

UNIVERSIDAD TÉCNICA DEL NORTE



Facultad de Ingeniería en Ciencias Aplicadas

Carrera de Ingeniería en Sistemas Computacionales

Desarrollar un conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte.

Trabajo de Grado previo a la obtención del título de Ingeniero en Sistemas
Computacionales

Autor:

Jonathan Javier Silva Morán

Director:

MSc. Vicente Alexander Guevara Vega

Ibarra – Ecuador

2022



UNIVERSIDAD TÉCNICA DEL NORTE BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	0401769989		
APELLIDOS Y NOMBRES:	SILVA MORAN JONATHAN JAVIER		
DIRECCIÓN:	EL CARMELO – CARCHI – ECUADOR		
EMAIL:	jxsilvam@utn.edu.ec		
TELÉFONO FIJO:	063014719	TELÉFONO MÓVIL:	0963929315

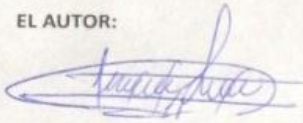
DATOS DE LA OBRA	
TÍTULO:	Desarrollar un conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte.
AUTOR (ES):	SILVA MORAN JONATHAN JAVIER
FECHA: DD/MM/AAAA	03/03/2022
SOLO PARA TRABAJOS DE GRADO	
PROGRAMA:	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
TÍTULO POR EL QUE OPTA:	INGENIERO EN SISTEMAS COMPUTACIONALES
ASESOR /DIRECTOR:	MSc. ALEXANDER GUEVARA

2. CONSTANCIAS

El autor (es) manifiesta (n) que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es (son) el (los) titular (es) de los derechos patrimoniales, por lo que asume (n) la responsabilidad sobre el contenido de la misma y saldrá (n) en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 03 días del mes de marzo de 2022

EL AUTOR:

A handwritten signature in blue ink, appearing to read 'Jonathan Silva', is written over a horizontal line.

Nombre: Jonathan Silva
C.I: 040176998-9



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CERTIFICACIÓN DEL DIRECTOR

Por medio del presente yo MSc. Alexander Guevara Vega, certifico que el Sr. Jonathan Javier Silva Morán portador de la cedula de ciudadanía número 0401769989, ha trabajado en el desarrollo del proyecto de grado **“Desarrollar un conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte”**, previo a la obtención del Título de Ingeniero en Sistemas Computacionales realizado con interés profesional y responsabilidad que certifico con honor de verdad.

Es todo en cuanto puedo certificar a la verdad

Atentamente

VICENTE
ALEXANDER
GUEVARA VEGA

Firmado digitalmente por
VICENTE ALEXANDER
GUEVARA VEGA
Fecha: 2022.03.08
11:14:45 -05'00'

.....

MSc. ALEXANDER GUEVARA
DIRECTOR DE TRABAJO DE GRADO



UNIVERSIDAD TÉCNICA DEL NORTE

Resolución No 173-SE-33- CACES – 2020

DEPARTAMENTO DE DESARROLLO TECNOLÓGICO E INFORMÁTICO

DIRECTOR DE LA DIRECCIÓN DE DESARROLLO TECNOLÓGICO E INFORMÁTICO

CERTIFICA

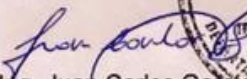
QUE: El señor JONATHAN JAVIER SILVA MORAN con cédula identidad 0401769989 estudiante de la Facultad de Ingeniería en Ciencias Aplicadas – de la Carrera de Ingeniería en Sistemas Computacionales, ha desarrollado con los datos entregados de la Dirección de Desarrollo Tecnológico e Informático, el Proyecto de Tesis "DESARROLLAR UN CONJUNTO DE TABLEROS VISUALES DE BUSINESS INTELLIGENCE (BI) MEDIANTE LA HERRAMIENTA DE TABLEAU QUE PERMITA FORTALECER LA INTERPRETACIÓN DE LA INFORMACIÓN DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DEL NORTE".

QUE: El señor Silva entrega el informe ejecutivo del proyecto de investigación, a la ingeniera Evelin Enríquez MSc. – funcionario de la Dirección de Desarrollo Tecnológico e Informático el 2 de marzo del 2022.

Es todo cuanto puedo certificar, facultando al interesado hacer uso de este certificado como estime conveniente, excepto para trámites judiciales.

Ibarra, 3 de marzo del 2022

Atentamente
CIENCIA Y TÉCNICA AL SERVICIO DEL PUEBLO


Ing. Juan Carlos García
DIRECTOR





UNIVERSIDAD TÉCNICA DEL NORTE

Ibarra-Ecuador

DEPARTAMENTO DE BIENESTAR UNIVERSITARIO

MIEMBROS DE LA COMISIÓN ASESORA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES
Presente

De nuestra consideración,

Por medio de la presente, tengo a bien informar que el proyecto de tesis: **“Conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte.”**, desarrollado por el estudiante de la Carrera de Ingeniería en Sistemas Computacionales, señor **JONATHAN JAVIER SILVA MORÁN** CI. 0401769989, se ha recibido satisfactoriamente. El estudiante realizó la socialización y entrega respectiva al personal del departamento de Bienestar Universitario. Los resultados presentados estuvieron a la altura, superando en varios aspectos las expectativas de nuestro equipo de trabajo. Por este motivo, el departamento de bienestar universitario decidió seguir utilizando la solución de BI como instrumento de visualización y análisis de datos relacionados con la deserción estudiantil.

Por lo expuesto anteriormente, agradecemos su cooperación y esperamos continuar trabajando en equipo para seguir fortaleciendo nuestras áreas de experticia.

Atentamente,



Escanea este código QR para:
**MARTHA ELIZABETH
MOLINA ORDÓNEZ**

Dra. Martha Molina O.
DIRECTORA (E) DEPARTAMENTO BIENESTAR UNIVERSITARIO

Av. 17 de Julio 5 - 21 y José María Córdova
Ciudadela Universitaria Barrio El Olivo
Teléfono: +593 6 299 78 00 Casilla 199
E-mail: info@utn.edu.ec
www.utn.edu.ec
Ibarra - Ecuador

DEDICATORIA

Por el esfuerzo, valor, amor y confianza que me ha brindado de manera incondicional en todo momento de mi vida y mi carrera universitaria y a quien ha sido mi soporte guiándome con su ejemplo y perseverancia. Por eso y más este proyecto de titulación va dedicado a ese ser maravilloso mi Madre.

AGRADECIMIENTOS

A Dios por guiarme por el camino correcto en cada paso de mi vida, por darme salud, bienestar y sabiduría para culminar esta etapa universitaria.

A mi madre, un agradecimiento eterno por su entrega, dedicación, apoyo y sacrificio en toda mi carrera universitaria y en este proceso de titulación.

A mi tutor de tesis, MSc. Alexander Guevara por el apoyo y dirección brindado, por los conocimientos impartidos dentro y fuera de clase, que sin duda han sido los factores clave para el desarrollo de este proyecto.

A la Universidad Técnica del Norte y a cada uno de los docentes de la Carrera de Ingeniería en Sistemas Computacionales que me han impartido sus conocimientos académicos y profesionales.

Al Departamento de Desarrollo Tecnológico e Informático y al Departamento de Bienestar Universitario por darme la apertura necesaria para el desarrollo de este proyecto.

A todos mis amigos y compañeros con los que he compartido risas y gratos momentos a lo largo de esta etapa de mi vida.

TABLA DE CONTENIDO

INTRODUCCIÓN	1
Tema	1
Planteamiento del Problema.....	1
Objetivos	2
Objetivo General	2
Objetivos Específicos.....	2
Alcance	2
Justificación.....	4
CAPÍTULO I	5
MARCO TEÓRICO	5
1.1 Deserción Académica	5
1.2 Los Datos	5
1.2.1 Clasificación de los datos	6
1.3 Sistemas de Apoyo a las Decisiones (DSS).....	6
1.3.1 Tipos de DSS.....	6
1.4 Business Intelligence	7
1.4.1 Definición.....	8
1.4.2 Arquitectura de un Sistema BI	8
1.4.3 Estrategias y Beneficios de Business Intelligence.....	11
1.4.4 Herramientas para Business Intelligence	12
1.5 Procesos ETL y Procedimientos Almacenados.....	13
1.5.1 Historia de los ETL.....	13
1.5.2 Definición de los ETL.....	14
1.5.3 Procedimientos Almacenados.....	15
1.6 Metodología Kimball	16
1.6.1 Ralph Kimball y su enfoque BI	16
1.6.2 Fases de la Metodología Kimball	17
1.7 Herramienta Tableau en la deserción estudiantil	19
1.7.1 Tableau.....	19
CAPÍTULO II	20
MATERIALES, METODOS Y PROCEDIMIENTOS	20
2 Desarrollo	20
2.1 Diagrama de Proceso del proyecto	20
2.2 Marco de trabajo Scrum.....	21

2.2.1	Sprint	21
2.2.2	Roles Scrum	21
2.3	Integración de Scrum y Kimball como base en el desarrollo del proyecto	21
2.4	Planificación y Gestión del Proyecto	22
2.4.1	Administración del Proyecto.....	22
2.4.2	Plan del Proyecto.....	23
2.4.3	Definición de Requerimientos del Negocio.....	24
2.4.4	Levantamiento preguntas (Necesidades).....	31
2.5	Sprint del Proyecto.....	32
2.5.1	Descripción y Ejecución de Sprint.....	33
2.6	Elaboración del Data Warehouse con la metodología Kimball	35
2.6.1	Exploración de los datos de Origen	35
2.6.2	Creación de la réplica en el área de desembarco n.....	37
2.6.3	Modelamiento Dimensional.....	37
2.6.4	Diseño Físico	38
2.6.5	Extracción, transformación y carga de los datos (ETL, SP).....	39
2.7	Elaboración de la Data Warehouse y Data Marts.....	54
2.7.1	Creación del modelo del área de análisis.....	54
2.7.2	Tablas Dimensión	54
2.7.3	Diseño y elaboración de Data Marts	55
2.7.4	Llenado de Data Warehouse (DW)	56
2.8	Diseño de Aplicación BI	58
2.8.1	Construcción de tableros visuales en Tableau	60
2.8.2	Identificación de las métricas a implementar en el tablero	60
2.8.3	Despliegue.....	68
2.8.4	Crecimiento y Mantenimiento.....	73
CAPÍTULO III.....		74
VALIDACIÓN DE RESULTADOS		74
3	Análisis y Validación de Resultados.....	74
3.1	Evaluación de Producto con metodología CSAT.....	74
3.2	Calificación de Producto con metodología NPS	77
3.3	Análisis de impactos	78
CONCLUSIONES		80
RECOMENDACIONES.....		81
REFERENCIAS Y BIBLIOGRAFIA		83
ANEXOS.....		87

Índice de Figuras

Fig. 1. Diagrama de Problemas.....	1
Fig. 2. Arquitectura de Funcionamiento.....	3
Fig. 3. Capas del Sistema Inteligencia	9
Fig. 4. Componentes de un Sistema BI.....	10
Fig. 5. Proceso ETL - Fuente propia basado en (Bernabeu & Mattío, 2017)	14
Fig. 6. Ciclo de Vida de Kimball – Fuente propia basado en (M. & P., 2014).....	18
Fig. 7. Diagrama de Procesos.....	20
Fig. 8. Metodología de desarrollo del proyecto.....	22
Fig. 9. H1 Levantamiento de necesidades	24
Fig. 10. H2 Gráfico estadístico de deserción.....	25
Fig. 11. H3 Gráfico estadístico de deserción.....	25
Fig. 12. H4 Grafico estadístico de deserción por provincia.....	25
Fig. 13. H5 Gráfico estadístico de deserción por tipo de institución.....	26
Fig. 14. H6 Gráfico estadístico de deserción por rango de ingreso	26
Fig. 15. H7 Gráfico estadístico de deserción por facultad	26
Fig. 16. H8 Gráfico estadístico de deserción por etnia	27
Fig. 17. H9 Gráfico estadístico de deserción por estado civil	27
Fig. 18. H10 Gráfico estadístico de deserción por tipo de financiación.....	27
Fig. 19. H11 Gráfico estadístico de deserción por tipo de vivienda	28
Fig. 20. H12 Gráfico estadístico de deserción por nivel.....	28
Fig. 21. H13 Gráfico estadístico de deserción por tipo de convivencia.....	28
Fig. 22. H14 Gráfico estadístico de deserción si dispone internet	29
Fig. 23. H15 Gráfico estadístico de deserción por número de matrícula	29
Fig. 24. H16 Dashboards de deserción de persona	29
Fig. 25. H17 Dashboards de deserción socioeconómico.....	30
Fig. 26. H18 Dashboards de deserción académico	30
Fig. 27. H19 Dashboards General de deserción.....	30
Fig. 28. Organización de Sprint.....	35
Fig. 29. Base de datos importada.....	36
Fig. 30. Matriz Bus de las dimensiones.....	37
Fig. 31. Diagrama Entidad Relación.....	38
Fig. 32. Área de Trabajo	39
Fig. 33. SQL Procedimiento Almacenado.....	40
Fig. 34. Diagrama Área de análisis	54
Fig. 35. Carga de Datos al Data Warehouse.....	57
Fig. 36. Instalación de Tableau	58
Fig. 37. Conexión a SQL Server	59
Fig. 38. Importación de SQL a Tableau.....	59
Fig. 39. Construcción de DW	60
Fig. 40. Identificación de Métricas.....	61
Fig. 41. Métrica Edad.....	61
Fig. 42. Métrica Estado Civil.....	62
Fig. 43. Métrica Etnia	62
Fig. 44. Métrica Facultad.....	63
Fig. 45. Métrica Financiamiento	63
Fig. 46. Métrica Género	64
Fig. 47. Métrica Tipo Vivienda.....	64
Fig. 48. Métrica Conviviente.....	65
Fig. 49. Métrica Servicio de Internet.....	65

Fig. 50. Métrica Provincia de Procedencia	66
Fig. 51. Métrica Numero Matricula	66
Fig. 52. Métrica Nivel	67
Fig. 53. Métrica Ingreso Mensual	67
Fig. 54. Métrica Tipo Institución	68
Fig. 55. Tablero Académico	69
Fig. 56. Tablero Persona.....	69
Fig. 57. Tablero Socioeconómico 1	70
Fig. 58. Tablero Socioeconómico 2	70
Fig. 59. Reunión para presentar la versión final	71
Fig. 60. Deserción Estudiantil en base a la información Académica.....	71
Fig. 61. Deserción Estudiantil en base a la información de la ficha socioeconómica.....	72
Fig. 62. Deserción Estudiantil en base a la información de la ficha socioeconómica 2.....	73
Fig. 63. Pregunta de Validación N°1	75
Fig. 64. Pregunta de Validación N°2	76
Fig. 65. Pregunta de Validación N°3	76
Fig. 66. Pregunta de Validación N°4	77

Índice de Tablas

Tabla 1. Definiciones Business Intelligence	8
Tabla 2. Beneficios en el entorno académico de la UTN	11
Tabla 3. Estrategias de un Sistema BI	12
Tabla 4. Herramientas utilizadas en Business Intelligence	12
Tabla 5. Principios de la Metodología Kimball	17
Tabla 6. Roles Scrum del Proyecto	23
Tabla 7. Estimación de tareas técnica T-Shirt Sizes	24
Tabla 8. Product Backlog del Proyecto	31
Tabla 9. Variables y Tipos de Datos.....	32
Tabla 10. Desarrollo del Proyecto a través de Sprint	32
Tabla 11. Desarrollo de Sprint.....	33
Tabla 12. Exploración de Tablas	36
Tabla 13. Niveles de Granularidad del Proyecto	38
Tabla 14. Catálogo Edad	41
Tabla 15. Catálogo Genero	41
Tabla 16. Tabla Homologada Género	42
Tabla 17. Catálogo Provincia	42
Tabla 18. Catálogo Institución.....	43
Tabla 19. Tabla Homologadora Institución	43
Tabla 20. Catálogo Ingreso Mensual.....	43
Tabla 21. Catálogo Facultad	44
Tabla 22. Catálogo Etnia.....	44
Tabla 23. Tabla Homologadora Etnia.....	45
Tabla 24. Catalogo Estado Civil	45
Tabla 25. Tabla Homologadora Estado Civil	46
Tabla 26. Catálogo Financiamiento.....	46
Tabla 27. Tabla Homologadora Financiamiento	47
Tabla 28. Catálogo Tipo Vivienda	47
Tabla 29. Tabla Homologadora Tipo Vivienda.....	48
Tabla 30. Catalogo Nivel.....	48
Tabla 31. Catalogo Convivencia	49
Tabla 32. Catalogo Servicio Internet	50
Tabla 33. Tabla Homologadora Internet.....	51
Tabla 34. Catalogo Número Matrícula.....	51
Tabla 35. Tabla Homologadora Número Matrícula	51
Tabla 36. Tabla Maestra Estudiante.....	52
Tabla 37. Tabla Maestra Matrícula.....	52
Tabla 38. Tabla Maestra Socioeconómica	52
Tabla 39. Tabla Detalle Estudiante	53
Tabla 40. Tabla Detalle Matrícula.....	53
Tabla 41. Tabla Detalle Socioeconómica	53
Tabla 42. Dimensión Edad.....	54
Tabla 43. Atributos correspondientes al Datamart Persona.....	55
Tabla 44. Atributos correspondientes al Datamart Académica	56
Tabla 45. Atributos correspondientes al Datamart Socioeconómica.....	56
Tabla 46. Atributos que componen el DataWarehouse	56

Resumen

La deserción estudiantil en las Instituciones de Educación Superior (IES) es un fenómeno que afecta a la educación y que depende de un exitoso acoplamiento entre la institución y el individuo. A esta problemática se le debe dar un tratamiento oportuno a través de la Business Intelligence cuya finalidad es obtener soluciones, y reducir el alto índice de deserción. En el presente trabajo de titulación se utilizó la Business Intelligence donde se obtuvieron indicadores claves de deserción estudiantil a través de la combinación de análisis de datos, visualización de datos, herramientas y buenas prácticas para la toma de decisiones. Se construyó un Data Warehouse basado en la metodología Kimball con los datos adquiridos de la base transaccional de la Universidad Técnica del Norte alojada en los servidores de Oracle, con la finalidad de exportar toda esa información a la herramienta Tableau Public y proceder con el análisis de los tableros estadísticos. Por último, para validar la satisfacción con los tableros se realizaron encuestas al Departamento de Bienestar Universitario (DBU) y a la Dirección de Desarrollo Técnico e Informático (DDTI) validando los resultados con la metodología CSAT y demostrando así la satisfacción, funcionalidad y viabilidad del proyecto realizado. Con esto se fortaleció la interpretación de la información a través de los tableros visuales permitiendo tomar decisiones estratégicas acordes a la deserción estudiantil.

Palabras Clave: Deserción Estudiantil, Business Intelligence, Tableau Public, Kimball, Scrum

Abstract

Student desertion in Higher Education Institutions (IES) is a phenomenon that affects education and that depends on a successful coupling between the institution and the individual. This problem must be given timely treatment through Business Intelligence whose purpose is to obtain solutions and reduce the high rate of attrition.

In this degree work, Business Intelligence was used where key indicators of student dropout were obtained through the combination of data analysis, data visualization, tools and good practices for decision making.

A Data Warehouse based on the Kimball methodology was built with the data acquired from the transactional database of the Universidad Técnica del Norte hosted on Oracle servers, to export all this information to the Tableau Public tool and proceed with the analysis of statistical dashboards.

Finally, to validate the satisfaction of the boards, surveys were carried out to the University Welfare Department (DBU) and to the Directorate of Technical and Computer Development (DDTI), validating the results with the CSAT methodology and thus demonstrating the functionality and viability of the project carried out.

Keywords: Student Dropout, Business Intelligence, Tableau Public, Kimball, Scrum

INTRODUCCIÓN

Tema

“Desarrollar un conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte”.

Planteamiento del Problema

Al existir un bajo índice de visualización de los patrones de estudiantes que desertan en la UTN, la problemática de la deserción estudiantil sigue latente, la difícil comprensión por los usuarios finales y la falta de conocimiento de métodos que permitan interpretar estos datos demuestra que el problema sigue sin ser tratado y sigue afectando al estado ecuatoriano por ser una universidad pública que está financiada por el gobierno. La Universidad Técnica del Norte no cuenta con un entorno BI que permita interpretar de manera fácil y eficiente los patrones de deserción encontrados sobre la deserción estudiantil.

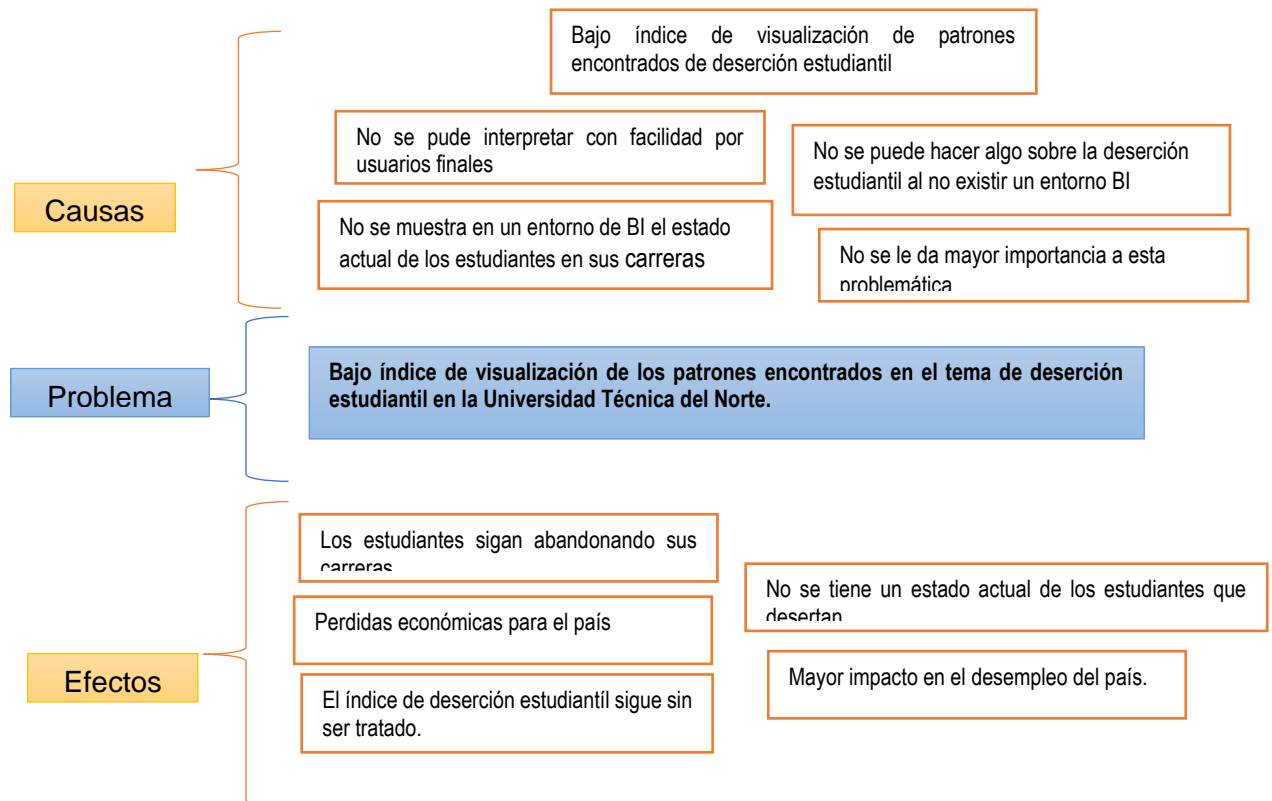


Fig. 1. Diagrama de Problemas

Objetivos

Objetivo General

Desarrollar un conjunto de tableros visuales de Business Intelligence (BI) mediante la herramienta de Tableau que permita fortalecer la interpretación de la información de deserción estudiantil en la Universidad Técnica del Norte.

Objetivos Específicos

- Elaborar un marco teórico que sustente la investigación de Inteligencia Empresarial (BI) para la interpretación de la información académica, socioeconómica y personas.
- Construir un Data Warehouse, a partir de información académica, socioeconómica y persona de los estudiantes de nivel de pregrado de la Universidad Técnica del Norte, mediante Kimball como metodología de BI.
- Desarrollar un conjunto de tableros de BI mediante la herramienta de Tableau que permita visualizar indicadores de deserción estudiantil en la Universidad Técnica del Norte, aplicando Scrum como marco de trabajo del proyecto.
- Validar los resultados de la investigación propuesta.

Alcance

El Proyecto consiste en construir un conjunto de tableros con la herramienta de análisis integral Tableau para visualizar los patrones de deserción estudiantil en la Universidad Técnica del Norte.

Se propone trabajar con la metodología Kimball ya que es una metodología que puede y ha sido utilizado en el desarrollo de bases de datos (Hart, 2016), esto para la implementación de Data Warehouse y posteriormente construir un entorno visual que permita ser interpretado por un usuario final que consta de un conjunto de tableros. La metodología Kimball parte de la base de datos transaccional de la Universidad Técnica del Norte de la cual se hará una copia idéntica en una base de datos que se llamará Storage, posteriormente mediante ETL y procedimientos almacenados se plantea construir una base de datos limpia

con la que nos permita trabajar o aplicar la metodología Kimball plantada, esta base de datos se llamará área de trabajo. La metodología Kimball construye primero los Data Marts y posteriormente construye el Data Warehouse (Hart, 2016) un fin que es la visualización de la información en un tablero que se pueda comprender por personas que no sea afines a la carrera de Ingeniería en Sistemas.

Si se siguen alimentando sets planos que sean de difícil interpretación no se estará buscando erradicar el problema de la deserción estudiantil que comúnmente es muy alto en el Ecuador, no se podrá tomar decisiones que sea acertadas y ayuden a disminuir esta problemática.

Al no existir un entorno visual se limita mucho el proceso de interpretación ya que se depende de un elemento muy técnico como es aprender un lenguaje de interpretación o técnicas para poder representar los datos que se quiere visualizar. Por ello los usuarios finales no podrán hacer uso eficiente de este recurso ya desarrollado.

Arquitectura de Funcionamiento

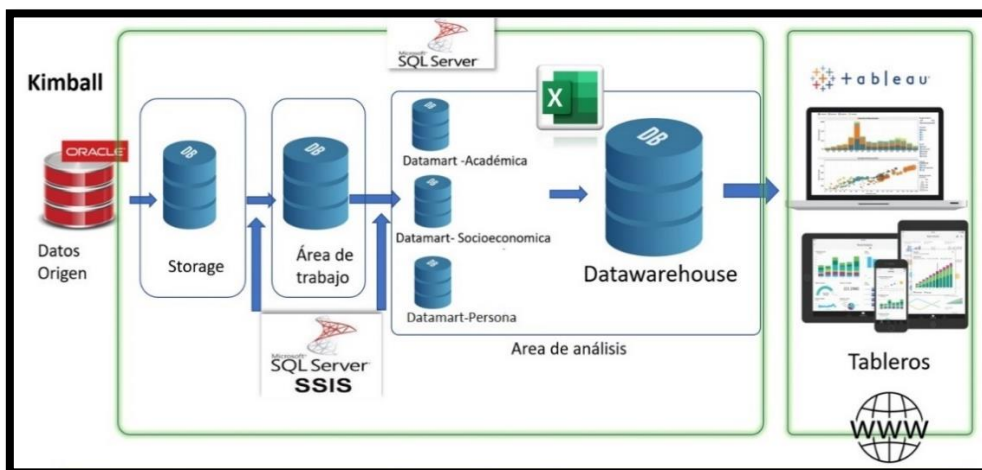


Fig. 2. Arquitectura de Funcionamiento

Justificación

El principal sustento para el proyecto es que está ligado al plan nacional toda una vida específicamente en el objetivo 1, donde el estado garantiza una vida digna con igualdad de oportunidades para todas las personas, haciendo énfasis en los servicios sociales como la educación de calidad y mayor oportunidad para todos (Plan toda una vida, 2017).

Otro sustento básicamente en uno de los ODS de Ecuador (ODS 4) que es educación de calidad que promueve oportunidades de aprendizaje durante toda la vida para todos (Naciones Unidas, 2016).

También existen otras áreas que justifican el proyecto:

Justificación Tecnológica. – Los avances tecnológicos hacen posible que se facilite el trabajo de las personas y por consecuencia se busca aprovechar al máximo la tecnología, habiendo un problema tan evidente como la deserción estudiantil y existiendo una forma de combatir ese problema con la ayuda de herramientas de BI, es una muy buena elección ayudar desde el área de informática y Tableau considerada una de las mejores herramientas del mercado (Ko & Chang, 2017).

Justificación Teórica. - Esta investigación se la realiza con el propósito de fortalecer el avance tecnológico y ser de referencia para futuras investigaciones que propongan el uso de BI en los negocios de la zona, por ello se busca investigar aportaciones anteriores validas y de calidad que serán de utilidad para la aplicación del proyecto que se va a desarrollar.

Justificación Metodológica. – Entre las metodologías de investigación que se usara están, investigación de campo, investigación documental que engloba lo que es libros, artículos y foros, también tiene un enfoque cualitativo ya que describe como se transforman los datos para que sean comprendidos por personas naturales y terminando con la exploración de temas poco tratados como es BI debido a que es nuevo en el medio (Alejandra & Claudia, 2014).

CAPÍTULO I

MARCO TEÓRICO

1.1 Deserción Académica

El término deserción hace referencia al abandono de actividades, causas, ideales, etc., dependiendo del enfoque que se le dé. Si hablamos de la Educación Superior la deserción es el alejamiento sucesivo de estudios académicos por parte del estudiante que culmina en el abandono de su carrera.

“El sistema de educación superior ecuatoriano se enfrenta a una paradoja: cada año 133.000 jóvenes que quieren ingresar a la universidad pública no lo logran, mientras que de los que sí lo lograron, 23.000 optan por desertar del sistema” (Machado, 2019).

Según la secretaria nacional de Educación Superior Ciencia y Tecnología (SENESCYT) existen varias causas para la deserción como la situación económica y la falta de conocimiento sobre las diversas carreras existentes. En Ecuador la oferta académica está compuesta por 2.037 carreras, pero el 60% de los aspirantes opta por 20 carreras tradicionales.

La deserción en Universidades depende de un exitoso acoplamiento entre la institución y el individuo. Este acoplamiento está mediado por factores individuales e institucionales. A esta problemática se le debe dar un tratamiento oportuno a través de la Business Intelligence para obtener soluciones, y reducir el alto índice de deserción.

1.2 Los Datos

Los datos son atributos, palabras o letras, que representan a una variable y no describen hechos que por sí solos, no tiene mayor relevancia. Sin embargo, almacenar una colección de datos permite examinar, organizar y tabular hechos que pueden llegarse a convertir en información útil.

“Los datos en bruto raramente son beneficiosos directamente”. Su verdadero valor se basa en la habilidad para extraer información, y la comprensión del fenómeno gobernante en la fuente de datos (Riquelme, Ruiz, & Gilbert, 2006).

1.2.1 Clasificación de los datos

Los datos se clasifican de acuerdo con la naturaleza que representan:

1.2.1.1 Cuantitativos

Representan valores numéricos o de magnitud:

- Discretos: toman un número limitado de valores diferentes.
- Continuos: para cualquier par de valores siempre existirá un intermedio

1.2.1.2 Cualitativos

Realizan representaciones de categorías:

- Nominales: Existe una asignación parcial de elementos de una categoría.
- Ordinales: Existe una relación directa con el orden entre las categorías.

1.3 Sistemas de Apoyo a las Decisiones (DSS)

Las Instituciones de Educación Superior (IES), deben estar en constantes mejoras, para obtener una acreditación a nivel nacional, con evaluaciones tanto internas como externas, esto implica tener informes sobre la situación académica de los estudiantes para buscar estrategias y tomar decisiones que permitan reducir el alto índice de deserción estudiantil.

Los DSS son sistemas de información desarrollados para solucionar problemas de decisiones utilizando datos; cuentan con herramientas para evaluar, analizar y comparar los datos con el fin de obtener información requerida. Ayudan en la toma de decisiones a través de la integración de diferentes fuentes de información (López de Munain et al., 2014).

1.3.1 Tipos de DSS

- a. Sistemas de información gerencial (MIS). Dan soporte a tareas organizacionales se encargan de brindar soporte y asesoramiento a clientes y usuarios.
- b. Sistemas de información ejecutiva (EIS). Combinan muchos de los atributos de los sistemas de información gerencial y de los sistemas de apoyo a las decisiones (Navarro & Perez, 2018), son los más empleados en Bussines Intelligence, debido a su facilidad de acceso a la información interna y externa de una compañía.

- c. Sistemas expertos basados en inteligencia artificial (SSEE). Utilizan redes neuronales para simular el conocimiento de un experto y utilizarlo de forma efectiva para resolver un problema concreto.
- d. Sistemas de apoyo a decisiones de grupo (GDSS). Es "un sistema basado en computadoras que apoya a grupos de personas que tienen una tarea (u objetivo) común, y que sirve como interfaz con un entorno compartido" ("Sistemas de Soporte a la Decisión (DSS)," n.d.).

1.4 Business Intelligence

De la misma manera que el cuerpo humano ingiere alimentos, los procesa, desecha los elementos que no son útiles ni valiosos y asimila los nutrientes necesarios para su supervivencia, las empresas requieren de información para nutrirse, alimentarse y crecer en el mundo de los negocios que avanza velozmente (López Benítez, 2018).

En los últimos años con los avances de la tecnología, la mayoría de las instituciones almacenan información de alto valor que les permite sacar ventaja competitiva, a través de la recolección, análisis y depuración de dicha información que les sirve como apoyo en la toma de decisiones, dejando a tras la intuición. Explotar la información ha pasado de ser una necesidad a una prioridad de máxima relevancia (Curto Díaz, 2016).

Tomando como ejemplo la cantidad de información que almacenan las empresas, se puede decir que las universidades no son la excepción, ya que generan grandes volúmenes de datos que son almacenados en espacios tanto tecnológicos como físicos, y que en su mayoría no se obtiene el máximo beneficio que brinda, desaprovechando dicha información.

En la Universidad Técnica del Norte (UTN), no se da un correcto tratamiento a la deserción estudiantil debido a que se basan en estadísticas y probabilidades. Es aquí donde se requiere el uso de Business Intelligence con el objetivo de ayudar y fortalecer la problemática existente aprovechando la cantidad de datos almacenados.

1.4.1 Definición

La Business Intelligence se la conoce por sus siglas en inglés BI (Business Intelligence), es un concepto usado desde mucho antes, y es considerado una evolución de los sistemas de soporte a las decisiones DSS (Decissions Suport Systems). En la Tabla 1 se establecen definiciones de varios autores:

Tabla 1. Definiciones Business Intelligence

Autor	Fecha	Definición
IBM	1958	La habilidad de aprehender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada
Howard Dresden	1989	Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos
(Conesa Caralt & Curto Díaz)	2013	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
(Camargo, Joyanes, & Giraldo)	2016	Conjunto de aplicaciones, estrategias y metodologías útiles para convertir los datos en información

Es así como el concepto de BI ha ido evolucionando, añadiendo diferentes términos y herramientas basándose en las definiciones establecidas anteriormente, estableciendo una definición de uso acorde al presente. Esta información se transforma en conocimiento con el objetivo de impulsar planes de acción rentables a nivel empresarial para gestionar contenidos.

La Business Intelligence es aplicada en cualquier ámbito donde se almacene gran cantidad de información y que a través de ella se pueda optimizar datos, generar impulsos y gestionar oportunidades transformadas en decisiones que permitan alcanzar el éxito como empresa, organización o institución.

1.4.2 Arquitectura de un Sistema BI

Un sistema de BI se compone de actividades relacionadas con los sistemas base de una organización, de los que se requiere optimizar procesos mediante una transformación

estructural, incluyendo aplicaciones analíticas, conocimientos, minería de datos, reportes y Data Warehouse.

Contiene una fase de extracción, transformación y carga de datos denominada ETL, esta fase se apoya en un Data Warehouse que sirve como almacén entre los sistemas fuente y destino evitando la saturación de servidores (Sinnexus, 2017).

Una vez que se realiza el proceso ETL, toda la información es unificada, depurada y almacenada en un Data Warehouse con el objetivo de construir diferentes Data Marts para el análisis de datos. Los datos almacenados se explotan utilizando herramientas de análisis o reporting (Thaís Balagueró, 2018).

Como se evidencia en la Fig.3. La arquitectura de estos sistemas se basa en tres capas que son:

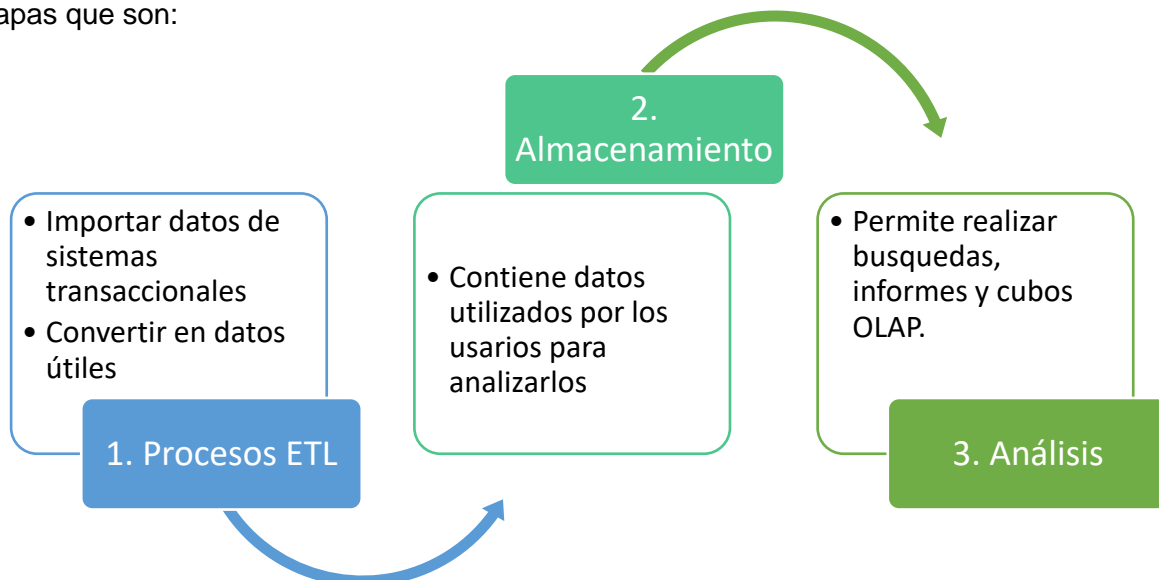


Fig. 3. Capas del Sistema Inteligencia
Fuente: (Barrios, 2019)

1.4.2.1 Componentes de un Sistema de Business Intelligence

La arquitectura de los sistemas de BI se compone de diferentes herramientas que se relacionan entre sí, facilitando la elaboración de estrategias para la toma de decisiones, presentando informes y cuadros de mando como se representa en la Fig. 4.

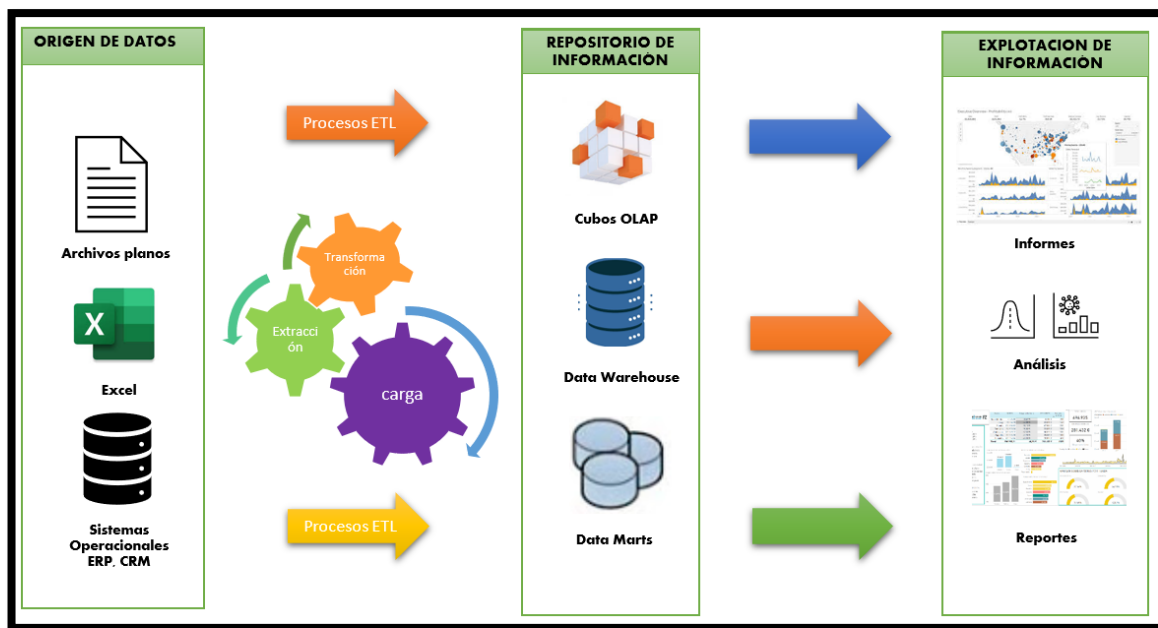


Fig. 4. Componentes de un Sistema BI

Cada componente cumple con un rol específico en el proceso de analizar datos almacenados y transformarlos en conocimiento.

- **Fuentes de Datos:** los datos contienen gran cantidad de información valiosa que son clave en el proceso de BI.
- **Procesos ETL:** conjunto de procesos que en sus siglas en inglés se denominan como Extract=extracción, Transform=transformación y Load=carga o almacenamiento de datos.
- **Data Warehouse:** es un almacén de datos construido a partir de diferentes fuentes de datos, está diseñado para organizar y optimizar los datos facilitando su análisis
- **Data Marts:** conjunto de datos pequeños, clasificados en diferentes contenedores específicos de cada área de análisis.
- **OLAP:** proceso analítico en línea, se los denomina cubos OLAP que permiten agrupar, agregar y ordenar los datos para su completa visualización.
- **Herramientas de visualización:** son informes que agrupan y relacionan gráficos de diferentes tipos para los usuarios finales.

1.4.3 Estrategias y Beneficios de Business Intelligence

El uso de Business Intelligence se vuelve un factor clave que aporta grandes beneficios en su aplicación a las diferentes instituciones u organizaciones permitiéndoles ser más competitivas.

1.4.3.1 *Business Intelligence en el entorno académico*

En el caso de la UTN la Business Intelligence presenta beneficios basados en tiempos, costos, estrategias y análisis de información que solventen las necesidades tanto de la Institución como de los estudiantes. En la Tabla 2 se muestra algunos de los beneficios.

Tabla 2. Beneficios en el entorno académico de la UTN

Beneficios de Business Intelligence	
Capacidad en la toma de decisiones	- Información actualizada y relevante, permiten tomar decisiones ajustadas a la realidad del estudiante.
Reducción de costos	- Análisis del desempeño académico estudiantil en base a su historial académico - Aprovechamiento de recursos a través de información clave disponible para su uso inmediato. - Menor infraestructura de almacenamiento aprovechando las nuevas tecnologías con información bien estructurada
Agilidad de tiempos de respuesta	- Permite la visualización de errores con formatos visuales atractivos. - Extracción de información relevante sobre la situación del estudiante a través de las fichas socioeconómicas
Comprensión de resultados	- Permitir una visión única, conformada, histórica, persistente y de calidad de toda la información académica.
Información flexible	- Asimilación de la información en cortos periodos de tiempo, a través de herramientas que brindar mejor entendimiento de los datos almacenados.
Métricas e indicadores	- Crear métricas e indicadores claves de rendimiento. - Manejar indicadores de deserción estudiantil para la implementación de estrategias con respuestas ágiles a los problemas académicos.
Mejora la Competitividad	- Acceso rápido a la información y calidad en la toma de decisiones, mejoran la estructura organizacional de la institución volviéndose una competencia con otras.

1.4.3.2 Estrategias para implementar un Sistema de Business Intelligence

Es necesario establecer estrategias que permitan implementar un sistema de BI útil y eficiente que genere resultados óptimos para conocer las fortalezas y debilidades de la organización, mejorando la toma de decisiones, transformando información en conocimiento.

En la Tabla 3 se describe las estrategias y su objetivo de implementación en un Sistema BI.

Tabla 3. Estrategias de un Sistema BI

N°	Estrategia	Objetivo
1	Comprender la diferencia entre datos, información y conocimiento.	<u>Datos</u> : valor simbólico <u>Información</u> : conjunto de datos organizados y estructurados <u>Conocimientos</u> : conocimientos que se obtiene después de analizar la información
2	Conocer las necesidades de la empresa para implementar herramientas BI	Comprender en qué estado se encuentra la compañía para implementar un sistema BI y tomar decisiones estratégicas.
3	Conocer diferentes herramientas de BI	permite realizar reportes de información a través de tableros de control, analizando datos en tiempo real.

1.4.4 Herramientas para Business Intelligence

En los últimos años la tecnología ha ido evolucionando, y el mercado de BI no es la excepción, ha presentado software con nuevas y mejores soluciones que cubren las necesidades de una organización para explotar información y transformarla en conocimiento.

En la Tabla 4 se presentan las soluciones más destacadas del mercado:

Tabla 4. Herramientas utilizadas en Business Intelligence

Nombre Herramienta	Descripción
Pentaho Community Edition	Pentaho es una de las suites más completas y maduras del mercado OSBI que existe desde el año 2006. Ofrece elementos básicos y útiles para BI, entre ellas están la visualización de gráficos de dispersión y los reportes en diferentes formatos (Millalen Alejandro, 2018).
Tableau Public	Plataforma gratuita para visualizar y compartir datos, permite importar y comprender datos de Excel, Azure y archivos CSV, también se puede crear mapas, dashboards en vivo y gráficos interactivos en poco tiempo (Tableau Software, 2020).

Rapid Miner	Plataformas de ciencia de datos que ofrece un impacto empresarial para generar ingresos, reducir costos y evitar riesgos. Ofrece versiones gratuitas para sus tres soluciones; RapidMiner Studio, RapidMiner Server y RapidMiner Radoop (RapidMiner, 2019).
Microsoft Power BI	Ofrece una versión simplificada y gratuita con capacidad de transformar datos CSV y hojas de cálculo de Excel, se actualiza a diario y puede obtener esta información a través de una aplicación móvil nativa. Permite la creación de informes memorables personalizados con KPI y marca (Microsoft, 2019).

1.5 Procesos ETL y Procedimientos Almacenados

Para analizar el historial académico y el motivo de la deserción estudiantil de la UTN, se requiere de indicadores claves localizados en la acumulación de datos. Algunos de los factores a tomar en cuenta son: la carrera, los semestres que tiene la carrera y los que ya han sido aprobados por el estudiante desde su inicio hasta su fin, es decir todo su paso por la institución, esto permite tener un punto de referencia histórica del estudiante para orientar la toma de decisiones.

Todos esos datos se incrementan constantemente por cada semestre y por cada estudiante, debido a eso se requiere buscar la forma eficaz de explorar esa cantidad de datos, a través de procesos ETL o procedimientos almacenados, con el fin de convertirlos en conocimiento útil y enfocarlo a la solución o disminución de la deserción estudiantil, estableciendo estrategias para el beneficio del estudiante y del sistema educativo.

1.5.1 Historia de los ETL

En algún tiempo atrás los ETL eran considerados como gestión de metadatos que describían información de estos, después de eso se empezó a gestionar dichos procesos por separado y en la década de 1970, llegó a ser popular cuando las organizaciones comenzaron a utilizar depósitos de datos múltiples para almacenar diferentes tipos de información (Trujillo, 2018).

En la actualidad los procesos ETL son métodos utilizados por la mayoría de las organizaciones para recopilar, importar y procesar datos que también han evolucionado, convirtiéndose en una necesidad para la gestión institucional, transformándolos en información útil a través de herramientas poderosas que permiten realizar estos procesos.

1.5.2 Definición de los ETL

Es una capa de software encargada de integrar y preparar diferentes fuentes de datos, su nombre se define por las siglas en ingles Extract – Transform – Load, las cuales comprenden las tres etapas del proceso. ETL es un término estándar con el que se identifica el movimiento y transformación de datos, su proceso es clave y permite trasladar datos de una fuente a otra, cargándolos en una bodega o almacén de datos conocidos como Data Mart o Data Warehouse, integrando los distintos sistemas para su posterior análisis.

1.5.2.1 Fases del proceso ETL

La institución cuenta con fuentes de datos que permiten realizar un proceso ETL a través de sus fases para la creación de una Data Warehouse donde se almacene la información que posteriormente será analizada como lo muestra la Fig.5.

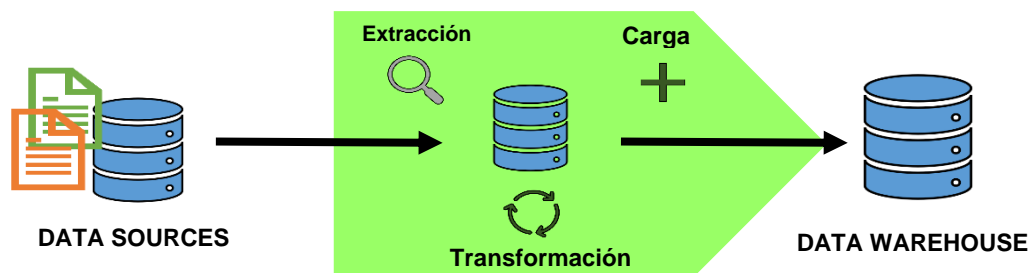


Fig. 5. Proceso ETL - Fuente propia basado en (Bernabeu & Mattío, 2017)

Un proceso ETL tiene tres fases que trabajan conjuntamente:

- Se extraen los datos relevantes
- Se integran y transforman los datos.
- Se cargan los datos al Data Warehouse.

a. Extracción

Esta etapa es donde se obtiene los datos útiles de la organización, de todas las fuentes de datos que existan, se debe identificar y determinar los datos que realmente aportan en la toma de decisiones. En la mayoría de los casos, se usan bases de datos relacionales, la extracción puede llevarse a cabo mediante consultas SQL o procedimientos almacenados (Bernabeu & Mattío, 2017).

b. Transformación

Esta fase consiste en transformar, depurar e integrar los datos obtenidos de los sistemas operacionales o externos. De todos los datos extraídos primero se debe transformar los datos modificando su estructura ya sea fusionando o dividiendo columnas, por ejemplo:

- Dividir la columna “Nombre: Chamorro Montenegro, Hugo” en dos columnas: “Nombre: Hugo” y “Apellidos: Chamorro Montenegro”.

Depurar los datos detectando y corrigiendo valores inconsistentes, así el almacén de datos se llenará con datos de calidad. Para finalizar se debe integrar los datos y cargarlos al almacén verificando que la información no esté duplicada (Chamorro, 2016).

c. Carga

La carga es la última fase en donde se transportan los datos de las diferentes plataformas para luego ser cargados en el sistema de destino correspondiente.

Existen dos formas para cargar los datos:

- Acumulación simple: se realiza un resumen de las transacciones comprendidas en un intervalo de tiempo y se transporta el resultado hacia el Data Warehouse, almacenando un valor que puede ser un sumatorio o promedio.
- Rolling: se almacena información a distintos niveles, que corresponden a agrupaciones de diferentes niveles jerárquicos en algunas dimensiones de magnitud almacenada. Este proceso es usado cuando se requiere tener varios niveles de granularidad.

desde las diferentes fuentes de datos que tiene la institución se puede realizar el proceso de extracción, carga y transformación que son las fases de los ETL para la creación de un Data Warehouse.

1.5.3 Procedimientos Almacenados

Un Procedimiento Almacenado, es un conjunto de sentencias estructuradas (SQL), que se almacenan en un sistema de bases de datos relacionales, por lo que puede ser reutilizado y compartido por múltiples programas.

Los procedimientos almacenados pueden acceder o modificar datos en una base de datos, tienen una capa de seguridad entre la base de datos y el usuario, a través de controles donde el usuario final pueda ingresar o cambiar datos preservando la integridad de estos

1.6 Metodología Kimball

En la actualidad la gran cantidad de información que generan las instituciones son útiles para generar conocimiento y apoyar a la toma de decisiones mejorando procesos. Centrándonos en la Educación Superior en el caso de la UTN se sabe que existe un alto índice de deserción estudiantil, si a toda esa información que se obtiene se le realiza un procedimiento ETL para transformar todos los datos en conocimiento, se obtendría indicadores clave sobre el problema y se mejoraría la calidad en la gestión académica y estudiantil.

Para ello se necesita de la creación de un Data Warehouse (DW) que permita la recolección y análisis de los datos a través de metodologías que facilitan su implementación. Existen varias metodologías para el diseño y construcción de DW. Según (Rivadera, 2010), cada fabricante de software de Business Intelligence busca imponer una metodología con sus productos. Sin embargo, para el desarrollo de este trabajo se escogió la metodología Kimball que desarrolla su estructura arquitectónica de manera ascendente empezando por la creación de Data Marts (DM) hasta la construcción de un Data Warehouse (DW).

1.6.1 Ralph Kimball y su enfoque BI

Kimball es considerado como el inventor del Modelo dimensional y pionero del Data Warehouse e Business Intelligence. Él define un almacén de datos como: " Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis".

La metodología que propuso Kimball se convirtió en apoyo a las decisiones, enfocándose en el diseño de la base de datos basada en la creación de tablas hechos para almacenar la información numérica o cuantitativa de los indicadores a analizarse (Carhuallanqui, 2017).

La filosofía de Kimball se enfoca en la construcción de un DW por interés de un departamento, creando así los DM departamentales. Conforme los departamentos requieren DM, estos se combinan con el primero creando así dimensiones conformadas. La metodología estandarizada que propone Kimball es compartir dichas dimensiones con diferentes DM garantizando su integridad, construyendo un DW conformado por un conglomerado de estructuras. (Dertiano, 2015)

La metodología Kimball se fundamenta en el ciclo de vida de los DW, que se denomina Ciclo de Vida Dimensional (Business Dimensional Lifecycle) y está basada en cuatro principios como se establece en la Tabla 5.

Tabla 5. Principios de la Metodología Kimball

N°	Principio	Definición
1	Centrarse en el negocio	identificación de los requerimientos de negocio y su valor asociado
2	Construir una infraestructura de información adecuada	diseño de una base de información única, integrada, fácil de usar
3	Realizar entregas e incrementos significativos	creación de un DW y entrega incremental de datos
4	Ofrecer la solución completa	proporcionar un DW sólido, de calidad y accesible, con herramientas de consulta para informes y análisis, con su respectiva documentación.

Fuente: (M. & P., 2014)

1.6.2 Fases de la Metodología Kimball

Kimball en su metodología plantea un ciclo de vida o más conocido como Lifecycle, para apoyar a la toma de decisiones, que ha sido aplicada en todas las industrias, áreas de aplicación, funciones empresariales y plataformas técnicas.

El enfoque del ciclo de vida de Kimball se ilustra en la Fig. 6. y proporciona una hoja de ruta general que representa la secuencia de tareas de alto nivel necesarias para proyectos de Business Intelligence.

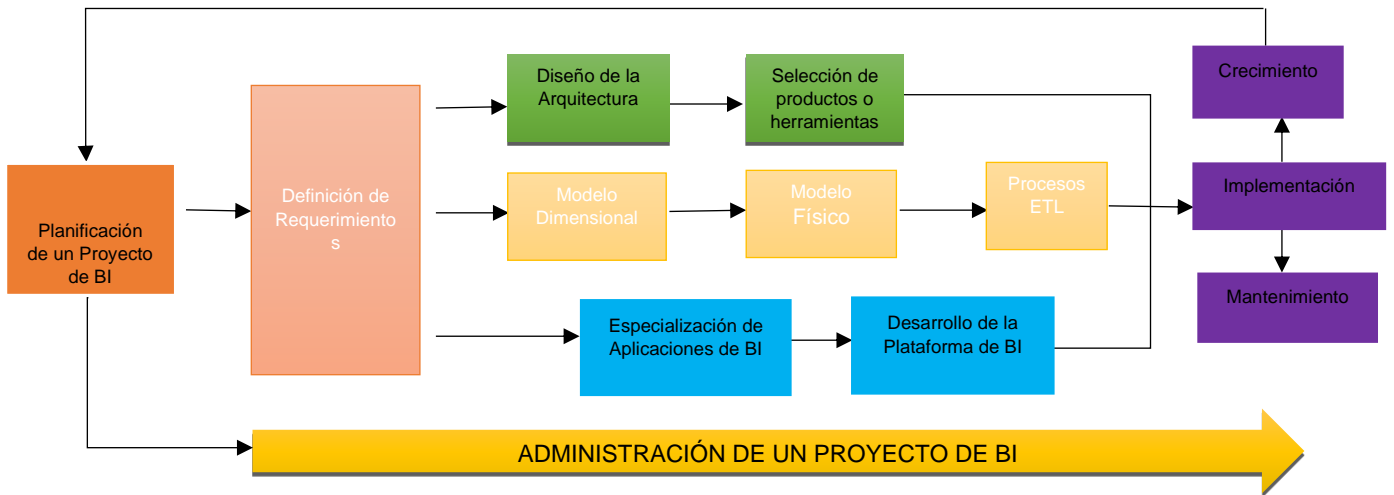


Fig. 6. Ciclo de Vida de Kimball – Fuente propia basado en (M. & P., 2014)

✓ **Planificación del Proyecto**

Identificar el propósito del proyecto, áreas involucradas, objetivos, alcance y los principales riesgos. Esta etapa identifica el escenario del proyecto para saber dónde surge la necesidad del Data Warehouse (Leon, 2014).

✓ **Definición de los requerimientos del negocio**

Definir requerimientos revisando los informes de la organización, los diseñadores de DW deben tener claro los factores claves que les permita crear diseños apropiados.

✓ **Modelado dimensional**

Matriz que determina la dimensionalidad, especificando los grados de detalle dentro de cada concepto, la granularidad y las diferentes jerarquías del modelo dimensional.

✓ **Diseño físico**

Selección de estructuras necesarias para soportar el diseño lógico, se prepara el entorno de base de datos con estándares de nombres y datos específicos

✓ **Diseño y desarrollo de presentación de datos**

Procesos ETL para la extracción de datos que permiten la carga del modelo físico. Estos procesos sirven para poblar el DW.

✓ **Implementación**

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio (M. & P., 2014).

✓ **Mantenimiento y Crecimiento**

El DW es un proceso que acompañara a la organización durante toda su historia, es necesario darle mantenimiento constante para seguir la evolución de su crecimiento manejando nuevos requerimientos para crecer y conseguir las metas.

1.7 Herramienta Tableau en la deserción estudiantil

El alto índice de deserción estudiantil en la UTN no es tratado de manera eficiente, debido a la dificultad que tienen los usuarios finales en la interpretación de datos. Para disminuir esta problemática se hace uso de la herramienta Tableau, que permite construir un conjunto de tableros para la visualización de patrones de deserción estudiantil, facilitando la exploración y administración de datos para la interpretación de resultados de manera eficaz, que ayude a tomar decisiones estratégicas y acertadas de la problemática.

1.7.1 Tableau

Tableau se fundó en 2003. Nació a partir de un proyecto de informática en la Universidad de Stanford cuyo objetivo es mejorar el flujo analítico y hacer que los datos sean más accesibles para las personas a través de la visualización interactiva de datos basados en el análisis y la inteligencia del negocio (Jose, 2019).

Esta herramienta permite tomar cualquier tipo de datos de cualquier sistema y convertirlos en información procesable con rapidez y facilidad, su misión es ayudar a las personas en la comprensión de datos (Tableau Software, 2020). Registra datos, añade metadatos, desarrolla cálculos y campos de datos para la manipulación de los usuarios dependiendo sus necesidades, a través de visualizaciones y controles de mando.

CAPÍTULO II

MATERIALES, METODOS Y PROCEDIMIENTOS

2 Desarrollo

2.1 Diagrama de Proceso del proyecto

Para cumplir con los objetivos planteados del proyecto de titulación se realizó un diagrama para representar de forma gráfica los principales procesos que intervienen en la obtención de información sobre la deserción estudiantil que es solicitada por el Departamento de Bienestar Universitario (DBU).

La Fig.7, representa la planificación que se tomó como base en el desarrollo del proyecto BI.

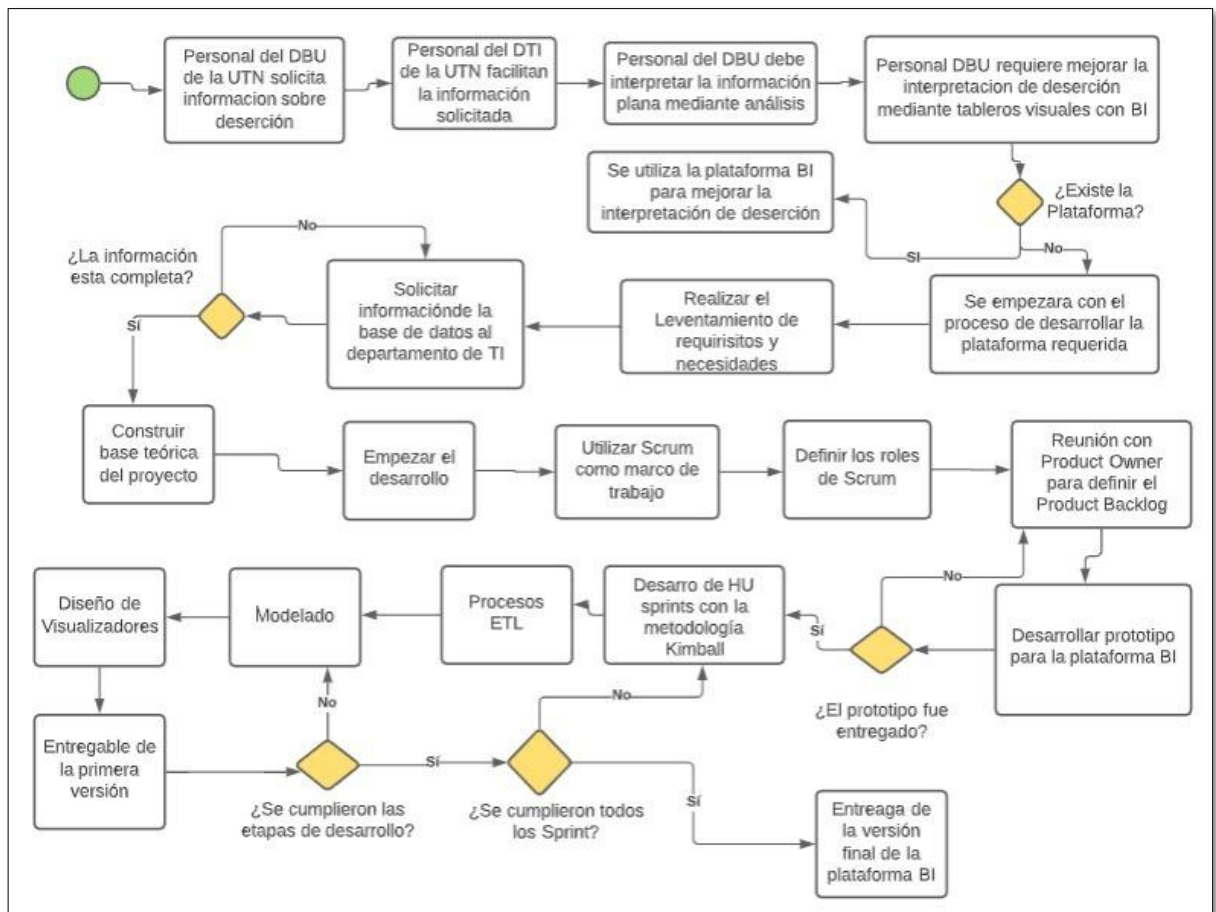


Fig. 7. Diagrama de Procesos

2.2 Marco de trabajo Scrum

En el desarrollo del capítulo II, se aplicó Scrum como marco de trabajo que permite administrar y gestionar el proyecto con procesos iterativos e incrementales realizando entregas parciales y regulares del producto a través de Sprint.

2.2.1 Sprint

Un Sprint es considerado un ciclo de trabajo dentro de un proyecto Scrum, cada sprint tiene una duración de cuatro semanas, una vez finalizado el sprint se obtienen los entregables del producto que aportan valor al cliente (Requena, 2018).

2.2.2 Roles Scrum

En Scrum un equipo puede estar compuesto entre 3 a 9 miembros, cada uno obtiene un rol con diferentes responsabilidades y distintas formas de rendir cuentas a la organización (Roche, 2017).

2.3 Integración de Scrum y Kimball como base en el desarrollo del proyecto

Para la construcción de un Data Warehouse y elaboración de tableros visuales con Tableau, se utilizó la metodología Kimball que se basa en la arquitectura Bottom-up, en el ciclo de vida de DW y en el modelado dimensional de manera ascendente, comenzando por la elaboración de Data Marts (DM) (Rivadera, 2010).

La integración de la metodología y el marco de trabajo han logrado que el desarrollo del proyecto cumpla con los objetivos establecidos. Tanto Scrum como Kimball comparten los mismos puntos en su estructura como es la Planificación y los requerimientos. Sin embargo, como se observa en la Fig. 8. Scrum fue usada para el desarrollo del proyecto en general realizando entregas en cortos periodos de tiempo, y Kimball se usó en la construcción del DW que se basa en el ciclo de vida dimensional del negocio.

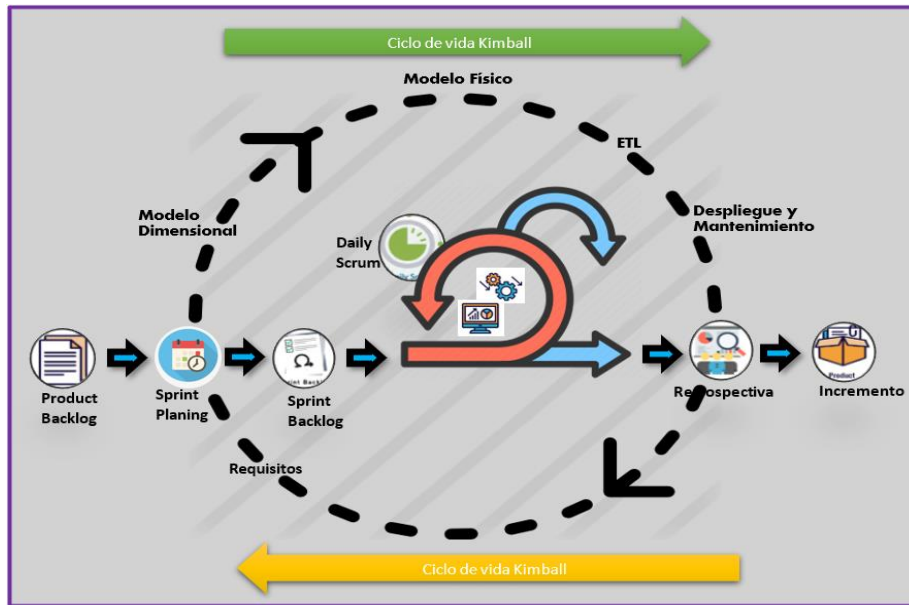


Fig. 8. Metodología de desarrollo del proyecto

2.4 Planificación y Gestión del Proyecto

La planificación y administración de un proyecto permite identificar el escenario y determinar el alcance del proyecto en base a los requerimientos de la institución. En este caso las partes involucradas tiene un interés en la obtención de información sobre la deserción estudiantil en el rendimiento académico de los estudiantes y en base a eso tomar decisiones estratégicas.

2.4.1 Administración del Proyecto

En la administración se consideran las reuniones con el equipo de trabajo, tutor y asesores, entrevistas con los altos jefes, monitoreo y entregables del proyecto, así también cada uno de los Sprint que se va a realizar.

- Roles Scrum

En la Tabla 6, se detallan los diferentes roles con su función y responsabilidad en el equipo Scrum.

Tabla 6. Roles Scrum del Proyecto

NOMBRE	ROL	RESPONSABILIDAD
Alexander Guevara	Scrum Master	Planificar reuniones para verificar el cumplimiento de los requisitos
Mercedes Navarrete	Product Owner	Facilitar los requisitos para el desarrollo del proyecto
Jonathan Silva	Desarrollador	Desarrollar los tableros BI
Evelyn Enríquez	Desarrollador	Facilitar los datos para la creación de los tableros BI
Martha Molina	Stakeholder	Solicitar entregables del proyecto para su revisión

2.4.2 Plan del Proyecto

El desarrollo del proyecto tiene la finalidad de construir un conjunto de tableros de análisis integral cuyo objetivo es la visualización de patrones de deserción estudiantil, a través de la elaboración de un Data Warehouse.

Para la estimación de cada tarea que se va a realizar con la denominada técnica T-Shirt Sizes, que se basa en asignar tallas de camiseta (XS, S, M L, XL) a la hora de estimar en vez de utilizar números (Deloitte, 2021).

Con la técnica de medición T-Shirt Sizes, se evalúan las tareas de acuerdo con el tamaño si es: extrapequeña, pequeña, mediana, grande, extragrande o doble extragrande, esto permite pensar de manera dinámica sobre el esfuerzo asociado a cada tarea (Mistry, 2017). Si es necesario a cada tamaño se le puede dar un valor numérico una vez finalizada la estimación de las tareas, como ejemplo en la siguiente Tabla 7, se muestra una posible manera de valoración numérica.

Tabla 7. Estimación de tareas técnica T-Shirt Sizes

Estimación	Nombre de Estimación	Valor Numérico Esfuerzo
S	Extrapequeña	5
S	Pequeña	15
M	Mediana	25
L	Grande	50
XL	Extragrande	100

2.4.3 Definición de Requerimientos del Negocio

En la definición de requerimientos se realizaron dos actividades para determinar las necesidades del usuario final y establecer las historias de usuario:

- Reunión con el Departamento de Desarrollo Tecnológico e Informático (DDTI),
- Entrevista con el Departamento de Bienestar, con la finalidad de recolectar variables requeridas para el posterior análisis

2.4.3.1 Historias de Usuario

De la entrevista se establecieron las historias de usuario requeridas por el cliente, las cuales se redactaron también en base a las variables de interpretación de datos.

HISTORIA DE USUARIO	
NUMERO: H1	USUARIO: Desarrollador
NOMBRE DE LA HISTORIA: Levantamiento preguntas (necesidades)	
PRIORIDAD: BAJA	
ESFUERZO: S	
DESCRIPCION: Como desarrollador necesito levantar todas las preguntas y necesidades para poder desarrollar el proyecto.	
PRUEBA DE ACEPTACIÓN: Tener un documeto transcrito de las entrevistas realizadas en línea	

Fig. 9. H1 Levantamiento de necesidades

HISTORIA DE USUARIO	
NUMERO: H2	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por rango de edades	
PRIORIDAD: MEDIA	
ESFUERZO: M	
DESCRIPCION: Como Usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por rangos de edad para saber que rango de edades hay el mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en tableau relacionado con el rango de edades	

Fig. 10. H2 Gráfico estadístico de deserción

HISTORIA DE USUARIO	
NUMERO: H3	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por genero	
PRIORIDAD: MEDIA	
ESFUERZO: L	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por género para saber que genero hay más porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en tableau relacionado con el género del estudiante.	

Fig. 11. H3 Gráfico estadístico de deserción

HISTORIA DE USUARIO	
NUMERO: H4	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por provincia de procedencia	
PRIORIDAD: MEDIA	
ESFUERZO: XL	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por tipo de institución secundaria para saber de qué tipo de institución secundaria hay más porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en tableau relacionado con el tipo de institución secundaria	

Fig. 12. H4 Grafico estadístico de deserción por provincia

HISTORIA DE USUARIO	
NUMERO: H5	USUARIO:
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por tipo de institución	
PRIORIDAD: MEDIA	
ESFUERZO: L	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por provincia de procedencia para saber de qué provincias hay el mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en tableau relacionado con la provincia de procedencia.	

Fig. 13. H5 Gráfico estadístico de deserción por tipo de institución

HISTORIA DE USUARIO	
NUMERO: H6	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por rango de ingreso mensual	
PRIORIDAD: MEDIA	
ESFUERZO: M	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por rango de ingreso mensual para saber de qué rangos de ingreso mensual en los hogares hay más porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en Tableau relacionado con el rango de ingreso mensual.	

Fig. 14. H6 Gráfico estadístico de deserción por rango de ingreso

HISTORIA DE USUARIO	
NUMERO: H7	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por facultad	
PRIORIDAD: MEDIA	
ESFUERZO: XL	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por facultades para saber de qué facultades hay más porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con la facultad	

Fig. 15. H7 Gráfico estadístico de deserción por facultad

HISTORIA DE USUARIO	
NUMERO: H8	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por etnia	
PRIORIDAD: MEDIA	
ESFUERZO: S	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por etnia para saber de qué etnias hay mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con la etnia	

Fig. 16. H8 Gráfico estadístico de deserción por etnia

HISTORIA DE USUARIO	
NUMERO: H9	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por estado civil	
PRIORIDAD: MEDIA	
ESFUERZO: S	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por estado civil para saber de qué estado civil del estudiante tienen los mayores porcentajes de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con el estado civil del estudiante.	

Fig. 17. H9 Gráfico estadístico de deserción por estado civil

HISTORIA DE USUARIO	
NUMERO: H10	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por tipo de financiamiento	
PRIORIDAD: MEDIA	
ESFUERZO: XL	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por el tipo de persona que financian los estudios para saber de qué tipo de financiamiento hay mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en Tableau relacionado con el tipo de financiamiento.	

Fig. 18. H10 Gráfico estadístico de deserción por tipo de financiación

HISTORIA DE USUARIO	
NUMERO: H11	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por tipo de vivienda	
PRIORIDAD: MEDIA	
ESFUERZO: M	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por tipo de vivienda del estudiante para saber de qué tipos de viviendas hay mayores porcentajes de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con el tipo de vivienda.	

Fig. 19. H11 Gráfico estadístico de deserción por tipo de vivienda

HISTORIA DE USUARIO	
NUMERO: H12	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por nivel	
PRIORIDAD: MEDIA	
ESFUERZO: S	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por nivel que cursa el estudiante para saber de qué niveles hay mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con el nivel.	

Fig. 20. H12 Gráfico estadístico de deserción por nivel

HISTORIA DE USUARIO	
NUMERO: H13	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por tipo de convivien	
PRIORIDAD: MEDIA	
ESFUERZO: M	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por tipo de conviviente vive el estudiante para saber de qué tipo de conviviente hay mayores porcentajes de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en tableau relacionado con el tipo de conviviente del estudiante.	

Fig. 21. H13 Gráfico estadístico de deserción por tipo de convivencia

HISTORIA DE USUARIO	
NUMERO: H4	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción por provincia de procedencia	
PRIORIDAD: MEDIA	
ESFUERZO: XL	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por tipo de institución secundaria para saber de qué tipo de institución secundaria hay más porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo en Tableau relacionado con el tipo de institución secundaria	

Fig. 22. H14 Gráfico estadístico de deserción si dispone internet

HISTORIA DE USUARIO	
NUMERO: H15	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Gráfico estadístico de deserción de deserción por número de matrícula	
PRIORIDAD: MEDIA	
ESFUERZO: S	
DESCRIPCION: Como usuario necesito un gráfico estadístico que me muestre el porcentaje de deserción estudiantil por número de matrícula para saber en qué número de matrícula hay mayor porcentaje de deserción.	
PRUEBA DE ACEPTACIÓN: Mostrar un gráfico interactivo sobre deserción en Tableau relacionado con el número de matrícula.	

Fig. 23. H15 Gráfico estadístico de deserción por número de matrícula

HISTORIA DE USUARIO	
NUMERO: H16	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Dashboard de deserción de persona	
PRIORIDAD: ALTA	
ESFUERZO: L	
DESCRIPCION: Como usuario necesito un dashboard donde pueda agrupar todos los gráficos que estén relacionados con la persona con sus respectivos filtros para tener una mejor interpretación	
PRUEBA DE ACEPTACIÓN: Mostrar un dashboard en tableau que muestre todos los gráficos relacionados con la persona	

Fig. 24. H16 Dashboards de deserción de persona

HISTORIA DE USUARIO	
NUMERO: H17	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Dashboard de deserción socioeconómico	
PRIORIDAD: ALTA	
ESFUERZO: L	
DESCRIPCIÓN: Como usuario necesito un dashboard donde pueda agrupar todos los gráficos que estén relacionados con la parte académica del estudiante con sus respectivos filtros, para tener una mejor interpretación	
PRUEBA DE ACEPTACIÓN: Mostrar un dashboard en tableau que muestre todos los gráficos relacionados con la persona.	

Fig. 25. H17 Dashboards de deserción socioeconómico

HISTORIA DE USUARIO	
NUMERO: H18	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Dashboard de deserción académico	
PRIORIDAD: ALTA	
ESFUERZO: M	
DESCRIPCIÓN: Como usuario necesito un dashboard donde pueda agrupar todos los gráficos que estén relacionados con la ficha socioeconómica del estudiante con sus respectivos filtros, para tener una mejor interpretación	
PRUEBA DE ACEPTACIÓN: Mostrar un dashboard en tableau que muestre todos los gráficos relacionados con la socioeconómica.	

Fig. 26. H18 Dashboards de deserción académico

HISTORIA DE USUARIO	
NUMERO: H19	USUARIO: Usuario
NOMBRE DE LA HISTORIA: Dashboard General de deserción	
PRIORIDAD: ALTA	
ESFUERZO: XXL	
DESCRIPCIÓN: Como usuario necesito un dashboard general con sus respectivos filtros para tener una interpretación de todas las métricas solicitadas.	
PRUEBA DE ACEPTACIÓN: Mostrar un tablero general en tableau para el cliente con todas las variables	

Fig. 27. H19 Dashboards General de deserción

2.4.3.2 Product Backlog

Es un listado de tareas que se van a realizar para el desarrollo del proyecto, las tareas se listan para tener visibilidad de todo lo que se espera realizar, este listado se

prioriza, de acuerdo con las historias de usuario redactadas. Como se observa en la Tabla 8.

Tabla 8. Product Backlog del Proyecto

TABLA PRODUCT BACKLOG			
PRIORIDAD	ID	HISTORIA DE USUARIO	ESFUERZO
1	H1	Levantamiento preguntas (necesidades)	S
2	H2	Gráfico estadístico de deserción por rango de edades	M
3	H3	Gráfico estadístico de deserción por genero	L
4	H4	Gráfico estadístico de deserción por provincia de procedencia	XL
7	H5	Gráfico estadístico de deserción por tipo de institución secundaria	L
5	H6	Gráfico estadístico de deserción por rango de ingreso mensual	M
6	H7	Gráfico estadístico de deserción por facultad	XL
11	H8	Gráfico estadístico de deserción por etnia	S
8	H9	Gráfico estadístico de deserción por estado civil	S
9	H10	Gráfico estadístico de deserción por tipo de financiamiento	M
10	H11	Gráfico estadístico de deserción por tipo de vivienda	M
12	H12	Gráfico estadístico de deserción por nivel	S
13	H13	Gráfico estadístico de deserción por tipo de conviviente	M
14	H14	Gráfico estadístico de deserción si dispone de servicio de internet o no	XS
15	H15	Gráfico estadístico de deserción de deserción por número de matrícula	S
16	H16	Dashboards de deserción de persona	L
17	H17	Dashboards de deserción socioeconómico	L
18	H18	Dashboards de deserción académico	M
19	H19	Dashboards General de deserción	XXL

2.4.4 Levantamiento preguntas (Necesidades)

Para el análisis de datos se realizó una entrevista con el Departamento de Desarrollo Tecnológico e Informático (DDTI), quienes proporcionaron diferentes requerimientos que son indicadores clave para el proceso de exploración y extracción de datos.

En la Tabla 9, se muestran las variables requeridas que permitirán fortalecer la interpretación de la información de la deserción estudiantil.

Tabla 9. Variables y Tipos de Datos

VARIABLES	TIPO DATO
EDAD	DATE
GENERO	VARCHAR (20)
PROVINCIA PROCEDENCIA	VARCHAR (100)
TIPO INSTITUCION	VARCHAR (100)
RANGO INGRESO MENSUAL	RANGO (50)
FACULTAD	VARCHAR (100)
ETNIA	VARCHAR (30)
ESTADO CIVIL	VARCHAR (20)
FINANCIAMIENTO	VARCHAR (100)
TIPO VIVIENDA	VARCHAR (100)
NIVEL	VARCHAR (10)
CONVIVENCIA	VARCHAR (100)
SERVICIO INTERNET	VARCHAR (10)
NUMERO MATRICULA	NUMBER

2.5 Sprint del Proyecto

Se establecieron 2 semanas de desarrollo por Sprint con el objetivo de cumplir con las fases de la metodología Kimball en base a las historias de usuario establecidas para la construcción de los tableros visuales.

En la Tabla 10 se detalla el número de historias de usuario correspondientes a cada Sprint.

Tabla 10. Desarrollo del Proyecto a través de Sprint

Id	NOMBRE	ESTIMACION	SPRINT	FECHA
1	HU1	S	1	
2	HU2	M	1,2,3	04/01/2021
3	HU3	L		A
4	HU4	XL		15/01/2021
5	HU5	L		

6	HU6	M	1,2,3	
7	HU7	XL		16/01/2021
8	HU8	S		A
9	HU9	S		29/01/2021
10	HU10	M		
11	HU11	M	1,2,3	
12	HU12	S		01/02/2021
13	HU13	M		A
14	HU14	XS		13/02/2021
15	HU15	S		
16	HU16	L	4	
17	HU17	L		15/02/2021
18	HU18	M		A
19	HU19	XXL		27/02/2021

2.5.1 Descripción y Ejecución de Sprint

Para este caso se dividió el proyecto en cuatro Sprint y dentro de él se establecieron las diferentes historias de usuario. A continuación, se presenta las Tabla 11 con la descripción de cada Sprint y el objetivo de este.

Tabla 11.Desarrollo de Sprint

N°	FECHA	DESCRIPCIÓN
1	04/01/2021 A 15/01/2021	Este Sprint tuvo la finalidad de realizar el levantamiento de necesidades y establecer los requisitos de desarrollo
2	16/01/2021 A 29/01/2021	En el segundo Sprint se realizaron los gráficos estadísticos de deserción
3	01/02/2021 A 13/02/2021	En el tercer Sprint se realizaron los gráficos estadísticos de deserción

4	15/02/2021	El último Sprint se crearon los Dashboards con la información
	A	obtenida y se presentó el proyecto en su versión Blanco y negro
	27/02/2021	

Para establecer el número de Sprint necesarios para el desarrollo del proyecto se realizó una entrevista de la que se obtuvo los requisitos necesarios para la elaboración de los tableros visuales. Se ejecutaron 4 Sprint y de cada uno se obtuvo un entregable que se presenta a continuación:

- Sprint 1.- Se presentó un documento con los requisitos de desarrollo y las historias de usuario establecidas.
- Sprint 2.- Se presentó los gráficos estadísticos de deserción estudiantil por rango de edad, genero, provincia de procedencia, institución secundaria, ingreso mensual, facultad, etnia.
- Sprint 3.- Se presentó los gráficos estadísticos de deserción estudiantil por estado civil, financiamiento, tipo de vivienda, nivel, conviviente, servicio de internet, numero de matrícula.
- Sprint 4.- Se entregó la versión Blanco y negro, con las posteriores mejoras del tablero con la deserción estudiantil por persona, socioeconómico y académico.

La gestión de cada Sprint y sus diferentes tareas se realizó con la herramienta Trello, como se observa en la Fig.28.

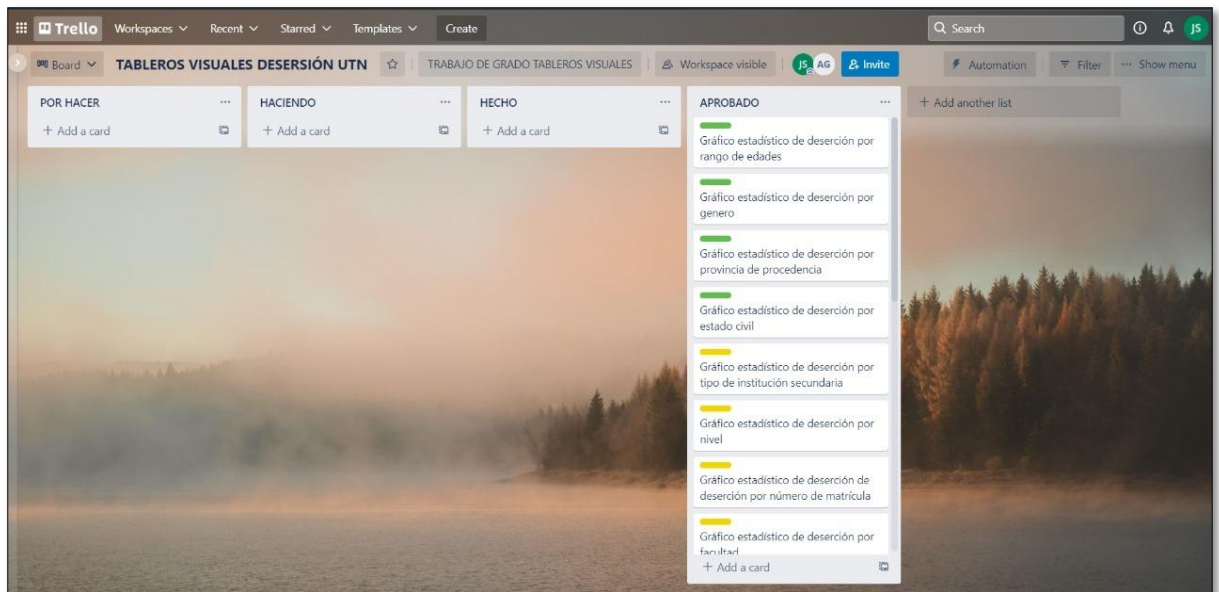


Fig. 28. Organización de Sprint

2.6 Elaboración del Data Warehouse con la metodología Kimball

2.6.1 Exploración de los datos de Origen

En el proceso de exploración se utilizó la base de datos de Oracle 12g proporcionada por el DTI, donde se encuentran almacenados los datos académicos, socioeconómicos, demográficos de cada estudiante.

Una vez que se obtuvo el archivo DMP de la base de datos, se procedió a importarla creando los tablespaces y usuarios necesarios para su funcionamiento, todos los datos se los puede visualizar en la herramienta TOAD de Oracle, como se observa en la Fig. 29.

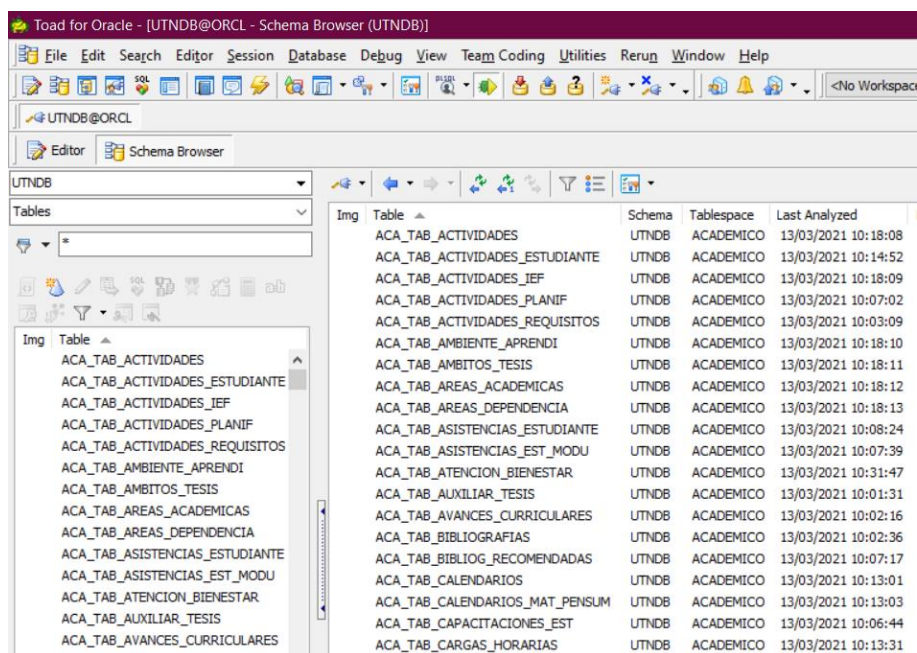


Fig. 29. Base de datos importada

Se realizó la exploración en la base de datos origen, encontrando alrededor de 20 posibles tablas que alojan los datos requeridos para la extracción de información que serán de utilidad en el proceso de análisis y obtención de variables. Las diferentes tablas, y campos encontrados se muestran en la Tabla 15.

Tabla 12. Exploración de Tablas

TABLA TRANSACCIONAL	NOMBRE_CAMPO	TIPO DATO
RHU_TAB_PERSONAS	FECHA_NACIMIENTO	DATE
RHU_TAB_GENERO	GENERO	VARCHAR (20)
INS_TAB_LOCALIDADES	DESCRIPCION	VARCHAR (100)
ACA_TAB_TIPOS_INSTITUCION	DESCRIPCION	VARCHAR (100)
DBU_TAM_INGRESO_MENSUAL	DESCRIPCION	RANGO (50)
ACA_TAB_DEPENDENCIAS	NOMBRE	VARCHAR (100)
INS_TAB_ETNIAS	ETNIA	VARCHAR (30)
RHU_ESTADO_CIVIL	ESTADO_CIVIL	VARCHAR (20)
DBU_TAB_PARENTESCO	DESCRIPCION	VARCHAR (100)
DBU_TAB_TIPOS_VIVIENDA	DESCRIPCION	VARCHAR (100)

2.6.2 Creación de la réplica en el área de desembarco n

Para crear la réplica de la base de datos se empleó SQL Server, que es la herramienta con la sé que trabajó todo el proceso de Extracción, transformación y limpieza de datos. Una vez culminada la exploración de datos, se realiza la importación de las tablas seleccionadas a SQL Server versión 2018.

2.6.3 Modelamiento Dimensional

Matriz Bus de Almacenamiento

Es una herramienta que permite un almacenamiento de datos incremental y el desarrollo de inteligencia empresarial para diseñar y comunicar la arquitectura de bus de almacén de datos (Kimball, 2013).

En este punto se establece la matriz bus con las dimensiones creadas y su conexión de las tablas con la base Transaccional u origen como se muestra en la Fig.30.

TABLA TRANSACCIONAL	DIMENSIONES														
	Edad	Género	Provincia	Procedencia	Tipo Institución	Ingreso Mensual	Facultad	Etnia	Estado Civil	Financiamiento	Tipo Vivienda	Nivel	Convivencia	Servicio Internet	Número Matricula
RHU_TAB_PERSONAS	X														
RHU_TAB_GENERO		X													
INS_TAB_LOCALIDADES			X												
ACA_TAB_TIPOS_INSTITUCION				X											
DBU_TAM_INGRESO_MENSUAL					X										
ACA_TAB_DEPENDENCIAS						X									
INS_TAB_ETNIAS							X								
RHU_ESTADO_CIVIL								X							
DBU_TAB_PARENTESCO									X						
DBU_TAB_TIPOS_VIVIENDA										X					
ACA_TAB_MATRICULA											X				
ACA_TAB_PARENTESCO												X			
DBU_TAB_FICHA_SOCIOEC													X		
ACA_TAB_MATRICULA														X	

Fig. 30. Matriz Bus de las dimensiones

2.6.4 Diseño Físico

2.6.4.1 Nivel de Granularidad

La granularidad representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se está analizando, para este caso se seleccionan los campos que permiten examinar y analizar los indicadores. En la Fig.31, se muestra el diagrama entidad relación con los campos seleccionados.

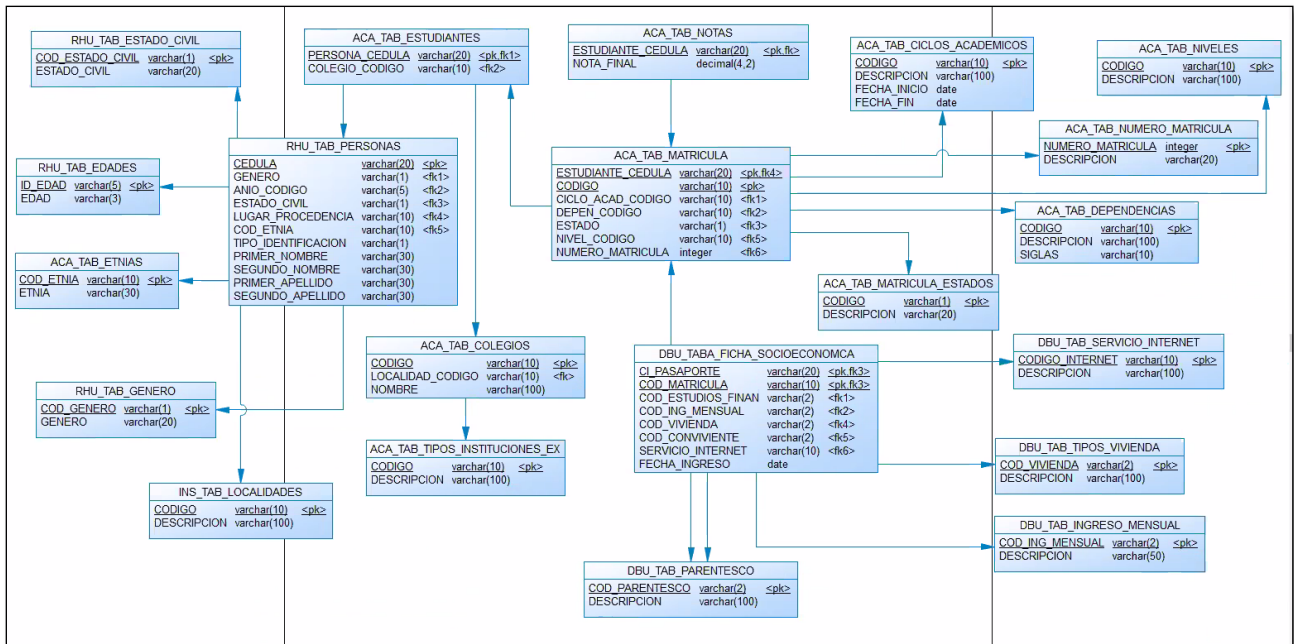


Fig. 31. Diagrama Entidad Relación

Se pretende averiguar el porcentaje de deserción académica de cada estudiante de la UTN de acuerdo con los diferentes campos y niveles de granularidad como se establece en la Tabla 16.

Tabla 13. Niveles de Granularidad del Proyecto

VARIABLES	GRANULARIDAD
EDAD	Rango de Edad
GENERO	Tipo de Género
PROVINCIA PROCEDENCIA	Tipo de Provincia correspondiente a las provincias del Ecuador
TIPO INSTITUCION	Pública - Particular - Extranjera
RANGO INGRESO MENSUAL	En base al Salario Básico Unificado
FACULTAD	Tipo de Facultad

2.6.5.1 **Llenado de tablas Homologadoras y Catálogos**

Del análisis realizado para la obtención de variables, se crearon los respectivos catálogos y tablas homologadoras que permiten enlazar la base transaccional con la base relacional denominada área de trabajo.

La creación de catálogos se basa en los datos extraídos de la base de datos origen para luego ser enlazados a través de las tablas homologadoras con la base de datos replicada, con el objetivo de almacenar información estructural a través de la colección de tablas especiales.

Para el llenado de las tablas catálogos y homologadoras se utilizó Procedimientos Almacenados (véase el Anexo 1), que permitieron limpiar y cargar los datos en la nueva base o área de trabajo como se evidencia en la Fig33.

```
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
ALTER PROCEDURE [DTM].[CARGA_DATAMART_ACADEMICA]
AS
BEGIN
    SET NOCOUNT ON;

    TRUNCATE TABLE ANALISIS.DTM.DTM_ACADEMICA

    INSERT INTO ANALISIS.DTM.DTM_ACADEMICA
    SELECT
        CAST(MAT.MATRICULA_ID AS INT) AS MATRICULA_ID,
        CAST(MAT.CODIGO_MATRICULA AS VARCHAR(15)) AS CODIGO_MATRICULA,
        CAST(ESTUDIANTE_ID AS INT) AS ESTUDIANTE_ID,
        CAST(MAT.CICLO_ACADEMICO_ID AS VARCHAR(10)) AS CICLO_ACADEMICO_ID,
        CAST(DET.FACULTAD_ID AS VARCHAR(5)) AS FACULTAD_ID,
        CAST(DET.NIVEL_ID AS VARCHAR(5)) AS NIVEL_ID,
        CAST(DET.NUMERO_MATRICULA_ID AS VARCHAR(5)) AS NUMERO_MATRICULA_ID,
        CAST(DET.ESTADO_MATRICULA_ID AS VARCHAR(2)) AS ESTADO_MATRICULA_ID,
        CAST(DET.NOTA_FINAL AS DECIMAL(4,2)) AS NOTA_FINAL
    FROM TRABAJO.MST.MST_MATRICULA MAT
    INNER JOIN TRABAJO.DET.DET_MATRICULA DET
    ON MAT.MATRICULA_ID=DET.MATRICULA_ID
END
```

Fig. 33. SQL Procedimiento Almacenado

Se establecieron los prefijos CTG y HMG para diferenciar a los catálogos de las tablas homologadoras. Por cada variable se creó los catálogos y tablas homologadoras como se muestra a continuación definiendo su nivel de granularidad:

- **EDAD**

Para la variable edad se tomaron valores con diferentes rangos que van desde los 18 años hasta los 34 en adelante como se especifica en la Tabla 17, con el fin de determinar la edad más propensa en la que desertan los estudiantes.

Tabla 14. Catálogo Edad

CTG_EDAD	
EDAD_ID	EDAD_DESCRIPCION
EME	0-18
EUN	19-21
EDO	22-24
ETR	25-27
EMA	28- 30
ECU	31-330
ECI	34-99
NOE	NO ESPECIFICADO

Elaboración Propia

- **GENERO**

La UTN en el año 2017 se encontraba conformada por el 49% de estudiantes masculinos y el 51% estudiantes femeninos, en base a estos datos se requiere conocer la tendencia de deserción por los diferentes géneros establecidos en la Tabla 18.

Tabla 15. Catálogo Genero

CTG_GENERO	
GENERO_ID	EDAD_DESCRIPCION
FEM	FEMENINO
MAS	MASCULINO
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 16. Tabla Homologada Género

HMG_GENERO		
GENERO_SIT_ID	GENERO_ID	GENERO_SIT_DESCRIPCION
F	FEM	FEMENINO
M	MAS	MASCULINO
	NOE	

Elaboración Propia

- **PROVINCIA**

La Universidad Técnica del Norte se encuentra ubicada en la provincia del Imbabura, con sede en la ciudad de Ibarra por esta razón la mayoría de los estudiantes que ingresan provienen de las provincias que conforman la Zona 1 del Ecuador con un 63,8% de Imbabura, 35,1% otras provincias y el 1,1% correspondiente a extranjeros. Para esto en la Tabla 20, se establecen las provincias de procedencia estudiantil.

Tabla 17. Catálogo Provincia

CTG_PROVINCIA	
PROVINCIA_ID	PROVINCIA_DESCRIPCION
IMB	IMBABURA
CAR	CARCHI
SUC	SUCUMBIOS
ESM	ESMERALDAS
PIC	PICHINCHA
OTR	OTROS
NOE	NO ESPECIFICADO

Elaboración Propia

- **TIPO INSTITUCION**

El tipo de institución de la que proceden los estudiantes es necesario para el análisis, cuyo objetivo es obtener resultados en porcentajes de deserción estudiantil dependiendo de la institución como se muestra en la Tabla 21.

Tabla 18. Catálogo Institución

CTG_INSTITUCION	
INSTITUCION_ID	INSTITUCION_DESCRIPCION
PUB	PUBLICA
PCO	PARTICULAR COFINANCIADA
PAU	PARTICULAR AUTOFINANCIADA
EXT	EXTRANJERO
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 19. Tabla Homologadora Institución

HMG_INSTITUCION		
INSTITUCIO_SIT_ID	INSTITUCION_ID	INSTITUCION_SIT_DESCRIPCION
01	PUB	PUBLICA
02	PCO	PARTICULAR COFINANCIADA
03	PAU	PARTICULAR AUTOFINANCIADA
04	EXT	EXTRANJERO

Elaboración Propia

- **RANGO INGRESO MENSUAL**

El ingreso mensual se categorizó en base al Salario Básico Unificado (RMU) del Ecuador del presente año que es \$400.00, la canasta básica familiar \$ 746.47 (INEC, 2018), cuyos valores se encuentran almacenados en la base de datos transaccional, de acuerdo con la Tabla 23.

Tabla 20. Catálogo Ingreso Mensual

CTG_INGRESO_MENSUAL	
INGRESO_MENSUAL_ID	INGRESO_MENSUAL_DESCRIPCION
CIE	1-100
DOC	101-200
TRE	201-300
CUA	301-400
QUI	401-500

OCH	501-800
MIL	801-1000
DMI	1001-2000
MAS	2001-MAS
NOE	NO ESPECIFICADO

Elaboración Propia

- **FACULTAD**

Cinco facultades conforman a la UTN cada una abarca alrededor de 10 carreras establecidas de acuerdo con su especialidad, en la Tabla 24, se muestran las diferentes facultades que serán analizadas posteriormente.

Tabla 21. Catálogo Facultad

CTG_FACULTAD	
FACULTAD_ID	FACULTAD_DESCRIPCION
FIC	FACULTAD DE INGENIERIA CIENCIAS APLICADAS
FCY	FACULTAD DE INGENIERIA CIENCIAS AGROPECUARIAS Y AMBIENTALES
FAC	FACULTAD DE CIENCIAS ADMINISTRATIVAS Y ECONOMICAS
FEC	FACULTAD DE EDUCACION CIENCIA Y TECNOLOGIA
FCS	FACULTAD DE CIENCIAS DE LA SALUD
NOE	NO ESPECIFICADO

Elaboración Propia

- **ETNIA**

La población del Ecuador está conformada por 15 etnias distribuidas entre Mestizos, Indígenas y Afroecuatorianos, de esas etnias el mayor ingreso de estudiantes a la UTN es de mestizos con el 81.2%, indígenas 8,8% y un 9,7% que hace referencia a otras etnias como se detalla en la Tabla 25.

Tabla 22. Catálogo Etnia

CTG_ETNIA	
ETNIA_ID	ETNIA_DESCRIPCION
MES	MESTIZO
AFR	AFROECUATORIANO

IND	INDÍGENA
BLA	BLANCO
ASI	ASIATICO
NEG	NEGRO
MUL	MULATO
MON	MONTUBIO
OTR	OTRAS
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 23. Tabla Homologadora Etnia

HMG_ETNIA		
ETNIA_SIT_ID	ETNIA_ID	ETNIA_SIT_DESCRIPCION
ME	MES	MESTIZO
AF	AFR	AFROECUATORIANO O AFRODESCENDIENTE
IN	IND	INDÍGENA
CA	BLA	BLANCO
AS	ASI	ASIATICO
NE	NEG	NEGRO
MU	MUL	MULATO
MO	MON	MONTUBIO
NO	NOE	DATO NO LEVANTADO

Elaboración Propia

- **ESTADO CIVIL**

Como se muestra en la Tabla 27, el estado civil se estableció de acuerdo con la ley de Registro Civil, Identificación y Cedulación.

Tabla 24. Catalogo Estado Civil

CTG_ESTADO_CIVIL	
ESTADO_CIVIL_ID	ESTADO_CIVIL_DESCRIPCION
UNI	UNION LIBRE
SEP	SEPARADO (A)
SOL	SOLTERO(A)
CAS	CASADO(A)

DIV	DIVORCIADO(A)
VIU	VIUDO(A)
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 25. Tabla Homologadora Estado Civil

HMG_ESTADO_CIVIL		
ESTADO_CIVIL_SIT_ID	ESTADO_CIVIL_ID	ESTADO_CIVIL_SIT_DESCRIPCION
U	UNI	UNION LIBRE
P	SEP	SEPARADO (A)
S	SOL	SOLTERO(A)
C	CAS	CASADO(A)
D	DIV	DIVORCIADO(A)
V	VIU	VIUDO(A)
	NOE	NO ESPECIFICADO

Elaboración Propia

- **FINANCIAMIENTO**

Tomando los datos almacenados de la base de datos transaccional se categorizaron las opciones con las que se financian sus estudios académicos por cada estudiante, como se detalla en la Tabla 29.

Tabla 26. Catálogo Financiamiento

CTG_FINANCIAMIENTO	
FINANCIAMIENTO_ID	FINANCIAMIENTO_DESCRIPCION
PAD	PADRES
ABU	ABUELOS
PRO	PROPIO
TIO	TIOS
CON	CONYUGUE
HER	HERMANOS
CRE	CREDITO
BEC	BECA
OTR	OTROS

NOE	NO ESPECIFICADO
-----	-----------------

Elaboración Propia

Tabla 27. Tabla Homologadora Financiamiento

HMG_FINANCIAMIENTO		
FINANCIAMIENTO_SIT_ID	FINANCIAMIENTO_ID	FINANCIAMIENTO_SIT_DESCRIPCION
01	PAD	Padre
02	PAD	Madre
03	PAD	Padres
04	CON	Cónyuge
05	OTR	Familiar
06	HER	Hermanos
07	ABU	Abuelos
08	OTR	Hijos
09	PRO	Sólo
10	OTR	Otro

Elaboración Propia

- **TIPO VIVIENDA**

Al igual que el financiamiento, en la Tabla 31 se muestra la categorización del tipo de vivienda de acuerdo con los datos almacenados en la base transaccional.

Tabla 28. Catálogo Tipo Vivienda

CTG_TIPO_VIVIENDA	
TIPO_VIVIENDA_ID	TIPO_VIVIENDA_DESCRIPCION
ANT	ANTICRESIS
PRO	PROPIA
PRE	PRESTADA
ARR	ARRENDADA
HIP	HIPOTECA
CON	CONCEDIDA POR TRABAJO
OTR	OTRAS
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 29. Tabla Homologadora Tipo Vivienda

HMG_TIPO_VIVIENDA		
TIPO_VIVIENDA_SIT_ID	TIPO_VIVIENDA_ID	TIPO_VIVIENDA_SIT_DESCRIPCION
01	ANT	ANTICRESIS
03	PRO	PROPIA
04	PRE	PRESTADA
05	NOE	
02	ARR	ARRENDADA
06	HIP	HIPOTECADA
07	CON	CONCEDIDA POR EL TRABAJO

Elaboración Propia

- **NIVEL**

El nivel es un atributo importante, con él se pueden obtener indicadores con un porcentaje de deserción acorde a la complejidad, en la Tabla 33 se establecen los 10 niveles correspondientes a cada carrera.

Tabla 30. Catalogo Nivel

CTG_NIVEL	
NIVEL_ID	NIVEL_DESCRIPCION
CER	CERO
PRI	PRIMERO
SEG	SEGUNDO
TER	TERCERO
CUA	CUARTO
QUI	QUINTO
SEX	SEXTO
SEP	SEPTIMO OCTAVO
OCT	OCTAVO
NOV	NOVENO
DEC	DECIMO
OTR	OTROS

Elaboración Propia

HMG_NIVEL		
NIVEL_SIT_ID	NIVEL_ID	NIVEL_SIT_DESCRIPCION
00	CER	CERO
10	DEC	DECIMO
09	NOV	NOVENO
08	OCT	OCTAVO
07	SEP	SÉPTIMO
06	SEX	SEXTO
05	QUI	QUINTO
04	CUA	CUARTO
03	TER	TERCERO
02	SEG	SEGUNDO
01	PRI	PRIMERO
AC	OTR	ACTUALIZACION
99	OTR	INSCRIPCION
I1	OTR	INTENSIVO INGLES
I2	OTR	INTENSIVO GLOBAL

Elaboración Propia

- **CONVIVENCIA**

Se establecieron 7 tipos de convivientes para obtener resultados concretos y evitar así la dispersión de estos, la Tabla 34 muestra la categorización de convivencia

Tabla 31. Catalogo Convivencia

CTG_CONVIVENCIA	
CONVIVENCIA_ID	CONVIVENCIA_DESCRIPCION
PAD	PADRES
ABU	ABUELOS
PDR	PADRINOS
TIO	TIOS
CON	CONYUGUE
HER	HERMANOS

SOL	SOLO
AMI	AMIGOS
OTR	OTROS
NOE	NO ESPECIFICADO

Elaboración Propia

HMG_TIPO_VIVIENDA		
CONVICENCIA_SIT_ID	CONVICENCIA_ID	CONVICENCIA_SIT_DESCRIPCION
01	PAD	Padre
02	PAD	Madre
03	PAD	Padres
04	CON	Cónyuge
05	OTR	Familiar
06	HER	Hermanos
07	ABU	Abuelos
08	OTR	Hijos
09	SOL	Sólo
10	OTR	Otro
11	NOE	

Elaboración Propia

- **SERVICIO INTERNET**

La conexión a internet es un servicio actualmente requerido para las clases virtuales, por el que se asigna también una beca estudiantil. El departamento de bienestar solicitó la variable que permita ver el grado de afectación del servicio de internet en la deserción estudiantil, como se muestra en la Tabla 35.

Tabla 32. Catalogo Servicio Internet

CTG_SERVICIO_INTERNET	
SERVICIO_INTERNET_ID	SERVICIO_INTERNET_DESCRIPCION
SII	SI

NOO	NO
OTR	OTRAS
NOE	NO ESPECIFICADO

Elaboración Propia

Tabla 33. Tabla Homologadora Internet

HMG_SERVICIO_INTERNET		
SERVICIO_INTERNET_SIT_ID	SERVICIO_INTERNET_ID	SERVICIO_INTERNET_SIT_DESCRIPCION
SI	SI	SI DISPONE DE INTERNET
NO	NO	NO DISPONE DE INTERNET

Elaboración Propia

- **NÚMERO DE MATRICULA**

Según los lineamientos del Reglamento de Régimen Académico de la UTN el número de matrículas se autoriza excepcionalmente hasta la tercera por asignatura, como se observa en la Tabla 34 se estableció el número de matrícula para su posterior análisis sobre el porcentaje de deserción en cada una.

Tabla 34. Catalogo Número Matrícula

CTG_NUMERO_MATRICULA	
NUMERO_MATRICULA_ID	NUMERO_MATRICULA_DESCRIPCION
PRI	PRIMERA
SEG	SEGUNDA
TER	TERCERA
OTR	OTROS
ESP	NO ESPECIFICADO

Elaboración Propia

Tabla 35. Tabla Homologadora Número Matrícula

HMG_NUMERO_MATRICULA		
NUMERO_MATRICULA_SIT_ID	NUMERO_MATRICULA_ID	NUMERO_MATRICULA_SIT_DESCRIPCION
1	PRI	PRIMERA
2	SEG	SEGUNDA
3	TER	TERCERA

Elaboración Propia

2.6.5.2 *Llenado de tablas maestras*

Una tabla maestra contiene datos reales que no varían con el tiempo y alimentan de información a las tablas transaccionales, convirtiéndose posteriormente en dimensiones. Para el presente proyecto se establecieron tres tablas maestras que contienen datos que rara vez cambian y son de utilidad para la creación de los Data Marts y Data Warehouse.

En la Tabla 39, se muestran los datos a usarse en la tabla maestra estudiantes.

Tabla 36. Tabla Maestra Estudiante

MST_ESTUDIANTE				
ESTUDIANTE_ID	IDENTIFICACION	PRIMER_NOMBRE	PRIMER_APELLIDO	ANIO_CODIGO
1	1234567890	Dayana	Vila	1995
2	1234567891	Jonathan	Silva	1994

Elaboración Propia

En la Tabla 40, se muestran los datos a usarse en la tabla maestra matricula

Tabla 37. Tabla Maestra Matrícula

MST_MATRICULA			
MATRICULA_ID	ESTUDIANTE_ID	CODIGO	CICLO_ID
1	1	351406	1008-0309
2	1	327698	0303-0803

Elaboración Propia

En la Tabla 41, se muestran los datos a usarse en la tabla maestra socioeconómica

Tabla 38. Tabla Maestra Socioeconómica

MST_SOCIOECONOMICA		
SOCIOECONOMICA_ID	MATRICULA_ID	FECHA_INGRESO
1	1	20/02/2019
2	2	13/02/2020

Elaboración Propia

2.6.5.3 *Llenado de tablas detalles*

Las tablas detalles o también denominadas transaccionales contienen datos que si cambian con frecuencia con el tiempo y se requiere registrar dichos cambios. En este caso

se establecen tres tablas detalle que derivan de las maestras cuyos datos probablemente cambien.

En las Tablas 42, 43, 44, se muestra los atributos que se consideraron como cambiantes, de estudiante, matrícula y socioeconómica.

Tabla 39. Tabla Detalle Estudiante

DET_ESTUDIANTE					
ESTUDIANTE_ID	ESTADO_CIVIL_ID	ETNIA_ID	GENERO_ID	PROVINCIA_ID	INSTITUCION_ID
1	SOL	AFR	FEM	IMB	PRI
2	SOL	MES	MAS	CAR	PUB

Elaboración Propia

Tabla 40. Tabla Detalle Matrícula

DET_MATRICULA				
MATRICULA_ID	FACULTAD_ID	ESTADO_MATRICULA_ID	NIVEL_ID	NOTA_FINAL
1	FIC	A	1	6
2	FIC	R	05	8.8

Elaboración Propia

Tabla 41. Tabla Detalle Socioeconómica

DET_SOCIOECONOMICA				
SOCIOECONOMICA_ID	FINANCIAMIENTO_ID	INGRESO_MENSUAL_ID	TIPO_VIVIENDA_ID	SERVICIO_INTERNET_ID
1	PAD	CIE	PRE	SII
2	PAD	CUA	PRO	NOO

Elaboración Propia

2.7 Elaboración de la Data Warehouse y Data Marts

2.7.1 Creación del modelo del área de análisis

Una vez realizado el llenado de tablas maestras y detalles, se procede a la creación del área de análisis, donde se integran las dimensiones y DM para construir el DW, almacenando toda la información estratégica para la toma de decisiones, como se muestra en la Fig.34.

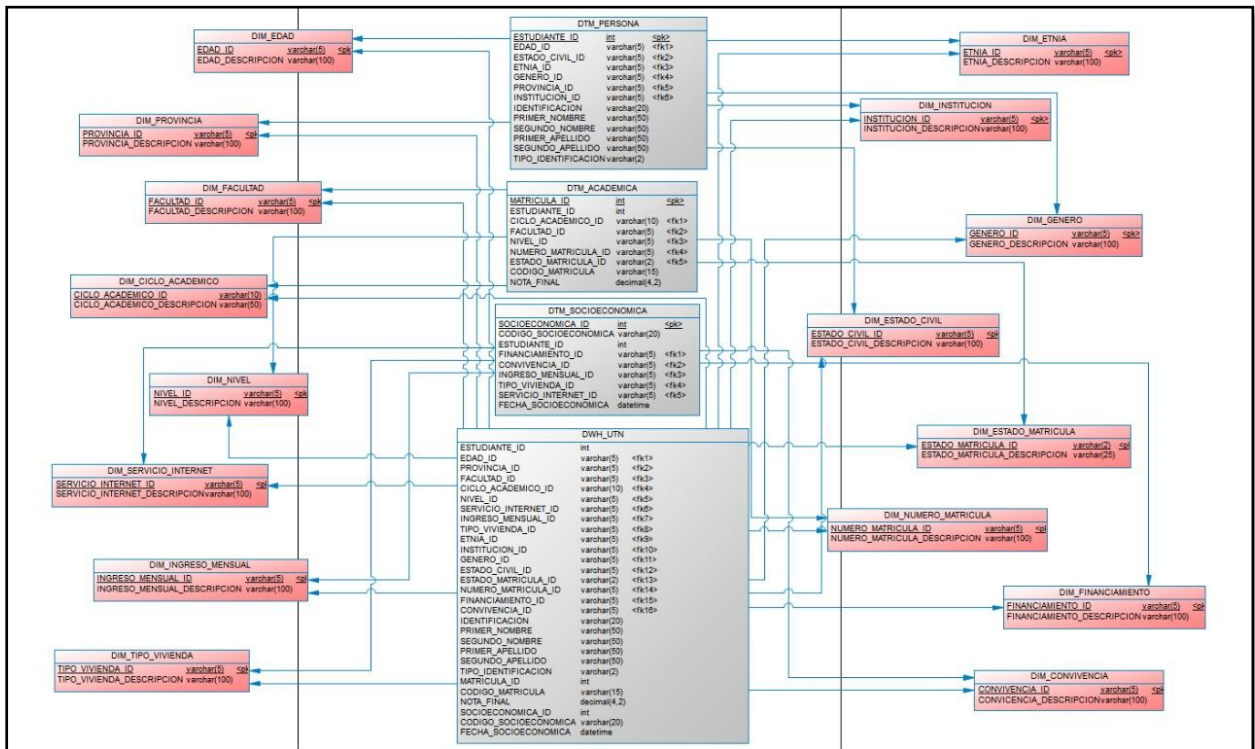


Fig. 34. Diagrama Área de análisis

2.7.2 Tablas Dimensión

Las tablas maestras permiten la creación de un modelo dimensional conformado por tablas de hecho centrales y dimensiones asociadas, con la finalidad de almacenar información descriptiva. Entonces una tabla dimensional contiene columnas o atributos que describen las filas, y se identifican con una clave primaria única que se enlaza a cualquier tabla de hechos a la que este juntada.

En la Tabla 45, se muestra la estructura de la tabla dimensional edad.

Tabla 42. Dimensión Edad

DIM_EDAD

EDAD_ID	Varchar (5)	Pk
EDAD_DESCRIPCION	Varchar (100)	

Elaboración Propia

2.7.3 Diseño y elaboración de Data Marts

El modelo dimensional consiste en una tabla de hechos normalizada, con tablas dimensionales desnormalizadas. Dependiendo del área para la creación de DM, se escogen las dimensiones con sus respectivos atributos que permitan obtener información relevante para su análisis.

Con las dimensiones establecidas la base multidimensional se conforma de los siguientes DM:

Data Mart Persona

Está conformado por las dimensiones edad, etnia, genero, provincia, institución, estado civil, integrando información básica de cada estudiante, que permitirá conocer por ejemplo el rango de edad con más deserción estudiantil. La Tabla 46 muestra los atributos del DM_Persona.

Tabla 43. Atributos correspondientes al Datamart Persona

DTM_PERSONA	
ESTUDIANTE_ID	Int <pk>
EDAD_ID	Varchar (5) <fk>
ESTADO_CIVIL_ID	Varchar (5) <fk>
ETNIA_ID	Varchar (5) <fk>
GENERO_ID	Varchar (5) <fk>
PROVINCIA_ID	Varchar (5) <fk>
INSTITUCION_ID	Varchar (5) <fk>
IDENTIFICACION	Varchar (20)
PRIMER_NOMBRE	Varchar (50)
SEGUNDO_NOMBRE	Varchar (50)
PRIMER_APELLIDO	Varchar (50)
SEGUNDO_APELLIDO	Varchar (50)
TIPO_IDENTIFICACION	Varchar (2)

Elaboración Propia

Data Mart Académica

Se conforma por las dimensiones: facultad, ciclo académico, nivel, numero matrícula, estado matrícula, toda esta información permitirá saber el nivel en el que más deserción estudiantil existe. La Tabla 47 muestra los atributos que componen el DM_ACADEMICA.

Tabla 44. Atributos correspondientes al Datamart Académica

DTM_ACADEMICA	
MATRICULA_ID	Int <pk>
ESTUDIANTE_ID	Int
CICLO_ACADEMICO_ID	Varchar (10) <fk>
FACULTAD_ID	Varchar(5) <fk>
NIVEL_ID	Varchar(5) <fk>
NUMERO_MATRICULA_ID	Varchar(5) <fk>
ESTADO_MATRICULA_ID	Varchar(2) <fk>
CODIGO_MATRICULA	Varchar(15)
NOTA_FINAL	Varchar(4,2)

Data Mart Socioeconómica

Por último, el DM Socioeconómica se conforma por las dimensiones: servicio de internet, tipo de vivienda, ingreso mensual, convivencia, financiamiento, cuya información almacenada será sobre las diferentes realidades sociales de cada estudiante, permitiendo así conocer el motivo de deserción. La Tabla 48 muestra los atributos que componen el DM_SOCIOECONOMICA.

Tabla 45. Atributos correspondientes al Datamart Socioeconómica

DTM_SOCIOECONOMICA	
SOCIOECONOMICA_ID	Int <pk>
CODIGO_SOCIOECONOMICA	Varchar(20)
ESTUDIANTE_ID	Int
FINANCIAMIENTO_ID	Varchar(5) <fk>
CONVIVECIA_ID	Varchar(5) <fk>
INGRESO_MENSUAL_ID	Varchar(5) <fk>
TIPO_VIVIENDA_ID	Varchar(5) <fk>
SERVICIO_INTERNET_ID	Varchar(5) <fk>
FECHA_SOCIOECONOMICA	Datetime

2.7.4 Llenado de Data Warehouse (DW)

Un DW toma los datos disponibles en una organización, los limpia y transforma, y los convierte en información estratégica para la toma de decisiones, provee una única versión de los parámetros y la habilidad para ver los datos de diferentes puntos de vista. En la Tabla 49 se muestra los campos que contiene el DW.

Tabla 46. Atributos que componen el DataWarehouse

DWH_UTN	
ESTUDIANTE_ID	Int

EDAD_ID	Varchar(5) <fk>
PROVINCIA_ID	Varchar(5) <fk>
FACULTAD_ID	Varchar(5) <fk>
CICLO_ACADEMICO_ID	Varchar(5) <fk>
NIVEL_ID	Varchar(5) <fk>
SERVICIO_INTERNET_ID	Varchar(5) <fk>
INGRESO_MENSUAL_ID	Varchar(5) <fk>
TIPO_VIVIENDA_ID	Varchar(5) <fk>
ETNIA_ID	Varchar(5) <fk>
INSTITUCION_ID	Varchar(5) <fk>
GENERO_ID	Varchar(5) <fk>
ESTADO_CIVIL_ID	Varchar(5) <fk>
ESTADO_MATRICULA_ID	Varchar(2) <fk>
NUMERO_MATRICULA_ID	Varchar(5) <fk>
FINANCIAMIENTO_ID	Varchar(5) <fk>
CONVIVENCIA_ID	Varchar(5) <fk>
IDENTIFICACION	Varchar(20)
PRIMER_NOMBRE	Varchar(50)
SEGUNDO_NOMBRE	Varchar(50)
PRIMER_APELLIDO	Varchar(50)
SEGUNDO_APELLIDO	Varchar(50)
TIPO_IDENTIFICACION	Varchar(2)
MATRICULA_ID	Int
CODIGO_MATRICULA	Varchar(15)
NOTA_FINAL	Varchar(4,2)
SOCIOECONOMICA_ID	Int
CODIGO_SOCIOECONOMICA	Varchar(20)
FECHA_SOCIOECONOMICA	Datetime

Haciendo uso de SQL Server, a través de consultas SQL con procedimientos almacenados se realizó el llenado del Data Warehouse cargando los datos de cada Data Mart. La Fig.35 muestra la consulta SQL realizada.

```

USE [ANALISIS] GO SET ANSI_NULLS ON
GO SET QUOTED_IDENTIFIER ON
GO ALTER PROCEDURE [DWH].[CARGA_DATAWAREHOUSE_UTN]
AS BEGIN
SET NOCOUNT ON;
INSERT INTO ANALISIS.DWH.DWH_UTN
SELECT
CAST(PER. ESTUDIANTE_ID AS int) AS ESTUDIANTE_ID,
CAST(PER. IDENTIFICACION AS VARCHAR(20)) AS IDENTIFICACION,
CAST(PER. TIPO_IDENTIFICACION AS varchar(2)) AS TIPO_IDENTIFICACION,
CAST(PER. GENERO_ID AS varchar(5)) AS GENERO_ID,
CAST(PER. ESTADO_CIVIL_ID AS VARCHAR(5)) AS ESTADO_CIVIL_ID,
CAST(PER. ETNIA_ID AS varchar(5)) AS ETNIA_ID,
CAST(PER. INSTITUCION_ID AS varchar(5)) AS INSTITUCION_ID,
CAST(PER. EDAD_ID AS varchar(5)) AS EDAD_ID,
CAST(PER. PROVINCIA_ID AS varchar(5)) AS PROVINCIA_ID,
CAST(ACA. MATRICULA_ID AS int) AS MATRICULA_ID,
CAST(ACA. CODIGO_MATRICULA AS VARCHAR(15)) AS CODIGO_MATRICULA,
CAST(ACA. FACULTAD_ID AS VARCHAR(5)) AS FACULTAD_ID,
CAST(ACA. CICLO_ACADEMICO_ID AS VARCHAR(10)) AS CICLO_ACADEMICO_ID,
CAST(ACA. NIVEL_ID AS varchar(5)) AS NIVEL_ID,
CAST(ACA. ESTADO_MATRICULA_ID AS varchar(5)) AS ESTADO_MATRICULA_ID,
CAST(ACA. NUMERO_MATRICULA_ID AS varchar(5)) AS NUMERO_MATRICULA_ID,
CAST(ACA. NOTA_FINAL AS DECIMAL(4,2)) AS NOTA_FINAL,
CAST(SOCI. SOCIOECONOMICA_ID AS int) AS SOCIOECONOMICA_ID,
CAST(SOCI. CODIGO_SOCIOECONOMICA AS VARCHAR(20)) AS CODIGO_SOCIOECONOMICA,
CAST(SOCI. SERVICIO_INTERNET_ID AS VARCHAR(5)) AS SERVICIO_INTERNET_ID,
CAST(SOCI. INGRESO_MENSUAL_ID AS VARCHAR(5)) AS INGRESO_MENSUAL_ID,
CAST(SOCI. TIPO_VIVIENDA_ID AS VARCHAR(5)) AS TIPO_VIVIENDA_ID,
CAST(SOCI. FINANCIAMIENTO_ID AS VARCHAR(5)) AS FINANCIAMIENTO_ID,
CAST(SOCI. CONVIVENCIA_ID AS VARCHAR(5)) AS CONVIVENCIA_ID,
CAST(SOCI. FECHA_SOCIOECONOMICA AS DATE) AS FECHA_SOCIOECONOMICA
FROM ANALISIS.DTM.DTM_PERSONA PER
INNER JOIN ANALISIS.DTM.DTM_ACADEMICA ACA
ON PER. ESTUDIANTE_ID=ACA. ESTUDIANTE_ID
INNER JOIN ANALISIS.DTM.DTM_SOCIOECONOMICA SOCI
ON PER. ESTUDIANTE_ID=SOCI. ESTUDIANTE_ID
END

```

Fig. 35. Carga de Datos al Data Warehouse

2.8 Diseño de Aplicación BI

Las Herramientas de BI son el primer medio de entrega de valores de un DW por esta razón en el diseño de la Aplicación BI se trabajó con la herramienta Tableau Desktop en su versión 2019.3 como se observa en la Fig.36. Esta herramienta permite mostrar la información a través de una interfaz amigable e interactiva para el usuario final.

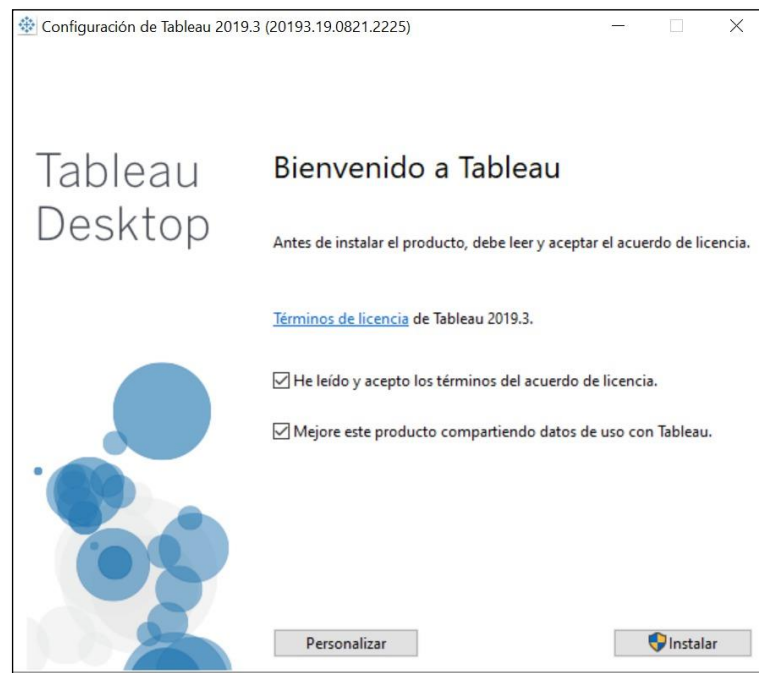


Fig. 36. Instalación de Tableau

En la Fig. 37 se muestra la interfaz de la herramienta instalada lista para configurar la conexión de Tableau a la base de datos SQL Server teniendo acceso a los datos.

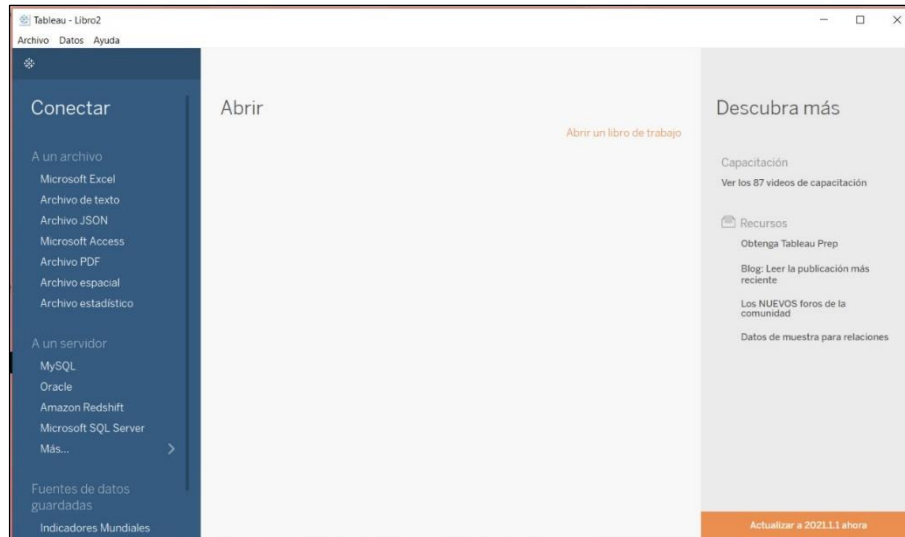


Fig. 37. Conexión a SQL Server

Una vez que se tiene acceso a los datos del Data Warehouse del SQL Server se importan las dimensiones creadas a la herramienta Tableau como se muestra en la Fig. 38

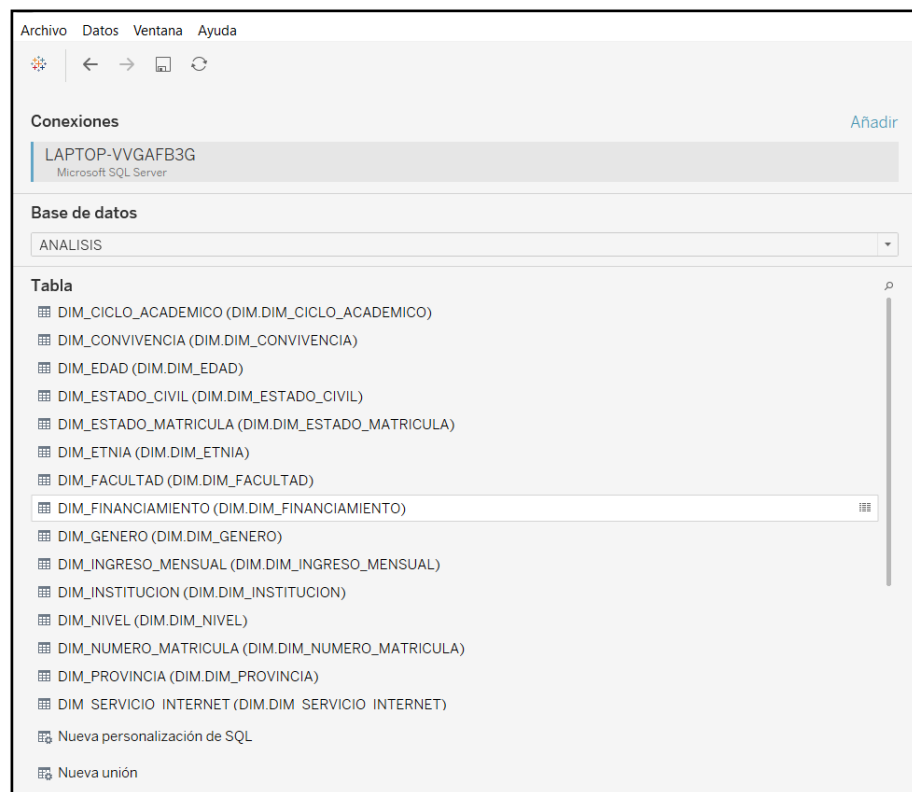


Fig. 38. Importación de SQL a Tableau

2.8.1 Construcción de tableros visuales en Tableau

Un tablero es una colección de vistas que amplía la variedad de datos simultáneamente, mostrando todas las vistas a la vez, estos tableros se actualizan con los últimos datos encontrados en la fuente de datos en este caso el DW (Tableau Software LLC., 2020).

Una vez realizada la conexión e importado los datos a la herramienta se procede con la construcción de tableros, uniendo el DW con sus dimensiones asociadas como se observa en la Fig.39.

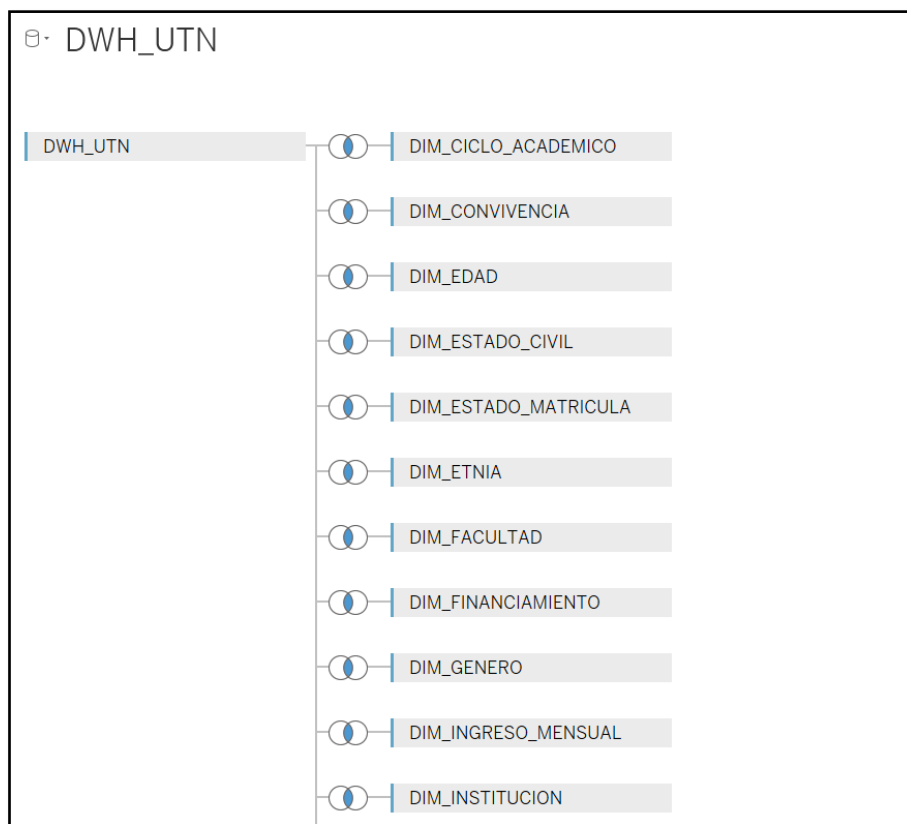


Fig. 39. Construcción de DW

2.8.2 Identificación de las métricas a implementar en el tablero

De los datos importados a Tableau se identifican donde se encuentran las métricas que se mostrarán en el tablero a través de graficas estadísticas y que servirán para su análisis e interpretación en la toma de decisiones. La Fig.40 muestra cómo se ordenaron los campos para su identificación.

Nombre de campo	Tabla	Nombre de campo remoto
ESTUDIANTE_ID	DWH_UTN	ESTUDIANTE_ID
ETNIA_ID	DWH_UTN	ETNIA_ID
FACULTAD_ID	DWH_UTN	FACULTAD_ID
FECHA_SOCIOECONOMICA	DWH_UTN	FECHA_SOCIOECONOMICA
FINANCIAMIENTO_ID	DWH_UTN	FINANCIAMIENTO_ID
GENERO_ID	DWH_UTN	GENERO_ID
IDENTIFICACION	DWH_UTN	IDENTIFICACION
INGRESO_MENSUAL_ID	DWH_UTN	INGRESO_MENSUAL_ID
INSTITUCION_ID	DWH_UTN	INSTITUCION_ID
MATRICULA_ID	DWH_UTN	MATRICULA_ID
NIVEL_ID	DWH_UTN	NIVEL_ID
Nota Final	DWH_UTN	NOTA_FINAL
NUMERO_MATRICULA_ID	DWH_UTN	NUMERO_MATRICULA_ID
PROVINCIA_ID	DWH_UTN	PROVINCIA_ID
SERVICIO_INTERNET_ID	DWH_UTN	SERVICIO_INTERNET_ID
SOCIOECONOMICA_ID	DWH_UTN	SOCIOECONOMICA_ID
TIPO_IDENTIFICACION	DWH_UTN	TIPO_IDENTIFICACION
TIPO_VIVIENDA_ID	DWH_UTN	TIPO_VIVIENDA_ID

Fig. 40. Identificación de Métricas

La métrica Edad, se encuentra en la tabla Dim_Edad con relación al Data Mart Persona y al DW, la Fig. 41 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

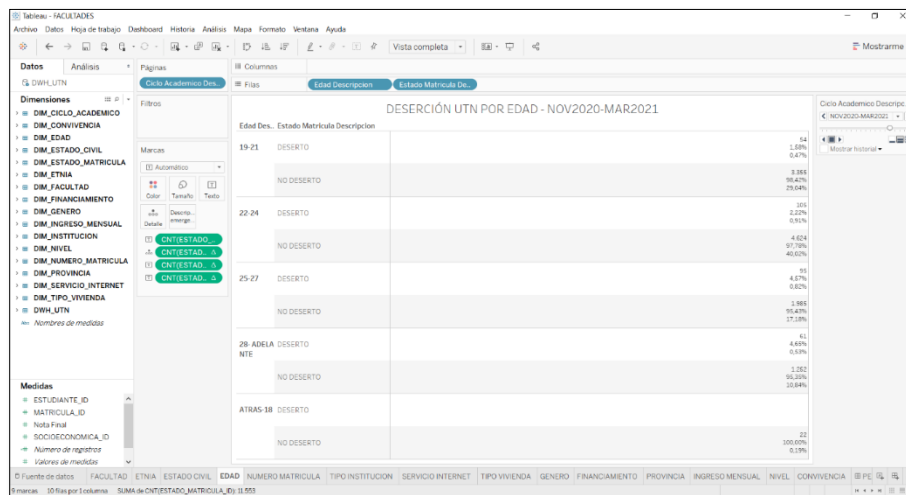


Fig. 41. Métrica Edad

La métrica Estado Civil, se encuentra en la tabla Dim_Estado_Civil con relación al Datamart Persona y al DW, la Fig.42 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

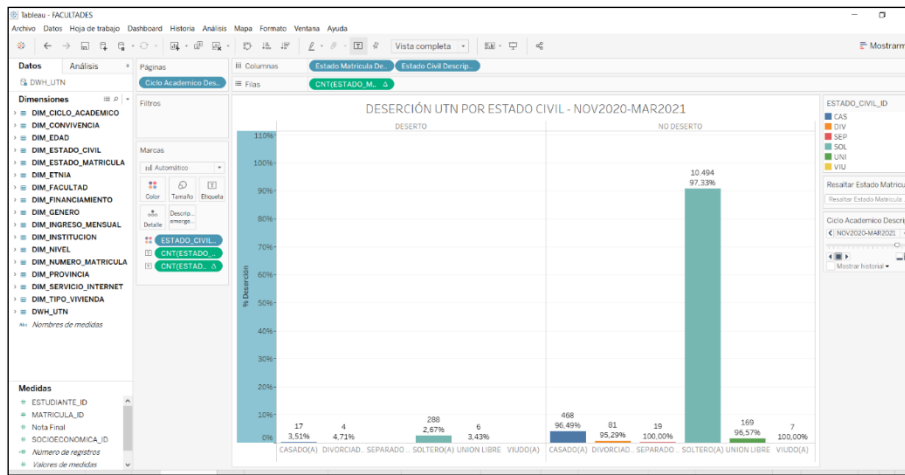


Fig. 42. Métrica Estado Civil

La métrica Etnia, se encuentra en la tabla Dim_Etnia con relación al Data Mart Persona y al DW, la Fig.43 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

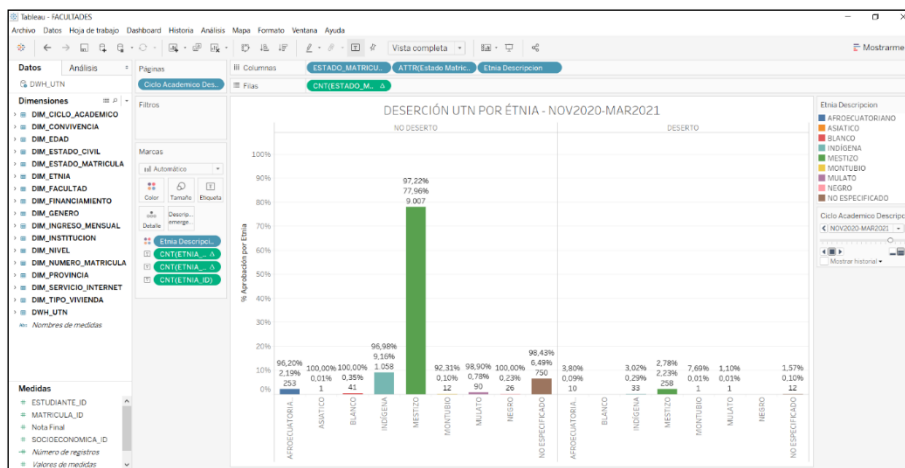


Fig. 43. Métrica Etnia

La métrica Facultad, se encuentra en la tabla Dim_Facultad con relación al Data Mart Académico y al DW, la Fig.44 muestra el porcentaje de deserción en el período Nov2020 Mar2021.

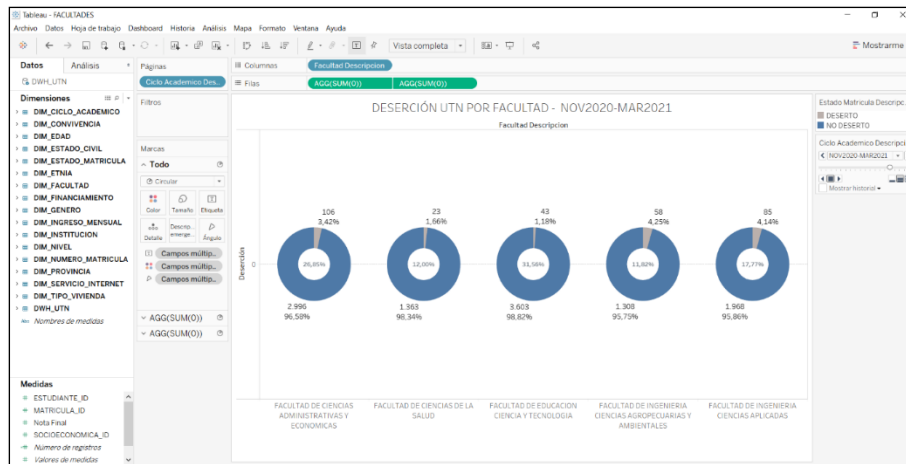


Fig. 44. Métrica Facultad

La métrica Tipo de Financiamiento, se encuentra en la tabla Dim_Tipo_Financiamiento con relación al Data art Socioeconómica y al DW, la Fig.45 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

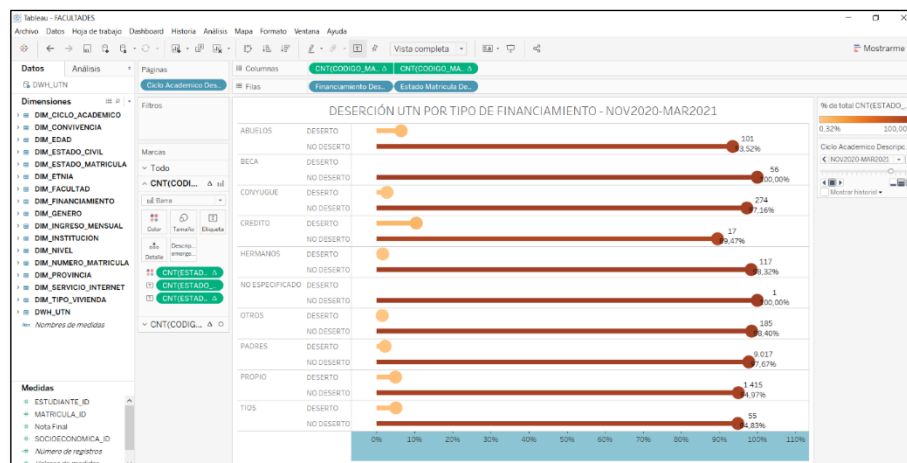


Fig. 45. Métrica Financiamiento

La métrica Género, se encuentra en la tabla Dim_Género con relación al Data Mart Persona y al DW, la Fig.46 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

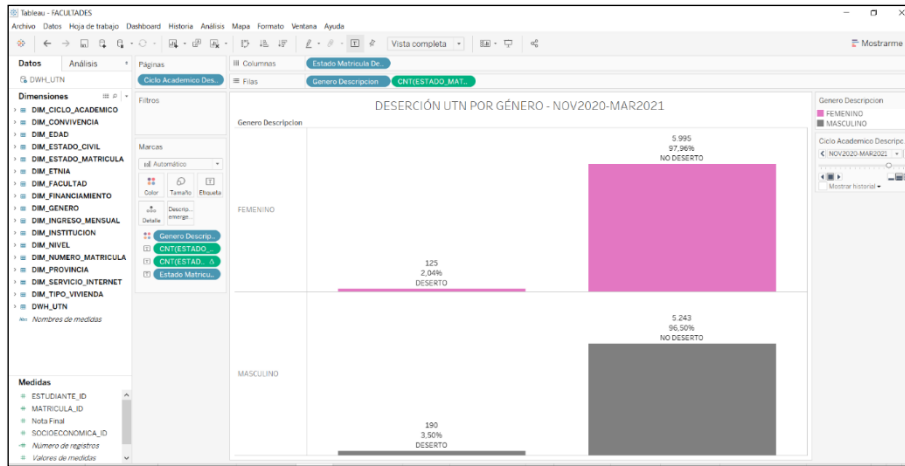


Fig. 46. Métrica Género

La métrica Tipo de Vivienda, se encuentra en la tabla Dim_Vivienda con relación al Data Mart Socioeconómica y al DW, la Fig.47 muestra el porcentaje de deserción en el período Nov2020 - Mar2021.

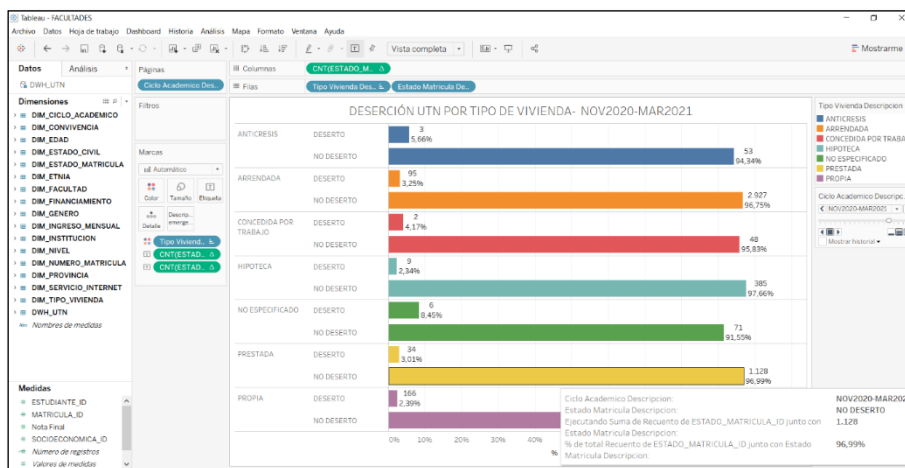


Fig. 47. Métrica Tipo Vivienda

La métrica Tipo Conviviente, se encuentra en la tabla Dim_Tipo_Conviviente con relación al Data Mart Socioeconómica y al DW, la Fig.48 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

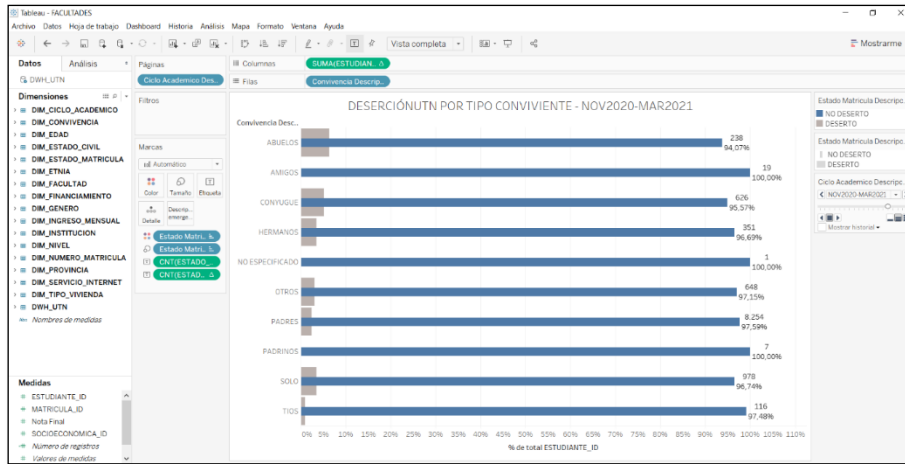


Fig. 48. Métrica Conviviente

La métrica Servicio de Internet, se encuentra en la tabla Dim_Servicio_Internet con relación al Data Mart Socioeconómica y al DW, la Fig.49 muestra el porcentaje de deserción en el período Sep2019-Mar2020.

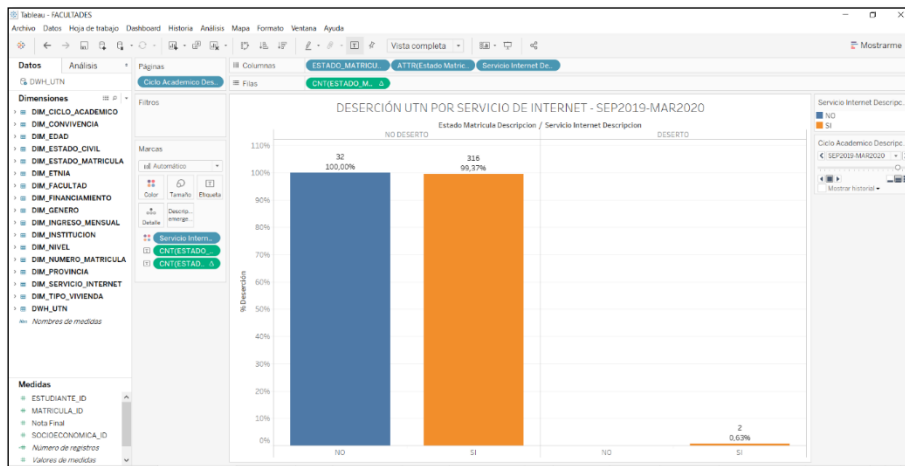


Fig. 49. Métrica Servicio de Internet

La métrica Provincia de Procedencia, se encuentra en la tabla Dim_Provincia con relación al Data Mart Persona y al DW, la Fig.50 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

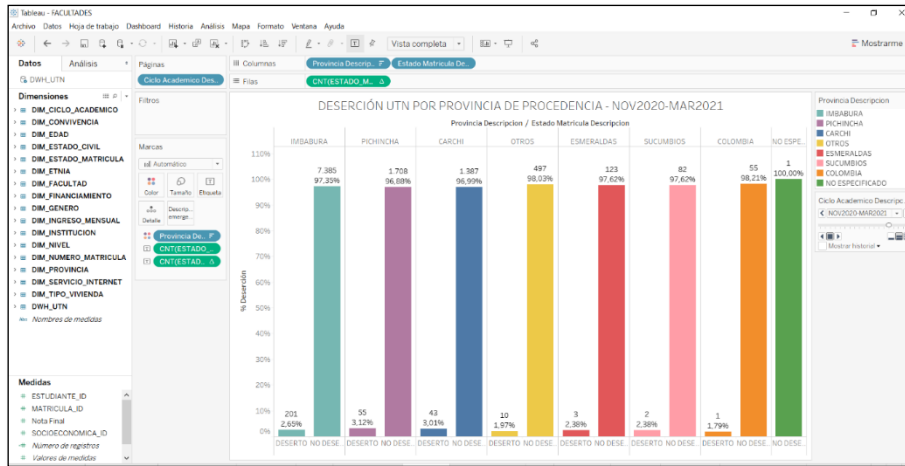


Fig. 50. Métrica Provincia de Procedencia

La métrica Número de Matrícula, se encuentra en la tabla Dim_Numero_Matricula con relación al Data Mart Académico y al DW, la Fig.51 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

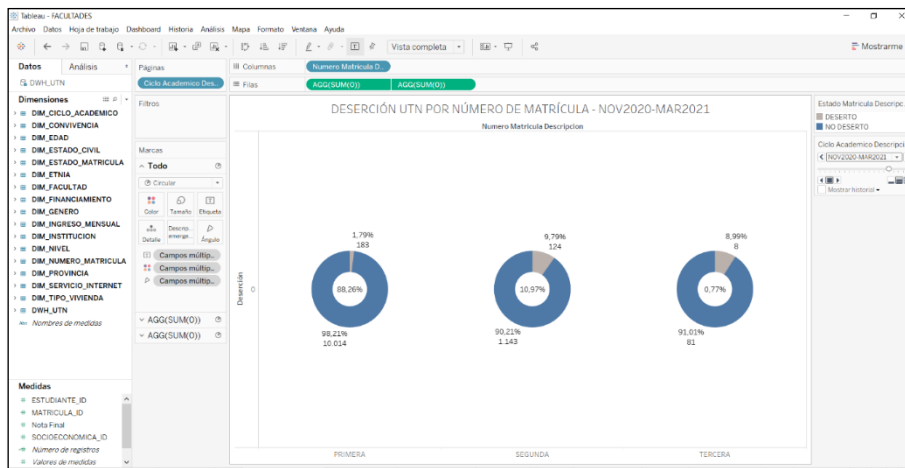


Fig. 51. Métrica Numero Matricula

La métrica Nivel, se encuentra en la tabla Dim_Nivel con relación al Data Mart Académico y al DW, la Fig.52 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

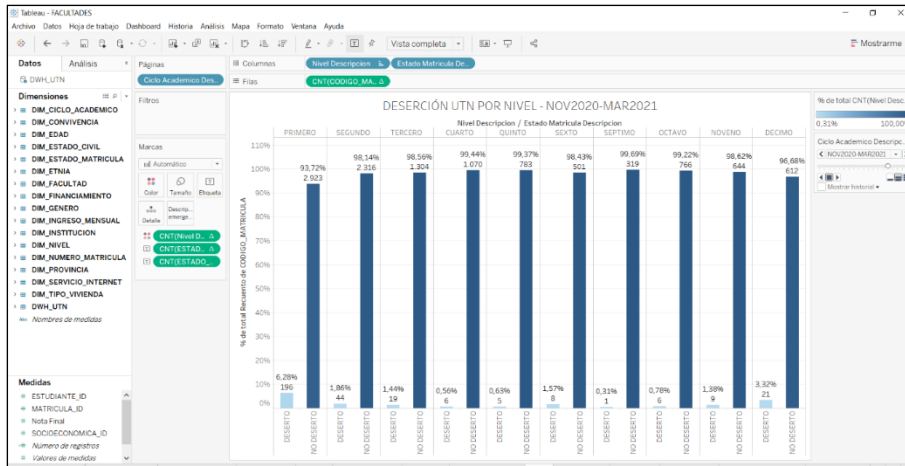


Fig. 52. Métrica Nivel

La métrica Ingreso Mensual, se encuentra en la tabla Dim_Ingreso con relación al Data Mart Socioeconómica y al DW, la Fig.53 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

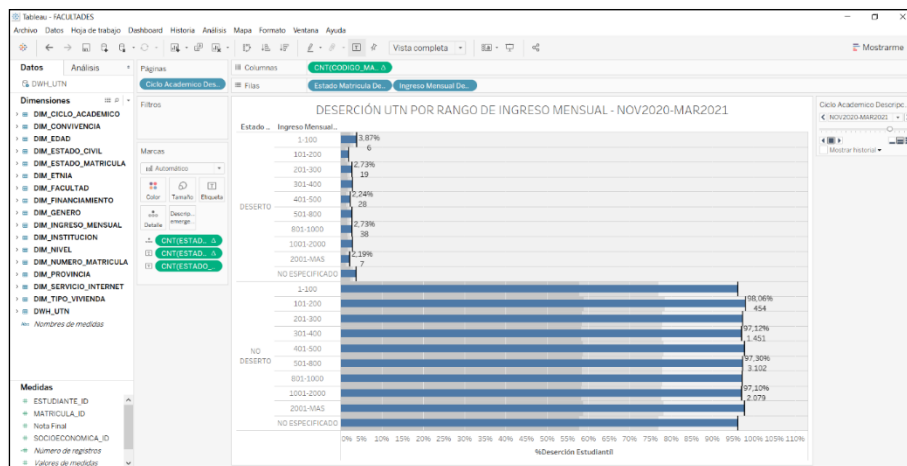


Fig. 53. Métrica Ingreso Mensual

La métrica Institución, se encuentra en la tabla Dim_Institucion2021 con relación al Data Mart Académico y al DW, la Fig.54 muestra el porcentaje de deserción en el período Nov2020-Mar2021.

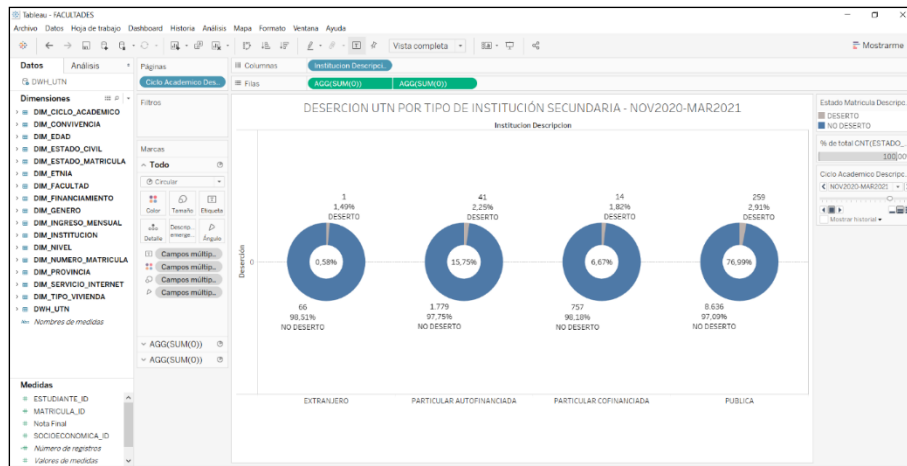


Fig. 54. Métrica Tipo Institución

2.8.3 Despliegue

En la fase de despliegue se unen todos los pasos realizados para presentar los tableros con la información obtenida de los datos extraídos de la base origen, importados en el área de trabajo y en Tableau, cuya finalidad es elaborar los tableros que serán presentados al usuario final.

2.8.3.1 Elaboración de la versión en blanco y negro del tablero

Se elaboraron tres tableros de acuerdo con cada Datamart, cada uno muestra información estadística sobre la deserción estudiantil permitiendo analizar y conocer el valor que aportan los indicadores encontrados en diferentes periodos académicos.

Para el Data Mart académico se utilizaron los indicadores: facultad, institución secundaria, número de matrícula y nivel, como se observa en la Fig.55 cada indicador muestra el porcentaje de deserción estudiantil en el periodo Nov2020-Mar2021.

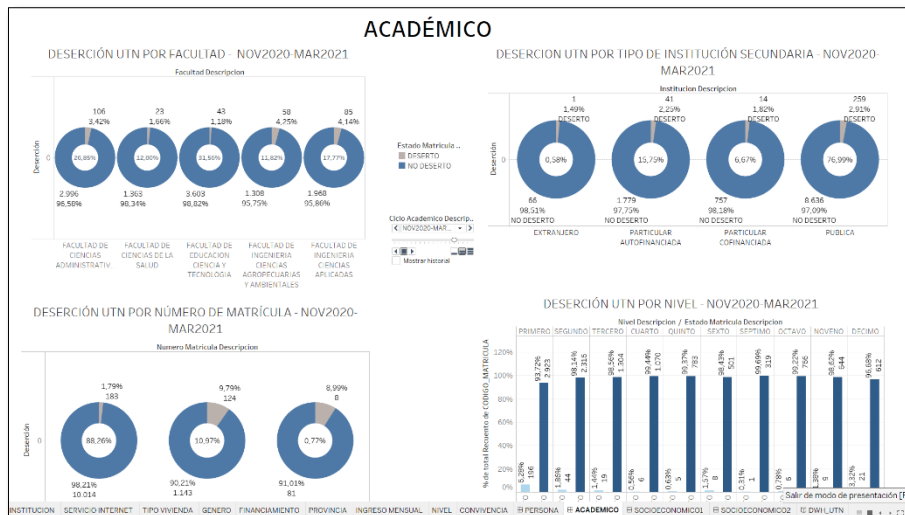


Fig. 55. Tablero Académico

En la Fig.56 se muestran los indicadores que conforman al Data Mart Persona que son: Estado Civil, Género, Provincia de Procedencia, Edad en el período Nov2019-Abr2020.

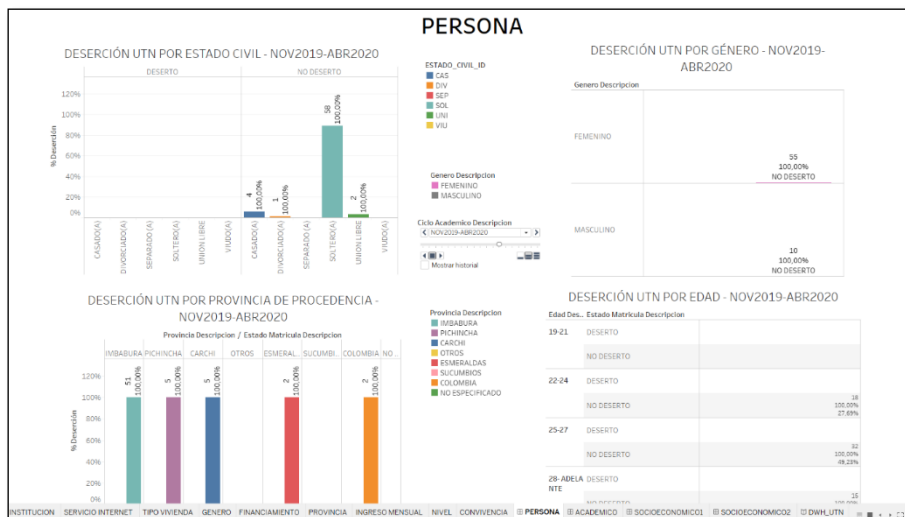


Fig. 56. Tablero Persona

El Data Mart Socioeconómico se conforma por los indicadores: Servicio de Internet, Etnia, Tipo Vivienda, Tipo Conviviente, Tipo de Financiamiento, Ingreso Mensual. Las Figuras 57 y 58, muestran los porcentajes de deserción en el período Nov2020-Mar2021.

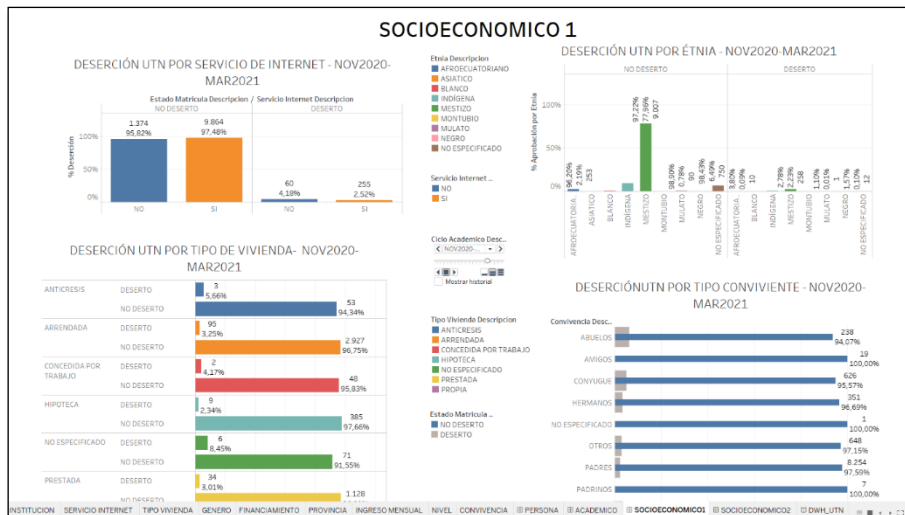


Fig. 57. Tablero Socioeconómico 1

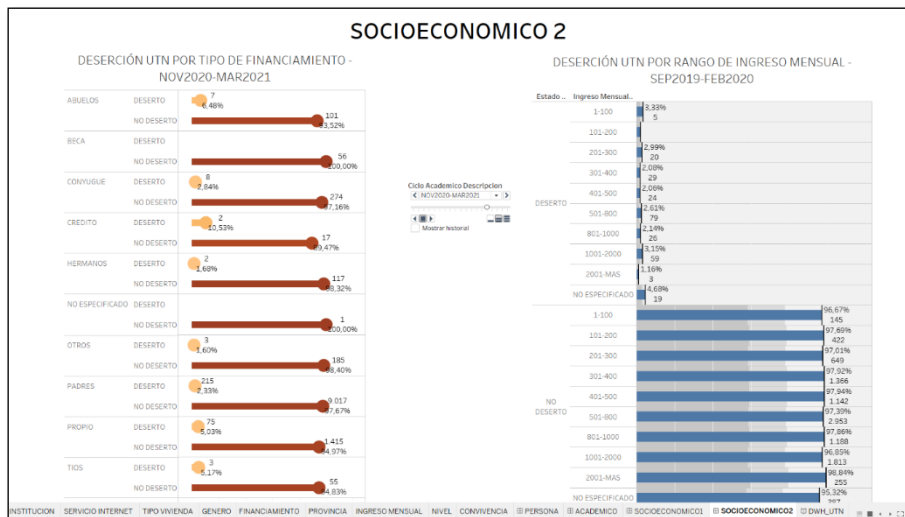


Fig. 58. Tablero Socioeconómico 2

2.8.3.2 Elaboración de la versión final del tablero

Una vez realizada la reunión con la persona responsable designada por el DDTI, y la personas Stakeholders del proyecto por parte del DBU de la Universidad Técnica del Norte, como se muestra en la Fig.59, presentados los tableros estadísticos de deserción estudiantil se procedió a realizar los cambios requeridos estableciendo mapas y barras estadísticas para cada indicador con la finalidad de tener una presentación mejorada y entendible de la información para la toma de decisiones.

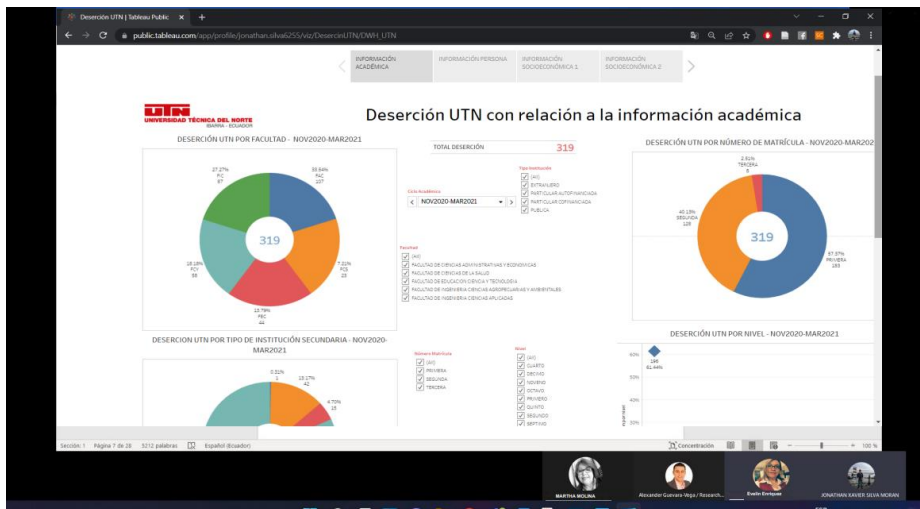


Fig. 59. Reunión para presentar la versión final

Con los indicadores identificados se establece el porcentaje de deserción con respecto a lo académico de cada estudiante por facultad a la que pertenece, el número de matrícula en la que se encuentra, por el tipo de institución y el nivel de carrera en el que se encuentra (véase el Anexo 2). La Fig.60 muestra la información mediante pasteles estadísticos.

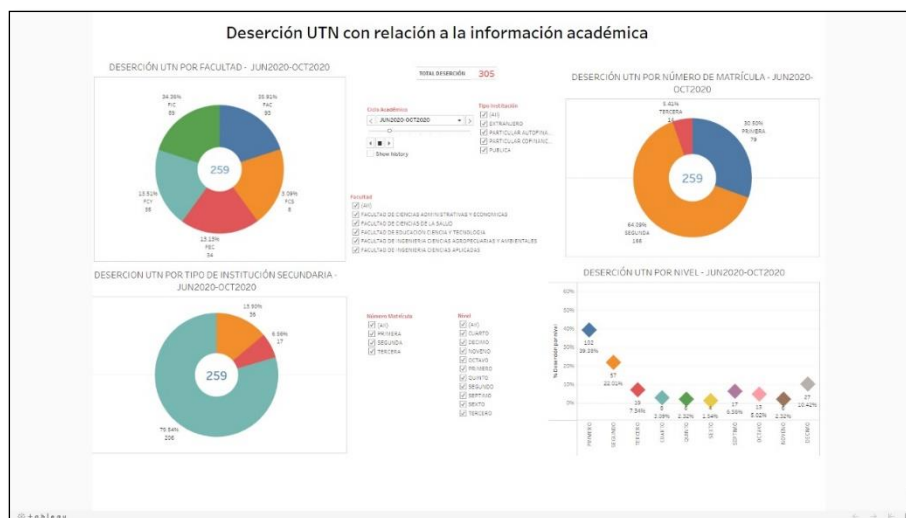


Fig. 60. Deserción Estudiantil en base a la información Académica

En la Fig.61 se muestra el porcentaje de deserción estudiantil por rango de edad, estado civil, genero a través de barras y el lugar de procedencia de los estudiantes con más deserción mediante el mapa territorial del Ecuador (véase el Anexo 3).

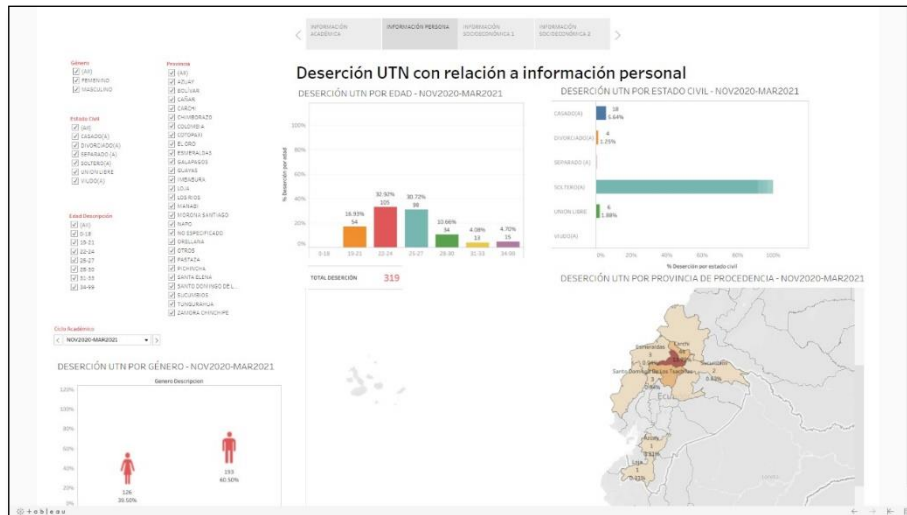


Fig. 61. Deserción Estudiantil en base a la información personal

En la Fig.62 se observa el porcentaje de deserción en base a la ficha socioeconómica del estudiante, tomando en cuenta los indicadores: Tipo conviviente, servicio de internet, tipo vivienda y etnia (véase el Anexo 4).

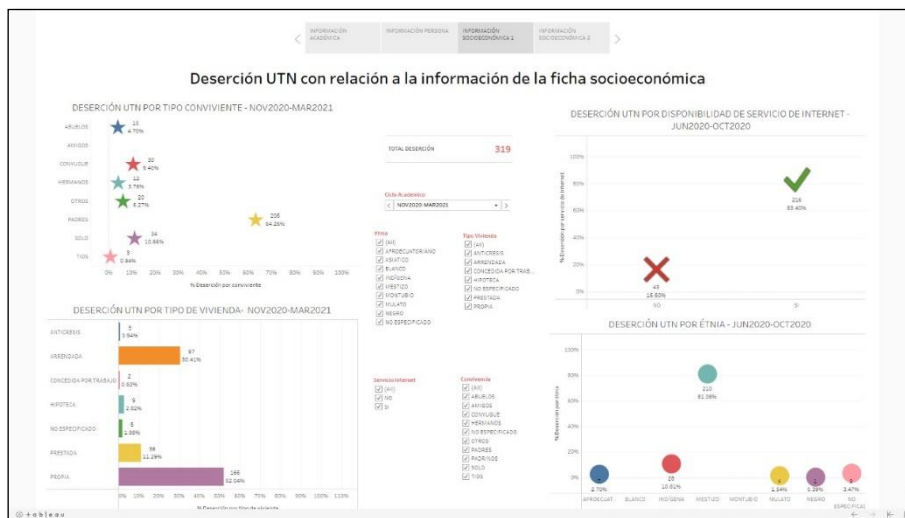


Fig. 61. Deserción Estudiantil en base a la información de la ficha socioeconómica

Por último, tomando la misma información de la ficha socioeconómica se requiere mostrar los indicadores de deserción por tipo de financiamiento e ingreso mensual, tal y como se ve en la Fig. 63 (véase el Anexo 5).



Fig. 62. Deserción Estudiantil en base a la información de la ficha socioeconómica 2

2.8.4 Crecimiento y Mantenimiento

El mantenimiento está enfocado en la retroalimentación entre el usuario final y el ciclo de vida, se verifica el funcionamiento correcto del desarrollo del DW, y el diagnóstico de los resultados.

CAPÍTULO III

VALIDACIÓN DE RESULTADOS

3 Análisis y Validación de Resultados

3.1 Evaluación de Producto con metodología CSAT

Para la validación de la satisfacción de los tableros se lo hizo con la metodología CSAT que es un indicador de satisfacción de cliente con el producto o servicio brindado a través de una encuesta que deberá ser llena posterior a la entrega del producto. CSAT son siglas en inglés es Customer Satisfaction Score, y en español significa Escala de satisfacción del cliente.

La metodología CSAT se ha popularizado en los ambientes virtuales, donde el valor numérico se representa con estrellas o emojis. En los medios virtuales sólo se consideran las evaluaciones que miden del uno al tres y del uno al cinco (Lazcano, 2009).

En el caso de este proyecto se ha partido por crear la encuesta en Microsoft forms (véase en el Anexo 6), con preguntas de validación y el posterior despliegue a través de correo electrónico a personal del Departamento de tecnología y el departamento de Bienestar Universitario de la Universidad Técnica del Norte

PV = Pregunta de validación.

La pregunta de validación 1 se muestra en la Fig.64 donde se evalúa el grado de cumplimiento por cada una de las historias de usuario.

PV1. ¿Cuál considera que es el nivel de cumplimiento de los requisitos de usuario?

[Más detalles](#)

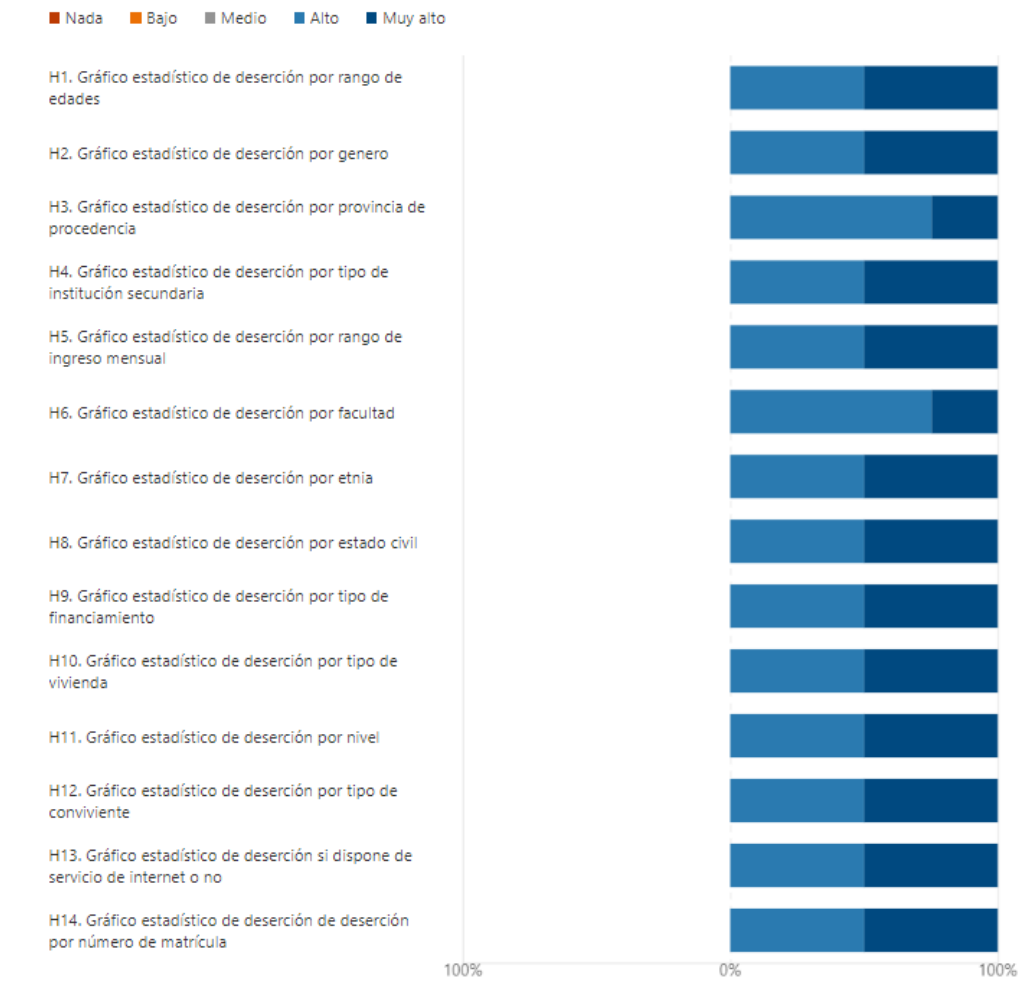


Fig. 63. Pregunta de Validación N°1

En la Fig. 65 se muestra la pregunta de validación 2 en donde se evalúa el nivel de satisfacción de los usuarios de la plataforma tomando en cuenta algunos ítems principales para el correcto desempeño en un entorno web.

PV2. ¿Cuál es su nivel de satisfacción con respecto a las siguientes variables relacionadas con la visualización de la solución de BI?

[Más detalles](#)

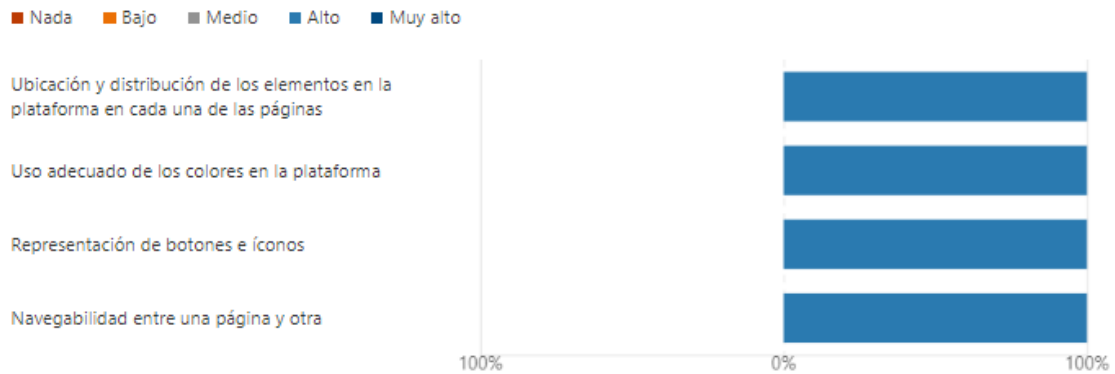


Fig. 64. Pregunta de Validación N°2

La Fig. 66 muestra la pregunta de validación 3 donde se evalúa el nivel de satisfacción relacionado con el cumplimiento de objetivos.

PV3. ¿Cuál es su nivel de satisfacción con respecto a las siguientes variables relacionadas con el cumplimiento de los objetivos de la solución?

[Más detalles](#)

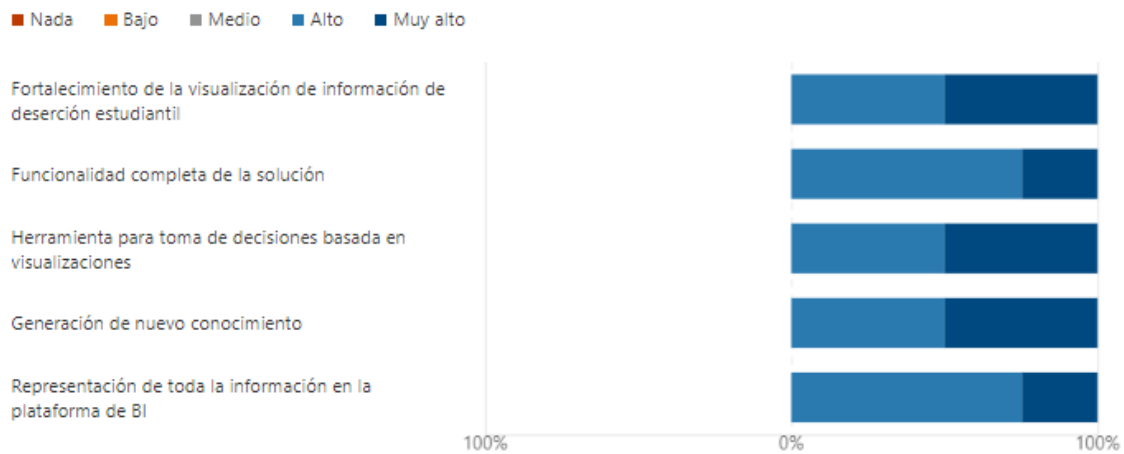


Fig. 65. Pregunta de Validación N°3

Pregunta de validación 4 la Fig. 67 muestra una calificación de manera general al producto final.

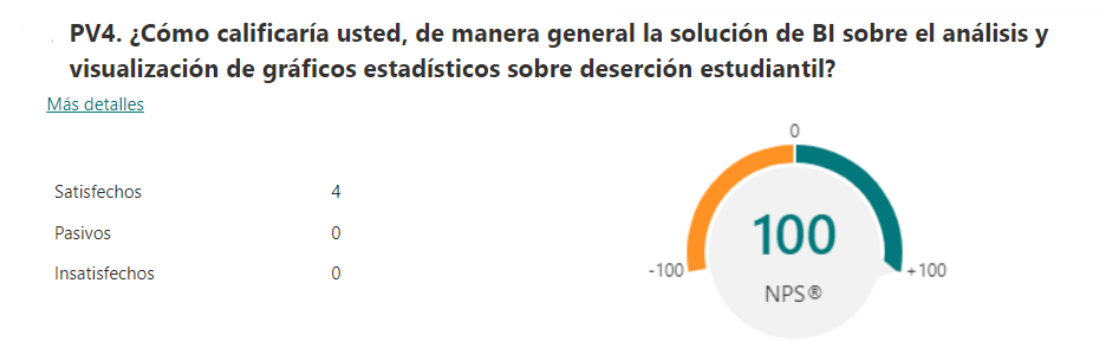


Fig. 66. Pregunta de Validación N°4

3.2 Calificación de Producto con metodología NPS

El NPS mide en una escala de 0 a 10 la satisfacción de un cliente con los productos y servicios. El objetivo es conocer su opinión a largo plazo, planteando la siguiente pregunta: ¿Cuál es el grado de satisfacción del cliente con el servicio?

Para aplicar esta metodología se sugiere tener una escala del uno al diez, en la cual se considere a quienes respondan entre seis y diez como clientes satisfechos con el producto o servicio y a quienes respondan entre uno y cinco, como insatisfechos (Lazcano, 2009).

Satisfecho: Cuando la calificación del producto obtiene una calificación de 9 o 10

Pasivo: Cuando la calificación se encuentra entre 8 y 7, esto significa que hay un pequeño grado de insatisfacción con el producto o servicio brindado

Insatisfecho: Califican al producto entre 5 y 0 lo que expresa un alto nivel de insatisfacción.

Una vez desplegado la encuesta se obtienen los siguientes resultados

Satisfecho 100%

Pasivo 0%

Insatisfecho 0%

Aplicando la fórmula para calcular NPS

$$\% \text{Satisfecho} - \% \text{Insatisfecho} = \text{NPS}$$

$$100\% - 0\% = 100\%$$

Una vez aplicada la fórmula para darle una calificación a la satisfacción con el servicio de Tableros Visuales para la interpretación de la deserción estudiantil en la Universidad Técnica del Norte se obtiene una calificación de 10 sobre 10 que se puede interpretar como una calificación pasiva tendiendo al grado de satisfacción perfecta.

3.3 Análisis de impactos

Este trabajo promueve el uso de nuevas técnicas y herramientas de Business Intelligence que ayudan a los directivos y personal encargado de velar por el bienestar de los estudiantes, tomen decisiones estratégicas basadas en datos reales sobre el índice de deserción estudiantil en los diferentes ámbitos encontrados a través del análisis de la información visualizada en tableros estadísticos.

Una vez realizados los tableros con la herramienta Tableau se procede con el análisis de impactos sobre el trabajo realizado definiendo posibles consecuencias estableciendo cuatro tipos de impactos detallados a continuación:

- **Impacto Tecnológico**

Utilizar nuevas tecnologías y a su vez explotarlas al máximo permiten mejorar habilidades y conocimientos en la toma de decisiones estratégicas, elaborando nuevos proyectos en base a los indicadores de deserción estudiantil con el objetivo de brindar ayuda a los estudiantes y así disminuir el alto índice de deserción.

- **Impacto Ambiental**

En toda Institución de Educación Superior se almacena grandes cantidades de datos ya sea en espacios físicos o virtuales, todos esos datos ocupan espacio y a mayor ingreso de estudiantes mayor consumo de papel y recursos tecnológicos. Sin embargo, al usar Business Intelligence y las diferentes herramientas de visualización de datos como Tableau se logra disminuir estos consumos realizando una valoración de los datos, limpiando datos no relevantes y procesándolos para convertirlos en información estratégica.

- **Impacto Educativo**

El proyecto permitirá realizar planes de acción abordando la problemática directamente, donde se garantice la educación de calidad para los estudiantes, logrando así subir los niveles de desempeño, a través de las pautas establecidas que generan una información relevante de los motivos a lo que se enfrentan los estudiantes al momento de desertar en sus estudios.

- **Impacto Económico**

El impacto económico para la institución se verá reflejado en el aprovechamiento de recursos ya que entre menos deserción estudiantil mayor explotación de recursos asignados a la institución.

En cuanto al impacto para los estudiantes, a través de la ayuda de los directivos con sus planes de acción, los estudiantes evitaran repetir materias y realizar pagos de segundas y terceras matrículas.

CONCLUSIONES

La construcción de la revisión bibliográfica permitió conocer y utilizar las técnicas y metodologías para trabajar con volúmenes extensos de datos centrándose únicamente en procesos de análisis de información que ayudaron a la toma de decisiones oportuna sobre el giro de negocio institucional, en este caso la calidad de educación, ayudando así a fortalecer el proceso y apoyando la mejora continua disminuyendo el índice de deserción estudiantil.

La implementación de la metodología Kimball permitió construir un almacén de datos partiendo desde lo más simple hasta lo más complejo adaptándose a cualquier tamaño obteniendo un almacén de datos versátil para el posterior análisis que esta presto al constante cambio y mantenimiento.

Tableau siendo una herramienta que ha tomado fuerza en la última década referente a el análisis y visualización de datos en forma interactiva con el usuario y no de forma plana, permitió la construcción y visualización de tableros dando al proyecto un realce y una mejor interpretación de la información, brindando facilidad de construcción de gráficos simples y detallados con vistas que posteriormente serán presentadas e interpretadas por el usuario final.

A través de la solución web de Tableau se conectó a la aplicación de escritorio que ayudó en la publicación de vistas y tableros gráficos de forma sencilla y casi instantánea facilitando el trabajo a los desarrolladores.

Para la validación de satisfacción de los tableros entre el cliente y el servicio brindado se aplicó la metodología CSAT calificando el servicio a través de una encuesta realizada al Departamento de Bienestar Universitario y al Departamento de Desarrollo Tecnológico e Informático con un rango del 1 al 10 y con una calificación de 8.5 se puede decir que en su mayoría los usuarios están conformes con el proyecto entregado.

RECOMENDACIONES

Para posteriores proyectos enfocados a la deserción estudiantil es concerniente definir claramente la palabra deserción ya que es un término que queda a la interpretación individual y no hay algo que especifique el termino deserción. En el presente proyecto de titulación se consideró como deserción cuyos estudiantes anularon la matrícula o los que realizaron el proceso de matrícula en línea, pero nunca legalizaron.

Para tener una correcta armonía con el desarrollo de proyectos de tecnología enfocados a dar apoyo a la Universidad Técnica del Norte es recomendable trabajar con infraestructura y software que dispongan en el departamento de tecnología, esto hace que el proyecto funcione correctamente, no haya problemas con licencias, soporte y o versiones de software, dando así la facilidad del mantenimiento y evolución del proyecto.

La base de datos transaccional de la Universidad Técnica del Norte cuenta con gran cantidad de información por lo que es de gran importancia tener una data limpia que facilite los posteriores proyectos orientados a el análisis de la información que allí se encuentra, en la actualidad aún hay información suelta que no tiene relación con nada y así mismos campos de relevancia que se encuentran vacíos, esto hace más complicada la tarea de la limpieza de información.

Se recomienda trabar con el mismo motor de base de datos para crear el área de análisis que facilite el trabajo de conexión de la base de datos transaccional y así poder trabajar con información actual aplicando métodos o procedimientos almacenados que estén programados en la extracción limpieza y carga de la nueva información en un determinado tiempo, con la finalidad de tener gráficos actualizados.

Disponer de personal para que trabaje con la problemática de deserción estudiantil, ya que se ha evidenciado un alto índice de deserción en la Universidad Técnica del Norte y a su vez darle continuidad al proyecto aplicando más variables para identificar ciertos patrones que no se han tomado en cuenta en este proyecto.

Por parte de los clientes se vieron interesados en saber el motivo de deserción, dicha información no se encuentra en la base de datos y se sugiere incorporar un campo donde se ingrese el motivo de la deserción que debería ser ingresado por el docente, secretaria de carrera o algún departamento que se encargue de dar el seguimiento a esta problemática.

REFERENCIAS Y BIBLIOGRAFIA

- Alejandra, O., & Claudia, B. (2014). *Metodología de la Investigación*. Retrieved from www.memoria.fahce.unlp.edu.ar
- Baaziz, A., & Quoniam, L. (2013). How to use Big Data technologies to optimize operations in Upstream Petroleum Industry. *International Journal of Innovation*, 1(1), 19–25. <https://doi.org/10.5585/iji.v1i1.4>
- Barr, J. R. (2019). Machine learning, A tutorial with R. *Proceedings - 2018 1st IEEE International Conference on Artificial Intelligence for Industries, AI4I 2018*, 120–121. <https://doi.org/10.1109/AI4I.2018.8665676>
- Barrios, J. (2019, December 24). Sistemas de Información. El Business Intelligence. Retrieved December 16, 2020, from BIGDATA website: <https://www.juanbarrios.com/sistemas-de-informacion-el-business-intelligence/>
- Bernabeu, D., & Mattío, G. (2017). *Introducción DATA WAREHOUSING : Marco Conceptual HEFESTO : Metodología Data Warehouse*. 182.
- Camargo, J., Joyanes, L., & Giraldo, L. (2016). La Business Intelligence como una herramienta en la gestión académica. *Revista Científica, Universidad Distrital Francisco José Caldas*, (24), 110–120. <https://doi.org/10.14483/udistrital.jour.RC.2016.24.a11>
- Carhuallanqui, J. L. (2017). Diseño de una solución de Business Intelligence como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma. Retrieved from <http://cybertesis.unmsm.edu.pe/handle/cybertesis/7208>
- Chamorro, J. (2016). *GRADO EN INGENIERÍA INFORMÁTICA APLICACIÓN WEB ELABORACIÓN Y*.
- Conesa Caralt, J., & Curto Díaz, J. (2013). *Introducción al Business Intelligence* (UOC). Retrieved from <https://elibro.net/es/ereader/utnorte/56524?page=19>
- Curto Díaz, J. (2016). *Introducción al business intelligence* (UOC, Ed.). Retrieved from <https://elibro.net/es/ereader/utnorte/101030?page=17>.
- Deloitte. (2021). Técnicas de estimación en SCRUM. Retrieved May 2, 2021, from

<https://www2.deloitte.com/es/es/pages/technology/articles/tecnicas-de-estimacion-en-scrum.html>

Dertiano, V. (2015). El enfoque de Ralph Kimball. Retrieved January 6, 2021, from 06/04/2015 website: <https://blog.bi-geek.com/arquitectura-el-enfoque-de-ralph-kimball/>

Jose, C. C. J. P. (2019). *Tableau para la Business Intelligence del Área de Análisis de Información TI LIMA- PERÚ Tableau para la Business Intelligence del Área de Análisis de Información TI LIMA- PERÚ*.

Kimball, R. (2013). *Enterprise Data Warehouse Bus Architecture*. 248–252. Retrieved from <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/kimball-data-warehouse-bus-architecture/>

Ko, I., & Chang, H. (2017). Interactive visualization of healthcare data using Tableau. *Healthcare Informatics Research*, 23(4), 349–354. <https://doi.org/10.4258/hir.2017.23.4.349>

Lazcano, J. A. B. (2009). *Métodos de Recolección de Información para el Análisis de la Satisfacción del Cliente*. Resumen Palabras clave Key words. 1–22.

Leon, M. (2014). *Inteligencia de Negocio: Ciclo de vida de Ralph Kimball*. Retrieved from <http://luisleonin.blogspot.com/2014/02/ciclo-de-vida-de-ralph-kimball.html>

López Benítez, Y. (2018). *Business Intelligence*. ADGG102PO. Retrieved from <https://elibro.net/es/ereader/utnorte/124393?page=10>.

López de Munain, C., Torrent, M., Molina, F., Schanz, S., Sandoval, A., Sfiligoy, J. P., & Sassaroli, M. L. (2014). Sistemas de apoyo a la toma de decisiones. Una aplicación en el área de gestión universitaria. *WICC 2014 XVI Workshop de Investigadores En Ciencias de La Computación*, (1), 178–182. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/42010>

M., D., & P., B. (2014). Business Intelligence: Metodología de Kimball. Retrieved January 6, 2021, from Blogspot website: <http://inteligenciadenegociosdiegobrito.blogspot.com/2014/01/la-metodologia-de-kimball.html>

- Machado, J. (2019). 23.000 universitarios abandonan cada año la educación superior. *Primicias*. Retrieved from <https://www.primicias.ec/noticias/sociedad/agustin-alban-senescyt-ser-bachiller-universidades/>
- Microsoft. (2019). What is Power BI Desktop? Retrieved December 16, 2020, from Website website: <https://powerbi.microsoft.com/en-us/what-is-power-bi/>
- Millalen Alejandro. (2018). Las 6 mejores soluciones Open Source para Business Intelligence | El Blog de Ale. Retrieved December 16, 2020, from <https://alejandromillalen.com/mejores-soluciones-open-source-para-inteligencia-de-negocios/>
- Mistry, A. (2017). Agile Story Point Estimation Techniques - T-Shirt Sizing. Retrieved May 2, 2021, from <https://www.c-sharpcorner.com/article/agile-story-point-estimation-techniques-t-shirt-sizing/>
- Navarro, M., & Perez, D. (2018). Sistema de información ejecutiva web para mejorar monitoreo de los estándares de calidad para la acreditación de una escuela profesional de una universidad de la región Lambayeque. *Interciencia*, 489(20), 65.
- Piedad, Z., Ruiz, M., & Álvarez, M. (2016). Deserción estudiantil universitaria en Ecuador y su influencia en la calidad del egresado. *Revista Magazine de Las Ciencias*, 1(4), 65–70. Retrieved from <https://revistas.utb.edu.ec/index.php/magazine/article/view/183>
- RapidMiner. (2019). Why RapidMiner. Retrieved December 16, 2020, from <https://rapidminer.com/why-rapidminer/>
- Requena, A. (2018). Qué es un Sprint de Scrum | OpenWebinars. Retrieved May 2, 2021, from <https://openwebinars.net/blog/que-es-un-sprint-scrum/>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial*, 10(29), 11–18.
- Rivadera, G. R. (2010). La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). *Cuadernos de La Facultad*, 5, 56–71. Retrieved from <http://www1.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>

- Roche, J. (2017). Scrum: roles y responsabilidades | Deloitte España. Retrieved May 2, 2021, from Deloitte website: <https://www2.deloitte.com/es/es/pages/technology/articles/roles-y-responsabilidades-scrum.html>
- Sinnexus. (2017). Arquitectura de una solución de Business Intelligence. Retrieved December 16, 2020, from https://www.sinnexus.com/business_intelligence/arquitectura.aspx
- Sistemas de Soporte a la Decisión (DSS). (n.d.). Retrieved January 28, 2021, from https://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx
- Tableau Software. (2020). Visualizar datos públicos | Explore, aprenda, comparte | Tableau Public. Retrieved December 16, 2020, from <https://public.tableau.com/en-us/s/about>
- Tableau Software LLC. (2020). Dashboards - Tableau. Retrieved May 4, 2021, from <https://help.tableau.com/current/pro/desktop/es-es/dashboards.htm>
- Thaís Balagueró. (2018, July 4). Big Data: Componentes del Business Intelligence | Deusto Formación. Retrieved December 16, 2020, from DEUSTO FORMACION website: <https://www.deustoformacion.com/blog/gestion-empresas/big-data-componentes-business-intelligence>
- Trujillo, T. M. (2018). *Gestión de datos empresariales utilizando procesos ETL*. 63. Retrieved from [http://ri.uaemex.mx/bitstream/handle/20.500.11799/95251/Gestion de datos empresariales utilizando procesos ETL.pdf?sequence=1&isAllowed=y](http://ri.uaemex.mx/bitstream/handle/20.500.11799/95251/Gestion_de_datos_empresariales_utilizando_procesos_ETL.pdf?sequence=1&isAllowed=y)
- Vila, D. (2019). *Detección De Patrones De Deserción Estudiantil Utilizando Técnicas Predictivas De Clasificación Y Regresión De Minería De Datos, Para La Gestión Académica De La Universidad Técnica Del Norte*. 104.

ANEXOS

Anexo 1. Procedimientos Almacenados: <https://bit.ly/3ufguGM>

Anexo 2. Tablero Tableau Deserción Académica: <https://tabsoft.co/34entoQ>

Anexo 3. Tablero Tableau Deserción Persona: <https://tabsoft.co/3hQDdqa>

Anexo 4. Tablero Tableau Deserción Socioeconómico 1: <https://tabsoft.co/34cNeFV>

Anexo 5. Tablero Tableau Deserción Socioeconómico 2: <https://tabsoft.co/3ulAreW>

Anexo 6. Encuesta de Validación de Tableros: <https://bit.ly/3wt3fUi>