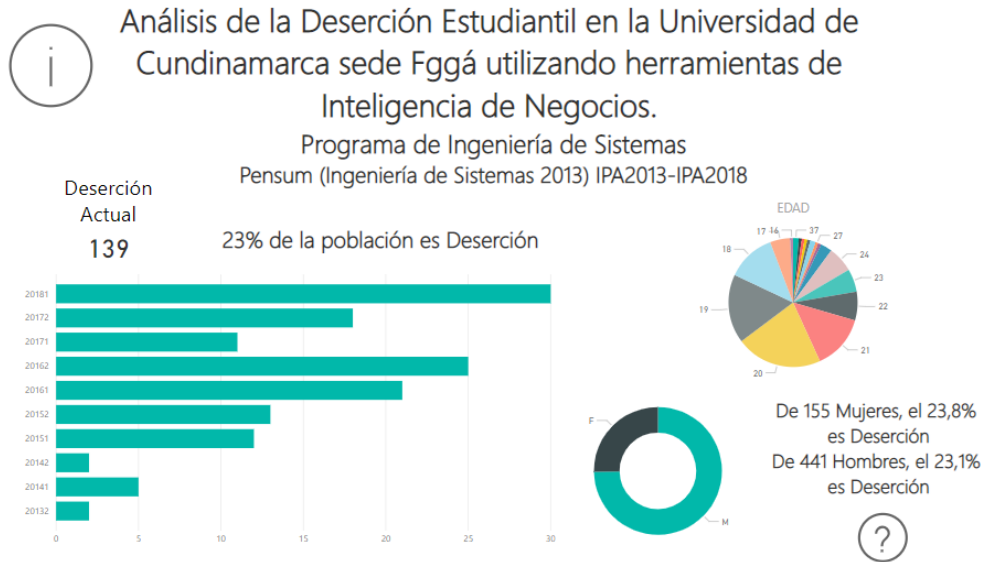


Figura 39 Total de Estudiantes Desertores Creación Propia.

En el tablero podemos observar la cantidad de estudiantes desertores en los periodos de muestra establecidos, destacando la edad y el género de esta población.

### Clasificación de la Deserción Estudiantil

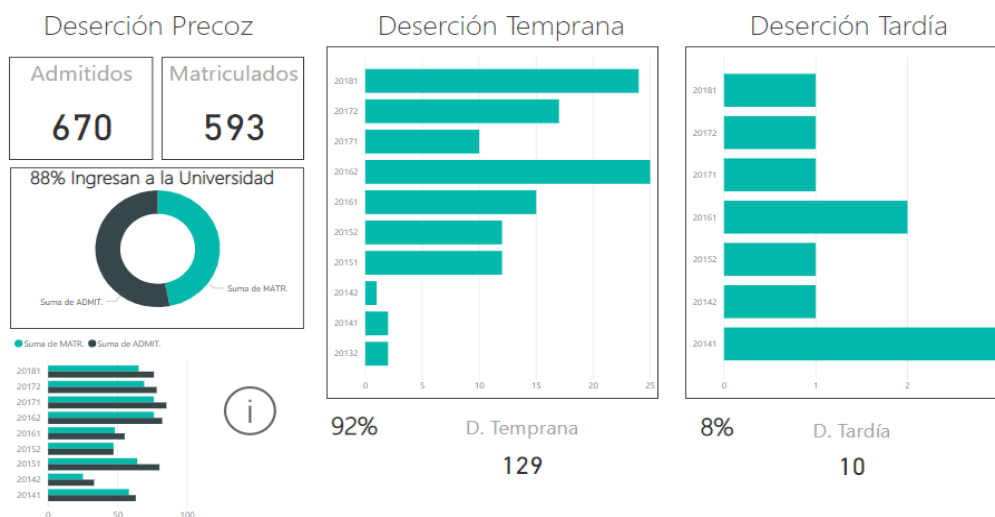


Figura 40 Clasificación de Deserciones Creación Propia.

En el tablero podemos encontrar los tres tipos de deserciones, destacando la cantidad de estudiantes en cada uno de los periodos.

### Deserción Temprana



Figura 41 Deserción Temprana Creación Propia.

En el tablero podemos observar la Deserción Temprana, cantidad de estudiantes, cantidad de materias perdidas y su clasificación en géneros y en edad.

### Deserción Tardía



Figura 42 Deserción Tardía Creación Propia.

En el tablero podemos observar la Deserción Tardía, cantidad de estudiantes, cantidad de materias perdidas y su clasificación en géneros y en edad.

### Calidad de los Datos Suministrados por el Área de Sistemas y Tecnología

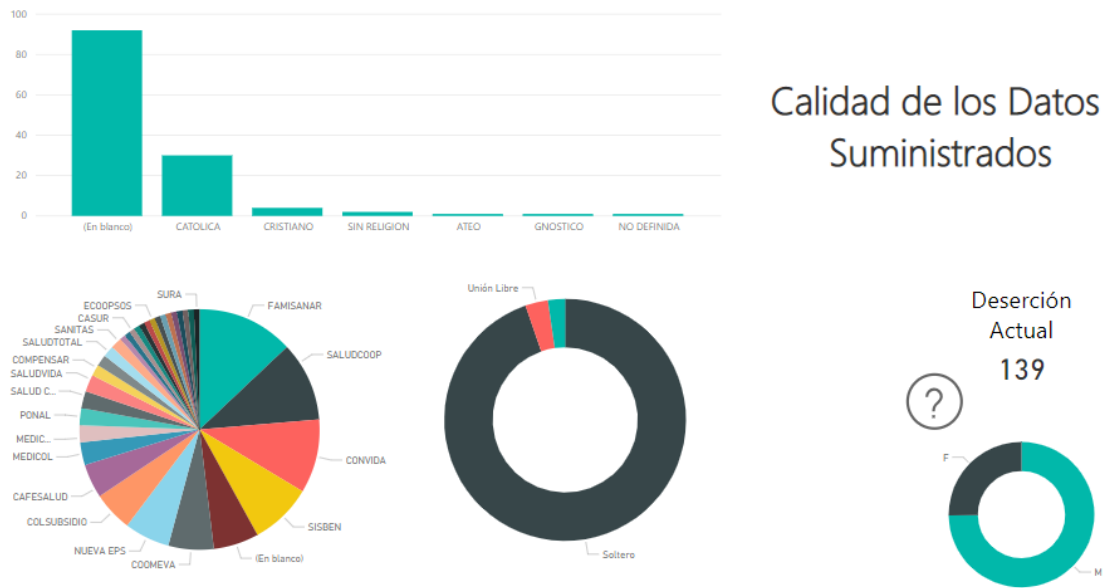


Figura 43 Calidad de Datos Suministrados Creación Propia.

En el tablero se puede observar las variables del estado civil, religión y EPS, donde se muestra la calidad de los datos por la cantidad de campos en blanco.

### Departamento de Domicilio de los Estudiantes Desertores

## Ubicación Geográfica

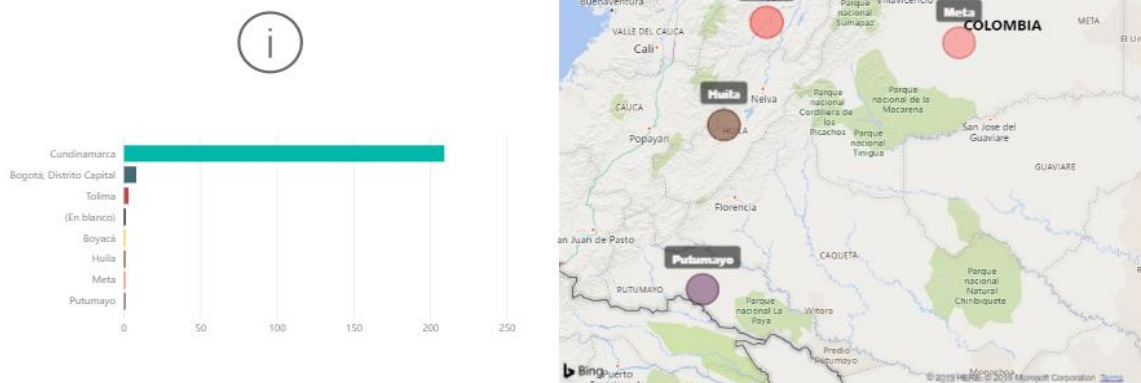


Figura 44 Ubicación Geográfica Creación Propia.

En el tablero se ilustra la ubicación geográfica de los estudiantes desertores.

### Posible Deserción en los Periodos del 2019

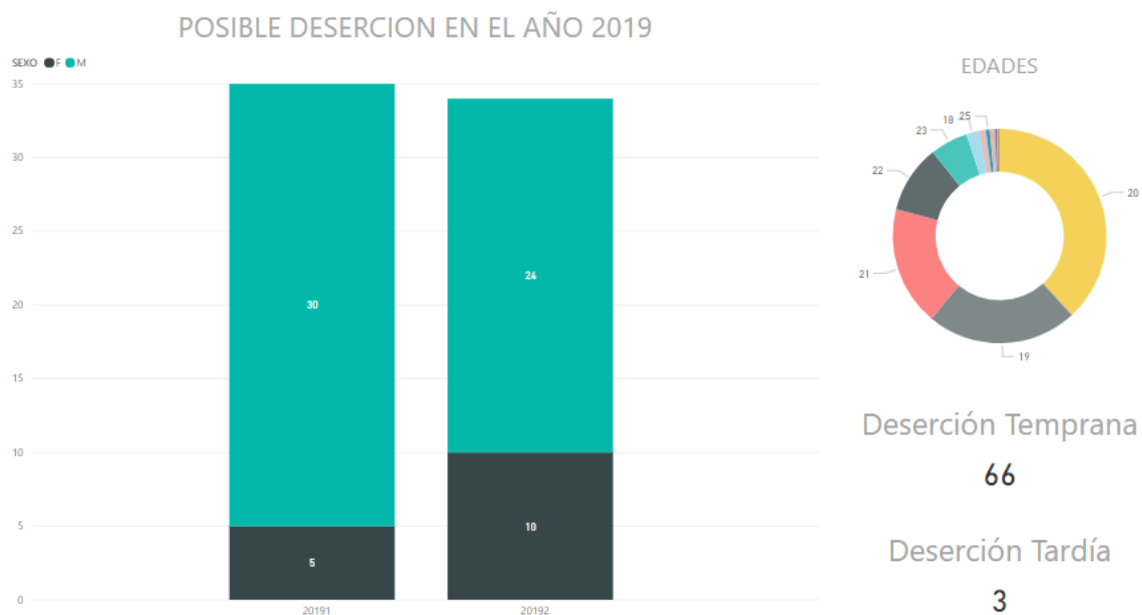


Figura 45 Posible Deserción del Año 2019 Creación Propia.

En el tablero se muestra la cantidad de mujeres y hombres posibles a desertar en el año 2019.

## FASE 5 EVALUACIÓN

En la evaluación se busca dar respuesta a la problemática planteada, mediante los gráficos anteriormente mostrados, donde se evalúa el proceso del estudiante que decide desertar y analizar cómo el historial académico puede afectar esta decisión.

### Evaluación de Resultados

- **¿Cuál es la población Desertora en el programa de Ingeniería de Sistemas entre los períodos ( IPA 2013- IPA 2018)?**

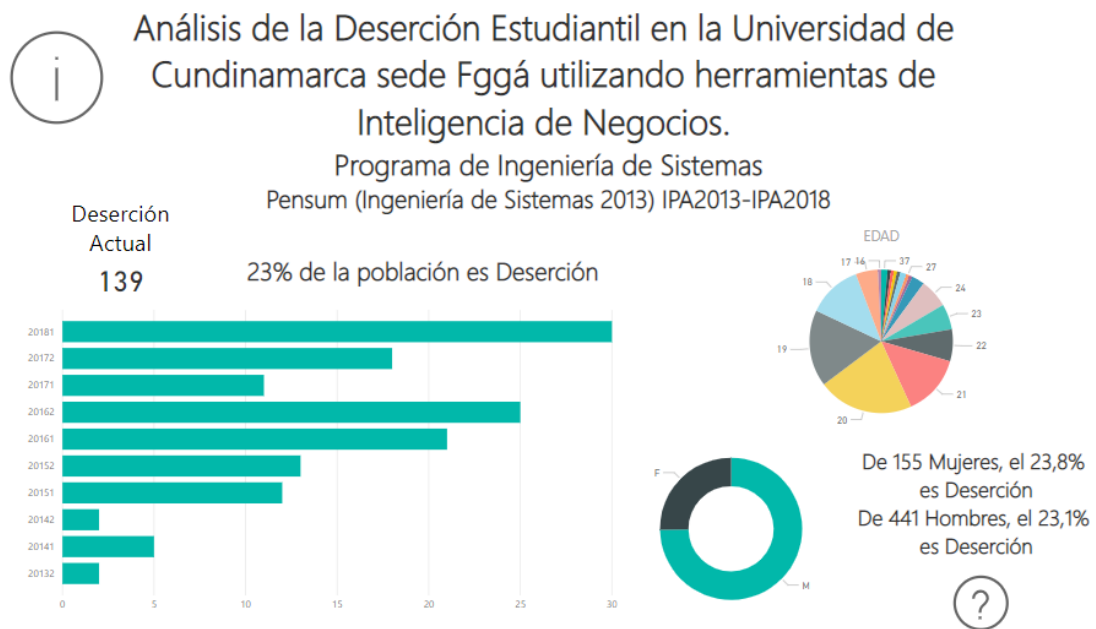


Figura 46 Cantidad de Estudiantes Desertores Creación Propia.

La población estudiantil entre los períodos ( IPA 2013- IPA 2018) es de 596 Estudiantes de la cual el 23% es desertora, de 441 Hombres el 23,1% son desertores y de 155 Mujeres el 23,8% son desertoras, lo que indica que el género no es una causa de la Deserción en el programa de Ingeniería de Sistemas.

- **¿De qué manera la Edad es un factor que está involucrado en la población desertora?**



## Análisis de la Deserción Estudiantil en la Universidad de Cundinamarca sede Fggá utilizando herramientas de Inteligencia de Negocios.

Programa de Ingeniería de Sistemas

Pensum (Ingeniería de Sistemas 2013) IPA2013-IPA2018

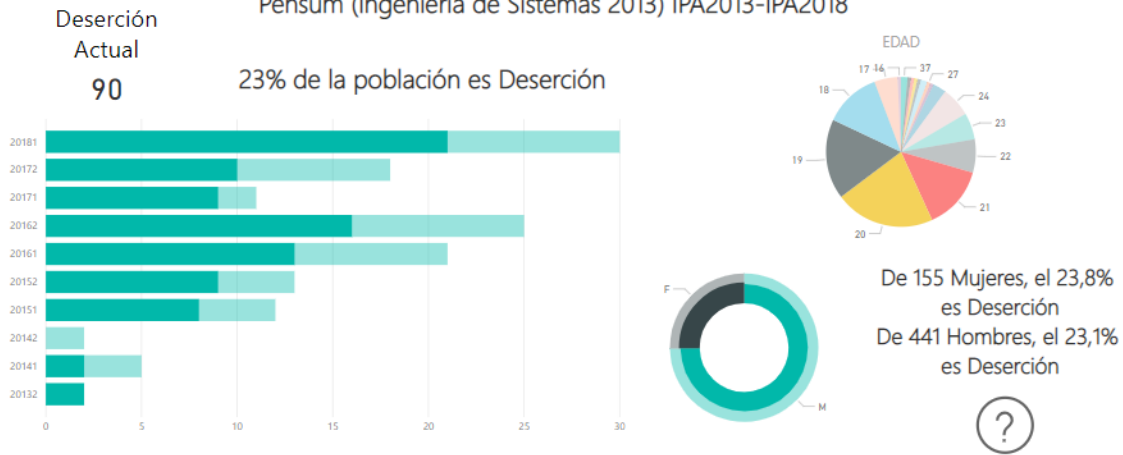


Figura 47 Edad Promedio de los Estudiantes Desertores Creación Propia.

Las Edades en que es frecuente que un estudiante desertor tome esta decisión es entre los 18 y 21 años de edad lo que indica que son los jóvenes la población más vulnerable y en peligro de ser posibles desertores, factores resaltados:

- Dificultad de aprendizaje o quizás no sea la carrera que desean estudiar.
- Poco éxito académico lo que los lleva repetir materias.
- No tener los medios económicos y tener que trabajar y estudiar al tiempo lo que puede llegar a causar una presión psicológica/física en el estudiante.

La Universidad de Cundinamarca clasifica la deserción en tres niveles.

- **¿Cuál es el porcentaje de Deserción Precoz?**

## Deserción Precoz

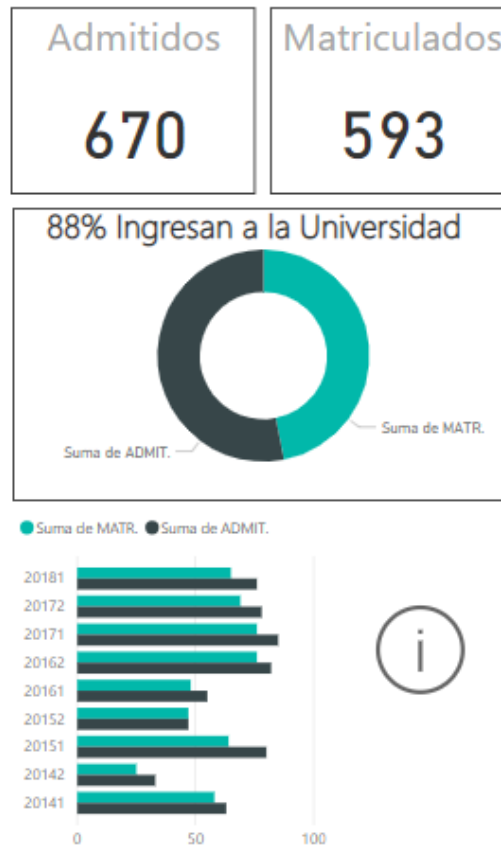
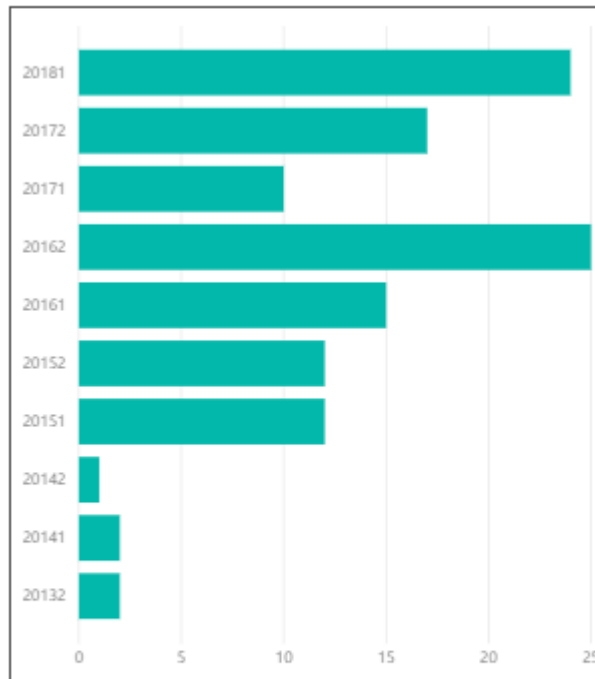


Figura 48 Deserción Precoz Creación Propia.

La Deserción Precoz la conforma la población de los estudiantes que fueron admitidos por la Universidad y no cancelaron su matrícula académica. Según el análisis realizado el 12% de la población es desertora ya que de 670 estudiantes que fueron admitidos, solo 593 ingresaron a la Universidad. Este tipo de Deserción ha llevado el mismo impacto en porcentaje desde el año 2014 hasta el año 2018, lo que indica que es bueno que no haya aumentado pero que se debe hacer algo para que este porcentaje disminuya.

- **¿Por qué La Deserción Temprana es el tipo de deserción con más población de estudiantes desertores?**

## Deserción Temprana



92%

D. Temprana

129

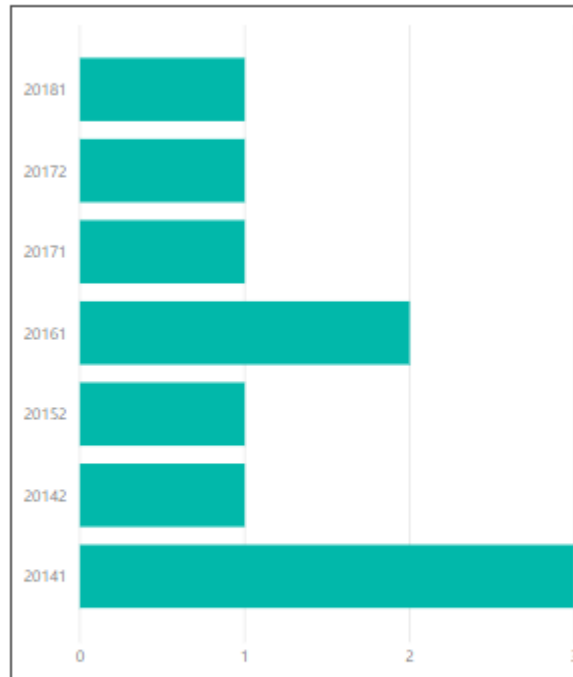
Figura 49 Deserción Temprana Creación Propia.

La Deserción Temprana está compuesta por un 92% de la población desertora, los cuales son los estudiantes que deciden abandonar su carrera en el transcurso de los primeros 5 semestres y se refleja en el tablero que en los periodos del IPA 2013 hasta IPA 2018, donde aumentó en cantidad de estudiantes en el periodo IIPA 2016 y el periodo IPA 2018 en comparación de los demás períodos.

- **¿Por qué la Deserción Tardía es el tipo de deserción con menos población de estudiantes desertores?**



## Deserción Tardía



8%

D. Tardía

10

Figura 50 Deserción Tardía Creación Propia.

La Deserción tardía la conforman los estudiantes que deciden abandonar sus estudios cuando ya llevan más de la mitad de su carrera cursada y solo hace parte de un 8% de la población desertora, lo cual es una buena noticia ya que esto indica que los estudiantes en su gran mayoría deciden culminar sus estudios cuando ya cursaron el 50% de su carrera.

- **Variables que se ven reflejadas en la Deserción Temprana**



Figura 51 Deserción Temprana Creación Propia.

La Deserción Temprana tiene un total de 129 estudiantes de los cuales el 7% retomo sus estudios, este tipo deserción en su gran mayoría está comprendida entre las edades de 18 a 21 años. Las materias que son cursadas en los primeros semestres de la carrera son de vital importancia para todo el proceso, en el tablero se puede observar las materias que los estudiantes desertores perdieron y que pueden ser una posible causa de deserción, ya que se resaltan las siguientes: Álgebra Lineal, Lógica y Algoritmia, Matemáticas Discretas, Programación II y Fundamentos de Ingeniería.

Las Materias resaltadas en este gráfico de barras, se consideran base en la carrera de Ingeniería de Sistemas, lo cual puede llegar a significar una decisión equivocada en la profesión escogida por las estudiante o simplemente malas bases de Matemáticas en su época de bachillerato.

- **Variables que se ven reflejadas en la Deserción Tardía**



Figura 52 Deserción Tardía Creación Propia.

La Deserción Tardía tiene una memoria de la población desertora que se compone por 9 personas de las cuales 6 retomaron sus estudios. Las edades se encuentran comprendidas entre los 22 a 37 años, lo que indica que ya son estudiantes que quizás no conviven con sus familias, también existe la posibilidad de que tengan ocupaciones laborales que no les permita continuar con sus estudios. Las materias perdidas resaltadas son pocas pero de vital importancia en la carrera.

- **Ubicación Geográfica de los Estudiantes Desertores**

# Ubicación Geográfica

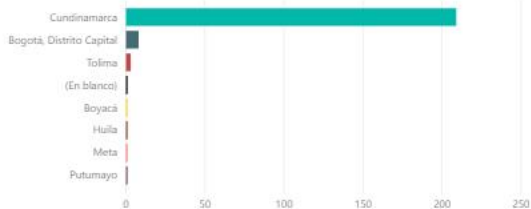
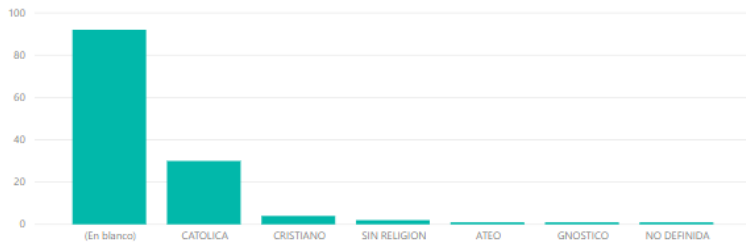


Figura 53 Ubicación Geográfica Creación Propia.

La ubicación geográfica de los estudiantes desertores se encuentra distribuida en 6 departamentos y en la capital del país, resaltando que una gran mayoría residen en el departamento de Cundinamarca.

## ● Calidad de los Datos Suministrados



## Calidad de los Datos Suministrados

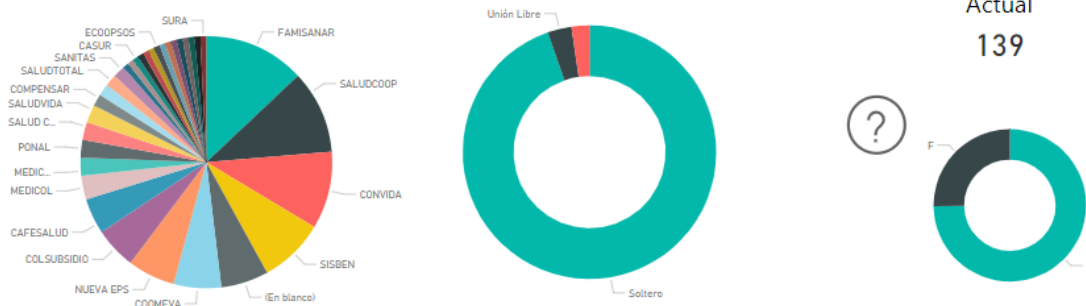


Figura 54 Calidad de Datos Suministrados

En el tablero se pueden observar 3 aspectos importantes a analizar en la población desertora las cuales son: Estado Civil, Religión, EPS. Una gran cantidad de datos de estos índices se encontraban vacíos, lo que nos lleva a analizar que por falta de una información completa no se puede llegar a una conclusión exacta pero sí cabe resaltar la completitud de los datos suministrados.

- **Proceso del Estudiante Desertor**

### PROCESO DE DESERCIÓN DEL PROGRAMA DE INGENIERÍA DE SISTEMAS EN LA UNIVERSIDAD DE CUNDINAMARCA SEDE FUSAGASUGÁ

De 130 estudiantes, solo el 11% realizó el proceso establecido en la UCundinamarca

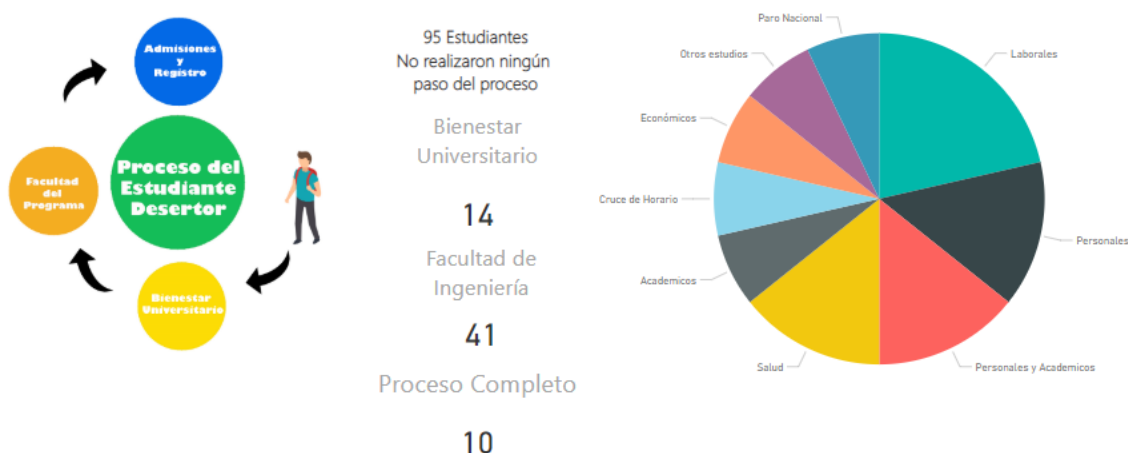


Figura 55 Seguimiento del Proceso de la UCundinamarca con respecto a la Deserción Estudiantil Creación Propia.

La Universidad de Cundinamarca cuenta con un proceso establecido para el estudiante que decide aplazar/cancelar su semestre, el cual se divide de la siguiente manera:

1. **Bienestar Universitario:** Es el primer lugar al que los estudiantes deben acudir para que bienestar pueda realizar un seguimiento y ofrecer una ayuda para evitar que el estudiante, suspenda sus estudios ya sea definida o indefinidamente, por esta dependencia de 130 estudiantes que desertaron entre los periodos IPA 2015

- IPA 2018, solo 15 estudiantes realizaron su solicitud mediante el formato 080 en Bienestar Universitario.

2. Facultad de Ingeniería: Como paso a seguir los estudiantes deben ir a Facultad y llevar su solicitud con un aval de ya haber pasado por Bienestar, de los 130 estudiantes que desertaron entre los periodos IPA 2015- IPA 2018, solo 41 estudiantes cumplieron este paso, resaltando que el 77% de los 41 fueron recibidas sus solicitudes sin haber cumplido el paso de Bienestar Universitario.
3. Proceso Completo: De 130 estudiantes solo 10 estudiantes siguieron los pasos establecidos.

Los motivos por los cuales los estudiantes toman la decisión de desertar, no se pueden generalizar, debido a que no se tiene el dato de todos los estudiantes en una totalidad dado a que el proceso no se realiza a cabalidad de una manera correcta.

- **Posible Deserción en el Año 2019**

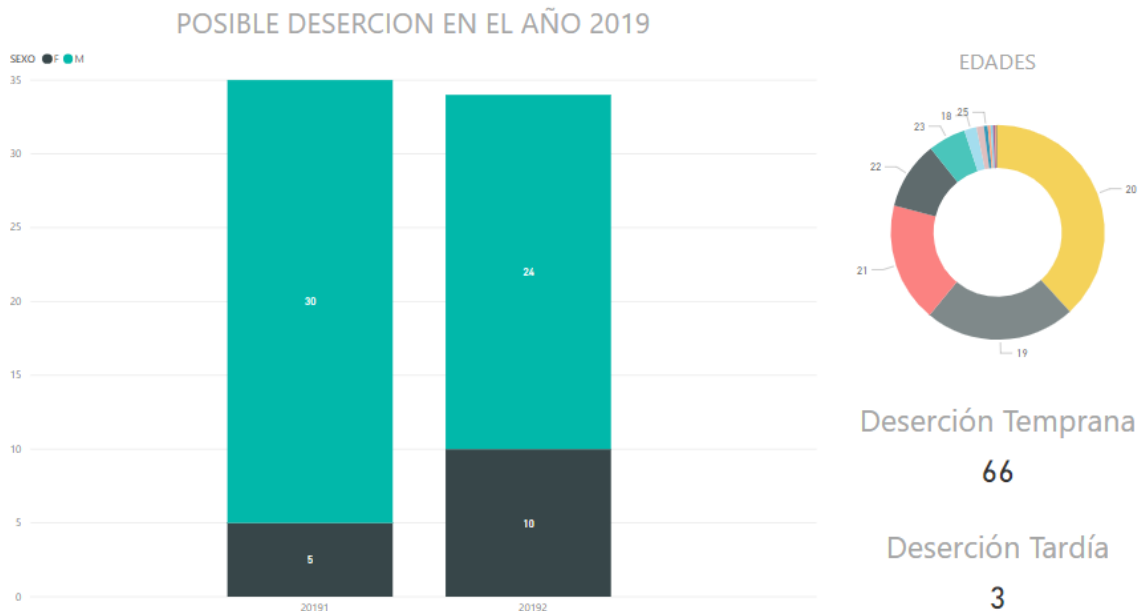


Figura 56 Posible deserción para el año 2019 Creación Propia.

La posible deserción en el año 2019 se divide en los dos periodos académicos del año, en el IPA 2019 30 Mujeres, 5 Hombres y en el IIPA 2019 10 Mujeres, 24 Hombres, actualmente no son estudiantes activos y son posibles a no retomar sus estudios, de los cuales 95% pertenecen al tipo de deserción temprana y el 5% a la deserción tardía. Sus Edades comprenden desde los 18 a 25 años.

### **Proceso de Revisión**

El Ingeniero Edilson Martínez, Gerente del Proyecto el día 30 de Mayo del 2019, aprobó el proyecto a satisfacción, como anexo se adjunta acta de aprobación.

Resaltando que en el proceso del proyecto ha existido un seguimiento por parte del Ingeniero Edilson y de las Psicólogas de Bienestar Universitario.

### **Determinar los próximos pasos**

La analítica realizada a la Deserción Estudiantil que se presenta en el programa de Ingeniería de Sistemas sede Fusagasugá, permite observar cómo la área académica puede influir en un estudiante desertor, resaltando género, edad y segmentando en los distintos tipos de deserción estudiantil que existe en la Universidad de Cundinamarca, así mismo destacando las materias que son dificultad para los estudiantes en el transcurso de su carrera.

El paso a seguir es que la Universidad implemente este análisis para mitigar la deserción estudiantil, realizando el mismo proceso en todos los programas de la Institución, mediante el uso de la Herramienta de Inteligencia de Negocios (Power BI) .

## **FASE 6 DESPLIEGUE**

### **Cartillas Digitales**

Se desarrollaron 2 cartillas digitales amigables para el usuario, en el uso correcto de la solución analítica.

- **Manual Técnico**

Este manual técnico permite al usuario conocer desde los conceptos básicos hasta los programas que se requieren para el proceso como: Power BI, Excel, Access entre otros.

Se especifican los requerimientos del equipo de cómputo donde se vaya a desarrollar el análisis y al momento de utilizarla funcione con éxito.

- **Manual de Usuario**

Este manual permite al usuario conocer el manejo de la solución analítica, desalentando paso a paso desde la obtención de datos hasta analizarlos en la herramienta de Inteligencia de Negocios, para que el usuario tenga un resultado exitoso a la hora de usarlo en la implementación.

### **Requerimientos para la implementación**

Para la implementación de esta solución analítica se requiere de los siguientes programas:

- Power BI, requiere de una licencia dependiendo su uso, si es para uso de muchos equipos al mismo tiempo tiene un costo de \$4,995 y si es para un usuario que superará el espacio que tiene power BI asignado a cada usuario debe cancelar un costo de \$9.99.
- Microsoft Office, requiere de una licencia para tener acceso a Excel y Access.
- Sophos SSL VPN Client

Requerimientos del Sistema



- 8 GB de RAM
- Desde Windows 7 en adelante.
- CPU: 1 gigahercio (GHz) o superior; se recomienda un procesador de x86 o x64 bits.
- Pantalla: Se recomienda al menos 1440 x 900 o 1600x900 (16:9). No se recomiendan las resoluciones inferiores a 1024 x 768 o 1280 x 800, ya que ciertos controles (por ejemplo, para cerrar la pantalla de inicio) solo se muestran en resoluciones superiores a esta.

## GLOSARIO

**Deserción Estudiantil:** Puede entenderse como el abandono del sistema educativo por parte de los estudiantes, provocado por la combinación de factores que se generan tanto al interior del sistema como en contextos de tipo social, familiar, individual y del entorno.

**Deserción Precoz:** Se refiere a las personas que aprueban el examen de admisión en la universidad pero que no pagan la matrícula, es decir no empiezan a estudiar.

**Deserción Temprana:** Estudiantes que abandonan sus estudios entre el primero (1ro) y cuarto (4to) semestre

**Deserción Tardía:** Estudiantes que abandonan sus estudios una vez que ya han cursado la mitad de la carrera.

**Business Intelligence:** Es un amplio conjunto de metodologías, herramientas, aplicaciones, entre otras, que permiten depurar y transformar los datos de tal que puedan ser analizados y utilizados para que generen valor.

**Data Warehouse:** Es una bodega que almacena grandes cantidades que se caracteriza por integrar y depurar información de una o más fuentes distintas

**Modelo Multidimensional:** Es una estructura para analizar los datos del negocio a través de varias dimensiones. De esta forma se logra aumentar el rendimiento de las consultas en los datos empresariales.

**Cubo OLAP:** Es una estructura de datos que supera a las bases de datos relacionales en el área de análisis de datos. Estos pueden manejar gran cantidad de datos y simultáneamente proporcionar acceso a esos datos.

**Dimensión:** Son tablas que almacenan los detalles acerca de los hechos, cada una de estas tablas tiene los campos y atributos que describen los procesos del negocio.

**Hecho:** Tabla que contiene las medidas que serán utilizados para el análisis de los datos.

**Extracción de datos:** Proceso por el cual se accede y se obtienen los datos provenientes de alguna fuente.

**Transformación de datos:** Proceso en el que se aplican reglas de negocio para convertir los datos que se cargaran en la bodega de datos.

**Carga de datos:** Proceso en el que se cargan los datos al sistema después de ser transformados.

**Calidad de Datos:** Proceso por el cual se corrobora la fiabilidad de los datos y se optimiza la comprensión de los mismos.

## **ANEXO 1**

Se anexa en la carpeta del archivo los manuales de usuario y técnico (Cartillas Digitales). De igual forma anexamos el artículo científico y las cartas correspondientes a la gestión del proyecto.

## ANEXO 2

Acta de Inicio del Proyecto.



ADOr010-V6

Página 1 de 2

21-1

### ACTA No. 001

**CLASE DE REUNIÓN:** ACTA DE INICIO PROYECTO ANALISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD DE CUNDINAMARCA SEDE FUSAGASUGA UTILIZANDO HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.

**CIUDAD Y FECHA:** FUSAGASUGÁ, 2018 -10 - 11

**HORA:** 9:40 a.m.

**LUGAR:** UNIVERSIDAD DE CUNDINAMARCA

**ASISTENTES:** Ingeniero(a) EDILSON MARTINEZ CLAVIJO  
DIRECTOR SISTEMAS Y TECNOLOGIA

Ingeniero(a) ANGELA PATRICIA ARENAS AMADO  
UNIVERSIDAD DE CUNDINAMARCA  
Director del Proyecto

ELIZABETH OJEDA MARÍN  
Estudiante Ingeniería de Sistemas

SANTIAGO GONZALEZ GOMEZ  
Estudiante Ingeniería de Sistemas

#### ORDEN DEL DIA:

1. LLAMADO A LISTA Y VERIFICACIÓN DEL QUÓRUM.
2. LECTURA Y APROBACIÓN DEL ORDEN DEL DÍA.
3. INICIO OFICIAL DEL PROYECTO ANALISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD DE CUNDINAMARCA SEDE FUSAGASUGÁ UTILIZANDO HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS.

#### DESARROLLO DE LA SESIÓN:

1. Se realiza llamado a lista de asistentes y se verifica que hay quórum para realizar la reunión.
2. Se aprueba el orden del día.
3. \*Se inicia de manera oficial el proyecto aprobado mediante Acta No. 007 del Comité de Trabajos de Grado programa de Ingeniería de Sistemas – Facultad Ingeniería de la Universidad de Cundinamarca, realizado el 04 de Octubre de 2018.  
\*Se acuerda que el seguimiento del proyecto se realizará de acuerdo al Anteproyecto (modalidad Trabajo de Grado), que fue aprobado en el Comité de Trabajos de Grado del programa de Ingeniería de Sistemas – Facultad Ingeniería de la Universidad de Cundinamarca, realizado el 04 de Octubre de 2018 mediante el Acta No. 007.

Diagonal 18 No. 20-29 Fusagasugá – Cundinamarca  
Teléfono (091) 8281483 Línea Gratuita 018000180414  
[www.ucundinamarca.edu.co](http://www.ucundinamarca.edu.co) E-mail: [info@ucundinamarca.edu.co](mailto:info@ucundinamarca.edu.co)  
NIT: 690.680.062-2



Documento controlado por el Sistema de Gestión de la Calidad

21-1

**ACTA No. 001**

\*Se acuerda, aprueba y firma el respectivo Acuerdo de Confidencialidad para el proyecto, el cual hace parte de la presente acta.  
\*Se acuerda usar una muestra de datos (sede Fusagasugá, facultad de Ingeniería, programa sistemas) para el respectivo análisis.

**COMPROMISOS Y RESPONSABLES:**


1. Se realizará como mínimo una reunión mensual para seguimiento y entrega de avances del proyecto, de acuerdo a los productos esperados aprobados en el Anteproyecto.
2. Se habilitará un repositorio Institucional para el proyecto.
3. El Ing. Edilson concreta reunión con las áreas interesadas (Autoevaluación y Acreditación, Bienestar Universitario).

**ANEXOS**

FECHA	CORREO ELECTRÓNICO	REMITENTE

Nota: Relación de comunicaciones recibidas y/o enviadas.

No siendo más el orden del día se da por terminada la reunión a las 10:10 a.m.

  
Ing. EDILSON MARTINEZ CLAVIJO  
UNIVERSIDAD DE CUNDINAMARCA  
DIRECTOR SISTEMAS Y TECNOLOGÍA

  
Ing. ANGELA PATRICIA ARENAS AMADO  
UNIVERSIDAD DE CUNDINAMARCA  
Director del Proyecto

  
ELIZABETH OJEDA MARÍN  
Estudiante Ingeniería de Sistemas

  
SANTIAGO GONZALEZ GOMEZ  
Estudiante Ingeniería de Sistemas

Copia: Archivo (repositorio institucional programa Ingeniería de Sistemas).

Anexos: Anteproyecto (Trabajo de Grado aprobado según Acta No. 007 Comité Trabajos de Grado), Acuerdo de Confidencialidad, Asistencia (ESGr015).

Elaboró: Elizabeth Ojeda Marín, Santiago González Gómez

21.1.3.39

# Acta de Entrega Final del Proyecto



AD0r010-V6

Página 1 de 3

21.1

## ACTA No. 010

**CLASE DE REUNIÓN:** ENTREGA FINAL DEL PROYECTO "ANÁLISIS DE LA DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD DE CUNDINAMARCA SEDE FUSAGASUGÁ UTILIZANDO HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS. Muestra de Datos: Facultad de Ingeniería, Programa de Ingeniería Sistemas, Plan de Estudios 2013, Periodos (IPA2013-IPA2018)".

**CIUDAD Y FECHA:** FUSAGASUGÁ, 2019 - 05 - 30

**HORA:** 9:00 a.m.

**LUGAR:** UNIVERSIDAD DE CUNDINAMARCA.

**ASISTENTES:** Ingeniero(a) EDILSON MARTINEZ CLAVIJO  
DIRECTOR SISTEMAS Y TECNOLOGÍA

Ingeniero(a) ANGELA PATRICIA ARENAS AMADO  
UNIVERSIDAD DE CUNDINAMARCA  
Director del Proyecto

ELIZABETH OJEDA MARÍN  
Estudiante Ingeniería de Sistemas

SANTIAGO GONZALEZ GOMEZ  
Estudiante Ingeniería de Sistemas

### ORDEN DEL DÍA:

1. LLAMADO A LISTA Y VERIFICACIÓN DEL QUÓRUM.
2. LECTURA Y APROBACIÓN DEL ORDEN DEL DÍA.
3. ENTREGA FINAL DEL PROYECTO.

### DESARROLLO DE LA SESIÓN:

1. Se realiza llamado a lista de asistentes y se verifica que hay quórum para realizar la reunión.
2. Se aprueba el orden del día.
3. Los estudiantes hacen entrega del Proyecto de acuerdo a los productos comprometidos mediante Anteproyecto aprobado en Comité de Trabajos de Grado de Octubre 4 del 2018 Acta N° 007.  
En su orden se hace entrega de:
  - Solución tecnológica para el Análisis de la Deserción Estudiantil a través de una herramienta de Inteligencia de Negocios

Diagonal 16 No 20-29 Fusagasugá - Cundinamarca  
Teléfono (091) 8281483 Línea Gratuita 018000180414  
[www.ucundinamarca.edu.co](http://www.ucundinamarca.edu.co) E-mail: [info@ucundinamarca.edu.co](mailto:info@ucundinamarca.edu.co)  
NIT: 800 680 062-2

Emc  
yf

**ACTA No. 010**

Mediante un escenario de negocio preparado se presentó al Ing. Edilson Martínez Gerente del Proyecto de parte de la Universidad de Cundinamarca, la solución tecnológica abarcando el proceso técnico y de usuario final, el cual se realizó con éxito.

- Manual de usuario
- Manual Técnico
- Capacitación
- Informe Técnico

Por lo anterior se hace entrega oficial del proyecto al Ing. Edilson Martínez Gerente del Proyecto de parte de la Universidad de Cundinamarca (cliente), quien en acuerdo manifiesta su aceptación del proyecto y sus productos a conformidad.

**COMPROMISOS Y RESPONSABLES:**


Ninguno.


**ANEXOS**

FECHA	CORREO ELECTRÓNICO	REMITENTE
AAAA-MM-DD		

Nota: Relación de comunicaciones recibidas y/o enviadas.

No siendo más el orden del día se da por terminada la reunión a las 12:30 p. m.

  
Ing. EDILSON MARTÍNEZ CLAVIJO  
UNIVERSIDAD DE CUNDINAMARCA  
DIRECTOR SISTEMAS Y TECNOLOGÍA

  
Ing. ÁNGELA PATRICIA ARENAS AMADO  
UNIVERSIDAD DE CUNDINAMARCA  
DIRECTOR DEL PROYECTO



**ACTA No. 010**

*Elizabeth Ojeda Marin*  
**ELIZABETH OJEDA MARÍN**  
Estudiante Ingeniería de Sistemas

*Santiago*  
**SANTIAGO GONZALEZ GOMEZ**  
Estudiante Ingeniería de Sistemas

## CONCLUSIONES

Implementar una solución de BI en la Universidad de Cundinamarca ha sentido personal se puede considerar un logro ya que no se cuenta con un área de analítica y por ende no se ha llevado a cabo en ningún proceso de la universidad hasta ahora.

El área de Inteligencia de Negocios abarca un proceso extenso del cual se aprendió durante el transcurso y las herramientas que se utilizaron cuentan con Dashboards que permiten entregar al cliente un análisis con información valiosa y entendible a simple vista, para así mismo construir estrategias y evaluar la situación analizada.

Los objetivos planteados se pueden concluir en que se identificó y se analizó con éxito la situación actual de la Deserción Estudiantil en el programa de Ingeniería de Sistemas sede Fusagasugá de la Universidad de Cundinamarca y se hizo entrega de una solución tecnológica a través de herramientas orientadas a la Inteligencia de Negocios.

La Deserción Estudiantil es una problemática que desafortunadamente seguirá existiendo en las Instituciones Educativas, por lo que resaltamos la importancia de tomar medidas en contra de esta.

Durante el transcurso se observó la importancia de los datos dentro de una organización, el valor que tienen y que pueden generar. Aún más cuando estos datos parecen estar sin uso alguno, y pueden ser convertidos en conocimiento valioso.

Finalmente podemos concluir con los resultados de la solución tecnológica:

- Detección de los estudiantes que han desertado.
- Cifra de desertores por cada tipo de deserción (Precoz, temprana, tardía)
- Cifra de desertores por cohorte.
- Identificación de las variables que son posibles causantes de la deserción estudiantil.

- Cifra real de desertores por género en la muestra evaluada (2013-1 hasta 2018-1, Ingeniería de Sistemas).

## **RECOMENDACIONES**

Junto a la entrega del análisis se hicieron recomendaciones que consideramos podrían ayudar a mitigar el índice de Deserción Estudiantil:

- Mejorar los controles que deben ser aplicados al proceso realizado por los estudiantes cuando aplican a un retiro, aplazamiento o cancelación de semestre.
- Crear o implementar un área de análisis en la Universidad de Cundinamarca.
- Implementar un examen vocacional para mitigar el riesgo de Deserción Temprana.
- Evaluar la posibilidad de ofrecer programas socio-económicos a los estudiantes admitidos que no tienen la capacidad económica para pagar la matrícula.
- Crear un software que apoye y mejore el proceso de retiro, aplazamiento o cancelación de semestre dentro de la Universidad de Cundinamarca.

## REFERENCIAS

Concepto Excel. (2018). *Concepto de Excel*. Obtenido de <https://concepto.de/excel/#ixzz5nuVdqeme>

Gartner. (2019). *Gartner*. Obtenido de <https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb>

DAVID, L. (2018). Tablas de dimensión vs tablas de hechos. Obtenido de <https://searchdatacenter.techtarget.com/es/consejo/Tablas-de-dimension-vs-tablas-de-hechos-Cual-es-la-diferenci>

DIAZ, J. (2010). INTRODUCCION AL BUSINESS INTELLIGENCE. UOC.

Microsoft. (2019). *Power BI Microsoft*. Obtenido de <https://powerbi.microsoft.com/en-us/>

Ministerio de Educación. (2009). Mineducacion. Obtenido de [https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702\\_libro\\_desercion.pdf](https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf)

Ministerio de Educación. (2010). Mineducacion. Obtenido de <https://www.mineducacion.gov.co/1621/article-217744.html>

Ministerio de Educación. (2019). *Mineducacion*. Obtenido de <https://www.mineducacion.gov.co/1621/article-196488.html>

Ministerio de Educación. (2019). *Mineducacion*. Obtenido de <https://www.mineducacion.gov.co/sistemasinfo/Informacion-Institucional/211868:Que-es-el-SNIES>

Office. (2019). *Productos Office*. Obtenido de <https://products.office.com/es-co/access>

RAMOS, S. (2017). Hechos y Dimensiones: Modelado Dimensional. Obtenido de <https://blogs.solidq.com/es/business-analytics/hechos-y-dimensiones-modelado-dimensional-12/>

Sinergia e Inteligencia de Negocio S.L.. (2019). Sinnexus. Obtenido de [https://www.sinnexus.com/business\\_intelligence/arquitectura.aspx](https://www.sinnexus.com/business_intelligence/arquitectura.aspx)

Sinergia e Inteligencia de Negocio S.L.. (2019). Sinnexus. Obtenido de [https://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](https://www.sinnexus.com/business_intelligence/olap_avanzado.aspx)

Sinergia e Inteligencia de Negocio S.L.. (2019). Sinnexus. Obtenido de [https://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](https://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx)

Statistical Analysis System. (2019). SAS. Obtenido de [https://www.sas.com/en\\_id/insights/data-management/what-is-etl.html](https://www.sas.com/en_id/insights/data-management/what-is-etl.html)

Universidad de Cundinamarca. (2016). *Ucundinamarca*. Obtenido de <https://www.ucundinamarca.edu.co/index.php/universidad/mision-y-vision>

Universidad de Cundinamarca. (2016). *Ucundinamarca*. Obtenido de <https://www.ucundinamarca.edu.co/index.php/noticias-ucundinamarca/84-institucional/746-ucundinamarca-presenta-baja-desercion-estudiantil>

Universidad de Cundinamarca. (2018). *Ucundinamarca*. Obtenido de <https://www.ucundinamarca.edu.co/index.php/noticias-ucundinamarca/84-institucional/746-ucundinamarca-presenta-baja-desercion-estudiantil>

Worldwide. (2019). Obtenido de <https://selecthub.com/Business-Intelligence-Software-Cost-and-Price-Guide.pdf>