

**UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN - TACNA**

**Escuela de Posgrado**

**Maestría en Ingeniería de Sistemas e Informática -  
Administración de Tecnologías de Información**

**ANÁLISIS PREDICTIVO DE LA DESERCIÓN ESTUDIANTIL  
UTILIZANDO DATA WAREHOUSE Y MINERÍA DE DATOS  
EN LA UNIVERSIDAD NACIONAL JORGE BASADRE  
GROHMANN - TACNA, 2012-2018**

**TESIS**

**PRESENTADA POR:**

**NELSON ABRAHAN PABLO MOLLO CONDORI**

**Para optar el Grado Académico de:**

**MAESTRO EN CIENCIAS (*MAGISTER SCIENTIAE*)  
CON MENCIÓN EN INGENIERÍA DE SISTEMAS E  
INFORMÁTICA - ADMINISTRACIÓN DE  
TECNOLOGÍAS DE INFORMACIÓN**

**TACNA – PERÚ**

**2018**





**UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN - TACNA**

**Escuela de Posgrado**

**MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA -  
ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN**

**ANÁLISIS PREDICTIVO DE LA DESERCIÓN ESTUDIANTIL  
UTILIZANDO DATA WAREHOUSE Y MINERÍA DE DATOS EN LA  
UNIVERSIDAD JORGE NACIONAL BASADRE GROHMANN - TACNA,  
2012-2018**

Tesis sustentada y aprobada el 28 de diciembre del 2018; estando el jurado calificador integrado por:

|            |   |   |
|------------|---|---|
| Presidente | : | <br>_____ |
|            |   | Dr. Roberto Encarnación Supo Hallasi  |
| Secretario | : | <br>_____ |
|            |   | M.Sc. Edgar Aurelio Taya Acosta   |
| Miembro    | : | <br>_____ |
|            |   | Dr. Edwin Antonio Hinojosa Ramos  |
| Asesor     | : | <br>_____  |
|            |   | Dr. Erbert Francisco Osco Mamani  |

## **AGRADECIMIENTO**

Gracias a Dios mi padre celestial que me brinda la fortaleza física y espiritual necesaria que me ha permitido tener la capacidad para poder culminar este trabajo y escalar un peldaño en mi carrera profesional.

Agradezco profundamente al Dr. Erbert Francisco Osco Mamani, por su apoyo como asesor en el desarrollo de esta investigación.

Mi reconocimiento a todos los docentes de la Maestría en Ingeniería de Sistemas e Informática - Administración Tecnología de Información por sus valiosas enseñanzas, experiencias, consejos y sugerencias, las cuales me han servido en el desarrollo de este trabajo.

Finalmente, mi agradecimiento a todas las personas que me apoyaron en forma directa e indirecta, las cuales me han permitido culminar la presente investigación.

## DEDICATORIA

A mis amados padres, Pablo Mollo Mita y Ana Condori Alaví los cuales me brindaron su apoyo incondicional e incentivaron en mí, el deseo de superación y progreso personal, igualmente, a mis hermanos Ada y Alex a quienes amo mucho.

## CONTENIDO

|  |       |
|--|-------|
| ÍNDICE DE FIGURAS                                | x     |
| ÍNDICE DE TABLAS                                 | xiv   |
| RESUMEN  | xvii  |
| ABSTRACT   | xviii |
| INTRODUCCIÓN                                     | 1     |
| CAPÍTULO I: PLANTEAMIENTO DE INVESTIGACIÓN       | 3     |
| 1.1. Descripción del problema                    | 3     |
| 1.2. Formulación del problema                    | 7     |
| 1.2.1. Formulación del problema general          | 7     |
| 1.2.2. Formulación de los problemas específicos  | 8     |
| 1.3. Justificación                               | 8     |
| 1.4. Alcances y limitaciones                     | 10    |
| 1.4.1. Alcances                                  | 10    |
| 1.4.2. Limitaciones                              | 10    |
| 1.5. Objetivos                                   | 11    |
| 1.5.1. Objetivo general                          | 11    |
| 1.5.2. Objetivos específicos                     | 11    |
| CAPÍTULO II: MARCO TEÓRICO                       | 12    |
| 2.1. Bases teóricas respecto al problema         | 12    |
| 2.1.1. Antecedentes de la investigación          | 12    |
| 2.2. Deserción: Definición                       | 14    |
| 2.2.1. Tipos de deserción                        | 14    |
| 2.2.2. Factores de la deserción                  | 16    |
| 2.3. Base de datos y modelo de datos             | 16    |
| 2.3.1. Historia de los sistemas de base de datos | 16    |
| 2.3.2. Base de datos                             | 17    |
| 2.3.3. Definición de sistema                     | 20    |

|            |  |    |
|------------|--|----|
| 2.3.4.     | Sistema de base de datos                     | 20 |
| 2.3.5.     | Concepto de información                      | 20 |
| 2.3.6.     | Modelo                                       | 21 |
| 2.3.7.     | Modelo de datos                              | 21 |
| 2.3.8.     | Modelo relacional                            | 21 |
| 2.3.9.     | Modelo entidad-relación                      | 24 |
| 2.4.       | Data Warehouse                               | 26 |
| 2.4.1.     | Definición de Data Warehouse                 | 26 |
| 2.4.2.     | Características de Data Warehouse            | 28 |
| 2.4.3.     | Arquitectura de Data Warehouse               | 29 |
| 2.5.       | Modelado multidimensional                    | 37 |
| 2.5.1.     | Modelado multidimensional conceptual         | 37 |
| 2.5.1.1.   | Propiedades multidimensionales estructurales | 38 |
| 2.5.1.1.1. | Dimensiones                                  | 39 |
| 2.5.1.1.2. | Hechos                                       | 42 |
| 2.5.1.1.3. | Cubo multidimensional                        | 45 |
| 2.5.1.2.   | Propiedades multidimensionales dinámicas     | 47 |
| 2.5.2.     | Modelado multidimensional lógico             | 49 |
| 2.5.2.1.   | Tipo de modelamiento de un Data Warehouse    | 51 |
| 2.5.2.1.1. | Esquema de estrella                          | 51 |
| 2.5.2.1.2. | Esquema de copa de nieve                     | 53 |
| 2.5.2.1.3. | Esquema de constelación                      | 55 |
| 2.6.       | Metodología Ralph Kimball                    | 57 |
| 2.6.1.     | Planificación del proyecto                   | 58 |
| 2.6.2.     | Definición de los requerimientos del negocio | 58 |
| 2.6.3.     | Diseño de la arquitectura                    | 59 |
| 2.6.4.     | Instalación y selección del producto         | 59 |
| 2.6.5.     | Modelamiento dimensional                     | 59 |
| 2.6.6.     | Diseño físico                                | 60 |

|            |  |    |
|------------|--|----|
| 2.6.7.     | Diseño e implementación del subsistema de extracción, transformación y carga (ETL) | 60 |
| 2.6.8.     | Especificación de aplicaciones de business intelligence (BI)                       | 60 |
| 2.6.9.     | Desarrollo de aplicación BI  | 61 |
| 2.6.10.    | Implementación   | 61 |
| 2.7.       | Descubrimiento de conocimiento en base de datos                                    | 62 |
| 2.7.1.     | Definición de descubrimiento de conocimiento en base de datos                      | 62 |
| 2.7.2.     | Fases del proceso de KDD   | 66 |
| 2.7.2.1.   | Fase de integración y recopilación   | 67 |
| 2.7.2.2.   | Fase de selección, limpieza y transformación                                       | 68 |
| 2.7.2.3.   | Fase de minería de datos   | 69 |
| 2.7.2.4.   | Fase de evaluación e interpretación  | 69 |
| 2.7.2.5.   | Fase de difusión, uso y monitorización   | 69 |
| 2.7.3.     | Cross-Industry Standard Process for Data Mining                                    | 70 |
| 2.7.3.1.   | Fase de comprensión del negocio  | 71 |
| 2.7.3.2.   | Fase de comprensión de los datos   | 71 |
| 2.7.3.3.   | Fase de preparación de los datos   | 71 |
| 2.7.3.4.   | Fase de modelado   | 72 |
| 2.7.3.5.   | Fase de evaluación   | 72 |
| 2.7.3.6.   | Fase de despliegue   | 73 |
| 2.8.       | Data Mining  | 73 |
| 2.8.1.     | Tipo de conocimiento   | 73 |
| 2.8.2.     | Definición de Data Mining  | 76 |
| 2.8.3.     | Técnicas de Data Mining  | 78 |
| 2.8.3.1.   | Predictivas  | 78 |
| 2.8.3.1.1. | Regresión  | 78 |
| 2.8.3.1.2. | Series temporales  | 78 |
| 2.8.3.1.3. | Redes bayesianas   | 79 |
| 2.8.3.1.4. | Algoritmos genéticos   | 79 |

|   |  |           |
|---|--|-----------|
| 2.8.3.1.5.                              | Redes neuronales   | 79        |
| 2.8.3.1.6.                              | Clasificación ad hoc árboles de decisión                   | 80        |
| 2.8.3.2.                                | Descriptivas   | 80        |
| 2.8.3.2.1.                              | Clustering   | 80        |
| 2.8.3.2.2.                              | Asociación   | 80        |
| 2.8.3.2.3.                              | Secuenciamiento  | 81        |
| 2.8.4.                                  | Técnicas de validación de algoritmos computacionales de DM | 81        |
| 2.8.4.1.                                | Validación cruzada de k iteraciones                        | 81        |
| 2.8.4.2.                                | Matriz de confusión  | 82        |
| 2.8.4.3.                                | La curva característica operativa del receptor             | 84        |
| 2.9.                                    | Definición de términos                                     | 84        |
| 2.9.1.                                  | Definiciones generales                                     | 84        |
| 2.9.2.                                  | Calidad  | 85        |
| 2.9.3.                                  | Metodología  | 85        |
| 2.9.4.                                  | Método   | 86        |
| 2.9.5.                                  | Técnica  | 86        |
| 2.9.6.                                  | Proceso  | 86        |
| 2.9.7.                                  | Calidad de software  | 87        |
| 2.9.8.                                  | Sistemas informáticos                                      | 88        |
| 2.9.9.                                  | Análisis y diseño de sistemas                              | 88        |
| 2.9.10.                                 | Análisis predictivo  | 89        |
| <b>CAPÍTULO III: MARCO METODOLÓGICO</b> |  | <b>90</b> |
| 3.1.                                    | Hipótesis  | 90        |
| 3.2.                                    | Variables  | 90        |
| 3.2.1.                                  | Identificación de variables                                | 90        |
| 3.2.2.                                  | Definición de variables                                    | 90        |
| 3.2.3.                                  | Operacionalización de variables                            | 91        |
| 3.3.                                    | Diseño de la investigación                                 | 92        |
| 3.3.1.                                  | Diseño experimental ó no experimental                      | 92        |

|        |   |            |
|--------|---|------------|
| 3.3.2. | Población y muestra   | 92         |
| 3.3.3. | Técnicas e instrumentos para la recolección de datos                    | 92         |
| 3.3.4. | Procesamiento y análisis de datos                                       | 93         |
|        | <b>CAPÍTULO IV: RESULTADOS</b>  | <b>95</b>  |
| 4.1.   | Resultados de la metodología de Ralph kimball en la construcción del DW | 95         |
| 4.1.1. | Factores individuales   | 95         |
| 4.1.2. | Factores académicos   | 99         |
| 4.1.3. | Factores Socioeconómicos  | 103        |
| 4.2.   | Resultados de la metodología Crisp-dm del desarrollo del proyecto DM    | 107        |
| 4.3.   | Constratación de los datos reales con los predichos.                    | 109        |
|        | <b>CAPÍTULO V: DISCUSIÓN</b>  | <b>111</b> |
|        | <b>CONCLUSIONES</b>   | <b>113</b> |
|        | <b>RECOMENDACIONES</b>  | <b>115</b> |
|        | <b>REFERENCIAS BIBLIOGRÁFICAS</b>                                       | <b>116</b> |
|        | <b>ANEXOS</b>   | <b>123</b> |
|        | <b>ANEXO 1: MATRIZ DE CONSISTENCIA</b>                                  | <b>124</b> |
|        | <b>ANEXO 2: DESARROLLO DEL DW Y DM</b>                                  | <b>125</b> |
|        | <b>ANEXO 4: CÓDIGO AUTOGENERADO DEL ERWIN DATA MODELE r7</b>            | <b>213</b> |
|        | <b>ANEXO 5: ETL PARA EL MODELO DIMENSIONAL</b>                          | <b>230</b> |
|        | <b>ANEXO 6: CONSTRUCCIÓN DEL MODELO DE MINERÍA DE DATOS</b>             | <b>242</b> |
|        | <b>ANEXO 7: IMPLEMENTACIÓN DE DW Y DM</b>                               | <b>253</b> |
|        | <b>ANEXO 8: VALIDEZ DEL INSTRUMENTO</b>                                 | <b>254</b> |

## ÍNDICE DE FIGURAS

|  |    |
|--|----|
| Figura 1. Deserción con respecto al tiempo.                            | 15 |
| Figura 2. Deserción con respecto al espacio.                           | 15 |
| Figura 3. Factores de deserción estudiantil.                           | 16 |
| Figura 4. Historia de los sistemas de base de datos.                   | 17 |
| Figura 5. Diagrama de entidad - relación.                              | 26 |
| Figura 6. Arquitectura lógica de un Data Warehouse.                    | 31 |
| Figura 7. Arquitectura de Data Warehouse.                              | 32 |
| Figura 8. Enfoque de inmon de un DW corporativo.                       | 33 |
| Figura 9. Enfoque de kimball - arquitectura bus del DW                 | 34 |
| Figura 10. Principales características del enfoque de inmon y kimball. | 34 |
| Figura 11. Arquitectura general del Data Warehouse.                    | 36 |
| Figura 12. Arquitectura de Data Warehouse de hefestio.                 | 36 |
| Figura 13. Cubo de datos.  | 38 |
| Figura 14. Jerarquía de grafos acíclicos dirigidos.                    | 41 |
| Figura 15. Tabla de hechos.  | 43 |
| Figura 16. Hechos básicos y derivados.                                 | 44 |
| Figura 17. Cubo dimensional.   | 47 |
| Figura 18. Aplicación de la operación roll-up.                         | 48 |
| Figura 19. Aplicación de la operación drill-across.                    | 49 |
| Figura 20. Aplicación de la operación pivoting.                        | 49 |
| Figura 21. Esquema de estrella.  | 52 |
| Figura 22. Esquema de copa de nieve.                                   | 54 |
| Figura 23. Esquema de constelación.                                    | 56 |
| Figura 24. Tareas de la metodología de Ralph Kimball.                  | 58 |
| Figura 25. Extracción de conocimiento a partir de datos.               | 64 |
| Figura 26. Dimensión sistémica del conocimiento.                       | 65 |
| Figura 27. Arquitectura de Data Warehouse.                             | 66 |
| Figura 28. Fase del proceso de KDD.                                    | 67 |
| Figura 29. Integración de un almacén de datos.                         | 68 |
| Figura 30. Discretización del atributo tamaño.                         | 68 |
| Figura 31. Ciclo de vida de minería de datos.                          | 70 |
| Figura 32. Tipo de conocimientos.                                      | 74 |
| Figura 33. Relación entre potencial de análisis y complejidad.         | 75 |
| Figura 34. Validación cruzada de k iteraciones.                        | 81 |

|  |     |
|--|-----|
| Figura 35. Matriz de confusión.  | 82  |
| Figura 36. Factores individuales del periodo 2012 hasta 2018.                                    | 95  |
| Figura 37. Factores individuales por periodo.  | 96  |
| Figura 38. Factores individuales por programa académico.   | 97  |
| Figura 39. Factores individuales por índice de masa corporal.                                    | 98  |
| Figura 40. Factores académicos del periodo 2012 hasta 2018.                                      | 99  |
| Figura 41. Factores académicos por periodo.  | 100 |
| Figura 42. Factores académicos por programa académico.   | 101 |
| Figura 43. Factores académicos por tipo de ingreso.  | 102 |
| Figura 44. Factores socioeconómicos del periodo 2012 hasta 2018.                                 | 103 |
| Figura 45. Factores socioeconómicos por periodo.   | 104 |
| Figura 46. Factores socioeconómicos por programa académico.                                      | 105 |
| Figura 47. Factores socioeconómicos por situación del padre.                                     | 106 |
| Figura 48. Predicción de la deserción estudiantil del periodo 2018 bajo los modelos predictivos. | 107 |
| Figura 49. Constratación de los desertores 2018 con el modelo predictivo de árboles de decisión. | 109 |
| Figura 50. Cronograma de actividades.  | 127 |
| Figura 51. Diagrama de la base de datos de registro académico.                                   | 133 |
| Figura 52. Diagrama de la base de datos de bienestar universitario.                              | 133 |
| Figura 53. Diagrama de la base de datos de grados y títulos.                                     | 134 |
| Figura 54. Diseño de la arquitectura.  | 148 |
| Figura 55. Instalación de Microsoft SQL Server 2017.   | 150 |
| Figura 56. Selección de las características de Microsoft SQL Server 2017.                        | 150 |
| Figura 57. Instalación de CA Erwin Data Modeler r7.3.  | 151 |
| Figura 58. Instalación de Microsoft SQL Server Management Studio 17.9.                           | 151 |
| Figura 59. Instalación de Microsoft SQL Server Data Tools 15.8.                                  | 152 |
| Figura 60. Diseño lógico de los ingresantes desertores.  | 166 |
| Figura 61. Diseño lógico de los estudiantes desertores.  | 167 |
| Figura 62. Diseño lógico de los egresados sin tramites de documentos académicos.                 | 167 |
| Figura 63. Diseño del subsistema de ETL.   | 169 |
| Figura 64. Proyecto multidimensional.  | 171 |
| Figura 65. Conexión con el origen de datos.  | 171 |
| Figura 66. Selección de tablas.  | 172 |
| Figura 67. Vista de datos de ingresante registro.  | 172 |

|   |     |
|---|-----|
| Figura 68. Vista de datos de ficha de matrícula.                                      | 173 |
| Figura 69. Vista de datos de egresado.  | 173 |
| Figura 70. Selección de método de creación.   | 174 |
| Figura 71. Selección de la tabla de hecho ingresante registro.                        | 174 |
| Figura 72. Selección de medidas de la tabla de hecho ingresante registro.             | 175 |
| Figura 73. Selección de dimensiones.  | 175 |
| Figura 74. Selección de atributos y jerarquías de la dimensión periodo.               | 176 |
| Figura 75. Selección de relaciones de atributos de la dimensión periodo.              | 176 |
| Figura 76. Selección de atributos y jerarquías de la dimensión centro de estudios.    | 176 |
| Figura 77. Selección de relaciones de atributos de la dimensión centro de estudios.   | 177 |
| Figura 78. Selección de atributos y jerarquías de la dimensión programa académicos.   | 177 |
| Figura 79. Selección de relaciones de atributos de la dimensión programa académicos.  | 177 |
| Figura 80. Selección de atributos y jerarquías de la dimensión estudiante.            | 177 |
| Figura 81. Selección de relaciones de atributos de la dimensión estudiante.           | 178 |
| Figura 82. Desertores de los ingresantes.   | 178 |
| Figura 83. Desertores potenciales de los ingresantes.                                 | 179 |
| Figura 84. Desertores de los matriculados.  | 179 |
| Figura 85. Desertores potenciales de los matriculados.                                | 180 |
| Figura 86. Egresados que no tramitaron sus documentos académicos.                     | 180 |
| Figura 87. Egresados que tramitaron sus documentos académicos.                        | 181 |
| Figura 88. Estudiantes desertores por año de ingreso.                                 | 191 |
| Figura 89. Estudiantes desertores por tipo de ingreso y año de ingreso.               | 191 |
| Figura 90. Estudiantes desertores por lugar de residencia y año de ingreso.           | 192 |
| Figura 91. Estudiantes desertores por nivel de categorización y año ingreso.          | 192 |
| Figura 92. Estudiantes desertores por centro de estudios y año de ingreso.            | 193 |
| Figura 93. Estudiantes desertores por uso de medios de comunicación y año de ingreso. | 193 |
| Figura 94. Estudiantes desertores por nivel educativo del padre y año de ingreso.     | 194 |
| Figura 95. Estudiantes desertores por lugar de nacimiento y año de ingreso.           | 194 |
| Figura 96. Estudiantes desertores por carrera profesional y año de ingreso.           | 195 |
| Figura 97. Cueva ROC de la técnica del árbol decisión.                                | 203 |



## ÍNDICE DE TABLAS

|   |     |
|---|-----|
| Tabla 1 Variable 1  | 91  |
| Tabla 2 Etapas del procesamiento de la información  | 93  |
| Tabla 3 Factores de deserción estudiantil   | 95  |
| Tabla 4 Valores predichos de los modelos predictivos del periodo 2018                         | 108 |
| Tabla 5 Constratación de los desertores 2018 con el modelo predictivo de árboles de decisión. | 110 |
| Tabla 6 Stakeholders  | 128 |
| Tabla 7 Recursos humanos del proyecto   | 128 |
| Tabla 8 Hardware requerido  | 129 |
| Tabla 9 Software requerido  | 129 |
| Tabla 10 Listado de requerimiento para los ingresantes  | 130 |
| Tabla 11 Listado de requerimiento para los matriculados                                       | 131 |
| Tabla 12 Listado de requerimiento para los egresados  | 131 |
| Tabla 13 Descripción de la tabla OPE_Entidad  | 134 |
| Tabla 14 Descripción de la tabla MAT_Matricula  | 135 |
| Tabla 15 Descripción de la tabla MAT_Ficha  | 136 |
| Tabla 16 Descripción de la tabla ACT_FichaDetalle   | 137 |
| Tabla 17 Descripción de la tabla ACT_Acta   | 138 |
| Tabla 18 Descripción de la tabla ACT_Aula   | 138 |
| Tabla 19 Descripción de la tabla ADS_Materia  | 139 |
| Tabla 20 Descripción de la tabla ADS_Plancurricular   | 140 |
| Tabla 21 Descripción de la tabla ADS_Especialidad   | 141 |
| Tabla 22 Descripción de la tabla MAT_Periodo  | 141 |
| Tabla 23 Descripción de la tabla Alumno   | 142 |
| Tabla 24 Descripción de la tabla FichaSocio   | 143 |
| Tabla 25 Descripción de la tabla Especialidad   | 144 |
| Tabla 26 Descripción de la tabla Especialidad   | 144 |
| Tabla 27 Descripción de la tabla CargaFamiliar  | 144 |
| Tabla 28 Descripción de la tabla Comunicacion   | 145 |
| Tabla 29 Descripción de la tabla RiesgoSalud_Familiar   | 145 |
| Tabla 30 Descripción de la tabla SituacionPadres  | 145 |
| Tabla 31 Descripción de la tabla Sostiene   | 145 |
| Tabla 32 Descripción de la tabla Familiar   | 146 |
| Tabla 33 Descripción de la tabla EgresadoCulminaron   | 146 |

|  |     |
|--|-----|
| Tabla 34 Columnas fundamentales de la arquitectura.                              | 149 |
| Tabla 35 Dimensiones encontradas en los procesos de negocios                     | 154 |
| Tabla 36 Descripción de la dimensión de estudiante                               | 155 |
| Tabla 37 Descripción de la dimensión de ubicación geográfica                     | 156 |
| Tabla 38 Descripción de la dimensión de centro de estudios                       | 156 |
| Tabla 39 Descripción de la dimensión de preparación de estudios                  | 157 |
| Tabla 40 Descripción de la dimensión de nivel educativo                          | 157 |
| Tabla 41 Descripción de la dimensión de ámbito de vivienda                       | 157 |
| Tabla 42 Descripción de la dimensión de tenencia de vivienda                     | 158 |
| Tabla 43 Descripción de la dimensión de material de construcción                 | 158 |
| Tabla 44 Descripción de la dimensión de servicio de vivienda                     | 158 |
| Tabla 45 Descripción de la dimensión de uso de medio de comunicación             | 159 |
| Tabla 46 Descripción de la dimensión de sostiene                                 | 159 |
| Tabla 47 Descripción de la dimensión de situación del padre                      | 159 |
| Tabla 48 Descripción de la dimensión de riesgo de salud familiar                 | 159 |
| Tabla 49 Descripción de la dimensión de carga familiar                           | 160 |
| Tabla 50 Descripción de la dimensión de grado educativo                          | 160 |
| Tabla 51 Descripción de la dimensión de situación de matrícula                   | 160 |
| Tabla 52 Descripción de la dimensión de tipo de ingreso                          | 161 |
| Tabla 53 Descripción de la dimensión de programa académicos                      | 161 |
| Tabla 54 Descripción de la dimensión de periodo                                  | 161 |
| Tabla 55 Descripción de la dimensión de fechas                                   | 162 |
| Tabla 56 Tablas de hechos encontradas en los procesos de negocios                | 162 |
| Tabla 57 Medidas de las tablas de hechos encontradas en los procesos de negocios | 162 |
| Tabla 58 Descripción de la tabla de hecho ingresante registro                    | 163 |
| Tabla 59 Descripción de la tabla de hecho ficha de matrícula                     | 164 |
| Tabla 60 Descripción de la tabla de hecho egresado                               | 165 |
| Tabla 61 Descripción de la tabla registro de ingresante                          | 187 |
| Tabla 62 Descripción de la tabla registro de serie temporal                      | 189 |
| Tabla 63 Selección de atributos y medidas de la tabla registro ingresante        | 196 |
| Tabla 64 Selección de atributos y medidas de la tabla registro serie temporal    | 197 |
| Tabla 65 Nuevos datos de la tabla registro de ingresante                         | 198 |
| Tabla 66 Nuevos datos de la tabla registro de serie temporal                     | 198 |
| Tabla 67 Diseño de evaluación de las técnicas de DM - Ingresante                 | 199 |
| Tabla 68 Diseño de evaluación de las técnicas de DM – serie temporal             | 200 |

|  |     |
|--|-----|
| Tabla 69 Matriz de confusión de árbol de decisión                    | 202 |
| Tabla 70 Matriz de confusión de la regresión logística               | 203 |
| Tabla 71 Matriz de confusión de las redes bayesianas                 | 205 |
| Tabla 72 Resultados de evaluación de las técnicas de DM - ingresante | 207 |
| Tabla 73 Validación cruzada de los errores de las técnicas de DM     | 208 |
| Tabla 74 Ingresantes desertores del 2018 por programas académicos    | 209 |

## RESUMEN

El presente trabajo de investigación tiene como objetivo efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y minería de datos en la Universidad Nacional Jorge Basadre Grohmann 2012-2018. Para ello, se ha construido un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann bajo la metodología de Ralph Kimball donde nos revela cuales son los indicadores más predominantes de la deserción estudiantil basadrina. Para efectos de la aplicación de técnicas de minería de datos, se eligieron tres de ellas, de las cuales la técnica elegida fue árbol de decisión por su precisión y entendimiento de sus resultados en las predicciones de datos. Como conclusión general, fue posible efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y minería de datos bajo metodologías permitiendo entender la información histórica e interpretación de sus resultados predichos.

**Palabras clave:** Análisis predictivo, deserción estudiantil, Data Warehouse y minería de datos.

## **ABSTRACT**

The objective of this research is to perform a predictive analysis of student desertion using Data Warehouse and data mining at the National University Jorge Basadre Grohmann 2012-2018. To do this, a Data Warehouse has been built for the analysis of student desertion data from the National University Jorge Basadre Grohmann under the Ralph Kimball methodology, where it reveals which are the most predominant indicators of basadrine desertion. For the purposes of the application of data mining techniques, three of them were chosen, of which the chosen technique was a decision tree for its accuracy and understanding of its results in the predictions of data. As a general conclusion, it was possible to carry out a predictive analysis of student desertion using Data Warehouse and data mining under methodologies allowing to understand the historical information and interpretation of its predicted results.

**Keywords:** Predictive analysis, student desertion, Data Warehouse and data mining.

## **INTRODUCCIÓN**

La presente investigación surge del interés de realizar un análisis predictivo de la deserción estudiantil utilizando data warehouse y minería de datos en la Universidad Nacional Jorge Basadre Grohmann - Tacna, 2012-2018”, ante las evidencias que se han dado en la educación peruana en los últimos años con respecto al crecimiento de la deserción por parte de los estudiantes de educación superior universitaria.

Cada día, las instituciones de educación superior universitaria, generan gran cantidad de datos individuales, académicos, socioeconómicos de los estudiantes; por tanto desde un punto de vista de las tecnologías de la información, se señala que las universidades necesitan sistematizar su información por lo que se ha generado diversas técnicas, herramientas y metodologías, tanto para la depuración y la selección de la información, como para el diseño y la implantación de Data Warehouse y el análisis de datos, mediante la minería de datos orientada, en su conjunto, a proveer información útil para la toma de decisiones en la universidad y sobre todo para tomar medidas preventivas.

A lo largo de los capítulos desarrollados en la presente tesis, encontraremos en el Capítulo I, el Planteamiento de la Investigación donde se señala la descripción del problema, la formulación del problema, la justificación, alcances y limitaciones y objetivos; en el Capítulo II, se tiene el marco teórico, donde se expone el fundamento que ha servido de soporte para la investigación; en el Capítulo III, se describe el marco metodológico para llevar a cabo la investigación, el cual contiene las hipótesis, las variables y el diseño de la investigación; en el Capítulo IV, se muestran los resultados de la investigación; posteriormente, se exponen las conclusiones ya que se arriban las recomendaciones sugeridas, para finalmente presentar la bibliografía.

También se han incluido anexos que ayudan a comprender el contenido de los diferentes capítulos de la presente tesis.

Se desea que el presente trabajo cumpla no sólo una función académica al considerar que se ha tenido una formación de gestor; se estima que ayude a participar activa y directamente en situaciones similares.

## **CAPÍTULO I**

### **PLANTEAMIENTO DE INVESTIGACIÓN**

#### **1.1. Descripción del problema**

Todos los días, las universidades generan una gran cantidad de información por sus distintas áreas, por lo que resulta imprescindible la utilización de las tecnologías de la información y la comunicación para que la información pueda ser almacenada, transformada, analizada y visualizada, como los investigadores mencionan:

“Se ha estimado que la cantidad de información en el mundo se duplica cada 20 meses. El tamaño y el número de bases de datos probablemente aumenta aún más rápido” (Frawley, Piatetsky-Shapiro, & Matheus, 1992, 57).

“La deserción estudiantil en cualquier etapa académica es un problema de índole mundial, motivo por el cual muchos países han dedicado tiempo y dinero para el estudio de tal fenómeno” (Rodríguez, 2012, 5).

“La deserción universitaria, que por décadas fue considerada como un fenómeno normal e incluso, como una muestra del nivel de exigencia de

la carrera y de la universidad, hoy se ve como un signo de ineficiencia y como un gran costo para el país, los estudiantes y las instituciones de educación superior” (Mateus, Perilla, & Herrera, 2010, 122).

“La deserción estudiantil universitaria es un fenómeno presente en todas las modalidades de estudio” (Universidad Abierta para Adultos, 2009, 199).

De acuerdo a lo citado por los respectivos autores la deserción universitaria se convierte en un problema para las universidades, que hay que entender para poder combatirlo.

En el presente trabajo de investigación, se ha encontrado estudios efectuados tanto en instituciones privadas como públicas, cuyos objetivos se relacionan con nuestra investigación, por lo cual, sirven como referente al presente estudio, dentro de estos trabajos se tienen:

En Colombia en el año 2010 la Universidad Católica de Colombia desde la Decanatura Académica realizó un estudio, titulado “El programa institucional de tutoría: de un programa de acompañamiento a un programa de retención estudiantil” (Robechi & Jácome, 2014, 22), el cual analiza los factores de deserción estudiantil.

Según Ministerio de Educación Nacional de Colombia (2015), el problema de la “deserción universitaria alcanza el 47,04%” a través de un estudio realizado por el Sistema de Información para la Prevención de la Deserción en Educación Superior (SPADIES), lo que significa que uno de cada dos estudiantes que ingresan a la educación superior no finaliza sus estudios.

Según Ministerio de Educación Nacional de Chile (2015), las cifras de “deserción en Chile son considerables cerca de un 30% de los estudiantes no continúa luego del primer año de estudios, lo que aumenta a un 43% al segundo año a través de un estudio realizado por el Servicio de Información de Educación Superior (SIES) en el año 2012”.

En el mundo hay países cuyas “tasas de deserción bordean entre el 10% o más. Por ejemplo, en el caso de Estados Unidos de Norte América la tasa de deserción en educación superior en el siglo XX ha sido alrededor del 50%, mientras que en los países de América Latina es del 57%, y en el caso de Perú de acuerdo a un estudio del Proyecto ALFA de la Comunidad Europea es del 57% en el año 2009” (Programa Nacional de Beca y Crédito Educativo, 2014,17).

“Hasta un 30% de alumnos se retira en el primer semestre de estudios en las universidades e institutos en el Perú, de acuerdo con un

estudio de la consultora en marketing Flanqueo” (Grupo Educación al Futuro, 2013, párr.1).

El estudio señala que las universidades orientadas al “segmento socioeconómico C registran la mayor deserción con 28%, seguidas de las universidades del segmento B con 21% de deserción y las universidades del segmento A con 15% de deserción” (Grupo Educación al Futuro, 2013, párr.2).

La principal razón para la deserción sería el “factor económico, que representa el 29%; seguida por la falta de vocación 26%, decepción de la carrera elegida 20%, mala enseñanza 15%, mala organización y la falta de preocupación por el alumno 10%” (Grupo Educación al Futuro, 2013, párr.4).

“Los estudiantes de la Universidad Católica Los Ángeles de Chimbote (ULADECH), modalidad a distancia, que han desertado son mayoritariamente varones jóvenes o adultos jóvenes, lo que hace suponer que, al hallarse en plena edad laboral, al decidirse por realizar estudios universitarios, tienen la perspectiva de mejorar sus niveles remunerativos o enfrentar el recorrido por el mercado laboral en mejores condiciones” (Universidad Abierta para Adultos, 2009, 219).

La deserción tiene una influencia marcada, en la Universidad Nacional Jorge Basadre Grohmann (UNJBG) de Tacna, donde según los registros de matriculados en el periodo 2012-2018 se observó una matrícula relativamente alta y que luego se ve reducido en cada semestre académico.

Por lo anteriormente descrito se hace necesario contar con un sistema de información que permita detectar, estimar, predecir el riesgo deserción estudiantil de la UNJBG, y reincorporar a los estudiantes a la vida académica

UNJBG de Tacna, se encuentra comprometida con el proceso de acreditación tal como lo establece en la Misión y Visión de su Plan Estratégico Institucional 2012- 2021.

## **1.2. Formulación del problema**

### **1.2.1. Formulación del problema general**

¿Es posible efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann - Tacna?

### **1.2.2. Formulación de los problemas específicos**

- a) ¿Es posible construir un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann?
- b) ¿Cuáles son los indicadores que predominan la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann?
- c) ¿Cuáles son las técnicas de minería de datos idóneas para el análisis predictivo de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann?

### **1.3. Justificación**

En la UNJBG tiene una influencia marcada en la deserción de sus estudiantes, donde según los registros de matriculados en el periodo 2012-2018 se observó una matrícula relativamente alta y que luego se ve reducido en cada semestre académico.

La investigación que se propone pretende dar una solución informática y computacional al análisis de datos vinculada a lo académico, socioeconómico y personal para la detección, estimación y predicción de riesgo deserción estudiantil mediante la implementación de un Data Warehouse y minería de datos, para el mejoramiento en el almacenamiento

y el análisis de grandes volúmenes de datos provenientes de las fuentes de datos operacionales de la UNJBG.

La UNJBG, no cuenta con un programa de retención estudiantil que permita impactar positivamente, en sus dimensiones académica, personal y social. En la que este programa permita vivir una experiencia de crecimiento personal, académica y social, detectar sus capacidades y, en consecuencia, se estimula su seguridad para desenvolverse en el contexto universitario, así mismo combatir este fenómeno denominado deserción estudiantil.

El principal motivo para realizar la presente investigación se relaciona con la experiencia del autor como analista de datos en instituciones relacionadas a la parte académica, socio-económico, en donde pudo detectar algunas deficiencias al momento de analizar cantidades considerables de datos, más de 500 000,00 registros. Las principales deficiencias se relacionaban con la ausencia de datos, la presencia de datos atípicos.

Esta investigación posee diversos motivos que la justifican, entre los cuales, que es un problema de la educación superior que afecta al desarrollo de la sociedad en gran manera. Es importante porque sus resultados de este estudio servirán como una alerta sobre la situación

académica de los estudiantes y les proporcionarán pautas e informaciones claves para que las autoridades de las universidades, docentes y padres de familias, puedan orientar a los jóvenes a tomar buenas, oportunas e inteligentes decisiones y por lo consiguiente a optar políticas y estrategias bien definidas que apunten a superar la situación actual y disminuir los índices de deserción académica detectados.

#### **1.4. Alcances y limitaciones**

##### **1.4.1. Alcances**

Como caso de estudio se tomará la base de datos de Registro Académico (REAC), Bienestar Universitario (BUN) y Grados y Títulos (GyT) de la Universidad Nacional Jorge Basadre Grohmann (UNJBG) - TACNA 2012-2018.

Este proyecto se realizará en la Universidad Nacional Jorge Basadre Grohmann y beneficiará a la comunidad basadrina.

##### **1.4.2. Limitaciones**

Este proyecto se limita hacer uso por los usuarios afiliados al SYSA, BUN, GyT de la UNJBG.

## **1.5. Objetivos**

### **1.5.1. Objetivo general**

Efectuar el análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna.

### **1.5.2. Objetivos específicos**

- a) Construir un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.
- b) Determinar los indicadores predominantes de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.
- c) Determinar las técnicas de minería de datos idóneas para el análisis predictivo de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.

## **CAPÍTULO II**

### **MARCO TEÓRICO**

#### **2.1. Bases teóricas respecto al problema**

##### **2.1.1. Antecedentes de la investigación**

En primer lugar, se tiene que, en el mes de mayo del 2012 fue presentado en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile en Santiago - Chile, con el título “MODELO PARA LA AUTOMATIZACIÓN DEL PROCESO DE DETERMINACIÓN DE RIESGO DE DESERCIÓN EN ALUMNOS UNIVERSITARIOS” como trabajo de Grado Académico por Erwin Sergio Fischer Angulo, para optar el grado de Magister en Tecnologías de la Información.

Esta investigación plantea “desarrollar un indicador que permita clasificar en forma automática a los alumnos con mayor riesgo de deserción de las carreras de Ingeniería de la Universidad de Las Américas, para trabajar un programa docente y de tutorías, focalizado con el objeto de mejorar los índices de retención” (Fischer, 2012, 3).

En segundo lugar, se tiene que, el 26 de septiembre del 2016 fue presentado en la Facultad de Ingeniería de la Universidad Nacional de Colombia en Bogota - Colombia, con el título “ANÁLISIS DE LA

DESERCIÓN Y PERMANENCIA ACADÉMICA EN LA EDUCACIÓN SUPERIOR APLICANDO MINERÍA DE DATOS” como trabajo de Grado Académico por Alvaro Agustin Otañe Bowen, para optar el grado de Magister en Ingeniería de Sistemas y Computación.

Esta investigación plantea “aplicar modelos predictos a los datos de los estudiantes de la universidad Popular del Cesar para identificar características similares entre los los grupos como los socioeconomicos” (Otañe Bowen, 2016, 45).

En tercer lugar, se tiene que, el 1 de septiembre del 2015 fue presentado en la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de la Paz en La Paz - México, con el título “MODELO DE MINERÍA DE DATOS PARA IDENTIFICACIÓN DE PATRONES QUE INFLUYEN EN EL APROVECHAMIENTO ACADEMICO” como trabajo de Grado Académico por Jaime Ángel Hernández Cedano, para optar el grado de Magister en Sistemas Computacionales.

Esta investigación plantea “diseñar y generar un modelo de minería de datos para la identificación de patrones de comportamiento relacionados con el desempeño académico de alumnos en una institución de educación media superior” (Hernández J., 2015, 5).

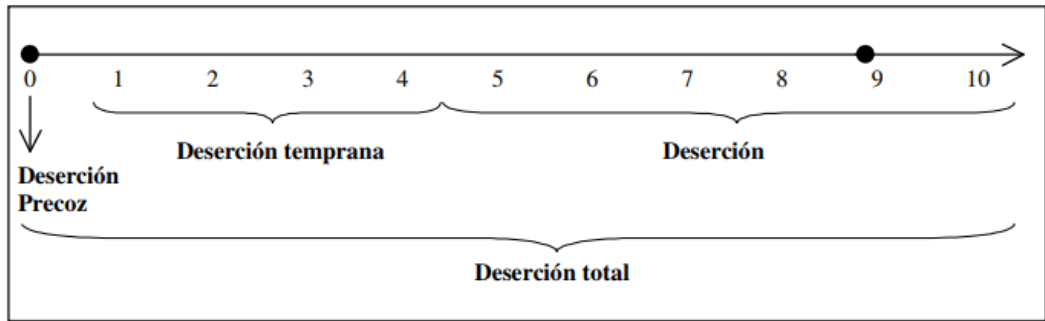
Estas investigaciones detalladas con anterioridad, proponen un modelo de minería de datos para la detección de deserción estudiantil, por esta razón se pretende, a través de un modelo de Data Warehouse y minería de datos, estimar, predecir el riesgo de deserción estudiantil de la UNJBG.

## **2.2. Deserción: Definición**

Himmel (2002), define la deserción como el “(...) abandono prematuro de un programa de estudios antes de alcanzar el título o grado, y considera un tiempo suficientemente largo como para descartar la posibilidad de que el estudiante se reincorpore”. (p. 94-95).

### **2.2.1. Tipos de deserción**

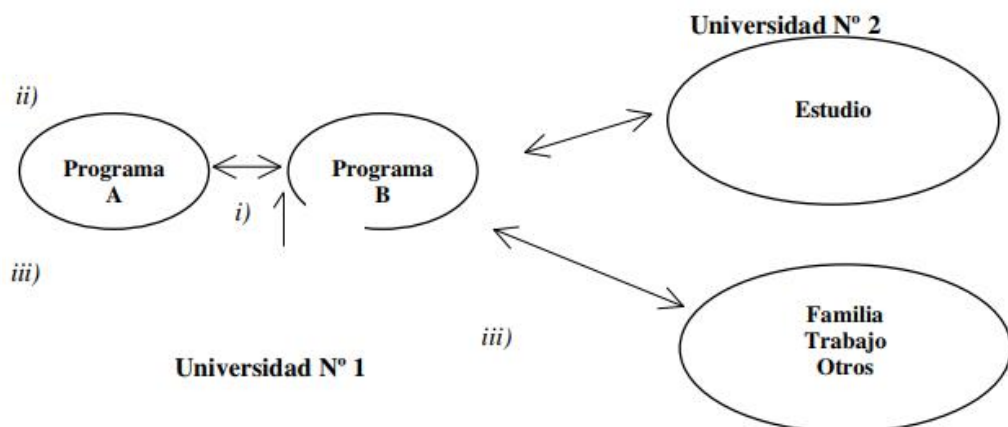
De acuerdo con los autores Castaño, Gallón, Gómez & Vásquez (2004) mencionan que existe dos tipos de abandonos en estudiantes universitarios: con respecto al tiempo y al espacio. “La deserción con respecto al tiempo se clasifica a su vez en: i) deserción precoz: individuo que habiendo sido aceptado por la universidad no se matricula, ii) deserción temprana: individuo que abandona sus estudios en los cuatro primeros semestres de la carrera, iii) deserción tardía: individuo que abandona los estudios en los últimos seis semestres” (p.46). Como se muestra en la figura 1.



**Figura 1.** Deserción con respecto al tiempo.

Fuente: Castaño, Gallón, Gómez & Vásquez (2004), “Deserción estudiantil universitaria: una aplicación de modelos de duración”. (p. 46).

Con respecto al espacio se divide en: i) deserción interna o del programa académico: se refiere al alumno que decide cambiar su programa académico por otro que ofrece la misma institución universitaria, ii) deserción institucional: caso en el cual el estudiante abandona la universidad, y iii) la deserción del sistema educativo. Como se muestra en la figura 2.

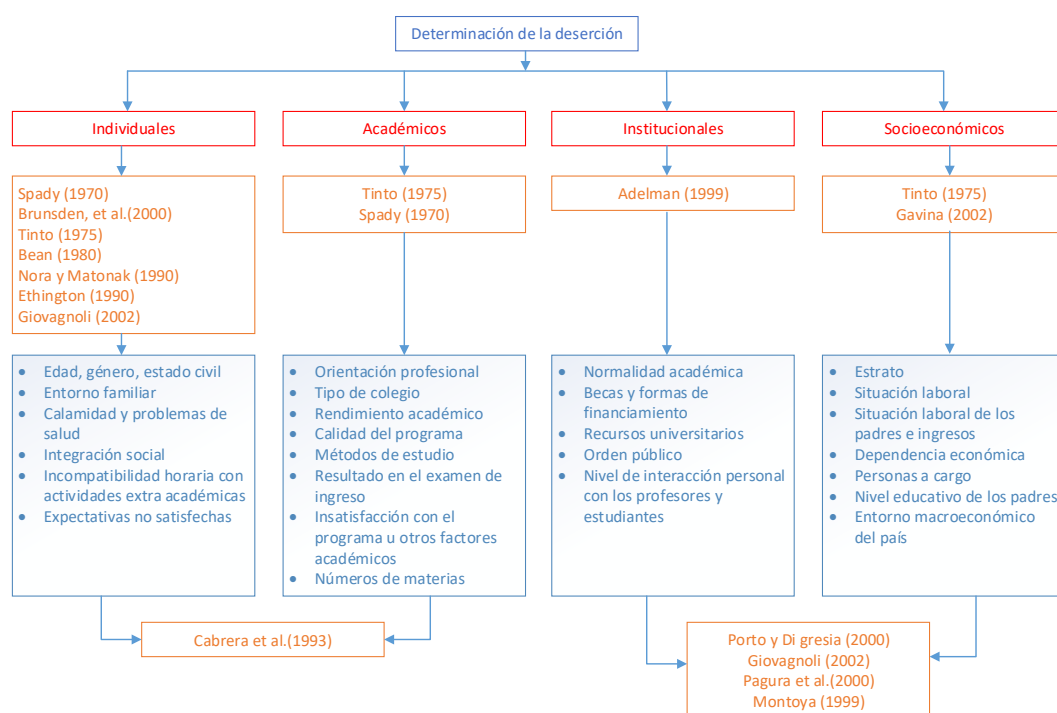


**Figura 2.** Deserción con respecto al espacio.

Fuente: Castaño, Gallón, Gómez & Vásquez (2004), “Deserción estudiantil universitaria: una aplicación de modelos de duración”. (p. 46).

## 2.2.2. Factores de la deserción

De acuerdo con los autores Castaño, Gallón, Gómez & Vásquez (2004) después de haber realizado una revisión literaria se enfatizan en un conjunto de factores “institucionales, socioeconómicas, académicas y individuales” (p.44). Como se muestra en la figura 3.



**Figura 3.** Factores de deserción estudiantil.

Fuente: Castaño, Gallón, Gómez & Vásquez (2004), “Deserción estudiantil universitaria: una aplicación de modelos de duración”. (p. 46).

## 2.3. Base de datos y modelo de datos

### 2.3.1. Historia de los sistemas de base de datos

Silberschatz et al. (2002), describe una breve reseña historia de los sistemas de bases de datos, la cual se resume en la Figura 4.

| Periodo   | Descripción de la historia de los sistema de Base de Datos   |
|---|--|
| Principios del Siglo XX.                          | <ul style="list-style-type: none"> <li>✓ Hollerith inventa las tarjetas perforadas, las cuales fueron usadas para el registro de datos del censo de los Estados Unidos de América. Las tarjetas perforadas posteriormente se usaron ampliamente como medio para introducir datos en los computadores.</li> </ul>   |
| Década de 1950 y principios de la década de 1960. | <ul style="list-style-type: none"> <li>✓ Se desarrollaron las cintas magnéticas para el almacenamiento de datos.</li> <li>✓ Las tareas de procesamiento de datos fueron automatizadas con los datos almacenado en cintas.</li> <li>✓ Las cintas y los paquetes de tarjetas perforadas se leían en forma secuencial, la cantidad de datos superaba la capacidad de la memoria principal por lo que los programas de procesamiento procesaban los datos siguiendo un orden determinado.</li> </ul>                                   |
| Finales de la década de 1960 y la década de 1970. | <ul style="list-style-type: none"> <li>✓ A finales de la década de 1960 se intensificó el uso de discos fijos que permitieron un acceso directo a los datos.</li> <li>✓ Los discos permitieron el desarrollo de las bases de datos de red y jerárquica, dando lugar a las estructuras de datos como listas y árboles pudieran ser almacenadas.</li> <li>✓ En 1970 Codd definió el modelo relacional y formas no procedimentales de consultar los datos, lo que representó el origen de las bases de datos relacionales.</li> </ul> |
| Década de 1980.                                   | <ul style="list-style-type: none"> <li>✓ System R un proyecto innovador condujo al primer producto de bases de datos relacionales IBM: SQL/DL capaz de competir en rendimiento con las bases de datos jerárquicas y de red.</li> <li>✓ Aparecen las primeras bases de datos relacionales como DB2 de IBM, Oracle, Ingres y Rdb de DEC.</li> <li>✓ Se realizan investigaciones en las bases de datos paralelas y distribuidas, así como el inicio de los trabajos sobre las bases de datos orientadas a objetos.</li> </ul>         |
| Principios de la década de 1990.                  | <ul style="list-style-type: none"> <li>✓ Se comienza a utilizar el lenguaje SQL para las aplicaciones de ayuda a la toma de decisiones.</li> <li>✓ Se incrementó el número de herramientas para analizar grandes cantidades de datos.</li> <li>✓ Los vendedores de bases de datos ofrecían productos de bases de datos paralelas y relacionales orientadas a objetos.</li> </ul>   |
| Finales de la década de 1990.                     | <ul style="list-style-type: none"> <li>✓ Se da un crecimiento del World Wide Web y con esto las bases de datos evolucionaron desarrollando interfaces web a los datos.</li> </ul>  |

**Figura 4.** Historia de los sistemas de base de datos.

Fuente: Silberschatz A. (2002), "Fundamentos de la base de datos". (p. 13).

### 2.3.2. Base de datos

Elmasri & Navathe lo define como: "Una base de datos es una colección de datos relacionados. Con la palabra datos nos referimos a los

hechos (datos) conocidos que se pueden grabar y que tienen un significado implícito” (Elmasri & Navathe, 2007, 4).

Una base de datos (BD) tiene las siguientes propiedades implícitas:

- Una base de datos representa algún aspecto del mundo real, lo que en ocasiones se denomina mini mundo o universo de discurso (UoD, Universe of discourse). Los cambios introducidos en el mini mundo se reflejan en la base de datos.
- Una base de datos es una colección de datos lógicamente coherente con algún tipo de significado inherente. No es correcto denominar base de datos a un surtido aleatorio de datos.
- Una base de datos se diseña, construye y rellena con datos para un propósito específico. Dispone de un grupo pretendido de usuarios y algunas aplicaciones preconcebidas en las que esos usuarios están interesados.

Date considera a la “propia base de datos como una especie de armario electrónico para archivar; es decir, es un depósito o contenedor de una colección de archivos de datos computarizados” (Date, 2001, 2).

Reinosa, Maldonado, & Muñoz (2012), lo define a las bases de datos como: “Un conjunto de datos estructurados y definidos a través de un

proceso específico, que busca evitar la redundancia, y que se almacena en algún medio de almacenamiento masivo, como un disco” (p. 3).

Castaño afirma que las bases de datos es una "colección o depósito de datos integrados, con redundancia controlada y con una estructura que refleja las interrelaciones y restricciones existentes en el mundo real" (Castaño, 2001, 46). Los datos, que deben ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstos; y su definición y descripción, únicas para cada tipo de datos, deben almacenarse junto con los mismos. En ese sentido, los procedimientos de actualización y recuperación, comunes y bien determinados, deben ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de los datos.

Ricardo comenta que “en la actualidad las bases de datos se usan tan ampliamente que se pueden encontrar en organizaciones de todos los tamaños, desde grandes corporaciones y agencias gubernamentales, hasta pequeños negocios e incluso en hogares. Las actividades diarias con frecuencia lo ponen en contacto con las bases de datos, ya sea directa o indirectamente” (Ricardo, 2004, 2).

### **2.3.3. Definición de sistema**

Según Sommerville (2005) señaló que “Un sistema es la colección de componentes interrelacionados que trabajan conjuntamente para cumplir algún objetivo” (p. 20).

También Horacio (2002) define que: “Un sistema es un conjunto de elementos interrelacionados de modo tal que producen como resultado algo superior y distinto a la simple agregación de los elementos” (p. 26).

Según Horacio (2002) describe que: “Todo sistema se puede definir por sus entradas, sus procesos y sus salidas” (p. 27).

### **2.3.4. Sistema de base de datos**

Según Date lo describe como: “Un sistema de base de datos es básicamente un sistema computarizado para guardar registros; es decir, es un sistema computarizado cuya finalidad general es almacenar información y permitir a los usuarios recuperar y actualizar esa información con base en peticiones” (Date, 2001, 5).

### **2.3.5. Concepto de información**

De esta manera Horacio (2002) señaló que “La información es un dato o un conjunto de datos evaluados por un individuo concreto que trabaja, en un momento dado, sobre un problema específico, para alcanzar un objetivo determinado” (p. 24).

### **2.3.6. Modelo**

Según Sesento (2008) lo define como: “La representación de la realidad, explicación de un fenómeno, ideal digno de imitarse, paradigma, canon, patrón o guía de acción; idealización de la realidad; arquetipo, prototipo, uno entre una serie de objetos similares, un conjunto de elementos esenciales o los supuestos teóricos de un sistema social” (p. 12).

### **2.3.7. Modelo de datos**

Según Silberschatz (2002), lo describe como: “una colección de herramientas conceptuales para describir los datos, las relaciones entre los datos, la semántica de los datos y las restricciones de los datos” (p. 14).

### **2.3.8. Modelo relacional**

En 1970, Edgar F. Codd publicó un artículo revolucionario, argumentando que los datos deberían relacionarse mediante interrelaciones naturales, lógicas, inherentes a los datos. Es decir, las personas debían ser capaces de combinar los datos de fuentes diferentes, si la información lógica necesaria para hacer la combinada estaba presente en los datos.

“Edgar F. Codd propuso un modelo simple de datos en el que todos ellos se representarían en tablas constituidas por filas y columnas, A estas tablas se les se les dio el nombre matemático de relaciones, y por eso el modelo se denominó modelo relacional. Codd también propuso dos

lenguajes para manipular los datos en las tablas; el álgebra relacional y el cálculo relacional. Ambos lenguajes soportan la manipulación de los datos sobre la base de operadores lógicos” (Hansen & Hansen, 1997, 19), según el autor es un evento trascendental en los años 1970 para la evolución de las bases de datos.

Edgar F. Codd introdujo otra innovación revolucionaria, con respecto a la manipulación de datos sobre una base conceptual en vez de una base física. En los sistemas de base de datos relacionales, los archivos completos de datos se pueden procesar con instrucciones sencillas.

Según Date (2001) describe que los modelos relacionales se basan en “3 aspectos fundamentales estructural, integridad, manipulación” (p.58).

- Aspecto estructural, el usuario percibe la información de la base de datos como tablas y nada más que tablas;
- Aspecto de integridad, estas tablas satisfacen ciertas restricciones de integridad.
- Aspecto de manipulación, los operadores disponibles para que el usuario manipule estas tablas.

“Los sistemas de bases de datos relacionales fueron desarrollados inicialmente por su familiaridad y simplicidad. A pesar de la familiaridad y de la simplicidad de las bases de datos relacionales, existe también una

fuerte base matemática. Las matemáticas de las bases de datos relacionales incluyen la conceptualización de las tablas como conjuntos” (Mannino, 2007, 46).

Este proyecto de sistema de gestión de base de datos relacional, llamado System R patrocinado por IBM, se diseñó para demostrar la practicidad del modelo relacional al proporcionar una implementación de sus estructuras de datos y operaciones.

Ricardo lo describe como: “El modelo relacional que se basa en la noción matemática de relación. Codd y otros extendieron la noción para aplicarla al diseño de bases de datos. Por ende, fueron capaces de sacar ventaja del poder de la abstracción matemática y de la expresividad de la notación matemática para desarrollar una estructura simple, pero poderosa, para las bases de datos” (Ricardo, 2004, 124).

La estructura básica del modelo relacional es simple, lo que lo hace sencillo de entender desde el punto de vista intuitivo. Permite la separación de los niveles conceptual y físico, de modo que el diseño conceptual se puede realizar sin considerar estructuras de almacenamiento.

Según Silberschatz (2002) lo describe como “Una base de datos relacional consiste en un conjunto de tablas, a cada una de las cuales se le

asigna un nombre exclusivo. Cada fila de la tabla representa una relación entre un conjunto de valores. Dado que cada tabla es un conjunto de dichas relaciones, hay una fuerte correspondencia entre el concepto de tabla y el concepto matemático de relación, del que toma su nombre el modelo de datos relacional” (p. 53).

### **2.3.9. Modelo entidad-relación**

El modelo Entidad-Relación (ER) fue propuesto por Peter Chen en 1976 para facilitar el diseño de bases de datos al permitir al diseñador expresar las propiedades conceptuales de la base de datos en un esquema de empresa, organización o unidad de negocio.

Según Pons lo define como: “Es un mecanismo formal para representar y manipular información de manera general y sistemática” (Pons, Carrillo, & Marin, 2005, 100).

Según Elmasri & Navathe (2007) describe al Modelo ER como: “Un modelo de datos conceptual de alto nivel. Además, este modelo y sus variaciones se utilizan con frecuencia para el diseño conceptual de las aplicaciones de base de datos, y muchas herramientas de diseño emplean estos conceptos” (p.51).

Según Silberschatz (2002) se enfatiza que “El modelo de datos ER está basado en una percepción del mundo real consistente en objetos

básicos llamados entidades y de relaciones entre estos objetos. Se desarrolló para facilitar el diseño de bases de datos permitiendo la especificación de un esquema de la empresa que representa la estructura lógica completa de una base de datos” (p. 19).

El modelo ER es el modelo de datos más extendido para el diseño conceptual de una base de datos porque, siendo un modelo relativamente simple y fácil de emplear, posee una gran capacidad expresiva.

Los elementos del modelo Entidad-Relación son las siguientes categorías: entidades, atributos y relaciones.

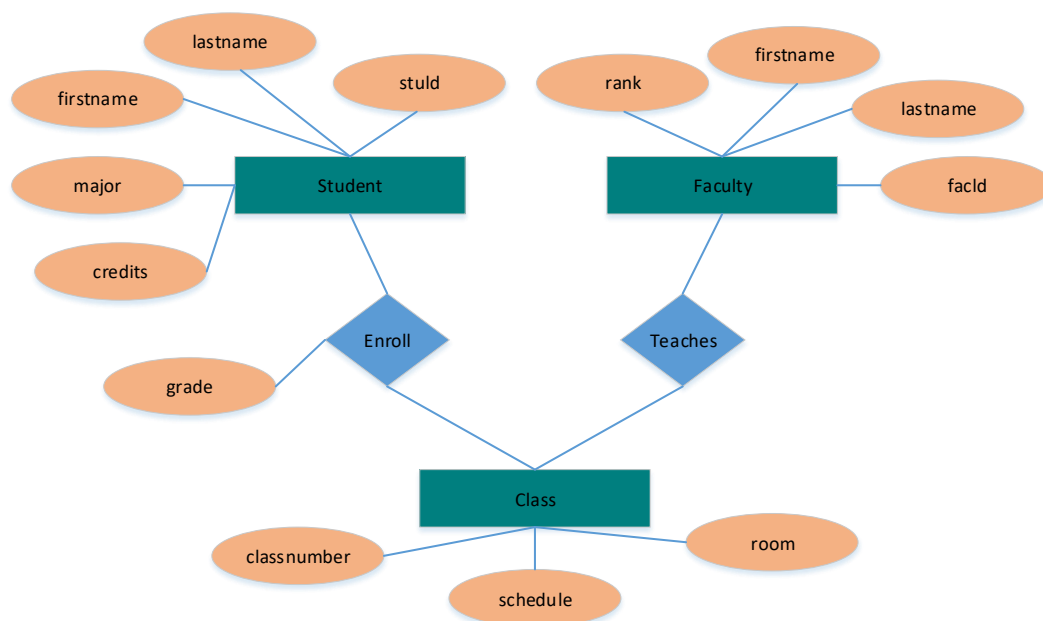
Pons define a una entidad “como un objeto que existe y que es distinguible de los demás” (Pons, Carrillo, & Marin, 2005, 101).

Pons define a los atributos como: “las propiedades que caracterizan un conjunto de entidades” (Pons, Carrillo, & Marin, 2005, 101).

Pons describe a una relación como: “una conexión a un conjunto de entidades consigo mismo” (Pons, Carrillo, & Marin, 2005, 104).

Ricardo (2004), describe al modelo ER como: “Un entorno de mini mundo en términos de entidades, atributos y relaciones”.

En la figura 5 muestra los símbolos básicos para diagramas ER. Un rectángulo se usa para representar una entidad, un óvalo para representar un atributo y un diamante para representar una relación. Estos elementos se conectan mediante líneas.



**Figura 5.** Diagrama de entidad - relación.  
Fuente: Ricardo, C. M. (2004), "Base de datos". (p. 89).

## 2.4. Data Warehouse

### 2.4.1. Definición de Data Warehouse

William Harvey Inmon, considerado como el padre del Data Warehouse quien introdujo este término lo define como: "A data warehouse is a subject-oriented, integrated, nonvolatile, and time variant collection of data in support of management's decisions. The data warehouse contains granular corporate data" (Inmon, 2005, 29).

Ricardo (2004), Date (2001), Krzysztof, Witold, Roman, & Lukasz (2007) y Elmasri & Navathe (2007), citan la definición de William Harver Inmmon del Data Warehouse (DW) como: “una colección de datos orientados al sujeto, no volátiles, integrados y variantes en el tiempo, usados principalmente en la toma de decisiones organizacionales” (p. 544, 852).

Reinosa, Maldonado, & Muñoz (2012), definen que, “El almacén de datos o Data Warehouse es una base de datos corporativa cuya característica principal es la integración y el filtrado de información de una o varias fuentes, que luego procesará para su análisis desde diferentes puntos de vista y con una gran velocidad de respuesta” (p. 194).

El almacén de datos es una colección de datos históricos e integrados diseñada para soportar el procesamiento informático. Para las organizaciones, el almacén de datos es una excelente herramienta para las tomas de decisiones estratégicas que no utilizan para la operatoria diaria.

Según Mannino (2007), afirma que William Harver Inmmon en la descripción de DW “se refiere a un depósito de datos central donde se integran, depuran y estandarizan los datos de bases de datos operacionales y otras fuentes para apoyar la toma de decisiones” (p. 554).

Ruiz, (2007), define al DW como: “Un repositorio de datos centralizado para apoyo de las actividades de análisis del negocio. Permite almacenar datos operacionales y eliminar inconsistencias entre los diferentes formatos existentes en los sistemas fuente. Además de integrar los datos de toda la empresa, no importando su ubicación, formato o requerimientos de comunicación; permite incorporar información adicional integrada por el experto del negocio” (p. 3).

Date (2001), aclara que las data warehouses surgieron por dos razones: “primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones; segundo, la necesidad de hacerlo sin afectar a los sistemas operacionales” (p. 709).

#### **2.4.2. Características de Data Warehouse**

De acuerdo con Inmon (2005), Mannino (2007), Ruiz (2007), Krzysztof, Witold, Roman, & Lukasz (2007), Bernabeu (2010) y Pérez (2015), concuerdan que existen cuatro características principales que describen un DW:

- Orientado a temas. Los datos en la base de datos están organizados de manera que todos los elementos de datos

relativos al mismo evento u objeto del mundo real queden unidos entre sí.

- Integrado. La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- No volátil. La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- Variante en el tiempo. Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

### **2.4.3. Arquitectura de Data Warehouse**

Jiawei, Micheline & Jian (2012), Elmasri & Navathe (2007) y Date (2001) concuerdan en que, desde el punto de vista de la arquitectura, existen tres modelos posibles para diseñar un Data Warehouse (DW): “el Data Warehouse Empresarial, los Data Marts y el Data Warehouse Virtual” (p. 132).

- El Data Warehouse Empresarial consiste en una colección de toda la información de la organización.
- Los Data Marts consisten subconjuntos de información agregada a partir de los requerimientos específicos de un grupo o

departamento de la organización; este puede ser categorizado como dependiente o independiente en función de su fuente de datos.

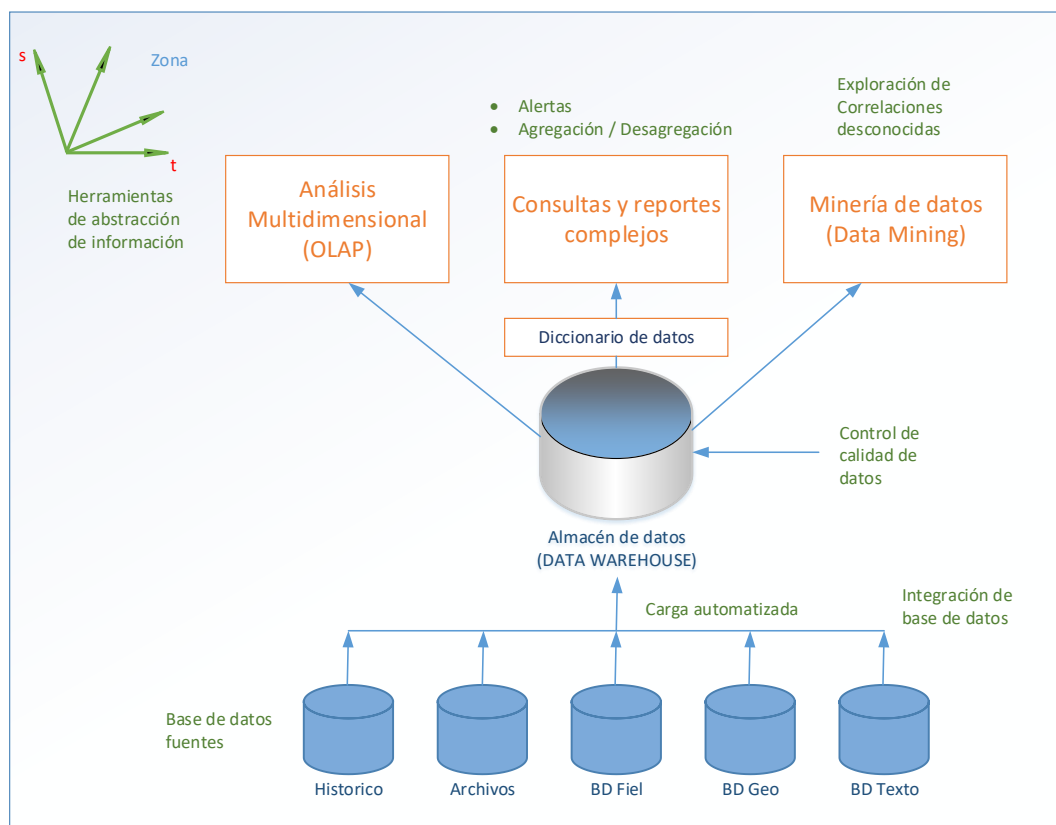
- El Data Warehouse Virtual consiste en un conjunto de vistas generadas a partir de la base de datos operacional para hacer más eficiente los procesos de consulta.

Según Reinoso, Maldonado, & Muñoz, (2012), afirma que “Un DW consta de tres niveles: el primero conformado por las bases de datos fuentes (de producción e históricos); el segundo, es una base de datos con datos resumidos extraídos de las bases de producción y, el tercero, constituido por las interfaces orientadas a usuarios que extraen información para la toma de decisiones” (p. 196). La arquitectura lógica de un sistema de almacenamiento de datos se representa en la siguiente figura 6.

Krzysztof, Witold, Roman, & Lukasz, (2007), también afirma que los DW “tienen una arquitectura de tres niveles” (p. 108), como se puede ver en la figura 7.

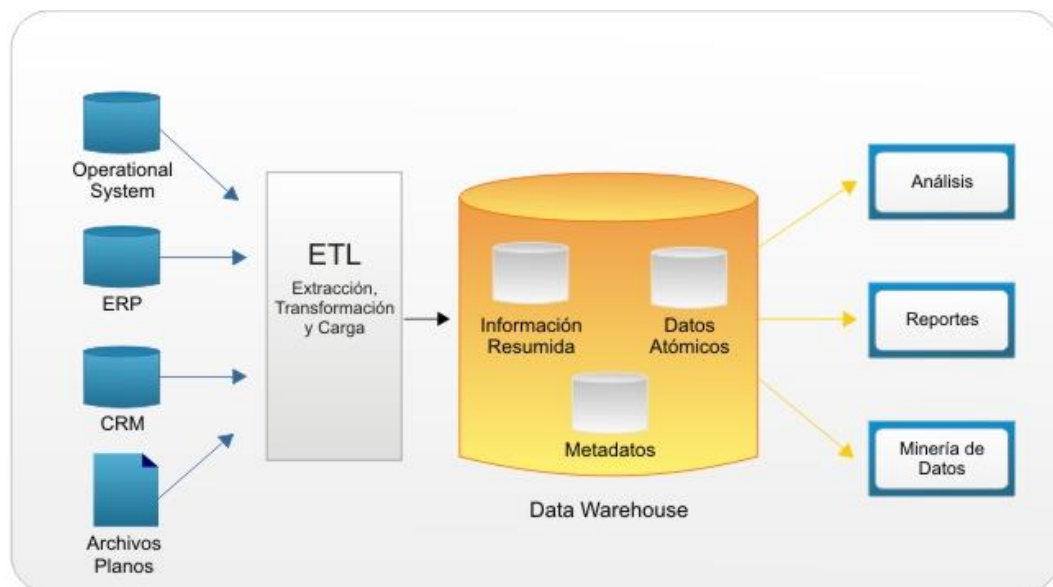
- Un nivel inferior que consiste en el servidor de DW, que puede incluir varios mercados de datos especializadas y un repositorio de metadatos

- Una etapa intermedia que consiste en un servidor OLAP (descrito más adelante) para la consulta rápida de los datos almacén
- Un nivel superior que incluye herramientas de front-end para la visualización de los resultados proporcionados por OLAP, así como herramientas adicionales para la extracción de datos de los datos OLAP generados



**Figura 6.** Arquitectura lógica de un Data Warehouse.

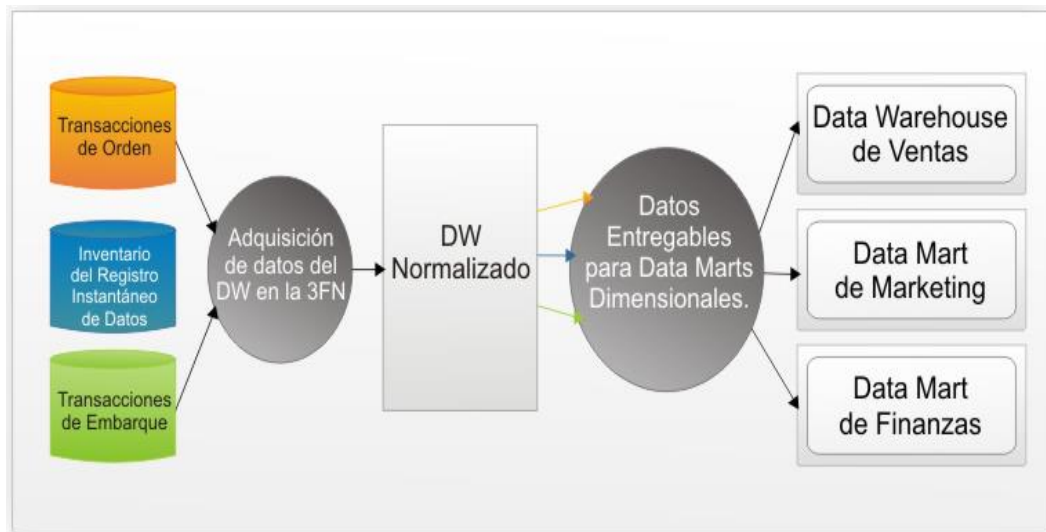
Fuente: Reinoso, E. J., Maldonado, C. A., & Muñoz, R. (2012). "Bases de datos" (Primera ed.). México: Alfaomega Grupo Editor. (p. 196).



**Figura 7.** Arquitectura de Data Warehouse.

Fuente: Pérez (2015). "Business Intelligence". (p. 2). Obtenida de <http://vortech-it.com/sets/cap-ii-bill-inmon.pdf>

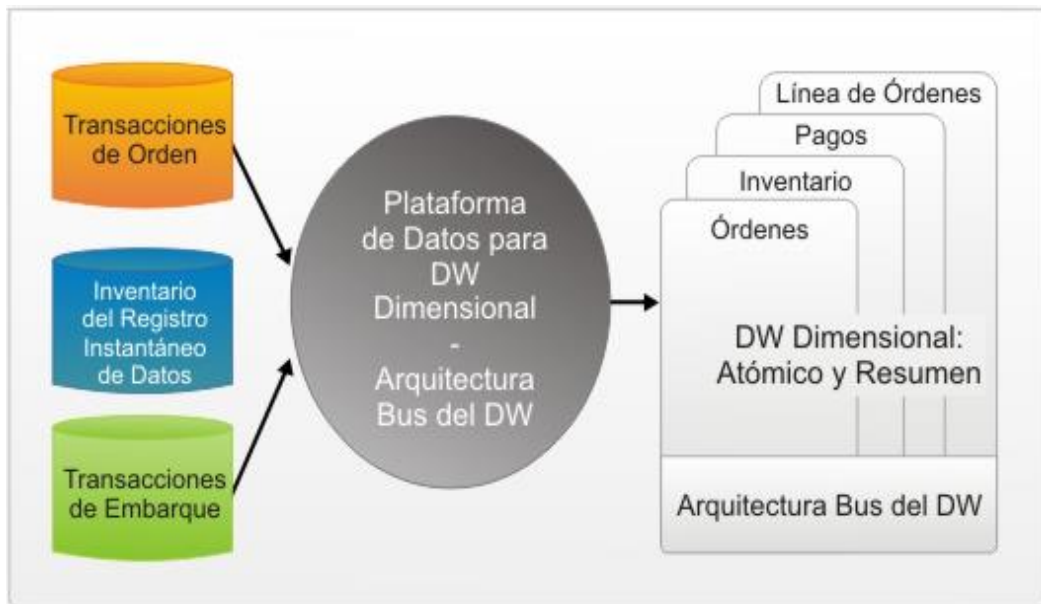
El enfoque Inmon también se referencia normalmente como Top-Down. Los datos son extraídos de los sistemas operacionales por los procesos Extracción, Transformación y Carga (ETL) y cargados en las áreas del escenario, donde son validados y consolidados en el DW corporativo, donde además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del DW. Una vez realizado este proceso, los procesos de refresco de los Data Mart departamentales obtienen la información de él, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido, como se visualiza en la figura 8.



**Figura 8.** Enfoque de inno de un DW corporativo.

Fuente: "Data Warehouse". (párr. 4). Obtenida de <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>

Este enfoque Ralph Kimball también se referencia como Bottom-Up, pues al final el DW Corporativo no es más que la unión de los diferentes datamarts, que están estructurados de una forma común a través de la bus structure. Esta característica le hace más flexible y sencillo de implementar, pues podemos construir un Data Mart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas. En este sistema, los procesos ETL extraen la información de los sistemas operacionales y los procesan igualmente en el área stage, realizando posteriormente el llenado de cada uno de los Data Mart de una forma individual, aunque siempre respetando la estandarización de las dimensiones, como veremos en la figura 9.



**Figura 9.** Enfoque de kimball - arquitectura bus del DW

Fuente: "Data Warehouse". (párr. 6). Obtenida de <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>

Las principales características tomadas de los dos enfoques se presentan en la figura 10.

| Característica                          | Inmon  | Kimball   |
|---|--|---|
| Diseño del DW                           | Toma mucho tiempo  | Toma menos tiempo   |
| Tiempo para poner en marcha el proyecto | Un largo periodo   | No requiere mucho tiempo  |
| Costo de implementación                 | Costos iniciales muy altos, subsecuente al desarrollo del proyecto los costos serán más bajos. | Bajos costos iniciales y en cada etapa subsecuente, los costos al menos serán los mismos. |
| Integración de datos                    | Toda la empresa  | Áreas específicas del negocio   |
| Mantenimiento                           | Fácil  | Difícil, siempre es redundante y sujeto a revisiones.                                     |

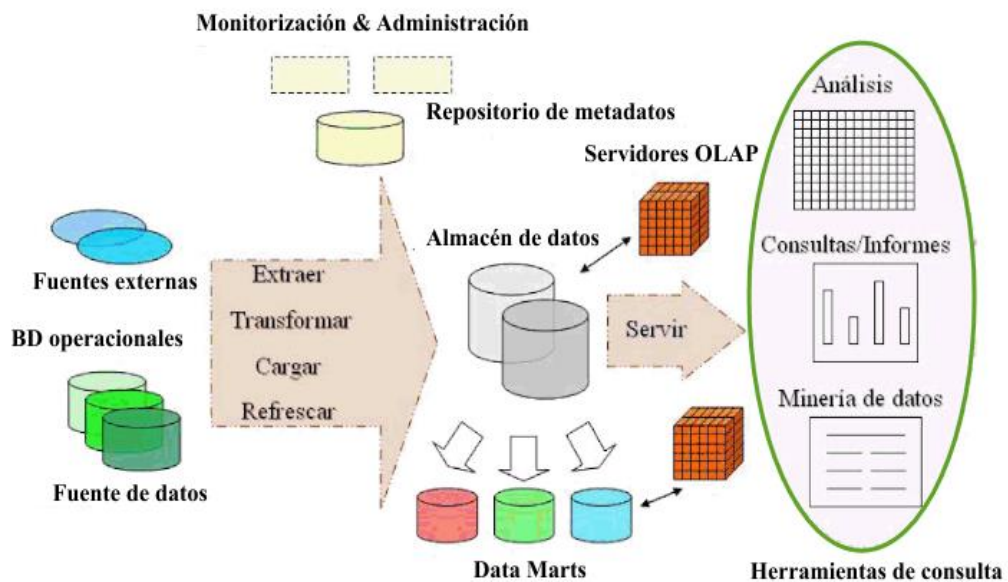
**Figura 10.** Principales características del enfoque de inmon y kimball.

Fuente: Florez, J. A. (2014). "Data Warehouse y minería de datos como alternativas al análisis de datos forestales". (p. 28). Obtenida de [http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2521/Flores\\_Cruz\\_JA\\_MC\\_Computo\\_Aplicado\\_2014.pdf](http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2521/Flores_Cruz_JA_MC_Computo_Aplicado_2014.pdf)

Según Trujillo (2013), el DW se estructura en cinco partes principales, como veremos en la figura 11:

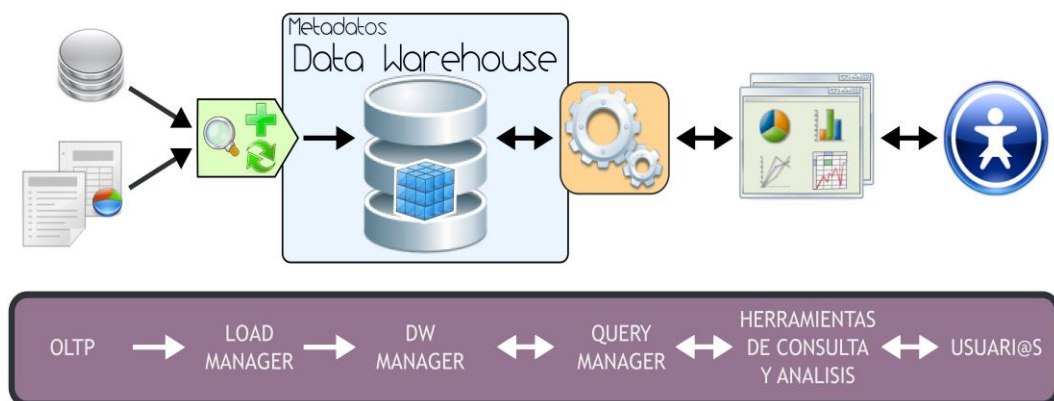
- Fuentes de datos externas (ya sean sistemas transaccionales o no) y los procesos responsables de realizar la extracción de los datos desde las fuentes, transformarlos y cargarlos en el repositorio.
- Repositorio del almacén de datos y data marts o almacenes de datos departamentales.
- Metadatos, que contienen datos sobre datos, es decir, información sobre todo lo concerniente al almacén de datos.
- Herramientas de consulta del usuario final (el también denominado front-end).
- Servidores On-Line Analytical Processing (OLAP), procesamiento analítico en línea que atienden a consultas de las herramientas de los usuarios finales y las lanzan al servidor del almacén de datos.

Según Bernabeu (2010), detalla que la estructura de un DW se describe en 6 componentes “On Line Transaction Processing (OLTP), Load Manager, DW Manager, Query Manager, Herramientas de Analisis de Consultas y Usuarios” (p. 19), como visualizaremos en la figura 12.



**Figura 11.** Arquitectura general del Data Warehouse.

Fuente: Trujillo, J. C. (2013). "Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional". (p. 15). España: Editorial Club Universitario.



**Figura 12.** Arquitectura de Data Warehouse de hefesto.

Fuente: Bernabeu, R. D. (2010). "HEFESTO: Metodología para la Construcción de un Data Warehouse". (p. 19). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

## **2.5. Modelado multidimensional**

Trujillo (2013), comenta que, “El modelado multidimensional requiere de técnicas de diseño especializadas que se asemejan a los métodos tradicionales de diseño de bases de datos” (p. 27).

En primer lugar, se desarrolla una fase de diseño conceptual con el fin de obtener un modelo multidimensional conceptual para el almacén de datos independiente de la implementación. Luego, una fase de diseño lógico se encarga de derivar un modelo dependiente de la tecnología a partir del modelo conceptual.

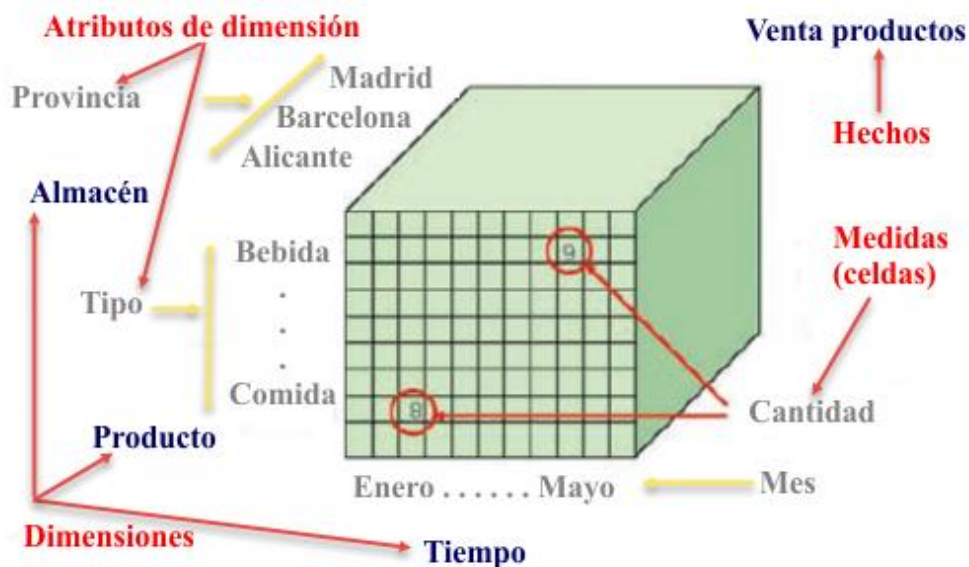
Cabe recalcar que el DW posibilitan una visión multidimensional de enormes cantidades de datos históricos provenientes de fuentes operacionales, suministrando la información necesaria para el apoyo a los procesos de toma de decisiones de una organización.

Un modelo multidimensional se representa en forma de cubo o hipercubo, como se muestra en la figura 13.

### **2.5.1. Modelado multidimensional conceptual**

De acuerdo con Trujillo (2013), “El modelo multidimensional conceptual debe representar aquellos elementos que permitan al usuario obtener la información necesaria para el apoyo a la toma de decisiones.

Estos elementos o propiedades multidimensionales pueden ser estructurales o dinámicos” (p. 28).



**Figura 13.** Cubo de datos.

Fuente: Trujillo, J. C. (2013). “Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional” (p. 30). España: Editorial Club Universitario.

### 2.5.1.1. Propiedades multidimensionales estructurales

Las propiedades estructurales consisten en la definición de estructuras de datos que faciliten el análisis multidimensional de los mismos. En conclusión, Trujillo (2013), afirma que “la información se estructura en hechos y dimensiones. Un hecho contiene medidas interesantes que son el objeto de análisis, mientras que las dimensiones representan diferentes perspectivas para analizar dichas medidas” (p. 29).

#### 2.5.1.1.1. Dimensiones

Bernabeu (2010), lo define como: “los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio” (p. 28).

Curto (2010), lo describe como: “la representación en el data warehouse de una vista para un cierto proceso de negocio” (p. 34).

Trujillo (2013), menciona que, “Una dimensión se compone de una serie de atributos organizados jerárquicamente. Estos atributos permiten analizar las medidas de los hechos a diferente nivel de detalle según se agreguen o desagreguen los datos” (p. 29).

Trujillo (2013), describe la cardinalidad en la relación entre estos niveles de detalle da lugar a diferentes tipos de jerarquías, son:

- a) Jerarquías estrictas y no-estrictas. Una relación entre dos niveles de jerarquía es estricta si para cada elemento de nivel más detallado, existe como máximo un elemento de nivel más general el cual se asocia con dicho elemento; de otra manera, se llama no estricta. En otras palabras, una asociación entre dos niveles de jerarquía es estricta si la cardinalidad es uno a muchos, si no, entonces se llama no-estricta. Por ejemplo, la asociación entre Semana y Mes es no-estricta ya que parte de

una misma semana puede estar en un mes dado y parte en el siguiente mes.

- b) Jerarquías simétricas y asimétricas. Una relación que involucra a un par de niveles de dimensión es simétrica si para cada elemento de nivel menos detallado, existe un elemento de nivel más detallado el cual se asocia con dicho elemento; de lo contrario la asociación es asimétrica. En otras palabras, una relación entre niveles es asimétrica si la cardinalidad es cero a uno; si no, se denomina simétrica. Por ejemplo, la asociación entre Producto y Tipo es asimétrica si puede haber tipos de productos que no contengan productos específicos. Una jerarquía asimétrica implica que las instancias de los niveles de la jerarquía no son obligatorias y pueden no estar instanciados.
- c) Jerarquías completas e incompletas. Una asociación entre dos niveles de jerarquía es completa si para cada elemento de un nivel más detallado, existe un elemento de nivel más general que se asocia con dicho elemento; si no, se llama incompleta. En otras palabras, una asociación entre dos niveles es incompleta si la asociación entre niveles de jerarquía es cero a muchos; si no, se denomina completa. Por ejemplo, la asociación entre

Producto y Categoría es incompleta si hay productos que no pertenecen a ninguna categoría.

- d) Jerarquías de generalización. En las dimensiones puede haber un alto grado de categorización, es decir, los atributos de la dimensión pueden tener sentido o no dependiendo de las instancias. Este tipo de jerarquías pueden representarse mediante relaciones de generalización/especialización. Por ejemplo, el volumen y el porcentaje de alcohol sólo tiene sentido si el producto es una bebida mientras que el tiempo y modo de preparación solo es válido para comidas.

Las jerarquías simples pueden representarse mediante árboles, las múltiples se suelen representar mediante grafos acíclicos dirigidos tal y como se observa en la figura 14.



**Figura 14.** Jerarquía de grafos acíclicos dirigidos.

Fuente: Trujillo, J. C. (2013). "Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional" (p. 31). España: Editorial Club Universitario.

De acuerdo con Bernabeu (2010), Los datos de las tablas de dimensiones, que proveen información del negocio o que describen alguna de sus características, son llamados datos de referencia.

Más detalladamente, cada tabla de dimensión podrá contener los siguientes campos:

- Clave principal o identificador único.
- Claves foráneas.
- Datos de referencia primarios: datos que identifican la dimensión.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión.

#### 2.5.1.1.2. Hechos

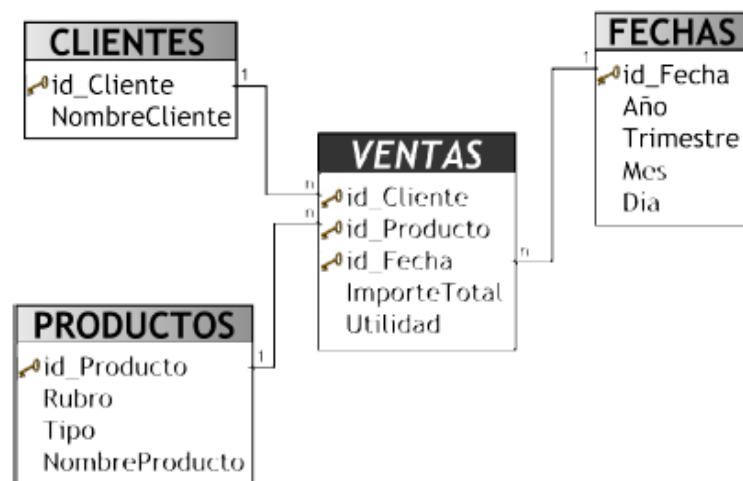
Según Bernabeu (2010), describe que, “Las tablas de hechos contienen, precisamente, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones”, también llamado dato agregado.

Elmasri & Navathe (2007), considera que, “Una tabla de hechos puede considerarse como una agrupación de tuplas, una por cada hecho registrado” (p. 855).

Kimball & Ross (2013), describe que, “Una tabla de hechos contiene las medidas numéricas producidas por una medida operativa evento en el mundo real” (p. 41).

Trujillo (2013), menciona que, “Los hechos contienen atributos de hecho o medidas a analizar” (p. 32). Éstos representan normalmente relaciones muchos a muchos con todas las dimensiones y muchos a uno con cada dimensión en particular.

El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este, como se visualiza en la figura 15.

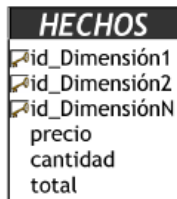


**Figura 15.** Tabla de hechos.

Fuente: Bernabeu, R. D. (2010). “HEFESTO: Metodología para la Construcción de un Data Warehouse”. (p. 31). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

Bernabeu (2010), afirma que, “Los hechos son aquellos datos que residen en una tabla de hechos y que son utilizados para crear indicadores, a través de sumalizaciones preestablecidas al momento de crear un cubo multidimensional” (p. 31).

Existen dos tipos de hechos, los básicos y los derivados, a continuación, se detallará cada uno de ellos, teniendo en cuenta para su ejemplificación la siguiente tabla de hechos, como se visualiza en la figura 16:



| HECHOS        |  |
|---------------|--|
| id_Dimensión1 |  |
| id_Dimensión2 |  |
| id_DimensiónN |  |
| precio        |  |
| cantidad      |  |
| total         |  |

**Figura 16.** Hechos básicos y derivados.

Fuente: Bernabeu, R. D. (2010). “HEFESTO: Metodología para la Construcción de un Data Warehouse”. (p. 31). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

Hechos básicos: son los que se encuentran representados por un campo de una tabla de hechos. Los campos “precio” y “cantidad” de la figura anterior son hechos básicos.

Hechos derivados: son los que se forman al combinar uno o más hechos con alguna operación matemática o lógica y que también residen en una tabla de hechos. Estos poseen la ventaja de almacenarse

previamente calculados, por lo cual pueden ser accedidos a través de consultas SQL sencillas y devolver resultados rápidamente, pero requieren más espacio físico en el DW, además de necesitar más tiempo de proceso en los ETL que los calculan. El campo “total” corresponde a una tabla hecho derivado, expresado en una multiplicación del precio y la cantidad. Cabe resaltar, que no es necesario que los hechos derivados se compongan únicamente con hechos pertenecientes a una misma tabla.

#### 2.5.1.1.3. Cubo multidimensional

Bernabeu (2010), contribuye que, “Un cubo multidimensional o hipercubo, representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones” (p. 34).

Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

Indicadores: sumarizaciones que se efectúan sobre algún hecho o expresiones basadas en sumarizaciones, pertenecientes a una tabla de hechos.

Atributos: campos o criterios de análisis, pertenecientes a tablas de dimensiones.

Jerarquías: representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”.

Relación representa la forma en que dos atributos interactúan dentro de una jerarquía. Existen básicamente dos tipos de relaciones:

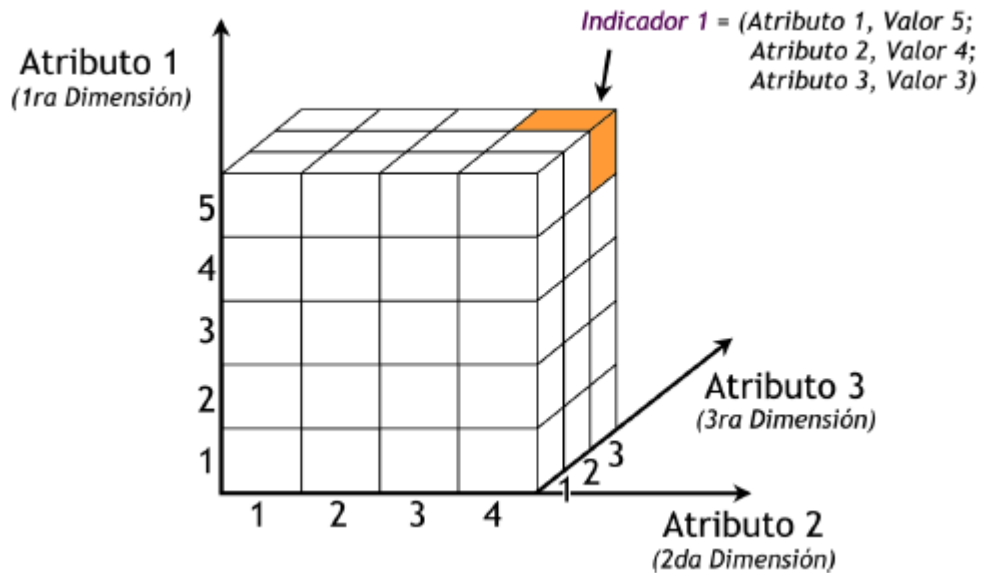
Explícitas: son las más comunes y se pueden modelar a partir de atributos directos y están en línea continua de una jerarquía.

Implícitas: son las que ocurren en la vida real, pero su relación no es de vista directa.

La granularidad: representa el nivel de detalle al que se desea almacenar la información sobre el negocio que se esté analizando.

De esta manera en un cubo multidimensional, los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando.

En la siguiente figura 17 la representación matricial se puede ver más claramente lo que se acaba de decir.



**Figura 17.** Cubo dimensional.

Fuente: Bernabeu, R. D. (2010). "HEFESTO: Metodología para la Construcción de un Data Warehouse". (p. 34). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

### 2.5.1.2. Propiedades multidimensionales dinámicas

Las propiedades dinámicas están relacionadas con la definición de los requisitos iniciales de usuario. Para contestar a estos requisitos se deben utilizar un tipo de operaciones con el fin de interrogar a la parte estructural del modelo multidimensional, realizando operaciones de consulta OLAP, a continuación, describiremos algunas de las operaciones importantes que son:

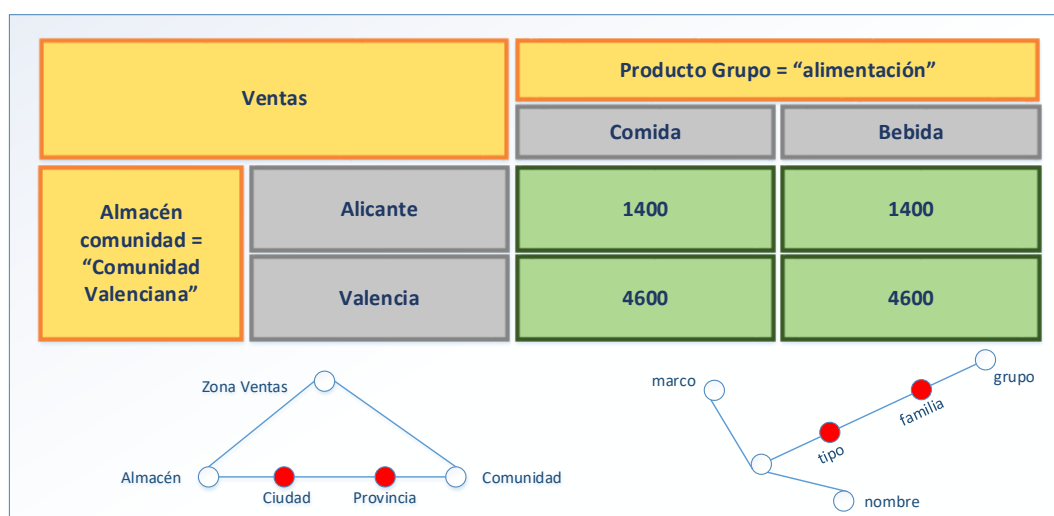
Roll-up: esta operación permite agregar valores de medidas a lo largo de los niveles de jerarquías de clasificación de las dimensiones. Un ejemplo de roll-up se observa en la figura 18.

Drill-down: esta operación permite desagregar valores de medidas a lo largo de jerarquías de clasificación.

Drill-across: esta operación permite navegar de un hecho a otro mediante el uso de dimensiones comunes.

Slice & dice: este par de operaciones permiten definir restricciones sobre niveles de jerarquías. Por ejemplo, analizar datos del año 1999 (ver figura 19).

Pivoting: esta operación permite reorientar la vista multidimensional de los datos, es decir, cambiar la distribución de filas o columnas. Algunos autores consideran también el intercambio de medidas y hechos como pivoting. Un ejemplo de pivoting se observa en la figura 20.



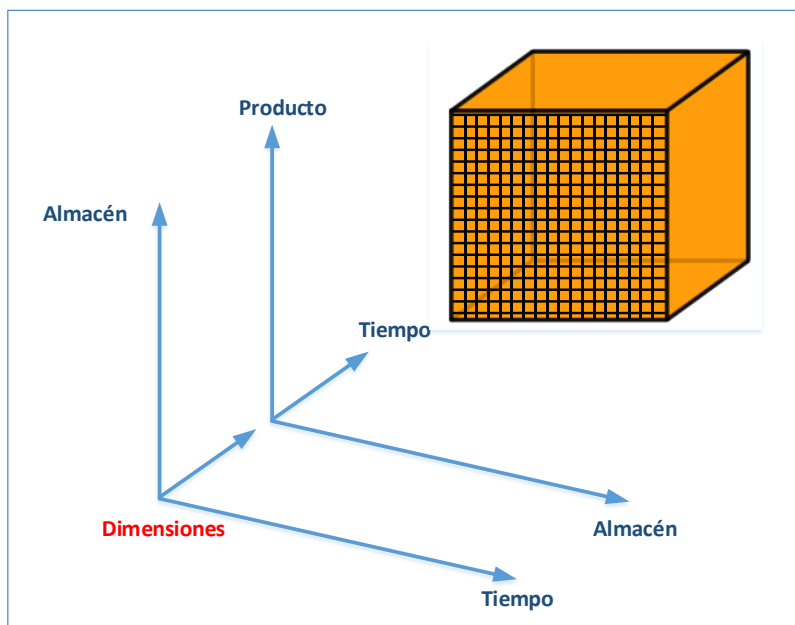
**Figura 18.** Aplicación de la operación roll-up.

Fuente: Trujillo, J. C. (2013). "Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional" (p. 33). España: Editorial Club Universitario.

|   |                 |  |               |     |
|---|-----------------|--|---------------|-----|
| <b>Ventas</b>                                     |                 | <b>Producto Grupo = "alimentación"</b> |               |     |
|   |                 | <b>Comida</b>                          |               |     |
|   |                 | <b>Congelada</b>                       | <b>Fresca</b> |     |
| <b>Almacén comunidad = "Comunidad Valenciana"</b> | <b>Alicante</b> | <b>Albatera</b>                        | 100           | 200 |
|   |                 | <b>Elche</b>                           | 500           | 600 |

**Figura 19.** Aplicación de la operación drill-across.

Fuente: Trujillo, J. C. (2013). "Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional" (p. 33). España: Editorial Club Universitario.



**Figura 20.** Aplicación de la operación pivoting.

Fuente: Trujillo, J. C. (2013). "Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional" (p. 34). España: Editorial Club Universitario.

### 2.5.2. Modelado multidimensional lógico

Según Curto (2010), se debe entender por OLAP, "al método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o

jerarquía de objetos como en un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes” (p. 96).

De acuerdo con el trabajo doctoral Trujillo (2001) y Trujillo (2013), establece que “Una vez definido un modelo conceptual multidimensional independiente de la plataforma de implementación, se debe desarrollar un modelo lógico. El modelado lógico dependerá del tipo de tecnología sobre el que se va a implementar el almacén de datos” (p. 33). Existen dos tendencias, arquitecturas para los sistemas OLAP principales:

Tecnología multidimensional. Se usan vectores o matrices multidimensionales para representar las diferentes estructuras multidimensionales. Este tipo de tecnología se denomina Multidimensional Online Analytical Processing (MOLAP).

Tecnología relacional. Se usan elementos relacionales como tablas, columnas, claves primarias, ajenas, etc., para poder definir las diferentes estructuras multidimensionales. Este tipo de tecnología se denomina Relacional Online Analytical Processing (ROLAP).

Reinosa, Maldonado, & Muñoz, (2012), hacen una diferenciación de los modelos OLAP de acuerdo a la forma en que se almacenan los datos.

En un ROLAP, los datos se almacenan físicamente en un Sistema Gestor de Bases de Datos Relacional; mientras que un MOLAP se almacenan en una base de datos multidimensional. Un híbrido generado a partir de estos dos modelos son los HOLAP, en los que se mezclan ambas arquitecturas, bases de datos multidimensionales y de datos relacionales.

### **2.5.2.1. Tipo de modelamiento de un Data Warehouse**

#### 2.5.2.1.1. Esquema de estrella

Según Elmasri & Navathe (2007), describe que, “El esquema estrella consiste en una tabla de hechos con una única tabla por cada dimensión” (p. 855).

Date (2001), menciona que, “el esquema de estrella involucra el uso de las tablas de dimensión para encontrar todas las combinaciones de clave externa que son de interés, y luego el uso de esas combinaciones para acceder a la tabla de hechos” (p. 712).

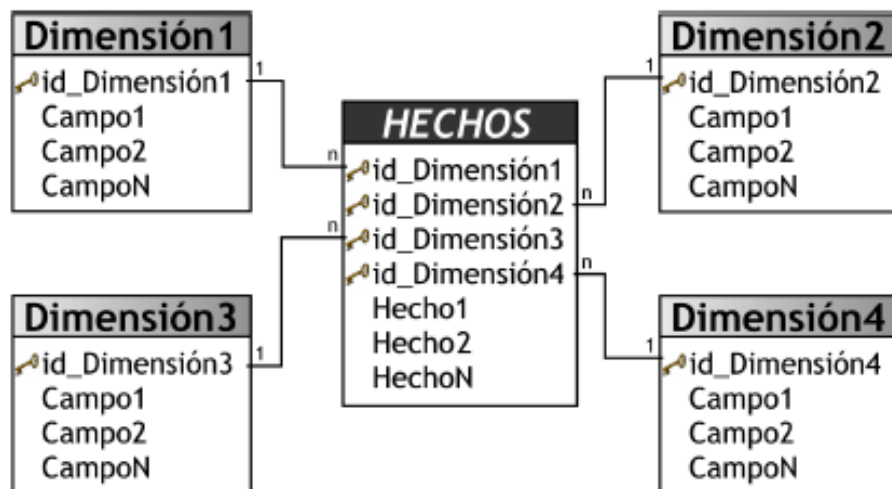
Trujillo (2013), contribuye que, “un esquema estrella consiste en una tabla central (tabla de hechos) y un conjunto de tablas de dimensión” (p. 34).

Las características principales de este tipo de esquema son:

- Cada tabla de dimensión contiene una clave primaria no compuesta.

- Cada una de las columnas de la tabla de hechos que forma parte de la clave primaria compuesta es una clave ajena que referencia a la clave primaria de cada dimensión.

Bernabeu (2010), afirma que, “El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves” (p. 37). En la siguiente figura 21 se puede apreciar un esquema en estrella estándar.



**Figura 21.** Esquema de estrella.

Fuente: Bernabeu, R. D. (2010). “HEFESTO: Metodología para la Construcción de un Data Warehouse”. (p. 37). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

Este modelo debe estar totalmente desnormalizado, es decir que no puede presentarse en tercera forma normal.

De acuerdo con Bernabeu (2010), destaca algunas características principales de este modelo, que ayudarán a comprender mejor el porqué de sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.
- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

A continuación, presentamos algunos inconvenientes del esquema de estrella.

- El aumento del tamaño de la tabla de hechos con datos agregados puede empeorar el rendimiento general.
- Las dimensiones tienen un tamaño enorme. Un valor normal según Kimball es de alrededor de 50 atributos.
- Es poco robusto o susceptible a cambios.
- Más lento de construir.

#### 2.5.2.1.2. Esquema de copa de nieve

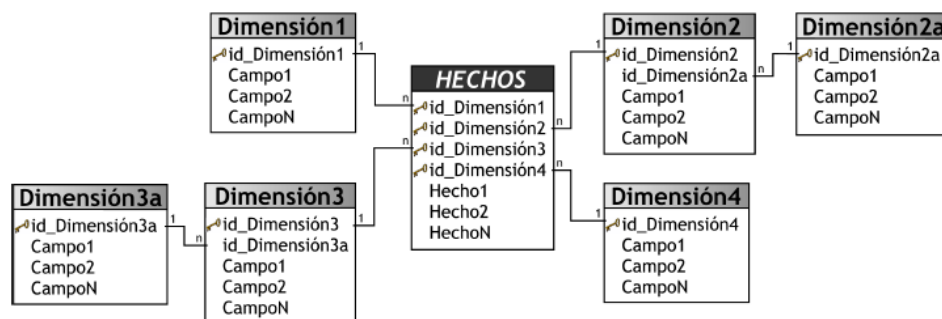
Según Elmasri & Navathe (2007), describe que, “El esquema en copo de nieve es una variación del esquema en estrella en el que las tablas dimensionales de éste están organizadas en una jerarquía para

normalizarlas” (p. 856). Este modelo puede estar normalizado hasta la tercera forma normal, a fin de tener un mejor nivel de detalle.

Date (2001), menciona que, “es una variante del esquema de estrella es el esquema de copo de nieve, el cual normaliza las tablas de dimensión” (p. 714).

Trujillo (2001), contribuye que, “En este esquema se normalizan las tablas de dimensiones por sus niveles de jerarquía” (p. 44). Este esquema lo utiliza para disminuir el tamaño de las tablas de dimensiones y no utilizar el atributo de nivel.

De acuerdo con Bernabeu (2010), menciona que, “representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones” (p. 39). En la siguiente figura 22 se puede apreciar un esquema de constelación.



**Figura 22.** Esquema de copa de nieve.

Fuente: Bernabeu, R. D. (2010). “HEFESTO: Metodología para la Construcción de un Data Warehouse”. (p. 39). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

Las características principales de este esquema y sus ventajas son:

- Más flexible y adecuado para requisitos.
- Carga de datos más rápida y sencilla.
- Directamente implementados por algunas herramientas.
- Mejora considerablemente el rendimiento cuando un gran número de requisitos solicita datos agregados o de niveles superiores de jerarquías ya que los requisitos escanean un reducido número de filas.

A continuación, presentamos algunos inconvenientes de este esquema.

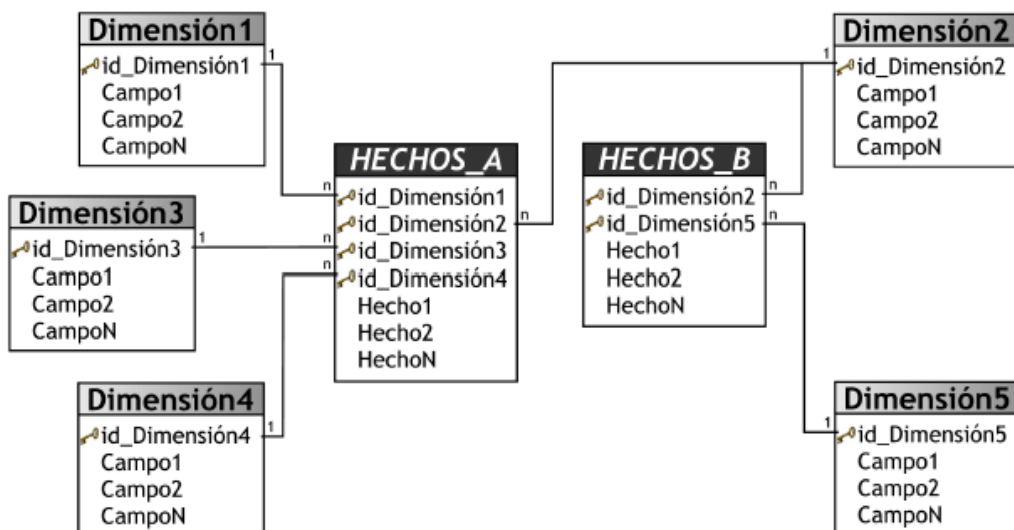
- Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

#### 2.5.2.1.3. Esquema de constelación

Según Elmasri & Navathe (2007), describe que, “Una constelación de hechos es un conjunto de tablas de hechos que comparten algunas tablas de dimensión” (p. 855).

Trujillo (2001), contribuye que, “En este esquema, además de la tabla de hechos central que contiene datos con respecto a los niveles de agregación más bajos, tenemos más tablas de hechos que representan agregados para los distintos niveles de jerarquías.

Bernabeu (2010), afirma que, “Este modelo está compuesto por una serie de esquemas en estrella, está formado por una tabla de hechos principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal” (p. 42). En la siguiente figura 23 se puede apreciar un esquema de constelación.



**Figura 23.** Esquema de constelación.

Fuente: Bernabeu, R. D. (2010). “HEFESTO: Metodología para la Construcción de un Data Warehouse”. (p. 37). Obtenida de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

Las características principales de este esquema y sus ventajas son:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.

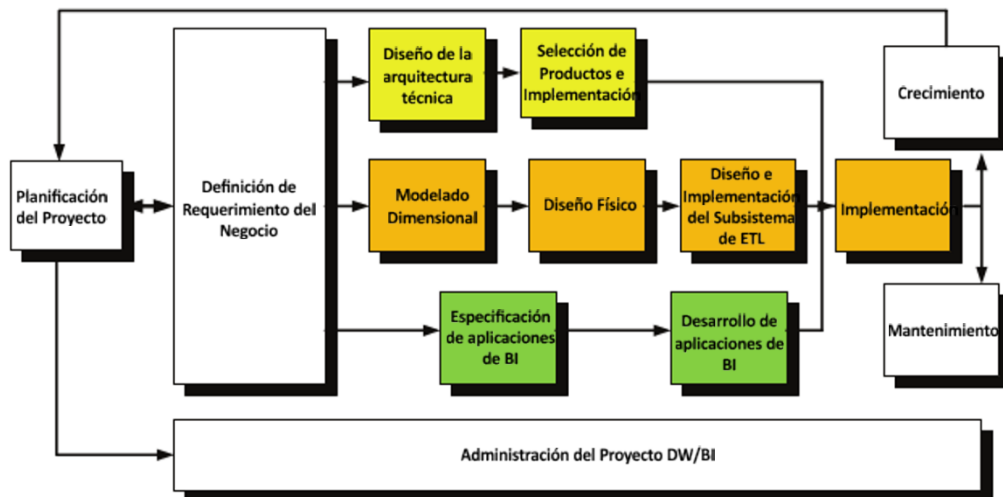
A continuación, presentamos algunos inconvenientes de este esquema.

- Un gran número de tablas de agregados por lo que se complica el mantenimiento de los metadatos y existe una necesidad del usuario de conocer la existencia de dichas tablas.
- Algunos requisitos pueden necesitar consultar datos de varias tablas y mermar el rendimiento general del sistema.

## **2.6. Metodología Ralph Kimball**

Es una metodología empleada para la construcción de un almacén de datos que no es más que, una colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo, que contribuye a tomar buenas decisiones en una organización.

Ralph Kimball propone una metodología mas explicada desde un enfoque de tareas para la construcción del DW. Ver la figura 24.



**Figura 24.** Tareas de la metodología de Ralph Kimball.

Fuente: Obtenida de <http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html>

### 2.6.1. Planificación del proyecto

En este proceso se determina el propósito del proyecto de DW, sus objetivos, alcances, riesgos, recursos necesarios y finalmente planificación de las actividades. Como se lista las siguientes tareas:

- Objetivos del Proyecto
- Alcance del proyecto
- Cronograma de Actividades
- Asignación de Recursos

### 2.6.2. Definición de los requerimientos del negocio

En esta etapa nos permite establecer los requerimientos de negocio a través de las entrevistas al personal de negocio y técnico, revisión de

documentos internos de la organización. Estos requerimientos deben estar claros y precisos antes de continuar a los siguientes procesos. Dado las existencias de relaciones entre procesos. A continuación, listamos las siguientes tareas:

- Requerimientos del negocio
- Entrevistas
- Base de datos transaccional
- Diccionario de datos de la Base de datos transaccional

### **2.6.3. Diseño de la arquitectura**

En esta etapa es la base de todo proyecto, idea o plan que con lleva a ser útil o funcional, construible y estéticamente bello. Según Ralph Kimball plantea las tres columnas fundamentales que son: Back Room, Front Room y la infraestructura.

### **2.6.4. Instalación y selección del producto**

En esta etapa se especifica las herramientas de software para el modelamiento dimensional, análisis de datos, exploración de resultados y la instalación del DW.

### **2.6.5. Modelamiento dimensional**

Es un conjunto de técnicas que permite diseñar modelos lógicos. Nos ayuda con las siguientes tareas:

- Elegir el proceso de negocio

- Establecer el nivel de granularidad
- Elegir las dimensiones
- Identificar medidas y las tablas de hechos
- Elegimos el esquema del modelo dimensional

#### **2.6.6. Diseño físico**

En esta etapa se establece la transformación del modelo lógico al modelo físico de una base de datos. También se debe prever el plan de indexación y particiones de algunas tablas si lo amerita y las características del software y hardware para la administración y el tipo almacenamiento.

#### **2.6.7. Diseño e implementación del subsistema de extracción, transformación y carga (ETL)**

En esta etapa de Extracción, Transformación y Carga (ETL) es la base sobre la cual se alimenta el DW. Si se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar la información en el DW para luego utilizar las herramientas de análisis para la exploración de resultados.

#### **2.6.8. Especificación de aplicaciones de business intelligence (BI)**

En esta etapa las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis

proporcionan información útil a los usuarios, como se lista las tareas a continuación:

- Informes estándar
- Aplicaciones analíticas
- Usuarios finales

#### **2.6.9. Desarrollo de aplicación BI**

En esta etapa se realiza el proceso del desarrollo de las aplicaciones de BI. Partiendo desde la creación de los modelos dimensionales, creación de cubos y finalmente explorando los resultados. Esto ayudara al usuario tomar buenas decisiones dentro de su organización. A continuación, listamos las tareas para el desarrollo BI:

- Elegir el proyecto multidimensional
- Elegir el origen de datos
- Crear vista de origen de datos
- Crear Cubos
- Explorar Resultados

#### **2.6.10. Implementación**

En esta etapa la implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Para asegurar el buen funcionamiento

se de tener en cuenta las capacitaciones, el soporte técnico, la comunicación y las estrategias de retroalimentación.

## **2.7. Descubrimiento de conocimiento en base de datos**

### **2.7.1. Definición de descubrimiento de conocimiento en base de datos**

Frawley, Piatetsky-Shapiro, & Matheus (1992), definen que, “El Descubrimiento de Conocimiento en Base de Datos (KDD) es la extracción de información interesante (no-trivial, implícita, previamente desconocida y potencialmente útil de datos)” (p. 58).

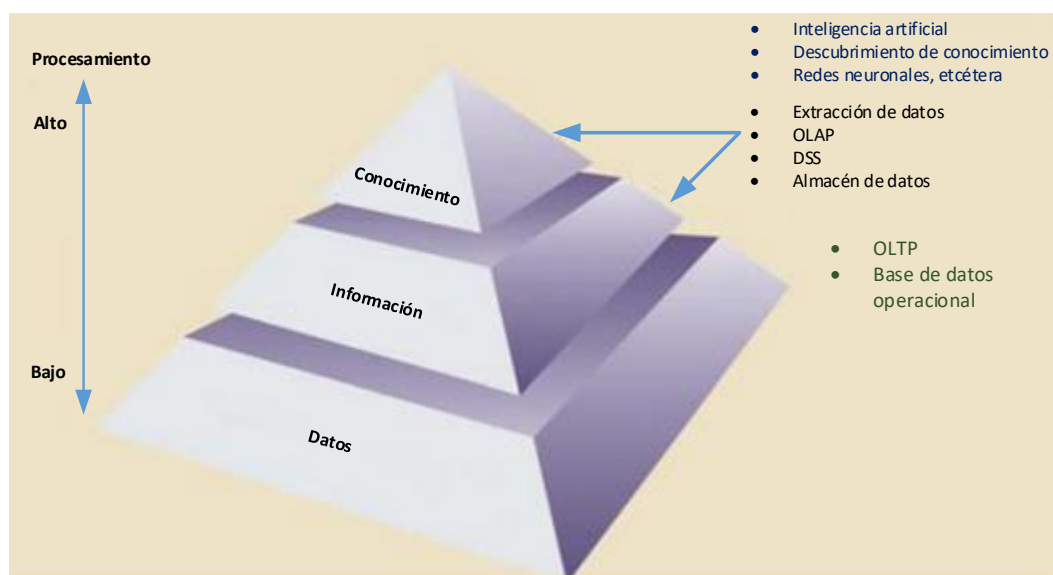
Además, Fayyad, Piatetsky-Shapiro, & Smyth (1996), definen el KDD como “el proceso no trivial de identificar patrones, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos” (p. 41).

Según Maimon & Rokach, (2010), describe que, “KDD es un análisis automático, exploratorio y modelización de los repositorios de datos de gran tamaño. KDD es el proceso organizado de identificación válida, novedosa, útil y patrones comprensibles de los conjuntos de datos grandes y complejos” (p. 1).

De acuerdo con Maimon & Rokach, (2010), se enfatiza que Data Mining es el núcleo o el paso esencial del proceso de KDD.

Hernández, Ramírez, & Ferri (2004), Afirman que, “existe una relación entre KDD y la minería de datos: el KDD es el proceso global de descubrir conocimiento útil desde la base de datos mientras que la minería de datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos” (p. 14).

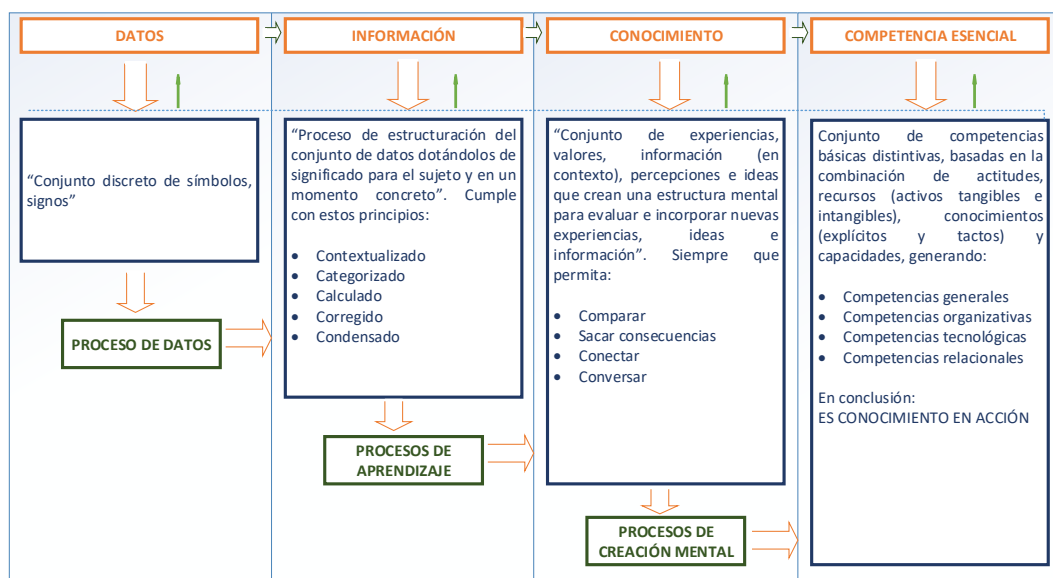
Para tener una idea de la extracción de conocimiento lo ilustraremos en la pirámide de la figura 25, que representa la forma en que se “extrae conocimiento a partir de datos, Los datos forman la base de la pirámide y representan lo que casi todas las organizaciones capturan en sus bases de datos operacionales. El segundo nivel contiene información que representa los datos refinados y procesados. La información constituye la base para la toma decisiones y comprensión del negocio. El conocimiento se encuentra en la cúspide de la pirámide y representa información altamente especializada” (Coronel & Morris & Rob, 2011, 554).



**Figura 25.** Extracción de conocimiento a partir de datos.

Fuente: Coronel, C., Morris, S., & Rob, P. (2011). "Bases de Datos, Diseño, Implementación y Administración" Pérez (2015). (p. 554).

Según Salmador (2005), citando a Bueno (1998), describe que, "los datos son un conjunto discreto de elementos, símbolos y signos. La información es un proceso de estructuración del conjunto de datos, dotándolos de significado para un sujeto y en un momento concretos. Este proceso cumple los requisitos de ser contextualizado, categorizado, calculado, corregido y condesado. El «output» del proceso es el conocimiento, como conjunto de experiencias, valores, información en contexto, percepciones e ideas que crean una estructura mental para evaluar e incorporar nuevas experiencias, ideas e información, siempre que permita comparar, sacar consecuencias, conectar y conversar" (p. 34), ver la figura 26.



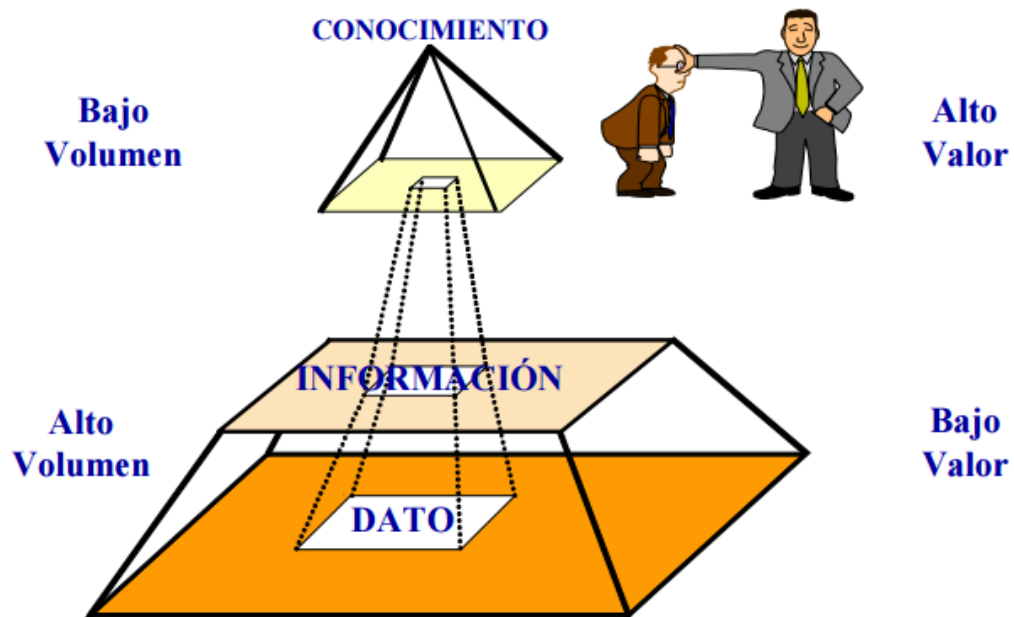
**Figura 26.** Dimensión sistémica del conocimiento.

Fuente: Salmador, M. P. (2005). “Raíces epistemológicas del conocimiento organizativo: estudio de sus dimensiones”. (p. 35). Obtenida de <http://eco.mdp.edu.ar/cendocu/repositorio/01047.pdf>

Según Davenport & Prusak (1998), describe que para la creación de conocimiento partimos por entender que, “dato es un conjunto de hechos objetivos discretos sobre eventos, la información a describirse como un mensaje que tiene un remitente y un receptor, y el conocimiento es una mezcla fluida de experiencia enmarcada, información contextual, los valores y la visión de expertos que ofrece un marco para evaluar e incorporar nuevas experiencias e información. Se puede encontrar en rutinas, procesos, prácticas y normas” (p. 3-4).

A continuación, ilustraremos en la figura 27 la jerarquía que existe en una base de datos entre datos, información y conocimiento, se observa

igualmente el volumen que presenta en cada nivel y el valor que los responsables de las decisiones le dan en esa jerarquía.

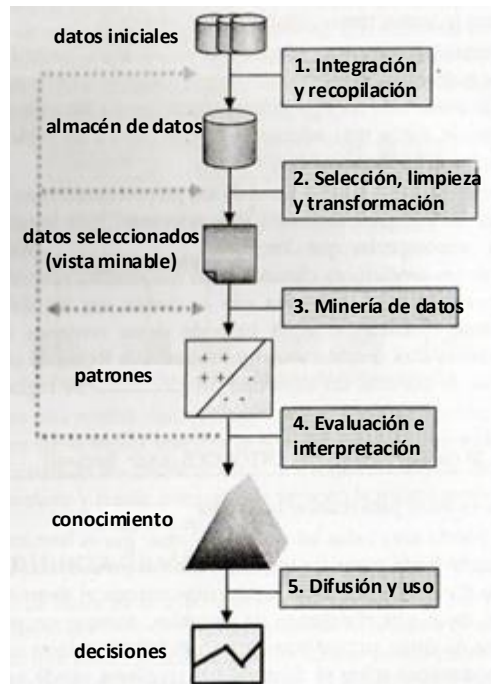


**Figura 27.** Arquitectura de Data Warehouse.

Fuente: Molina (2001). "Torturando a los datos hasta que confiesen". (párr. 4). Obtenida de <http://www.businessintelligence.info/docs/dss/torturando-datos-ppt.pdf>

### **2.7.2. Fases del proceso de KDD**

Antes de utilizar Minería de datos para el KDD es necesario contar con información confiable y adecuada para los modelos que se pretendan utilizar, de esta manera se abre paso al proceso iterativo del KDD, según Hernández, Ramírez, & Ferri (2004), organiza al proceso KDD en cinco fases, como se ilustra en la figura 28.

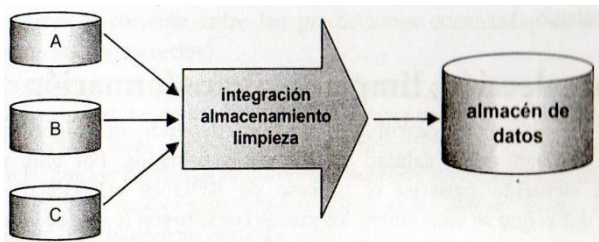


**Figura 28.** Fase del proceso de KDD.

Fuente: Hernández, J., Ramírez, J., & Ferri, C. (2004). "Introducción a la Minería de Datos". (p. 20) Madrid: Person Prentice Hall.

### 2.7.2.1. Fase de integración y recopilación

En esta fase se determina las fuentes de información que pueden ser útiles y dónde conseguirlas, según Hernández, Ramírez, & Ferri (2004), menciona a veces es necesario el uso de la información generada por otras organizaciones, a distintos departamentos de una misma entidad e incluso recurrir a la información en bases de datos públicas. Seguidamente se transforman todos los datos a un formato común, frecuentemente mediante un almacén de datos que permita unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias. A continuación, ilustraremos esta fase en la figura 29.

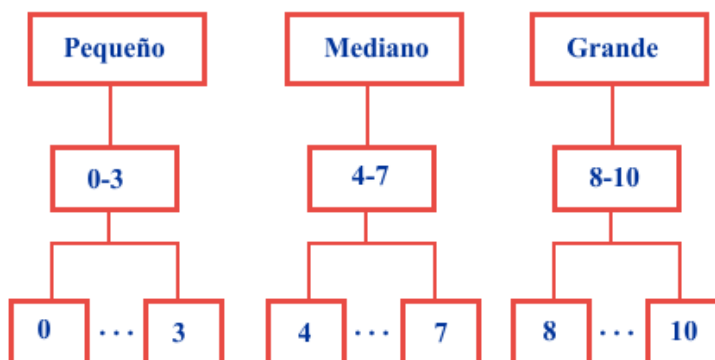


**Figura 29.** Integración de un almacén de datos.

Fuente: Hernández, J., Ramírez, J., & Ferri, C. (2004). "Introducción a la Minería de Datos". (p. 21) Madrid: Person Prentice Hall.

### 2.7.2.2. Fase de selección, limpieza y transformación

En esta fase se realiza una selección de la muestra de los datos más relevantes de esta manera se debe hacer cuidadosamente para asegurar la aleatoriedad. Seguidamente se eliminan o corrigen los datos incorrectos (datos anómalos, faltantes o perdidos) y se decide la estrategia a seguir con los datos incompletos. En el proceso de transformación consiste en discretizar los atributos continuos, es decir, transformar valores numéricos en atributos discretos o nominales. Como ilustraremos en la figura 30.



**Figura 30.** Discretización del atributo tamaño.

Fuente: Hernández, J., Ramírez, J., & Ferri, C. (2004). "Introducción a la Minería de Datos". (p. 24) Madrid: Person Prentice Hall.

### **2.7.2.3. Fase de minería de datos**

La fase de la minería de datos es la más importante del KDD. El objetivo de esta fase es producir un nuevo conocimiento que pueda utilizar el usuario. Para explicar mejor este proceso seguiremos los pasos siguientes:

- Determinar qué tipo de tarea de minería es el más apropiado.
- Elegir el tipo de modelo.
- Elegir el algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo que estamos buscando.

### **2.7.2.4. Fase de evaluación e interpretación**

En esta fase se evalúan los patrones descubiertos usando diferentes técnicas, estos patrones deben tener tres cualidades: ser precisos, comprensibles e interesante. La interpretación de resultados consiste en entender los resultados del análisis y sus implicaciones y se es necesario se vuelve a las fases anteriores para una nueva iteración.

### **2.7.2.5. Fase de difusión, uso y monitorización**

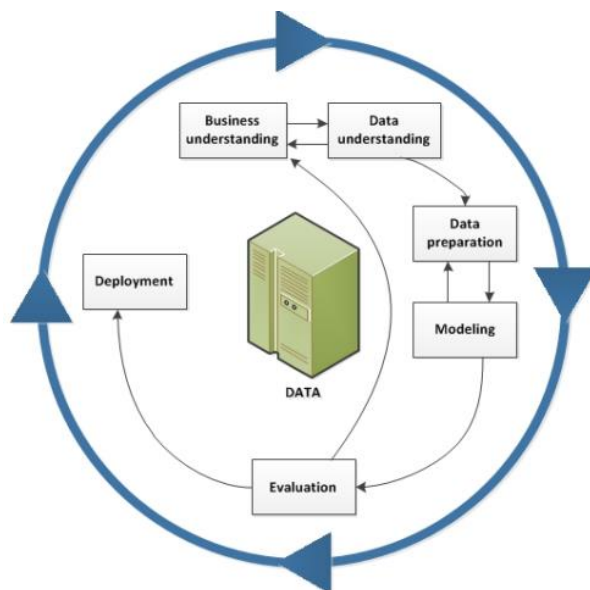
En esta fase el modelo construido y validado se usa principalmente con dos finalidades: para que un analista recomiende acciones basándose en el modelo y en sus resultados, o bien para aplicar el modelo a diferentes conjuntos de datos. Seguidamente es necesario que la difusión a todos los posibles usuarios de la organización. El nuevo conocimiento extraído debe integrar el know-how de la organización.

### 2.7.3. Cross-Industry Standard Process for Data Mining

Conocido también con el acrónimo CRISP-DM, “es un método probado para orientar sus trabajos de minería de datos” (International Business Machines Corporation, 2012, p.1). Desde sus dos enfoques podremos describirlos:

- Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.
- Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.

El ciclo vital del modelo contiene seis fases, algunas de estas fases son bidireccionales, lo que significa que algunas fases permitirán revisar parcial o totalmente las fases anteriores. Ver la figura 31.



**Figura 31.** Ciclo de vida de minería de datos.  
Fuente: Propia.

### **2.7.3.1. Fase de comprensión del negocio**

En esta etapa inicial se requiere establecer los objetivos y requerimientos desde una perspectiva no técnica, como se detalla las siguientes tareas:

- Establecimiento de los objetivos del negocio: Contexto inicial, objetivos, criterios de éxito.
- Evaluación de la situación: Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio.
- Establecimiento de los objetivos de la minería de datos: objetivos y criterios de éxito.
- Generación del plan del proyecto: plan de actividades, herramientas, equipo y técnicas.

### **2.7.3.2. Fase de comprensión de los datos**

En esta etapa es necesario familiarizarse con los datos teniendo presente los objetivos del negocio. Como se detalla en las siguientes tareas:

- Recopilación inicial de datos
- Descripción de los datos
- Exploración de los datos
- Verificación de calidad de datos

### **2.7.3.3. Fase de preparación de los datos**

En esta etapa es muy minuciosa por el tratamiento de los datos en la zona minable. Con lleva a realizar las siguientes tareas:

- Selección de los datos
- Limpieza de datos
- Construcción de datos
- Integración de datos
- Formateo de datos

#### **2.7.3.4. Fase de modelado**

En esta etapa se aplica las técnicas de minería de datos a los nuevos datos obtenidos de la zona minable, como se menciona en las siguientes tareas:

- Selección de la técnica de modelado
- Diseño de la evaluación
- Construcción del modelo
- Evaluación del modelo

#### **2.7.3.5. Fase de evaluación**

En esta etapa se evalúa los resultados de los modelos cuales son más precisos o eficaces y esto con lleve a determinar si son útiles a las necesidades del negocio, como se detalla en las siguientes tareas:

- Evaluación de resultados
- Revisar el proceso
- Establecimiento de los siguientes pasos o acciones

### **2.7.3.6. Fase de despliegue**

En esta etapa se explota la utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización.

- Planificación de despliegue
- Planificación de la monitorización y del mantenimiento
- Generación de informe final
- Revisión del proyecto

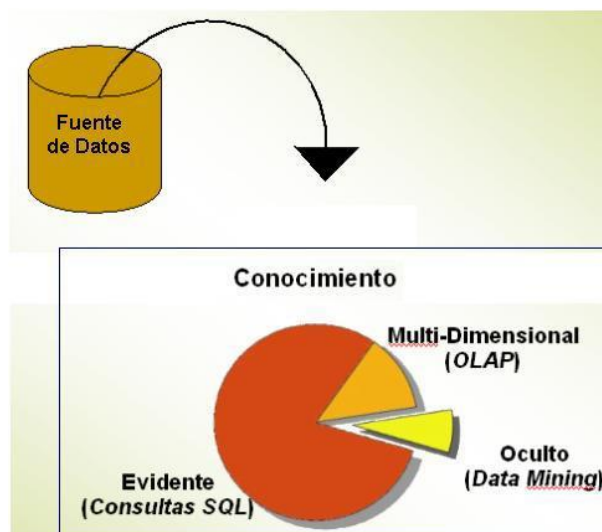
## **2.8. Data Mining**

Minería de datos no es un tema nuevo, a partir de los años sesenta los estadísticos manejaban términos como Data Fishing, Data Mining o Data Archaeology. A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro entre otros, empezaron a consolidar los términos de Minería de Datos y KDD. Y en la década de los noventa se afianzó el tema de Minería de Datos Inteligente y Almacén de Datos y Apoyo a las Decisiones.

### **2.8.1. Tipo de conocimiento**

De acuerdo con Pautsch (2009), en su trabajo de investigación menciona que, "Antes de comenzar a describir el tema de minería de datos, debemos comprender e identificar los tipos de conocimientos que podemos extraer de una Base de datos" (p.37). Podríamos clasificar a este conocimiento según las siguientes categorías:

- Evidente: esta información se puede obtener de la Base de Datos a través de consultas SQL.
- Multidimensional: modela una tabla con n atributos como un espacio de n dimensiones, lo que nos permite detectar varias situaciones difíciles de observar. Este tipo de análisis se logra utilizando herramientas OLAP.
- Oculto: es la información no evidente, desconocida hasta el momento, pero potencialmente útil, que puede obtenerse a través de técnicas de minería de datos. Esta información tiene un gran valor, ya que hasta el momento no se conocía, y descubrirla permite tener una nueva visión del problema y de su solución, ver en la Figura 32.

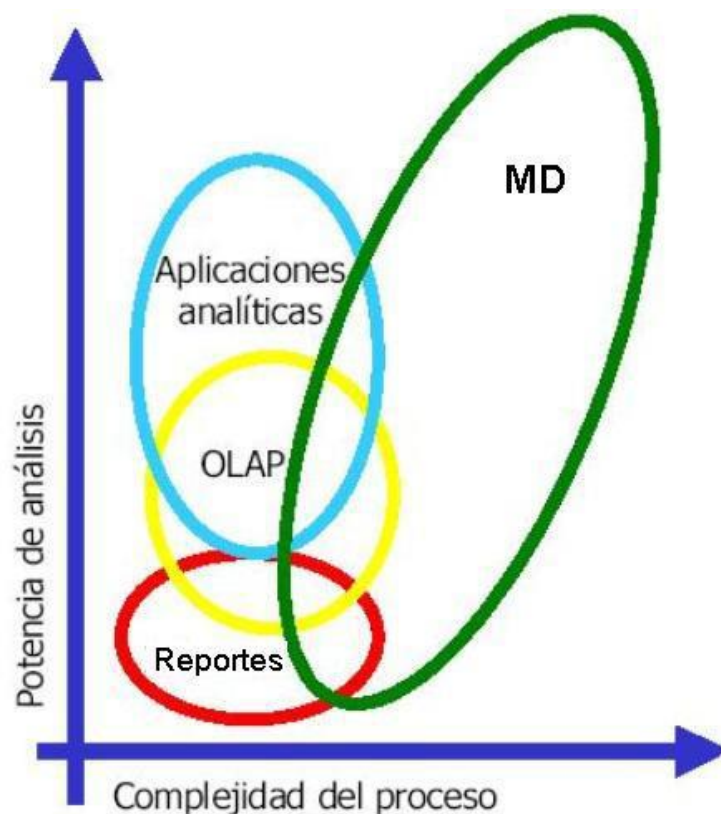


**Figura 32.** Tipo de conocimientos.

Fuente: Pautsch, J. (2009). "Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación". (p. 37). Obtenida de <http://exa.unne.edu.ar/informatica/SO/TFAGermanPAUTSCHFinal.pdf>

Pautsch (2009), estima que un 80% de la información contenida en una base de datos corresponde al conocimiento evidente (fácilmente recuperable). El otro 20% requiere de técnicas más complejas para su obtención ver en la figura 33.

Puede que esta cifra parezca despreciable, pero la información oculta en ese pequeño porcentaje puede ser de vital importancia para el éxito de la empresa u organización.



**Figura 33.** Relación entre potencial de análisis y complejidad.

Fuente: Pautsch, J. (2009). "Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación". (p. 38). Obtenida de <http://exa.unne.edu.ar/informatica/SO/TFAGermanPAUTSCHFinal.pdf>

### **2.8.2. Definición de Data Mining**

Según Jiawei, Micheline, & Jian (2012), comenta que a Data Mining (DM), es conocida también como: “la minería del conocimiento a partir de datos” (p. 6).

De acuerdo con Trujillo (2013) y Frawley, Piatetsky-Shapiro, & Matheus (1992), definen que, “La minería de datos es la extracción de información interesante (no-trivial, implícita, previamente desconocida y potencialmente útil), relaciones y/o patrones a partir de los datos en grandes bases de datos” (p. 82).

Reinosa, Maldonado, & Muñoz (2012), define que, “La minería de datos es un conjunto de técnicas que se utilizan para obtener información implícita de las grandes bases de datos. Es una tecnología poderosa que permite a las organizaciones la recolección de información desconocida de sus propias bases” (p. 20). Afirma también que, la minería de datos no es un modelo de almacenamiento: es una tecnología conformada por un conjunto de técnicas algorítmicas que se vinculan directamente con la programación avanzada.

Según Palma, Palma, & Pérez, (2009), describe que, “Minería de datos es conjunto de metodologías estadísticas y computacionales que, junto a un enfoque desde las ciencias de la conducta, permite el análisis de

datos y la elaboración de modelos matemáticos descriptivos y predictivos de la conducta del consumidor” (p. 43).

La minería de datos puede describirse como “análisis de datos exploratorio” (Date, 2001, 722).

Witten, Frank, & Hall (2011), define a DM como: “el proceso de descubrimiento de patrones en los datos” (p. 5).

DM, se debe entender como: “la exploración y análisis de grandes cantidades de los datos con el fin de descubrir patrones y reglas significativa” (Berry & Linoff, 2004, 7). Este proceso de DM, se le conoce como el descubrimiento de conocimiento o descubrimiento de conocimiento en bases de datos (KDD).

DM, considerado también como: “la extracción de información oculta y predecible de grandes bases de datos, es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información” (Presser, 2009, 4). Las herramientas de DM predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información.

### **2.8.3. Técnicas de Data Mining**

Los algoritmos de minería de datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o descriptivos o de descubrimiento del conocimiento.

#### **2.8.3.1. Predictivas**

Esta técnica está encargada de estimar o predecir los valores de salida.

##### **2.8.3.1.1. Regresión**

Es una técnica estadística, se usa para predecir valores y simular la relación existente entre dos o más variables, a continuación, describimos los tipos de regresión:

Regresión Lineal: Beltrán (2018), lo describe como: “Una técnica básica de Data Mining, se implementa identificando una variable dependiente ( $y$ ) y todas las variables independientes ( $X_1, X_2 \dots$ )” (p.53).

Regresión Logística: Beltrán (2018), lo describe como: “Una técnica básica de Data Mining, pueden trabajar con variables discretas y también requiere que todas las variables sean lineales” (p.53).

##### **2.8.3.1.2. Series temporales**

Beltrán (2018), lo menciona como: “El estudio de una variable a través del tiempo para, a partir de ese conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones”.

#### 2.8.3.1.3. Redes bayesianas

De acuerdo con Beltrán (2018), describe que: “Las redes bayesianas son una alternativa para minería de datos” (p. 55), la cual tiene varias ventajas:

- Permiten aprender sobre relaciones de dependencia y causalidad.
- Permiten combinar conocimiento con datos.
- Evitan el sobre-ajuste de los datos.
- Pueden manejar bases de datos incompletos.

#### 2.8.3.1.4. Algoritmos genéticos

Los algoritmos genéticos “representan la modelización matemática de como los cromosomas en un marco evolucionista alcanzan la estructura y composición más óptima en aras de la supervivencia. Entendiendo la evolución como un proceso de búsqueda y optimización de la adaptación de las especies que se plasma en mutaciones y cambios en los genes o cromosomas” (Beltrán Martínez, 2018, 54).

#### 2.8.3.1.5. Redes neuronales

Las redes neuronales son una recreación artificial del comportamiento de nuestro cerebro aprendiendo de la experiencia y el pasado y aplicando tal conocimiento a la resolución de problemas nuevos. Además, “son capaces de detectar y aprender patrones y características dentro de los datos” (Beltrán Martínez, 2018, 53).

#### 2.8.3.1.6. Clasificación ad hoc árboles de decisión

Es un conjunto de condiciones organizadas en forma jerárquica. Representada “mediante la ruptura y subdivisión sistemática de la información contenida en el conjunto de datos” (Beltrán Martínez, 2018, 53).

#### **2.8.3.2. Descriptivas**

Esta técnica está encargada de describir un conjunto de datos.

##### 2.8.3.2.1. Clustering

Conocida también como agrupamiento o segmentación, “permite la identificación de tipologías o grupos donde los elementos guardan similitud entre sí y diferencias con aquellos de otros grupos” (Beltrán Martínez, 2018, 48). Para alcanzar las distintas tipologías o grupos existentes en una base de datos, estas técnicas requieren, como entrada, información sobre el colectivo a segmentar.

##### 2.8.3.2.2. Asociación

Este tipo de técnica establece “las posibles relaciones o correlaciones entre distintas acciones o sucesos aparentemente independientes, pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros” (Beltrán Martínez, 2018, 48). Normalmente este tipo de técnicas se fundamenta en técnicas estadísticas como los análisis de correlación y de variación.

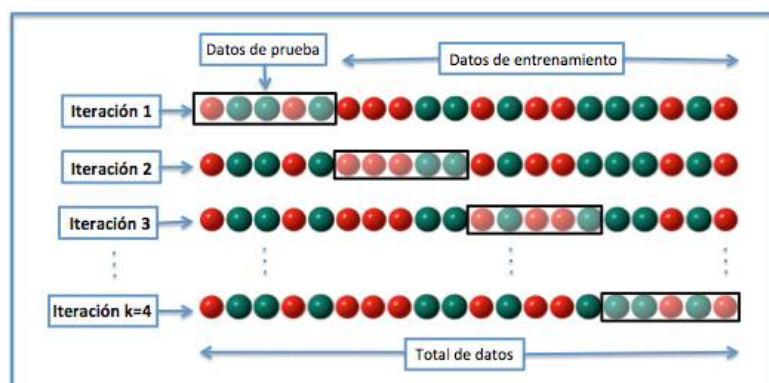
### 2.8.3.2.3. Secuenciamiento

Esta técnica permite identificar como, “en el tiempo, la ocurrencia de una acción desencadena otras posteriormente. Es muy similar a la anteriormente analizada si bien, en este caso, el tiempo es una variable crítica e imprescindible a introducir en la información a analizar” (Beltrán Martínez, 2018, 48).

## 2.8.4. Técnicas de validación de algoritmos computacionales de DM

### 2.8.4.1. Validación cruzada de k iteraciones

En la validación cruzada de K iteraciones los datos de muestra se dividen en K subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Como se puede visualizar en la siguiente figura 34.



**Figura 34.** Validación cruzada de k iteraciones.

Fuente: Obtenida de [https://es.wikipedia.org/wiki/Validaci%C3%B3n\\_cruzada](https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada)

El error de la validación cruzada de k iteraciones se calcula como la media aritmética de los errores de cada iteración para obtener un único resultado, y está dado por la siguiente formula:

$$E = \frac{1}{K} \sum_{i=1}^k E_i \quad (1)$$

Donde:

$E_i$  denota el error en la iteración i-ésima.

**K**: Número de grupos

#### 2.8.4.2. Matriz de confusión

La matriz de confusión de la información de los reales y predichos se interrelaciona a través de un sistema de clasificación. Como se puede ver en la siguiente figura 35.

|      |          | Predicho |          |
|------|----------|----------|----------|
|      |          | Negativo | Positivo |
| Real | Negativo | <b>a</b> | <b>b</b> |
|      | Positivo | <b>c</b> | <b>d</b> |

**Figura 35.** Matriz de confusión.

Fuente: Propia.

Las entradas en la matriz de confusión tienen el siguiente significado en el contexto de nuestro estudio:

- **a** es el número de predicciones correctas de que una instancia es negativa.

- **b** es el número de predicciones incorrectas de que una instancia es positiva.
- **c** es el número de predicciones incorrectas que una instancia es negativa.
- **d** es el número de predicciones correctas de que una instancia es positiva.

Se definen fórmulas para la matriz de confusión como se detalla a continuación:

$$AC = \frac{a + d}{a + b + c + d} \quad (2)$$

La precisión global (AC) es la proporción del número total de predicciones que fueron correctas.

$$TE = \frac{b + c}{a + b + c + d} \quad (3)$$

La tasa de error global (TE) es la proporción del número total de predicciones que fueron incorrectas.

$$TP = \frac{d}{c + d} \quad (4)$$

La tasa positiva verdadera (TP) es la proporción de casos positivos que se identificaron correctamente.

$$TN = \frac{a}{a + b} \quad (5)$$

La tasa negativa verdadera (TN) es la proporción de casos negativos que se clasificaron correctamente.

$$PF = \frac{b}{a + b} \quad (6)$$

La tasa de falsos positivos (PF) es la proporción de casos negativos que se clasificaron incorrectamente como positivos.

$$FN = \frac{c}{c + d} \quad (7)$$

La tasa de falsos negativos (FN) es la proporción de casos positivos que se clasificaron incorrectamente como negativos.

$$P = \frac{d}{b + d} \quad (8)$$

La precisión (P) es la proporción de casos positivos pronosticados que fueron correctos.

#### **2.8.4.3. La curva característica operativa del receptor**

La curva Receiver Operating Characteristic (ROC) es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación.

### **2.9. Definición de términos**

#### **2.9.1. Definiciones generales**

- UNJBG: Universidad Nacional Jorge Basadre Grohmann.
- VIAC: Vicerrectorado Académico.
- DASA: Dirección Académica de Actividades y Servicios Académicos.
- URC: Unidad de Registro Central.

- OGPL: Oficina General de Planificación.
- SYSA: Sistema de Registros Académicos.
- REAC: Base de Datos de Registro Académicos.
- BUN: Base de Datos de Bienestar Universitarios.
- GyT: Base de Datos de Grados y Títulos.
- DW: Data Warehouse.
- DM: Data Mining.

### **2.9.2. Calidad**

Según Hoyer (2001), la definición de calidad se divide en dos niveles.

El nivel uno de la calidad significa trabajar en las especificaciones, mientras que el nivel dos significa satisfacer al cliente.

La Real Academia española (2014) lo define como: “Propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor”.

### **2.9.3. Metodología**

Según Sierra Guzmán (2014), lo describe como: “El estudio o tratado de los métodos”. Es la rama de lógica que se encarga del estudio de los diferentes métodos para llegar al conocimiento crítico y reflexivo que permita la fundamentación de la ciencia.

#### **2.9.4. Método**

Según Sierra Guzmán (2014), lo describe como: “Serie o conjunto de pasos ordenados y sistematizados que tienen como fin llegar a la obtención del conocimiento”.

#### **2.9.5. Técnica**

Según Sierra Guzmán (2014), lo describe como: “Conjunto de habilidades, reglas y operaciones para el manejo de los instrumentos que auxilian al individuo en la aplicación de métodos”.

#### **2.9.6. Proceso**

Según García, Quispe, & Luis (2003) indican que según la NTP-ISO 9000:2001 Sistemas de gestión de la calidad - Fundamentos y Vocabulario, un proceso se define como "conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman elementos de entrada en resultados, y en el caso de un producto la misma norma lo define como resultado de un conjunto de actividades mutuamente relacionadas o que interactúan, las cuales transforman entradas en salidas. Los Insumos y productos terminados pueden ser tangibles e intangibles” (p. 89-94).

Según Perugachi (2004) define los procesos como: “actividades claves que se requieren para manejar y/o dirigir una organización; acción que describe un conjunto de pasos a seguirse para lograr una determinada

acción; conjunto de actividades que convierten insumos (inputs) en productos (outputs) de mayor valor para el cliente” (p. 18-19).

Los procesos tienen una jerarquía: macro-procesos, procesos, sub-procesos, actividades y tareas, cada uno de los cuales puede ser diferenciado de acuerdo al tamaño, tanto los macro procesos y subprocesos están compuestos por actividades, cada actividad consta de un determinado número de tareas.

Características de un proceso:

- Son independientes de la localidad o del grupo de personas que lo ejecutan.
- Se describe “que” se hace, no “como” se hace
- Son sinónimos con actividades o tareas.
- Se descomponen en una jerarquía, donde los componentes del proceso están contruidos por más elementos del mismo.

### **2.9.7. Calidad de software**

Pressman (2010) lo define como: “Proceso eficaz de software que se aplica de manera que crea un producto útil que proporciona valor medible a quienes lo producen y a quienes lo utilizan” (p.340).

### **2.9.8. Sistemas informáticos**

La Real Academia (2014) define el vocablo Sistema como un "conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objetivo".

Senn (2000) menciona que "un sistema es un conjunto de componentes que interaccionan entre sí para lograr un objetivo común" (p.19).

Pressman (2002) indica que un sistema basado en computadora es "un conjunto u ordenación de elementos organizados para llevar a cabo algún método, procedimiento o control mediante el procesamiento de información" (p.140).

Sobre la base de la noción de sistemas, se considera que un sistema de computación e información es el conjunto de componentes físicos (hardware), lógicos (software), de comunicación (redes de cualquier tipo o tipo Internet) y humanos que actuando en forma sistémica permite el tratamiento de la información.

### **2.9.9. Análisis y diseño de sistemas**

El análisis y diseño de sistemas permite examinar la situación de una empresa con el propósito de mejorarla con métodos y procedimientos más adecuados.

Senn (2000) describe: “El análisis de sistemas es el proceso de clasificación e interpretación de hechos, diagnóstico de problemas y empleo de la información para recomendar mejoras al sistema” (p.12). Éste es el trabajo del Analista de Sistemas.

El analista, debe comprender cómo trabaja el sistema actual y, de manera más específica, cuál es el flujo de información en todo el sistema.

El diseño de sistemas, es el proceso de planificar, reemplazar o complementar un sistema organizacional existente. Pero antes de ello, es necesario comprender, en su totalidad, el viejo sistema y determinar la mejor forma en que se pueden, si es posible, utilizar las computadoras para hacer la operación más eficiente.

#### **2.9.10. Análisis predictivo**

Siegel (2013), describe al análisis predictivo como: “la tecnología que aprende de la experiencia para predecir el futuro comportamiento de individuos para tomar mejores decisiones” (p.11).

El análisis predictivo es parte de la minería de datos, que consiste extraer la información existente y utilizarlo para la predicción de tendencias o patrones de comportamiento.

## **CAPÍTULO III**

### **MARCO METODOLÓGICO**

#### **3.1. Hipótesis**

Para una investigación, que tiene un alcance descriptivo, no se formula hipótesis. (Hernández, Fernández y Baptista, 2014, p.104).

#### **3.2. Variables**

##### **3.2.1. Identificación de variables**

**Variable 1.** Análisis predictivo de la deserción estudiantil.

##### **3.2.2. Definición de variables**

###### **Análisis predictivo de la deserción estudiantil.**

El análisis predictivo es una tarea de la minería de datos que tiene relación con la predicción de las probabilidades y tendencias futuras. En cambio, la deserción estudiantil es el abandono prematuro de estudios antes de alcanzar el título o grado académico.

De esta manera lo definimos como el estudio detallado de la probabilidad de riesgo que tiene los estudiantes al abandonar sus estudios prematuramente en el tiempo.

### 3.2.3. Operacionalización de variables

Tabla 1 Variable 1

| Variable  | Factores    | Dimensiones                   | Indicadores  |
|---|-------------|-------------------------------|--|
| Análisis predictivo de la deserción estudiantil | Predictivos | <b>Individuales</b>           | Edad   |
|   |             |                               | Estado civil                                       |
|   |             |                               | Sexo   |
|   |             |                               | Peso   |
|   |             |                               | Talla  |
|   |             |                               | Imc  |
|   |             |                               | Lugar de Residencia                                |
|   |             |                               | Lugar de Nacimiento                                |
|   |             | <b>Académicos</b>             | Tipo de Ingreso                                    |
|   |             |                               | Centro de Estudios                                 |
|   |             |                               | Tipo de Centro Estudios                            |
|   |             |                               | Programas Académicos                               |
|   |             |                               | Nota del Promedio Final                            |
|   |             |                               | Total de asignaturas desaprobadas                  |
|   |             |                               | Asignaturas desaprobadas mas del 50 % del semestre |
|   |             |                               | Preparación de Estudio                             |
|   |             |                               | Grado Educativo                                    |
|   |             |                               | Periodo lectivo de matrícula                       |
|   |             |                               | Periodo lectivo de ingreso                         |
|   |             |                               | <b>Socioeconómicos</b>                             |
|   |             | Nivel educativo del madre     |  |
|   |             | Ambito de vivienda            |  |
|   |             | Tenencia de vivienda          |  |
|   |             | Material de construcción      |  |
|   |             | Servicio de vivienda          |  |
|   |             | Uso del medio de comunicación |  |
|   |             | Sostiene                      |  |
|   |             | Situación del padre           |  |
|   |             | Riesgo de Salud familiar      |  |
|   |             | Carga familiar                |  |
| Ingreso mensual familiar                        |             |                               |  |
| Nivel de categorización                         |             |                               |  |
| Tipo de categorización                          |             |                               |  |
| Puntaje de ficha social                         |             |                               |  |

Fuente: Elaboración Propia.

### **3.3. Diseño de la investigación**

#### **3.3.1. Diseño experimental ó no experimental**

El presente estudio de acuerdo con lo que expresa Hernández, Fernandez & Baptista (1997), es no experimental porque la investigación se centra en el análisis de la variable de estudio; sin manipular deliberadamente ninguna variable, lo que se hace es observar el fenómeno tal como se da en su contexto natural para después analizarlos. Asimismo, es transaccional descriptivo, porque permite conocer la variable, a través de una exploración inicial en un momento específico las variables.

#### **3.3.2. Población y muestra**

La población comprende todos los registros de la base de datos de REAC, BUN y GyT del periodo 2012 hasta el 2018, cuyas cifras son: 11 058 registros de la base de datos de REAC, 10 632 registros de la base de datos de BUN y finalmente de 1 588 registros de la base de datos de GyT.

La muestra comprende todos los registros que contengan información sobre los estudiantes vinculada con los datos académicos, socioeconómicos y individuales.

#### **3.3.3. Técnicas e instrumentos para la recolección de datos**

En el presente estudio se utilizó las siguientes técnicas para la investigación en el campo como:

- Técnicas de recojo de información tales como la memoria de gestión de la UNJBG, Boletín estadístico de la UNJBG, reportes estadísticos por periodo de DASA y DBUN.
- Técnicas de recolección de datos, se recabó a través de entrevistas, análisis documental y la observación del objeto de estudio.

Asimismo, el instrumento a utilizarse es la ficha de registro de datos como se detalla en el anexo 8.

### 3.3.4. Procesamiento y análisis de datos

Para el procesamiento de la información se trabajó con la metodología Ralph Kimball para la construcción de un DW como se detalla en el anexo 2 y para la predicción de los datos se trabajó con la metodología Crisp-dm como se describe en el anexo 3. A continuación en la tabla 2 detallamos las etapas del procesamiento de información.

**Tabla 2** Etapas del procesamiento de la información

| Etapa | Descripción   |
|-------|---|
| 1     | Revisión y análisis de la información recolectada. <ul style="list-style-type: none"> <li>✓ Para la construcción del Data Warehouse se escogió la metodología de Ralph Kimball.</li> <li>✓ Para el desarrollo de técnicas de minería de datos se escogió la metodología de Crisp-dm.</li> </ul> |

| <b>Etapa</b> | <b>Descripción</b>   |
|--------------|--|
| <b>2</b>     | <p>Análisis de la base de datos de REAC, BUN y GyT del periodo 2012 hasta el 2018.</p> <ul style="list-style-type: none"> <li>✓ Cantidad de 11058 registros de la base de datos de REAC.</li> <li>✓ Cantidad de 10632 registros de la base de datos de BUN.</li> <li>✓ Cantidad de 1588 registros de la base de datos de GyT.</li> </ul> |
| <b>3</b>     | <p>Desarrollo del paquete para la extracción, transformación y carga de datos.</p> <ul style="list-style-type: none"> <li>✓ Con la herramienta de Visual studio 2017 se procedió a desarrollar el proceso ETL a los modelos dimensionales del DW.</li> </ul>   |
| <b>4</b>     | <p>Construcción de los modelos dimensionales.</p> <ul style="list-style-type: none"> <li>✓ Se realizó bajo el modelamiento de tipo estrella.</li> </ul>  |
| <b>5</b>     | <p>Construcción de estructuras de modelos de minería de datos.</p> <ul style="list-style-type: none"> <li>✓ Se eligió tres técnicas de minería de datos los cuales son: arboles de decisión, redes bayesianas y regresión logística para los datos predichos.</li> </ul>   |
| <b>6</b>     | <p>Elaboración de informes estándares y aplicaciones analíticas.</p> <ul style="list-style-type: none"> <li>✓ Con la herramienta Power BI, se elaboró informes estándares y reportes analíticos.</li> </ul>  |
| <b>7</b>     | <p>Interpretación de los resultados.</p> <ul style="list-style-type: none"> <li>✓ Con la herramienta Power BI, se creó las interfaces para la interpretación de resultados.</li> </ul>   |

Fuente: Elaboración Propia.

Para el análisis de los datos, se realizó usando las herramientas del servicio de análisis de SQL Server 2017 y para las interfaces de visualización, análisis ha sido desarrollada usando los softwares de Visual Studio 2017 y Power BI.

## CAPÍTULO IV

### RESULTADOS

#### 4.1. Resultados de la metodología de Ralph kimball en la construcción del DW

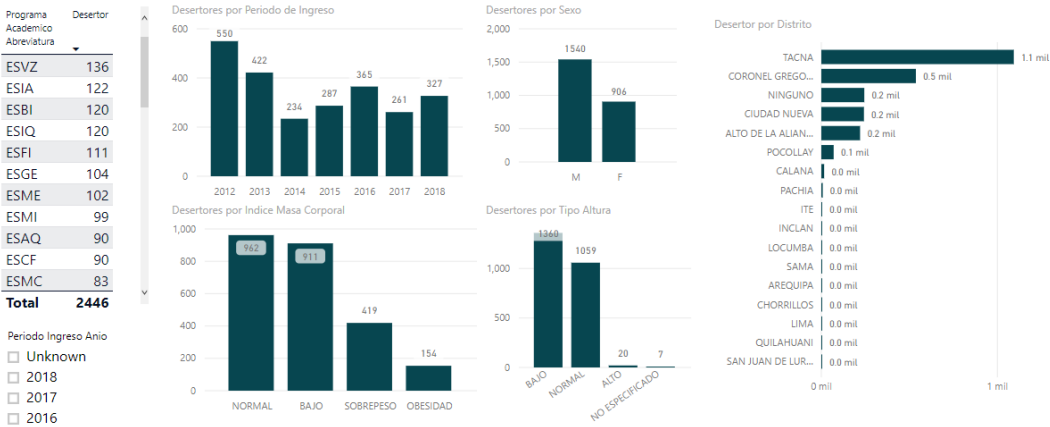
**Tabla 3** Factores de deserción estudiantil

| Factores               | Indicadores  |
|------------------------|--|
| <b>Individuales</b>    | Edad, Estado civil, Sexo, Talla, Imc, Lugar de Residencia, Lugar de Nacimiento.  |
| <b>Académicos</b>      | Tipo de Ingreso, Tipo de Centro Estudios, Centro de Estudios, Programas Académicos, Preparación de Estudio, Grado Educativo, Periodo lectivo de ingreso.   |
| <b>Socioeconómicos</b> | Nivel educativo del padre, Nivel educativo de la madre, Ambito de vivienda, Tenencia de vivienda, Material de construcción, Servicio de vivienda, Uso del medio de comunicación, Sostiene, Situación del padre, Riesgo de Salud familiar, Carga familiar, Nivel de categorización, Tipo de categorización. |

Fuente: Elaboración Propia.

Para todas las figuras de los factores de deserción se está visualizando las 10 primeras carreras profesionales.

#### 4.1.1. Factores individuales

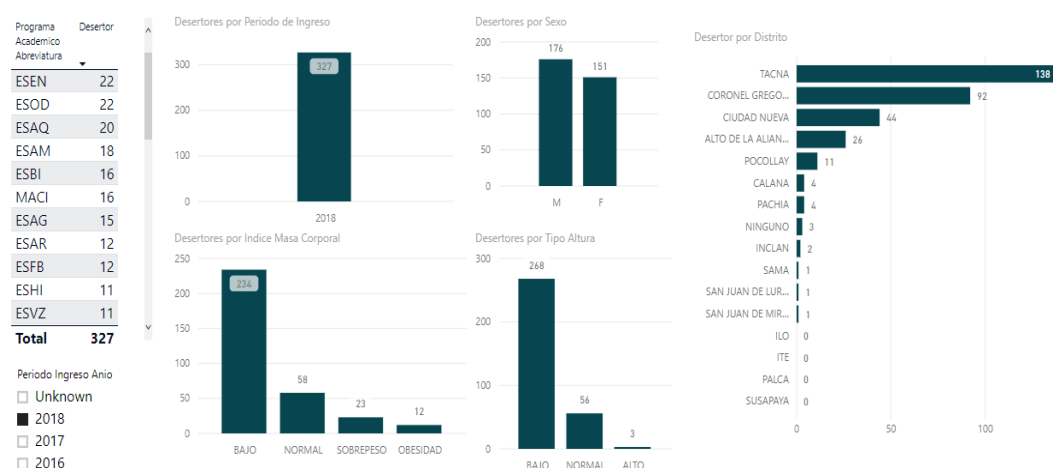


**Figura 36.** Factores individuales del periodo 2012 hasta 2018.

Fuente: Propia.

## Interpretación

De la figura 36, se puede apreciar la existencia de 2446 desertores desde el periodo 2012 hasta 2018. Dentro de eso el periodo de ingreso con mayor deserción fue el 2012 con un 22,49%, ahora en el indicador sexo se tiene que los hombres son los que más desertan con un índice de 62,96 %, del otro lado el indicador lugar de residencia describe que la ciudad de Tacna tiene más desertores con índice de 44,69 %, ahora los estudiantes que más desertan son de baja estatura con un índice de 55,60% y finalmente existe un sector con un índice de masa corporal bajo con una representación del índice del 37,24 %.



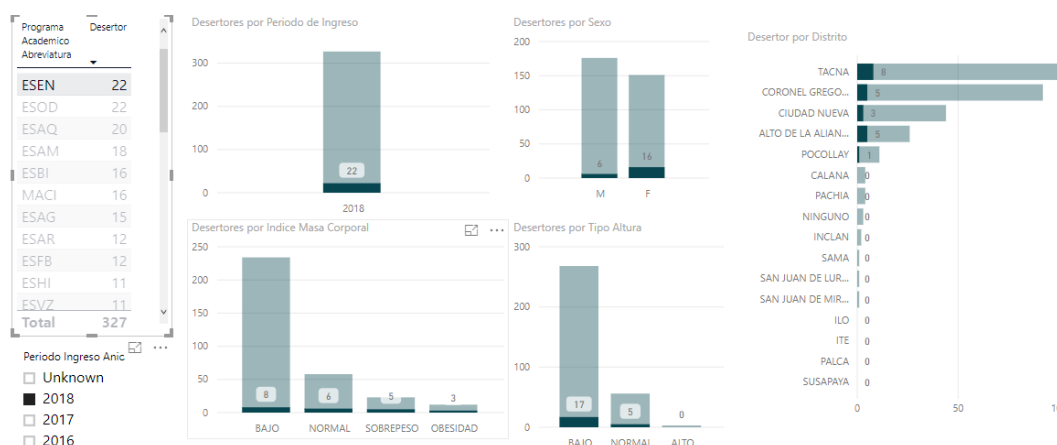
**Figura 37.** Factores individuales por periodo.

Fuente: Propia.

## Interpretación

De la figura 37, se puede apreciar la existencia de 327 desertores del periodo 2018. Ahora en el indicador sexo se tiene que los hombres son los que más desertan con un índice de 53,82 %, del otro lado el indicador

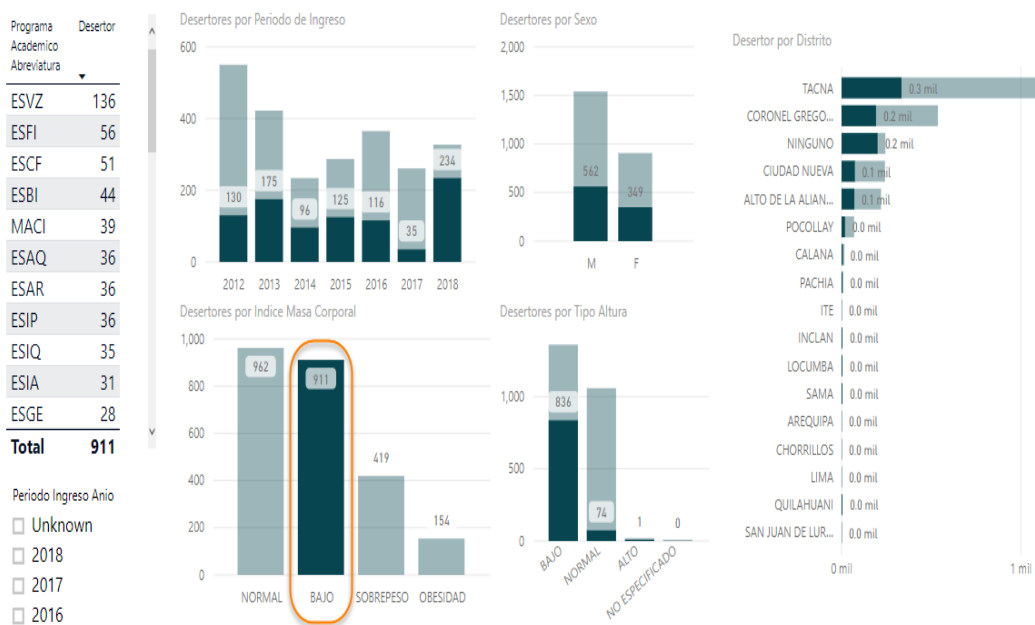
lugar de residencia describe que la ciudad de Tacna tiene más desertores con índice de 42,20 %, ahora los estudiantes que más desertan son de baja estatura con un índice de 81,96% y finalmente existe un sector con un índice de masa corporal bajo con una representación del índice del 71,56 %.



**Figura 38.** Factores individuales por programa académico.  
Fuente: Propia.

## Interpretación

De la figura 38, se puede apreciar la existencia de 22 desertores de la carrera profesional de enfermería y del periodo 2018. Ahora en el indicador sexo se tiene que las mujeres son los que más desertan con un índice de 72,72 %, del otro lado el indicador lugar de residencia describe que la ciudad de Tacna tiene más desertores con índice de 36,36 %, ahora los estudiantes que más desertan son de baja estatura con un índice de 77,27% y finalmente existe un sector con un índice de masa corporal bajo con una representación del índice del 36,36 %.



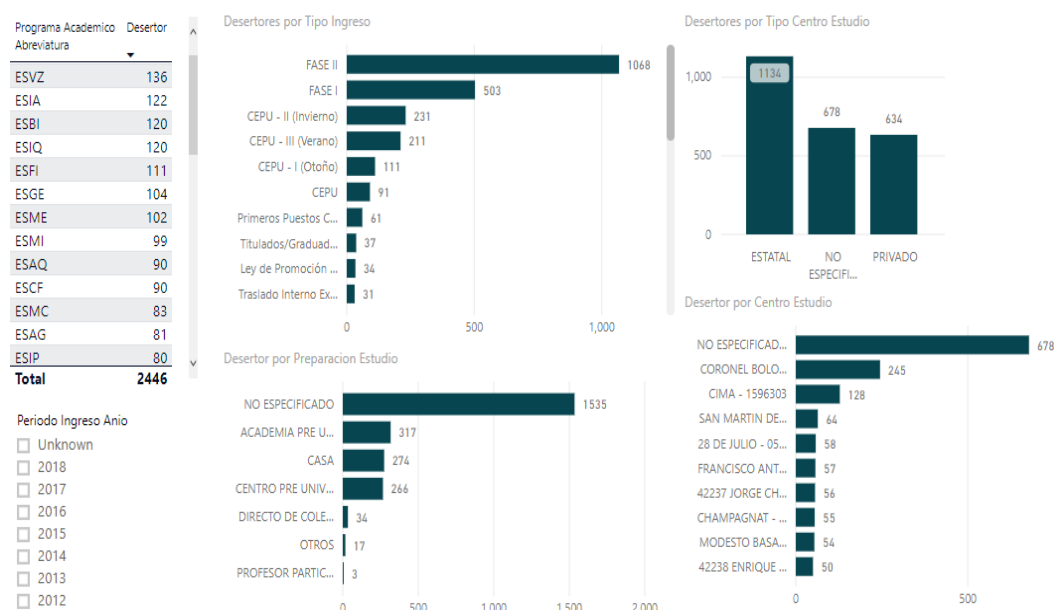
**Figura 39.** Factores individuales por índice de masa corporal.

Fuente: Propia.

### Interpretación

De la figura 39, se puede apreciar el indicador más predominante del factor individual es el índice de masa corporal que refleja la existencia de 911 estudiantes desertores con una condición bajo desde el periodo 2012 hasta 2018. Dentro de eso el periodo de ingreso con mayor deserción es el 2018 con un índice de 25,69 %, a su vez el género masculino alcanza un índice de deserción de 61,69 %, de otro lado el lugar de residencia que tiene más desertores es la ciudad de Tacna alcanzado un índice de 36,66 %, los estudiantes de baja estatura poseen un mayor índice deserción de 91,77 % y finalmente la carrera profesional de medicina veterinaria tienen el mayor índice deserción del 14,93 %.

## 4.1.2. Factores académicos

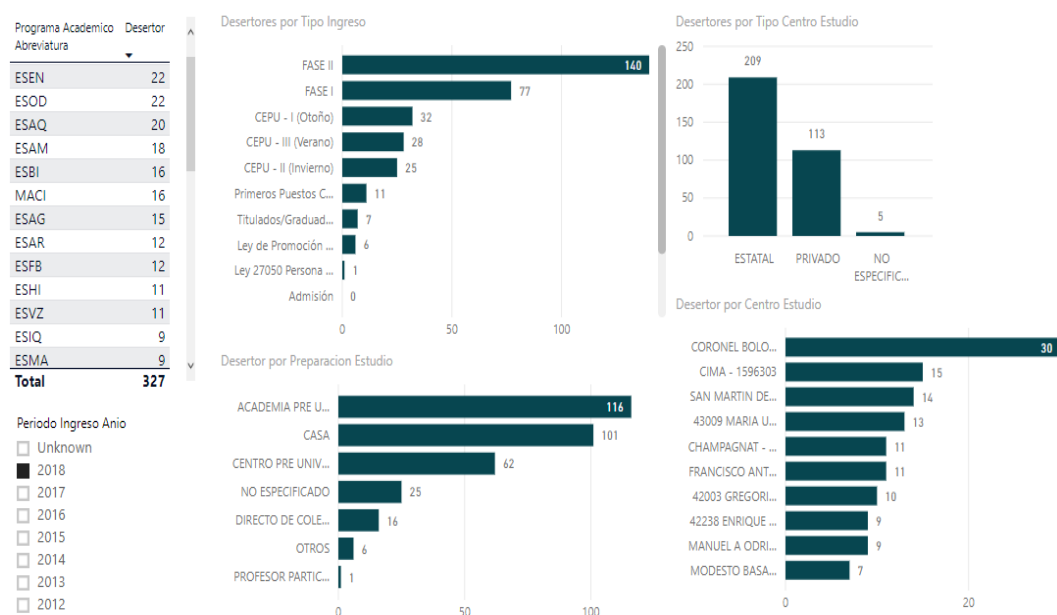


**Figura 40.** Factores académicos del periodo 2012 hasta 2018.

Fuente: Propia.

### Interpretación

De la figura 40, se puede apreciar la existencia de 2446 desertores desde el periodo 2012 hasta 2018. Dentro de eso el tipo de ingreso con mayor deserción es la Fase II con índice de 43,66%, ahora en el indicador preparación de estudio se observa que la preparación de estudiantes en las academias pre universitaria generan deserción a un índice de 12,96 %, del otro lado el indicador tipo de centro de estudio describe que los colegios estatales poseen más desertores con índice de 46,36 % y finalmente los estudiantes que provienen del colegio coronel Bolognesi de Tacna son los que más desertan con un índice de 10,02%.

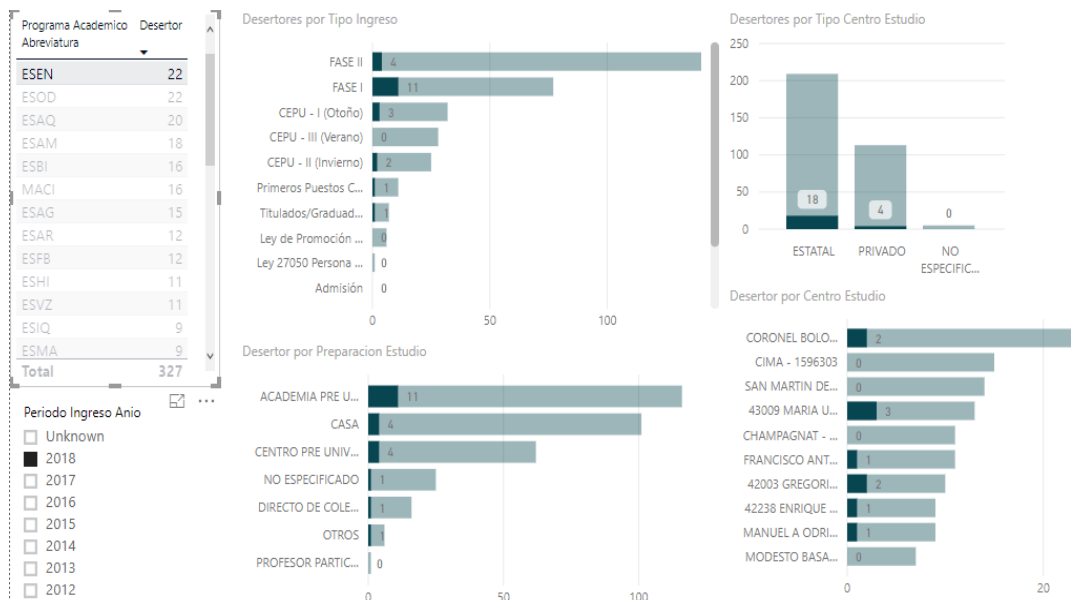


**Figura 41.** Factores académicos por periodo.

Fuente: Propia.

### Interpretación

De la figura 41, se puede apreciar la existencia de 327 desertores del periodo 2018. Dentro de eso el tipo de ingreso con mayor deserción es la Fase II con índice de 42,81%, ahora en el indicador preparación de estudio se observa que la preparación de estudiantes en las academias pre universitaria generan deserción a un índice de 35,47 %, del otro lado el indicador tipo de centro de estudio describe que los colegios estatales poseen más desertores con índice de 63,91 % y finalmente los estudiantes que provienen del colegio coronel Bolognesi de Tacna son los que más desertan con un índice de 9,17%.

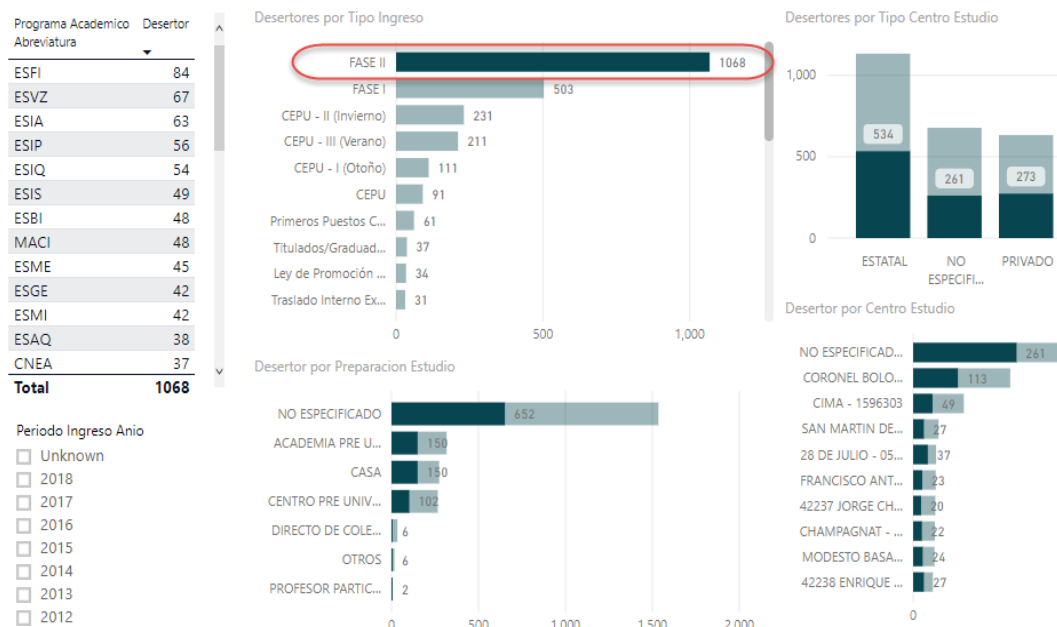


**Figura 42.** Factores académicos por programa académico.

Fuente: Propia.

### Interpretación

De la figura 42, se puede apreciar la existencia de 22 desertores de la carrera profesional de enfermería y del periodo 2018. Dentro de eso el tipo de ingreso con mayor deserción es la Fase I con índice de 50,00%, ahora en el indicador preparación de estudio se observa que la preparación de estudiantes en las academias pre universitaria generan deserción a un índice de 50,50 %, del otro lado el indicador tipo de centro de estudio describe que los colegios estatales poseen más desertores con índice de 81,81 % y finalmente los estudiantes que provienen del colegio Maria Ugarteche de Maclean de Tacna son los que más desertan con un índice de 13,63 %.



**Figura 43.** Factores académicos por tipo de ingreso.

Fuente: Propia.

### Interpretación

De la figura 43, se puede apreciar el indicador más predominante del factor académico es el tipo de ingreso que refleja la existencia de 1068 estudiantes desertores de la fase II desde el periodo 2012 hasta 2018. Dentro de eso las academias pre universitarias generan una deserción a un índice de 14,04 %, por otro lado, los colegios estatales son los que poseen más desertores con índice de 50,00 % de esta manera los estudiantes provenientes del colegio Coronel Bolognesi de Tacna son los que más desertan con un índice de 10,58 % y finalmente la carrera profesional de Física Aplicada es la que tienen el mayor índice deserción al 7,87 %.

### 4.1.3. Factores Socioeconómicos

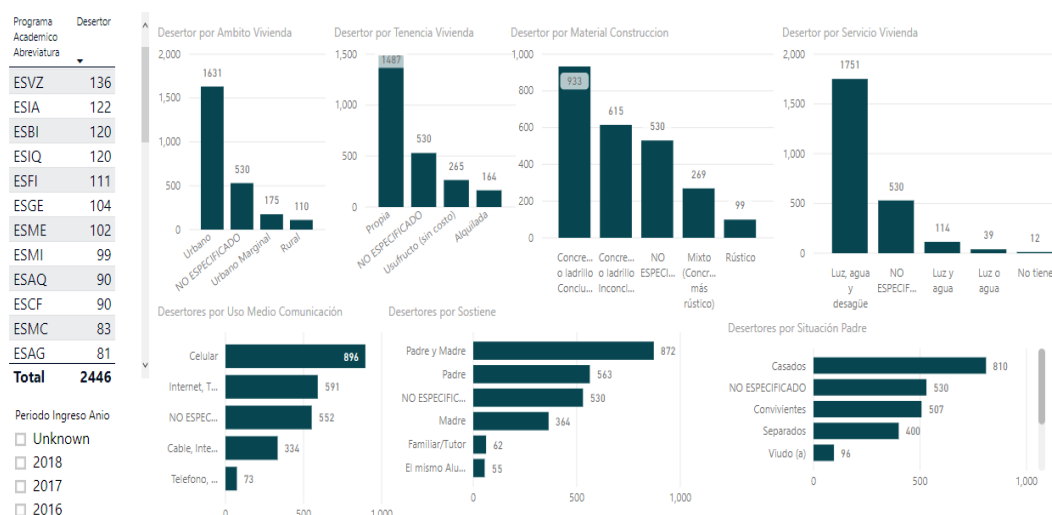
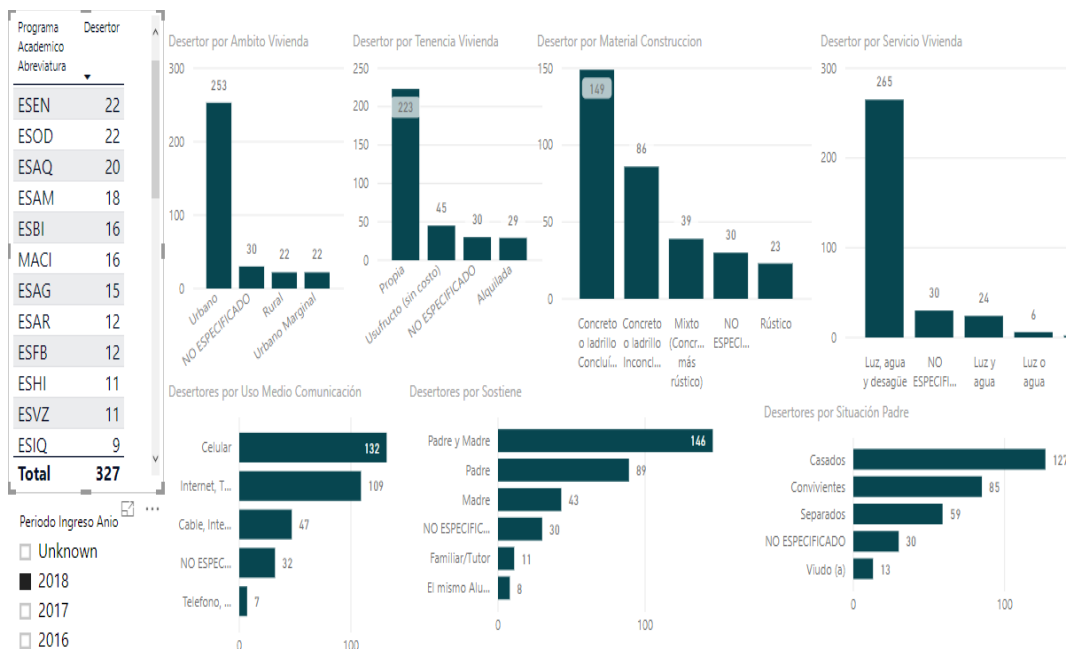


Figura 44. Factores socioeconómicos del periodo 2012 hasta 2018.

Fuente: Propia.

### Interpretación

De la figura 44, se puede apreciar la existencia de 2446 deserciones desde el periodo 2012 hasta 2018. Dentro de eso la zona urbana tiene la mayor deserción del 2018 con un índice de 66,68 %, ahora los que tienen vivienda propia son los que más desertan con un índice de 60,79 %, por otro lado, los estudiantes que tienen los servicios básicos son los que también desertan con un índice de 71,59 %, se hace evidente la deserción en la familia de padres casados con un índice de 33,12 %, luego también en los padres y madres que sostienen el hogar se acrecienta el índice deserción a un 35,65 % y finalmente los estudiante que poseen algún artículo tecnológico como el celular se registra la mayor índice de deserción con un 36,63 %.

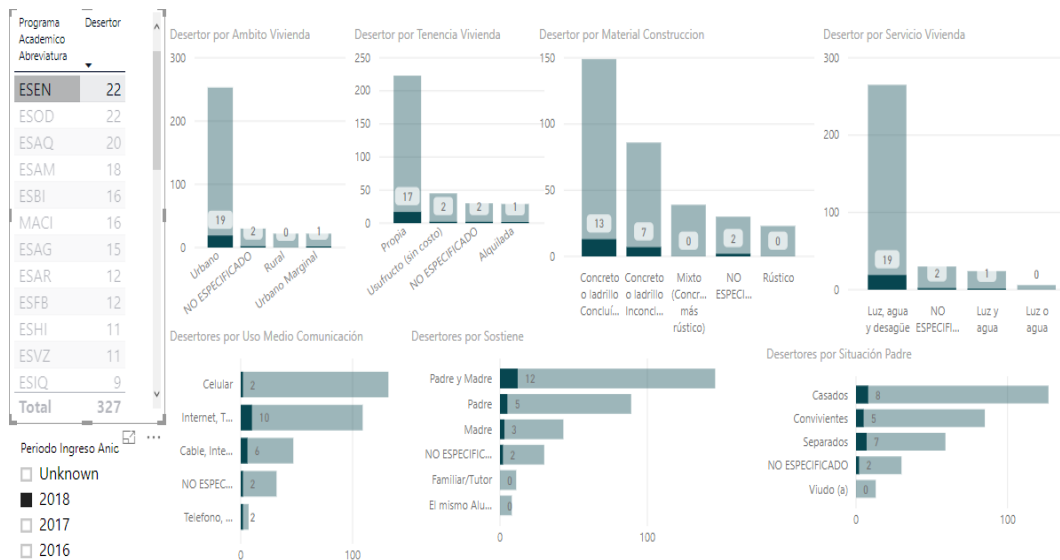


**Figura 45.** Factores socioeconómicos por periodo.

Fuente: Propia.

### Interpretación

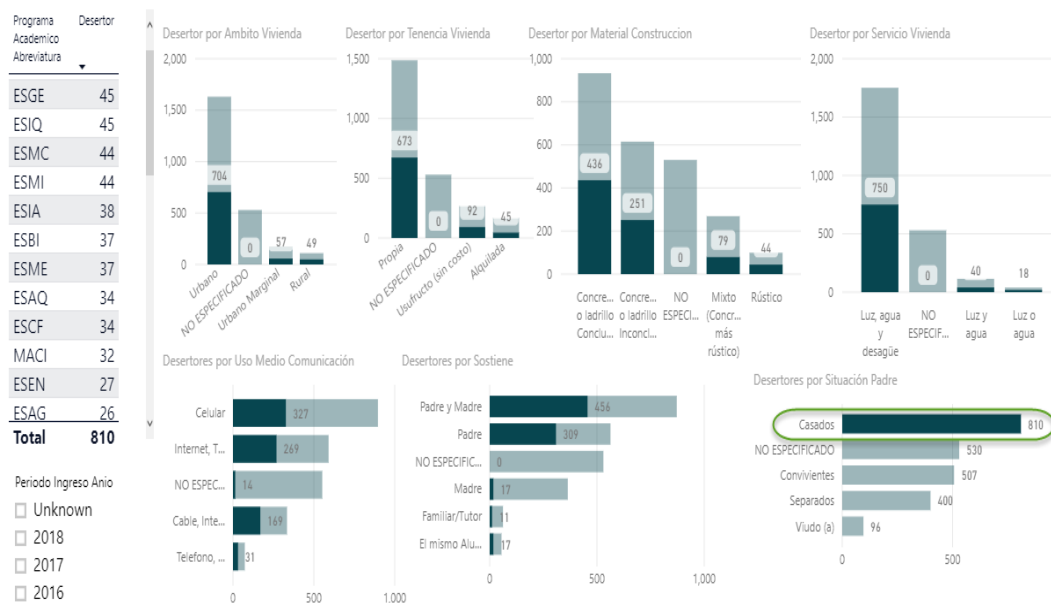
De la figura 45, se puede apreciar la existencia de 327 desertores del periodo 2018. Dentro de eso la zona urbana tiene la mayor deserción del 2018 con un índice de 77,37 %, ahora los que tienen vivienda propia son los que más desertan con un índice de 68,20 %, por otro lado, los estudiantes que tienen los servicios básicos son los que también desertan con un índice de 81,04 %, se hace evidente la deserción en la familia de padres casados con un índice de 38,84 %, luego también en los padres y madres que sostienen el hogar se acrecienta el índice deserción a un 44,65 % y finalmente los estudiante que poseen algún artículo tecnológico como el celular se registra la mayor índice de deserción con un 40,37 %.



**Figura 46.** Factores socioeconómicos por programa académico.  
Fuente: Propia.

### Interpretación

De la figura 46, se puede apreciar la existencia de 22 desertores de la carrera profesional de enfermería y del periodo 2018. Dentro de eso la zona urbana tiene la mayor deserción del 2018 con un índice de 86,36 %, ahora los que tienen vivienda propia son los que más desertan con un índice de 77,27 %, por otro lado, los estudiantes que tienen los servicios básicos son los que también desertan con un índice de 86,36 %, se hace evidente la deserción en la familia de padres casados con un índice de 36,86 %, luego también en los padres y madres que sostienen el hogar se acrecienta el índice deserción a un 54,54 % y finalmente los estudiante que poseen algún artículo tecnológico como el celular y acceso al internet se registra la mayor índice de deserción con un 45,45 %.

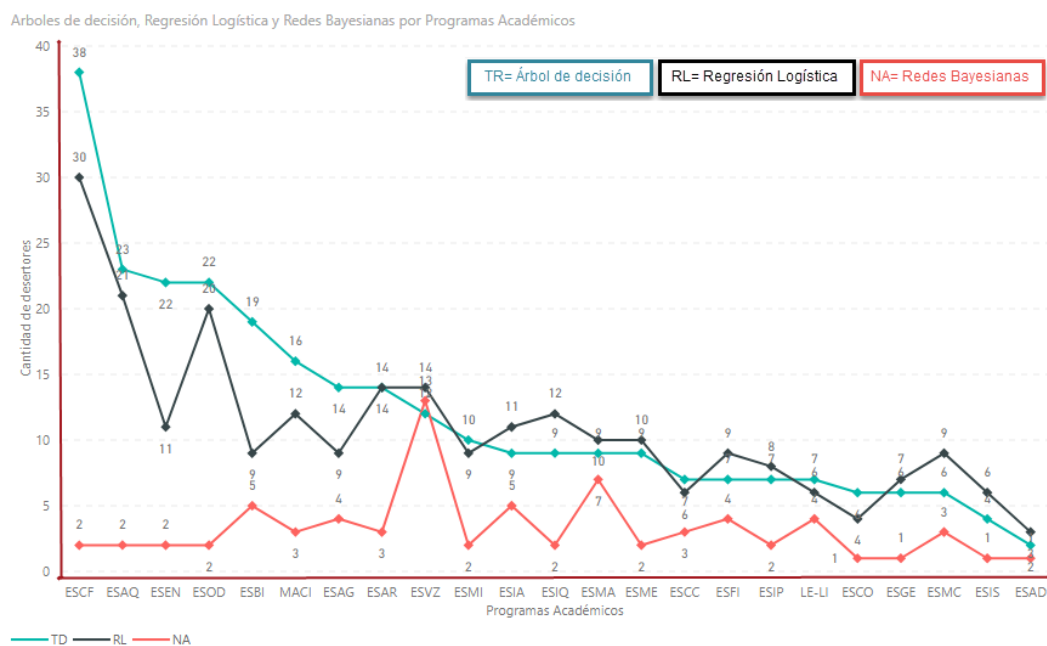


**Figura 47.** Factores socioeconómicos por situación del padre.  
Fuente: Propia.

### Interpretación

De la figura 47, se puede apreciar el indicador más predominante del factor socioeconómico es la situación del padre que refleja la existencia de 810 estudiantes desertores dentro de un núcleo familiar de padres casados. A pesar de que los padres sostienen el hogar siguen generando un índice elevado de deserción de 56,30 %, teniendo los servicios básicos, vivienda propia en una zona urbana todavía persiste la deserción con índice elevados y finalmente unas de las carreras profesionales es la Medicina Veterinaria y Zootecnia que mayor índice de deserción posee al del 5,56 %.

## 4.2. Resultados de la metodología Crisp-dm del desarrollo del proyecto DM



**Figura 48.** Predicción de la deserción estudiantil del periodo 2018 bajo los modelos predictivos.

Fuente: Propia.

### Interpretación

De la figura 48, se observa el comportamiento de las líneas de los tres modelos predictivos arboles de decisión, regresión logística y redes bayesianas. De acuerdo a la fase de evaluación del modelo del anexo 2 el más preciso, mínima tasa de error y mejor entendimiento de resultados es el modelo de árboles de decisión por lo tanto se elige a este modelo para las predicciones de los desertores.

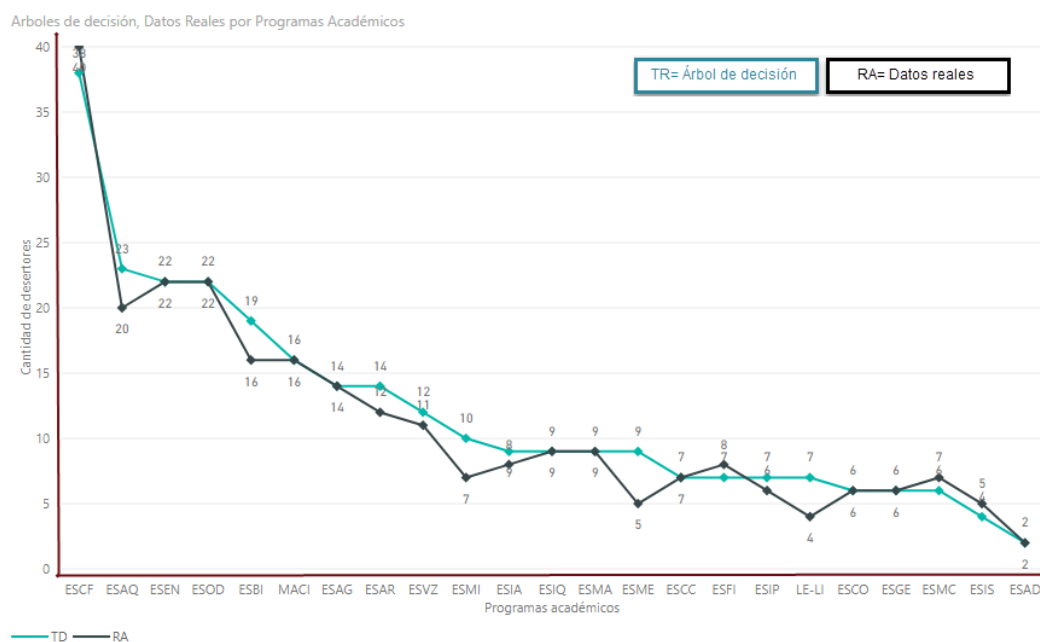
**Tabla 4** Valores predichos de los modelos predictivos del periodo 2018

| N° | Programas Académicos   | Arbol<br>Decisión | Regresión<br>Logística | Redes<br>Bayesianas |
|----|--|-------------------|------------------------|---------------------|
| 1  | Carrera Profesional de Ciencias Contables y Financieras      | 38                | 30                     | 2                   |
| 2  | Carrera Profesional de Arquitectura                          | 23                | 21                     | 2                   |
| 3  | Carrera Profesional de Odontología                           | 22                | 20                     | 2                   |
| 4  | Carrera Profesional de Enfermería                            | 22                | 11                     | 2                   |
| 5  | Carrera Profesional de Biología - Microbiología              | 19                | 9                      | 5                   |
| 6  | CAR.PROF. Educación - Matemática, Computación e Informática  | 16                | 12                     | 3                   |
| 7  | Carrera Profesional de Artes                                 | 14                | 14                     | 3                   |
| 8  | Carrera Profesional de Agronomía                             | 14                | 9                      | 4                   |
| 9  | Carrera Profesional de Medicina Veterinaria y Zootecnia      | 12                | 14                     | 13                  |
| 10 | Carrera Profesional de Ingeniería de Minas                   | 10                | 9                      | 2                   |
| 11 | Carrera Profesional de Ingeniería Química                    | 9                 | 12                     | 2                   |
| 12 | Carrera Profesional de Ingeniería en Industrias Alimentarias | 9                 | 11                     | 5                   |
| 13 | Carrera Profesional de Matemática                            | 9                 | 10                     | 7                   |
| 14 | Carrera Profesional de Ingeniería Metalúrgica                | 9                 | 10                     | 2                   |
| 15 | Carrera Profesional de Física Aplicada                       | 7                 | 9                      | 4                   |
| 16 | Carrera Profesional de Ingeniería Pesquera                   | 7                 | 8                      | 2                   |
| 17 | CAR.PROF. Educación - Lengua, Literatura                     | 7                 | 6                      | 4                   |
| 18 | Carrera Profesional de Ciencias de la Comunicación           | 7                 | 6                      | 3                   |
| 19 | Carrera Profesional de Ingeniería Mecánica                   | 6                 | 9                      | 3                   |
| 20 | Carrera Profesional de Ingeniería Geológica - Geotecnia      | 6                 | 7                      | 1                   |
| 21 | Carrera Profesional de Ingeniería Comercial                  | 6                 | 4                      | 1                   |
| 22 | Carrera Profesional de Ingeniería en Informática y Sistemas  | 4                 | 6                      | 1                   |
| 23 | Carrera Profesional de Ciencias Administrativas              | 2                 | 3                      | 1                   |

Fuente: Elaboración Propia.

De la tabla 4, se observa los valores predichos por los tres modelos predictivos como los arboles de decisión, regresión logística y redes bayesianas agrupados por carreras profesionales.

### 4.3. Constratación de los datos reales con los predichos.



**Figura 49.** Constratación de los desertores 2018 con el modelo predictivo de árboles de decisión.

Fuente: Propia.

#### Interpretación

De la figura 49, se observa el comportamiento de las líneas del modelo predictivo árboles de decisión y los desertores reales del periodo 2018, en contraste con los datos reales existe una aproximación de los resultados predichos. Dado que este modelo garantiza las predicciones futuras de los nuevos posibles desertores.

**Tabla 5** Constratación de los desertores 2018 con el modelo predictivo de árboles de decisión.

| <b>N°</b> | <b>Programas Académicos</b>                                  | <b>Arbol Decisión</b> | <b>Datos Reales</b> |
|-----------|--|-----------------------|---------------------|
| 1         | Carrera Profesional de Ciencias Contables y Financieras      | 38                    | 40                  |
| 2         | Carrera Profesional de Arquitectura                          | 23                    | 20                  |
| 3         | Carrera Profesional de Odontología                           | 22                    | 22                  |
| 4         | Carrera Profesional de Enfermería                            | 22                    | 22                  |
| 5         | Carrera Profesional de Biología - Microbiología              | 19                    | 16                  |
| 6         | CAR.PROF. Educación - Matemática, Computación e Informática  | 16                    | 16                  |
| 7         | Carrera Profesional de Artes                                 | 14                    | 12                  |
| 8         | Carrera Profesional de Agronomía                             | 14                    | 14                  |
| 9         | Carrera Profesional de Medicina Veterinaria y Zootecnia      | 12                    | 11                  |
| 10        | Carrera Profesional de Ingeniería de Minas                   | 10                    | 7                   |
| 11        | Carrera Profesional de Ingeniería Química                    | 9                     | 9                   |
| 12        | Carrera Profesional de Ingeniería en Industrias Alimentarias | 9                     | 8                   |
| 13        | Carrera Profesional de Matemática                            | 9                     | 9                   |
| 14        | Carrera Profesional de Ingeniería Metalúrgica                | 9                     | 5                   |
| 15        | Carrera Profesional de Física Aplicada                       | 7                     | 8                   |
| 16        | Carrera Profesional de Ingeniería Pesquera                   | 7                     | 6                   |
| 17        | CAR.PROF. Educación - Lengua, Literatura                     | 7                     | 4                   |
| 18        | Carrera Profesional de Ciencias de la Comunicación           | 7                     | 7                   |
| 19        | Carrera Profesional de Ingeniería Mecánica                   | 6                     | 7                   |
| 20        | Carrera Profesional de Ingeniería Geológica - Geotecnia      | 6                     | 6                   |
| 21        | Carrera Profesional de Ingeniería Comercial                  | 6                     | 6                   |
| 22        | Carrera Profesional de Ingeniería en Informática y Sistemas  | 4                     | 5                   |
| 23        | Carrera Profesional de Ciencias Administrativas              | 2                     | 2                   |

Fuente: Elaboración Propia.

De la tabla 5, se observa los valores predichos del modelo de árboles de decisión y los datos reales de los desertores 2018, además agrupados por carreras profesionales.

## **CAPÍTULO V**

### **DISCUSIÓN**

La presente investigación, tuvo un alcance descriptivo partiendo con el objetivo de efectuar el análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna; mediante la investigación se logró construir el Data Warehouse utilizando la metodología Ralph Kimball y en la minería de datos se basó a la metodología de CRISP-DM donde también se aplicaron las técnicas de los árboles de decisión, regresión logística y redes bayesianas lo que permitió determinar que los indicadores más predominantes en la deserción estudiantil del periodo 2012 al 2018 de la Universidad Nacional Jorge Basadre Grohmann fueron el Índice de masa corporal (factores individuales), tipo de ingreso (factores académicos) y situación del padre (factores socioeconómicos).

Si bien es cierto, para la presente investigación no se planteó hipótesis por lo que cada accionar en la presente investigación fue dar una respuesta a cada uno de los objetivos planteados.

Otañe Bowen (2016), analiza la información socioeconómica, historial académico y concluye que el algoritmo que presento mejor desempeño en las predicciones de datos fueron los arboles de decisión en comparación con las redes bayesianas esto nos lleva afirmar que la técnica más apropiada para este estudio de esta investigación es el árbol de decisión.

Dentro del trabajo de investigación de Frutos Ortega (2017), determina que el indicador más predominante es la edad, estudiante de 21 años son los que más desertan en comparación a nuestro trabajo de investigación la edad no es un indicador predominante.

Por otro lado, los resultados obtenidos del estudio nos conllevan afirmar lo señalado por Fischer (2012) y Hernandez (2015) que es posible el análisis predictivo de los estudiantes a pesar que ambos establecen diferentes objetivo y diseño de investigación.

Finalmente, se considera que está investigación es importante porque deja planteado un modelo predictivo el cual parte de un análisis real mejorando a un simple modelo predictivo tradicional. Asimismo, esta investigación sirve como un aporte que permite contribuir a futuras investigaciones y nuevos métodos de abordaje en el tema de análisis predictivo de la deserción estudiantil.

## **CONCLUSIONES**

### **PRIMERA**

Se efectuó el análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna; tal como se puede observar en el anexo 2.

### **SEGUNDA**

Se construyó el Data Warehouse utilizando la metodología Ralph Kimball para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann; tal como se ha descrito en la sección 5.2 del anexo 2.

### **TERCERA**

Se determinó que los indicadores más predominantes en la deserción estudiantil del periodo 2012 al 2018 de la Universidad Nacional Jorge Basadre Grohmann fueron el Índice de masa corporal (factores individuales), tipo de ingreso (factores académicos) y situación del padre (factores socioeconómicos); tal como se puede observar en la figura 39; figura 43 y figura 47.

## **CUARTA**

Las técnicas idóneas aplicadas para la minería de datos que permitió el análisis predictivo de la deserción estudiantil del periodo 2012 al 2018 de la Universidad Nacional Jorge Basadre Grohmann fueron los arboles de decisión, regresión logística y redes bayesianas, tal como se puede observar en la sección 5.4.4 del anexo 2.

## **RECOMENDACIONES**

### **PRIMERA**

Las universidades deben darle importancia al tema de la deserción estudiantil proponiendo programas de retención estudiantil y esto a minore la tasa de deserción.

### **SEGUNDA**

La etapa de recolección de información es importante para esto debería revisar cada cierto periodo los formatos estandarizado de los factores individuales, académicos y socioeconómicos.

### **TERCERA**

La presente investigación puede servir como base o el inicio de otras investigaciones futuras.

## REFERENCIAS BIBLIOGRÁFICAS

- Beltrán Martínez, B. (2018). Minería de Datos. Recuperado el 5 de Noviembre de 2018, de <http://bbeltran.cs.buap.mx/NotasMD.pdf>
- Bernabeu, R. D. (2010). HEFESTO: Metodología para la Construcción de un Data Warehouse. Córdoba. Recuperado el 25 de Noviembre de 2015, de <http://www.businessintelligence.info/docs/hefesto-v2.pdf>
- Berry, M. J., & Linoff, G. S. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (Segunda ed.). Indianapolis-Indiana: Wiley Publishing.
- Bueno, E. (1998). El capital intangible como clave estratégica en la competencia actual. Recuperado el 25 de Noviembre de 2015, de <http://www.sedic.es/bueno.pdf>
- Castaño, A. d. (2001). Concepción y diseño de Bases de Datos. Madrid: Hispamerica.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (8 de Junio de 2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. Obtenido de <http://www.redalyc.org/pdf/1552/155217798002.pdf>
- Coll, C., & Martín, E. (1996). La evaluación de los aprendizajes en el marco de la reforma: una perspectiva de conjunto. Signos N°18.
- Coronel, C., Morris, S., & Rob, P. (2011). Bases de Datos, Diseño, Implementación y Administración (Novena ed.). Santa Fe: Cengage Learning Editores S.A.
- Curto, J. (2010). Introducción al Business Intelligence (Primera ed.). España: UOC.
- Date, C. J. (2001). Introducción a los Sistemas de Base de Datos (Septima ed.). México: Pearson Educación S.A.
- Davenport, T. H., & Prusak, L. (1998). Working Knowledge: How Organizations Manage What They Know. Boston: Harvard Business School Press.
- Elmasri, R., & Navathe, S. B. (2007). Fundamentos de Sistemas de Base de Datos (Quinta ed.). Madrid: Pearson Educación S.A.

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. doi:<http://dx.doi.org/10.1609/aimag.v17i3.1230>
- Fischer, E. (Mayo de 2012). Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios. (Tesis de Maestría). Recuperado el 25 de Noviembre de 2015, de [http://repositorio.uchile.cl/bitstream/handle/2250/111188/cf-fischer\\_ea.pdf?sequence=1&isAllowed=y](http://repositorio.uchile.cl/bitstream/handle/2250/111188/cf-fischer_ea.pdf?sequence=1&isAllowed=y)
- Florez, J. A. (2014). Data Warehouse y minería de datos como alternativas al análisis de datos forestales. INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS, POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA, Montecillo.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview, 13(3), 57. Recuperado el 25 de Noviembre de 2015, de <http://aaai.org/journals/aimagazine/article/viewFile/1011/929>
- Frutos Ortega, C. A. (15 de Enero de 2017). Análisis de la tasa de retención y su incidencia en la detección de patrones. Obtenido de [http://repositorio.uta.edu.ec/bitstream/123456789/25797/1/Tesis\\_%20t1266mbd.pdf](http://repositorio.uta.edu.ec/bitstream/123456789/25797/1/Tesis_%20t1266mbd.pdf)
- García, M., Quispe, C., & Luis, R. (Agosto de 2003). Mejora continua de la calidad de procesos. Recuperado el 2 de Septiembre de 2014, de <http://www.redalyc.org/pdf/816/81606112.pdf>
- Grupo Educación al Futuro. (10 de Mayo de 2013). Alta deserción de estudiantes en la universidad. Recuperado el 25 de Noviembre de 2015, de <http://educacionalfuturo.com/noticias/alta-desercion-de-estudiantes-en-la-universidad/>
- Hansen, G. W., & Hansen, J. V. (1997). Diseño y Administración de Base de Datos (Segunda ed.). Prentice Hall.
- Hernández, J. (2015). Modelo de minería de datos para identificación de patrones que influyen en el aprovechamiento académico. (Tesis de Maestría). Instituto Tecnológico de la Paz, División de Estudios de Posgrado e Investigación, La Paz. Recuperado el 25 de Noviembre de 2015, de <http://posgrado.itlp.edu.mx/uploads/55f7167f2302e.pdf>

- Hernández, J., Ramírez, J., & Ferri, C. (2004). *Introducción a la Minería de Datos*. Madrid: Person Prentice Hall.
- Hernández, R., Fernandez, C., & Baptista, P. (1997). *Metodología de la investigación científica*. México: McGraw Hill / Interamericana.
- Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación. *Revista Calidad en la Educación*. Recuperado el 25 de Noviembre de 2015, de [http://www.cned.cl/public/secciones/seccionpublicaciones/doc/35/cse\\_articulo141.pdf](http://www.cned.cl/public/secciones/seccionpublicaciones/doc/35/cse_articulo141.pdf)
- Horacio, S. (2002). *Sistemas De Información de La Era Digital*. Argentina: Fundación OSDE.
- Hoyer, R. (2001). *Calidad*. Quality Progress.
- Inmon, W. H. (2005). *Building the Data Warehouse* (Cuarta ed.). Indianapolis: Wiley Publishing.
- International Business Machines Corporation. (8 de Octubre de 2012). *Manual CRISP-DM de IBM SPSS Modeler*. Obtenido de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- Jiawei, H., Micheline, K., & Jian, P. (2012). *Data Mining Concepts and Techniques* (Tercera ed.). San Francisco-United States of America: Morgan Kaufmann-Elsevier.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, (Tercera ed.). Indianapolis: John Wiley & Sons, Inc.
- Krzysztof, J. C., Witold, P., Roman, W. S., & Lukasz, A. K. (2007). *Data Mining A Knowledge Discovery Approach*. New York: Springer Science+Business Media,.
- Kubski, M. (2004). *Minería de Datos*. Universidad Nacional del Nordeste, Corrientes. Recuperado el 2015 de Noviembre de 2015, de <http://exa.unne.edu.ar/informatica/SO/MineriaDatosKub.pdf>
- Leonard Brizuela , E. I., & Castro Blanco, Y. (2013). Metodologías para desarrollar Almacén de Datos. *Arquitectura e Ingeniería*, 7(3), 1-12. Obtenido de <http://www.redalyc.org/pdf/1939/193930080003.pdf>

- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. New York: Springer Science+Business Media. doi:10.1007/978-0-387-09823-4
- Mannino, M. V. (2007). *Administración de base de datos Diseño y desarrollo de aplicaciones* (Tercera ed.). México: Mc Graw Hill Interamericana Editores S.A. de C.V.
- Martinez, M. (2000). *Diseño de Base de Datos Relacionales*. Mexico.
- Mateus, J., Perilla, C., & Herrera, C. (2010). Factores presentes en la deserción universitaria en la facultad de psicología de la Universidad de San Buenaventura en el periodo comprendido entre 1998 - 2009. Bogotá. Obtenido de <http://www.redalyc.org/articulo.oa?id=297224114011>
- Ministerio de Educación Nacional de Chile. (25 de Noviembre de 2015). Servicio de Información de Educación Superior. Recuperado el 25 de Noviembre de 2015, de [http://www.mifuturo.cl/images/Estudios/Estudios\\_SIES\\_DIVE SUP/desercion\\_reingreso\\_a\\_educacin\\_superior.pdf](http://www.mifuturo.cl/images/Estudios/Estudios_SIES_DIVE_SUP/desercion_reingreso_a_educacin_superior.pdf)
- Ministerio de Educación Nacional de Colombia. (25 de Noviembre de 2015). Sistema de Información para la Prevención de la Deserción en Educación Superior. Recuperado el 25 de Noviembre de 2015, de [http://spadies.mineducacion.gov.co/spadies/consultas\\_predefinidas.html?2](http://spadies.mineducacion.gov.co/spadies/consultas_predefinidas.html?2)
- Moine, J. M. (20 de Agosto de 2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo. Recuperado el 13 de Noviembre de 2018, de [http://sedici.unlp.edu.ar/bitstream/handle/10915/29582/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/29582/Documento_completo.pdf?sequence=1)
- Mondy, R., & Noé, R. (1997). *Administración de Recursos Humanos* (6ta ed.). México: Prentice-Hall Hispanoamericana.
- Otañe Bowen, A. A. (2016). *Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería De Datos*. Obtenido de <http://bdigital.unal.edu.co/53635/1/alvaroagustino%C3%B1atebowen.2016.pdf>
- Palma, C., Palma, W., & Pérez, R. (2009). *Data mining: el arte de anticipar* (Primera ed.). Santiago: RIL editores.
- Pautsch, J. (2009). *Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación*. Universidad Nacional

de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, Corrientes. Recuperado el 25 de Noviembre de 2015, de <http://exa.unne.edu.ar/informatica/SO/TFAGermanPAUTSCHFinal.pdf>

Pérez, C. (4 de Agosto de 2015). Business Intelligence. México. Recuperado el 25 de Noviembre de 2015, de <http://vortech-it.com/assets/cap-ii-bill-inmon.pdf>

Perugachi, M. L. (2004). Optimización de Procesos la concesión de radiofrecuencias en el Ecuador. Ecuador: ABYAYALA.

Pons, O., Carrillo, S. A., & Marin, N. (2005). Introducción a las bases de datos: el modelo relacional. Madrid: Paraninfo S.A.

Presser, C. (2009). Data mining. Santa Fe: El Cid Editor. Obtenido de <http://site.ebrary.com/lib/bibliounjbgsp/detail.action?docID=10311813>

Pressman, R. S. (2002). Ingeniería de Software (Tercera ed.). México: Mc Graw – Hill.

Pressman, R. S. (2010). Ingeniería del Software. un enfoque práctico (Séptima ed.). New York: The McGraw-Hill.

Programa Nacional de Beca y Crédito Educativo. (Septiembre de 2014). Plan de Gestión Institucional del Programa Nacional de Becas y Crédito Educativo del Ministerio de Educación 2014 - 2018. Recuperado el 25 de Noviembre de 2015, de [http://www.pronabec.gob.pe/modPublicaciones/descarga/plan\\_de\\_gestion\\_institucional\\_del\\_pronabec\\_minedu\\_2014\\_2018.pdf](http://www.pronabec.gob.pe/modPublicaciones/descarga/plan_de_gestion_institucional_del_pronabec_minedu_2014_2018.pdf)

Real Academia, E. (15 de Diciembre de 2014). Definición de Calidad, Sistema. Obtenido de <http://lema.rae.es/drae/>

Reinosa, E. J., Maldonado, C. A., & Muñoz, R. (2012). Bases de datos (Primera ed.). México: Alfaomega Grupo Editor.

Ricardo, C. M. (2004). Base de datos. México: McGraw Hill Interamericana Editores S.A. de C.V.

Robechi, O. T., & Jácome, J. O. (2014). Propuesta de un modelo de gestión del riesgo aplicando lógica difusa para la universidad católica de Colombia en el proceso de la deserción estudiantil en la facultad de ingeniería. Bogotá. Recuperado el 25 de Noviembre de 2015, de <http://repository.ucatolica.edu.co/jspui/handle/10983/1733>

- Rodríguez, C. Y. (2012). Diseño de un modelo estocástico usando cadenas de markov para pronosticar la deserción académica de estudiantes de ingeniería. caso: escuela colombiana de ingeniería julio garavito. (Tesis de Maestría). Bogotá. Recuperado el 25 de Noviembre de 2015, de <http://repository.javeriana.edu.co/bitstream/10554/13688/1/RodriguezRiosClaudiaYadira2012.pdf>
- Ruiz, M. K. (2007). GUÍAS Y TEXTOS DE CÓMPUTO: DATA WAREHOUSE y minería de datos. Universidad Nacional Autónoma de México, Dirección General de Servicios de Cómputo Académico, México. Recuperado el 25 de Noviembre de 2015, de <http://es.scribd.com/doc/270535671/DATA-WAREHOUSE-y-Mineria-de-Datos#scribd>
- Salmador, M. P. (2005). Raíces epistemológicas del conocimiento organizativo: estudio de sus dimensiones. Universidad Autónoma de Madrid. Recuperado el 25 de Noviembre de 2015, de <http://eco.mdp.edu.ar/cendocu/repositorio/01047.pdf>
- Senn, J. A. (2000). Análisis y Diseño de Sistemas de Información (Segunda ed.). México: Mc Graw – Hill.
- Sesento, L. (1 de Septiembre de 2008). MODELO SISTÉMICO BASADO EN COMPETENCIAS. Morelia, Michoacán, Mexico: Centro de Investigación y Desarrollo del Estado de Mochoacán.
- Siegel, E. (2013). Predictive Analytics. United States of America: John Wiley & Sons, Inc., Hoboken, New Jersey.
- Sierra Guzmán, M. P. (20 de Diciembre de 2014). Conceptos generales. Obtenido de [http://www.uaeh.edu.mx/docencia/P\\_Presentaciones/prepa3/conceptos\\_generales\\_inv.pdf](http://www.uaeh.edu.mx/docencia/P_Presentaciones/prepa3/conceptos_generales_inv.pdf)
- Silberschatz, A. (2002). Fundamentos de Base de Datos (Cuarta ed.). México: McGraw – Hill.
- Sommerville, I. (2005). Ingeniería de software. España: Pearson Educación, S.A.
- Trujillo, J. C. (2001). El modelo GOLD: un modelo conceptual orientado a objetos para el diseño de aplicaciones OLAP. Universidad de Alicante, Lenguajes y Sistemas Informáticos, Alicante. Recuperado el 25 de Noviembre de 2015, de <http://rua.ua.es/dspace/bitstream/10045/4103/1/Trujillo%20Mondajar,%20Juan%20Carlos.pdf>

Trujillo, J. C. (2013). Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional. España: Editorial Club Universitario.

Universidad Abierta para Adultos. (2009). Deserción en las instituciones de educación superior a distancia en América Latina y el Caribe. Santiago, República Dominicana: Ediciones UAPA. Recuperado el 25 de Noviembre de 2015, de [http://virtualeduca.org/documentos/observatorio/oevalc\\_2009\\_\(desercion\).pdf](http://virtualeduca.org/documentos/observatorio/oevalc_2009_(desercion).pdf)

Witten, I., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (Tercera ed.). Burlington-Massachusetts: Elsevier.

## **ANEXOS**

## ANEXO 1: MATRIZ DE CONSISTENCIA

| ANÁLISIS PREDICTIVO DE LA DESERCIÓN ESTUDIANTIL UTILIZANDO DATA WAREHOUSE Y MINERÍA DE DATOS EN LA UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN - TACNA, 2012-2018.  |   |  |   |   |                               |   |
|---|---|--|---|---|-------------------------------|---|
| OPERACIONALIZACIÓN DE VARIABLES   |   |  |   |   |                               |   |
| PROBLEMA  | OBJETIVOS   | HIPÓTESIS  | Variable  | Definición  | Dimensiones                   | Indicadores   |
| <p><b>GENERAL:</b></p> <p>¿Es posible efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann - Tacna?</p>  | <p><b>GENERAL:</b></p> <p>Efectuar el análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna.</p>  | <p><b>GENERAL:</b></p> <p>Si es posible efectuar un análisis predictivo de la deserción estudiantil utilizando Data Warehouse y la minería de datos en la Universidad Nacional Jorge Basadre Grohmann – Tacna.</p> |   | <p>El análisis predictivo es una tarea de la minería de datos que tiene relación con la predicción de las probabilidades y tendencias futuras.</p>  | <p><b>Personales</b></p>      | <ul style="list-style-type: none"> <li>✓ Edad</li> <li>✓ Estado civil</li> <li>✓ Sexo</li> <li>✓ Índice de masa corporal</li> <li>✓ Lugar de Residencia</li> <li>✓ Lugar de Nacimiento</li> </ul>   |
| <p><b>ESPECÍFICOS:</b></p> <p>¿Es posible construir un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann?</p> <p>¿Cuáles son los indicadores que predominan la deserción estudiantil de la Universidad Jorge Nacional Basadre Grohmann?</p> <p>¿Cuáles son las técnicas de minería de datos idóneas para el análisis predictivo de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann?</p> | <p><b>ESPECÍFICOS:</b></p> <p>Construir un Data Warehouse para el análisis de datos de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.</p> <p>Determinar los indicadores predominantes de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.</p> <p>Determinar las técnicas de minería de datos idóneas para el análisis predictivo de la deserción estudiantil de la Universidad Nacional Jorge Basadre Grohmann.</p> |  | <p><b>Análisis predictivo de la deserción estudiantil</b></p> | <p>En cambio, la deserción estudiantil es el abandono prematuro de estudios antes de alcanzar el título o grado académico.</p> <p>De esta manera lo definimos como el estudio detallado de la probabilidad de riesgo que tiene los estudiantes al abandonar sus estudios prematuramente en el tiempo.</p> | <p><b>Académicos</b></p>      | <ul style="list-style-type: none"> <li>✓ Tipo de Ingreso</li> <li>✓ Tipo de Centro Estudios</li> <li>✓ Centro de Estudios</li> <li>✓ Programas Académicos</li> <li>✓ Preparación de Estudio</li> <li>✓ Grado Educativo</li> <li>✓ Periodo lectivo de ingreso</li> </ul>   |
|   |   |  |   |   | <p><b>Socioeconómicas</b></p> | <ul style="list-style-type: none"> <li>✓ Nivel educativo del padre</li> <li>✓ Nivel educativo de la madre</li> <li>✓ Ámbito de vivienda</li> <li>✓ Tenencia de vivienda</li> <li>✓ Material de construcción</li> <li>✓ Servicio de vivienda</li> <li>✓ Uso del medio de comunicación</li> <li>✓ Sostiene</li> <li>✓ Situación del padre</li> <li>✓ Riesgo de Salud familiar</li> <li>✓ Carga familiar</li> <li>✓ Nivel de categorización</li> <li>✓ Tipo de categorización</li> </ul> |

## **ANEXO 2: DESARROLLO DEL DW Y DM**

### **5.1. Selección de una metodología para la construcción del Data Warehouse**

Existe varias metodologías para la construcción de Data Warehouse como los describen los autores Leonard Brizuela & Castro Blanco (2013) en la publicación de su revista denominada “Metodologías para desarrollar almacén de datos” (p.1).

Metodología de Ralph Kimball: denominada también como ciclo de vida Dimensional del negocio, representada como un marco de trabajo que tiene etapas en todo el proceso de la construcción del DW.

Hefesto: es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Consta de cuatro fases: análisis de requerimientos, análisis de los OLTP, modelo lógico del almacén de datos y proceso ETL.

Data Warehouse Engineering Process: fue propuesta por SAS Institute como una metodología iterativa y basada en el desarrollo incremental del proyecto, pero presenta poca documentación, lo que dificulta su aplicación integral.

Para este trabajo de investigación se está escogiendo la metodología de Ralph Kimball, por las etapas y procesos a seguir dentro de su marco de trabajo establecido.

## **5.2. Metodología Ralph Kimball**

### **5.2.1. Planificación del proyecto**

#### **5.2.1.1. Objetivos del proyecto**

- Implementar una solución de DW para determinar indicadores predominantes de los factores de la deserción estudiantil de la UNJBG.

#### **5.2.1.2. Alcance del proyecto**

Consiste en desarrollar una solución de DW para la Dirección Académica de Actividades y Servicios Académicos (DASA) en los cuales nos permita encontrar a los desertores y así nos permita tomar buenas decisiones para esta problemática.

Para este proyecto se tomó la base de datos de Registro Académico, Bienestar Universitario y Grados y Títulos de la UNJBG – TACNA.

### 5.2.1.3. Cronograma de actividades

|  |          |              |              |
|--|----------|--------------|--------------|
| ▸ Metodología Ralph Kimball  | 139 days | Tue 1/05/18  | Fri 9/11/18  |
| ▸ Planificación del Proyecto   | 9 days   | Tue 1/05/18  | Fri 11/05/18 |
| Objetivos del Proyecto   | 4 days   | Tue 1/05/18  | Fri 4/05/18  |
| Alcance del proyecto   | 2 days   | Mon 7/05/18  | Tue 8/05/18  |
| Cronograma de Actividades  | 2 days   | Wed 9/05/18  | Thu 10/05/18 |
| Asignación de Recursos   | 1 day    | Fri 11/05/18 | Fri 11/05/18 |
| ▸ Definición de los Requerimientos del Negocio                               | 14 days? | Mon 14/05/18 | Thu 31/05/18 |
| Requerimientos del negocio   | 2 days   | Mon 14/05/18 | Tue 15/05/18 |
| Entrevistas  | 3 days?  | Wed 16/05/18 | Fri 18/05/18 |
| Base de datos transaccional  | 5 days?  | Mon 21/05/18 | Fri 25/05/18 |
| Diccionario de datos de la Base de datos transaccional                       | 4 days   | Mon 28/05/18 | Thu 31/05/18 |
| Diseño de la Arquitectura  | 3 days   | Fri 1/06/18  | Tue 5/06/18  |
| Instalación y Selección del Producto   | 3 days   | Wed 6/06/18  | Fri 8/06/18  |
| ▸ Modelamiento Dimensional   | 20 days  | Mon 11/06/18 | Fri 6/07/18  |
| Elegir el proceso de negocio   | 5 days   | Mon 11/06/18 | Fri 15/06/18 |
| Establecer el nivel de granularidad  | 3 days   | Mon 18/06/18 | Thu 20/09/18 |
| Elegir las dimensiones   | 2 days?  | Thu 21/06/18 | Fri 22/06/18 |
| Identificar medidas y las tablas de hecho                                    | 5 days   | Mon 25/06/18 | Fri 29/06/18 |
| Elegimos el esquema del modelo dimensional                                   | 5 days   | Mon 2/07/18  | Fri 6/07/18  |
| Diseño Físico  | 10 days  | Mon 9/07/18  | Fri 20/07/18 |
| Diseño e Implementación del subsistema de Extracción, Transformación y Carga | 7 days   | Mon 23/07/18 | Tue 31/07/18 |
| ▸ Especificación de aplicaciones de Business Intelligence                    | 19 days  | Wed 1/08/18  | Mon 27/08/18 |
| Informes estándar  | 8 days   | Wed 1/08/18  | Fri 10/08/18 |
| Aplicaciones analíticas  | 10 days  | Mon 13/08/18 | Fri 24/08/18 |
| Usuarios finales   | 1 day    | Mon 27/08/18 | Mon 27/08/18 |
| Desarrollo de Aplicación de Business Intelligence                            | 47 days  | Tue 28/08/18 | Wed 31/10/18 |
| Implementación   | 7 days   | Thu 1/11/18  | Fri 9/11/18  |

**Figura 50.** Cronograma de actividades.

Fuente: Elaboración Propia.

### 5.2.1.4. Asignación de recursos

En esta sección se especifica los Stakeholders y personal de proyecto de investigación.

**Tabla 6 Stakeholders**

| <b>Persona</b>                  | <b>Cargo</b>                         | <b>Función que desempeña</b>  |
|---------------------------------|--------------------------------------|---|
| <b>Dr. Jorge Lozano Cervera</b> | Vicerrector Académico                | - Supervisar el cumplimiento de actividades académica con la finalidad de garantizar de la misma y logro meta establecidas. |
| <b>Lic. Elba Plata Ticona</b>   | Asesora del Vicerrectorado Académico | - Asesora la política general de la formación académica de la institución.  |
| <b>Dr. Oscar Mamani Aguilar</b> | Director de DASA.                    | - Propone al Vicerrector Académico las políticas relacionadas con los servicios Académicos que brinda la universidad.       |
| <b>Sap. Alicia Montes Roque</b> | Jefa de Unidad de Registro Central   | - Supervisa la aplicación de la normas, reglamentos, directivas de la parte académica.                                      |

Fuente: Elaboración Propia.

**Tabla 7 Recursos humanos del proyecto**

| <b>Persona</b>                       | <b>Cargo</b>   | <b>Función que desempeña</b>  |
|--------------------------------------|--|---|
| <b>Ing. Nelson Mollo Condori</b>     | Administrador del Proyecto                               | - Encargado de supervisar, controlar las etapas programadas del proyecto.   |
| <b>Bach. Israel Chaparro Cruz</b>    | Analista de Sistemas                                     | - Desarrollar los documentos de análisis y especificación de requerimientos |
| <b>Bach. Airtonn Melendez Quispe</b> | Programador de Sistemas y Administrador de Base de Datos | - Desarrollar interfaces visuales para la presentación de datos.            |

| Persona                              | Cargo           | Función que desempeña   |
|--------------------------------------|-----------------|---|
|                                      |                 | - Diseñador del Modelo Dimensional del DW.                                    |
|                                      |                 | - Encargado de la preparación y carga de datos de distintas fuentes de datos. |
| <b>Bach. Silvana Cabana Yupanqui</b> | Analista Tester | - Es responsable por realizar el Control de Calidad del Producto de Software  |

Fuente: Elaboración Propia.

En esta sección especificamos los requerimientos de Software y Hardware.

**Tabla 8** Hardware requerido

| Equipos                | Características                                   | Costo                              |
|------------------------|---|------------------------------------|
| <b>01 SERVIDOR</b>     | Servidor HP ProLiant DL160 Gen9, Xeon E5-2609v4   | - Proporcionada por la Institución |
| <b>04 COMPUTADORAS</b> | HP PRODESK G4 I5-6500 8GB 1TB DW 4USB SFF W10-PRO | - Proporcionada por la Institución |

Fuente: Elaboración Propia.

**Tabla 9** Software requerido

| Software                         | Características  | Costo                              |
|----------------------------------|--|------------------------------------|
| <b>Microsoft SQL Server 2017</b> | Almacén de Datos Relacionales, Dimensionales.                | - Proporcionada por la Institución |
| <b>Power BI</b>                  | Permite modelar la información de los Modelos Dimensionales. | - Proporcionada por la Institución |

Fuente: Elaboración Propia.

## 5.2.2. Definición de los requerimientos del negocio

### 5.2.2.1. Requerimientos del negocio

Los requerimientos que mostraremos a continuación, están enfocados en la problemática de la deserción estudiantil de la UNJBG.

**Tabla 10** Listado de requerimiento para los ingresantes

| Código | Descripción   | Tipo      | Prioridad |
|--------|---|-----------|-----------|
| R01    | Cantidad de ingresantes que no se matricularon en el periodo actual.  | Funcional | Alta      |
| R02    | Cantidad de ingresantes que desaprobaron más del 50 % de las asignaturas matriculas del periodo actual.                                   | Funcional | Alta      |
| R03    | Cantidad de ingresantes que ingresaron con nota desaprobatoria y desaprobaron en el periodo actual.                                       | Funcional | Alta      |
| R04    | Cantidad de ingresantes que no se matricularon dentro de 2 años.  | Funcional | Alta      |
| R05    | Cantidad de Ingresantes que no matricularon dentro de 4 semestres consecutivos o alternados.  | Funcional | Alta      |
| R06    | Cantidad de ingresantes que ingresaron por las distintas modalidades que desaprobaron más del 70 % de las asignaturas cursadas.           | Funcional | Alta      |
| R07    | Cantidad de ingresantes que desaprobaron el 1 semestre por su situación vinculada a los datos académicos, socioeconómicos y individuales. | Funcional | Alta      |

Fuente: Elaboración Propia.

**Tabla 11** Listado de requerimiento para los matriculados

| <b>Código</b> | <b>Descripción</b>   | <b>Tipo</b> | <b>Prioridad</b> |
|---------------|--|-------------|------------------|
| <b>R08</b>    | Cantidad de matriculados que desaprobaron las asignaturas de tercera matricula.                            | Funcional   | Alta             |
| <b>R09</b>    | Cantidad de matriculados que desaprobaron las asignaturas de cuarta matricula.                             | Funcional   | Alta             |
| <b>R10</b>    | Cantidad de matriculados que no registran matriculas dentro de los 04 semestres consecutivos o alternados. | Funcional   | Alta             |
| <b>R11</b>    | Cantidad de matriculados que regresaron por amnistía general que desaprobaron el primer semestre.          | Funcional   | Alta             |
| <b>R12</b>    | Cantidad de matriculados que desaprobaron dentro de 4 semestres más del 70 % de las asignaturas cursadas.  | Funcional   | Alta             |
| <b>R13</b>    | Cantidad de matriculados que desaprobaron dentro de 2 semestres más del 90 % de las asignaturas cursadas.  | Funcional   | Alta             |

Fuente: Elaboración Propia.

**Tabla 12** Listado de requerimiento para los egresados

| <b>Código</b> | <b>Descripción</b>                                       | <b>Tipo</b> | <b>Prioridad</b> |
|---------------|--|-------------|------------------|
| <b>R14</b>    | Cantidad de egresados que culminaron su plan de estudios | Funcional   | Alta             |

---

|            |   |           |      |
|------------|---|-----------|------|
|            | pero no tramitaron su constancia de egresado a partir de dos años.  |           |      |
| <b>R15</b> | Cantidad de egresados que no tramitaron su bachillerato a partir de tres años de su trámite de egresado.          | Funcional | Alta |
| <b>R16</b> | Cantidad de bachilleres que no tramitaron su título profesional a partir de cinco años del trámite del bachiller. | Funcional | Alta |

---

Fuente: Elaboración Propia.

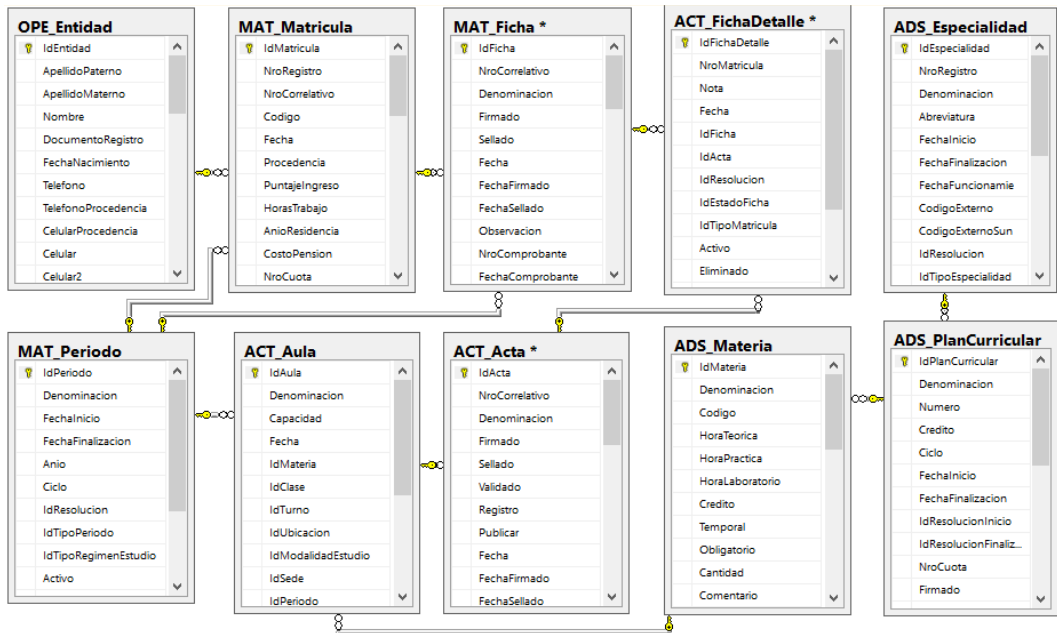
### **5.2.2.2. Entrevistas**

Las entrevistas se realizaron a las siguientes personas:

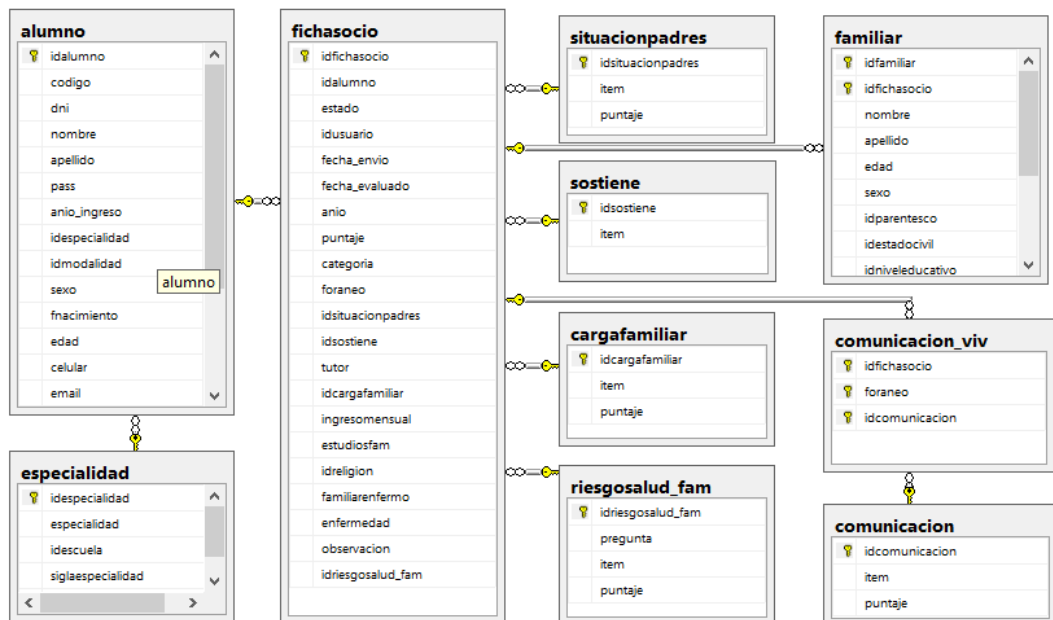
Dr. Oscar Mamani Aguilar, Director de DASA, SAP Alicia Angélica Montes Roque, Jefa de la Unidad de Registro Central (URC), Lic. Elba Virginia Plata Ticona, asesora del VIAC, Ing. Olga Veronica Huaman Guzman, Especialista administrativo de la Oficina General de Planificación (OGPL).

### **5.2.2.3. Base de datos transaccional**

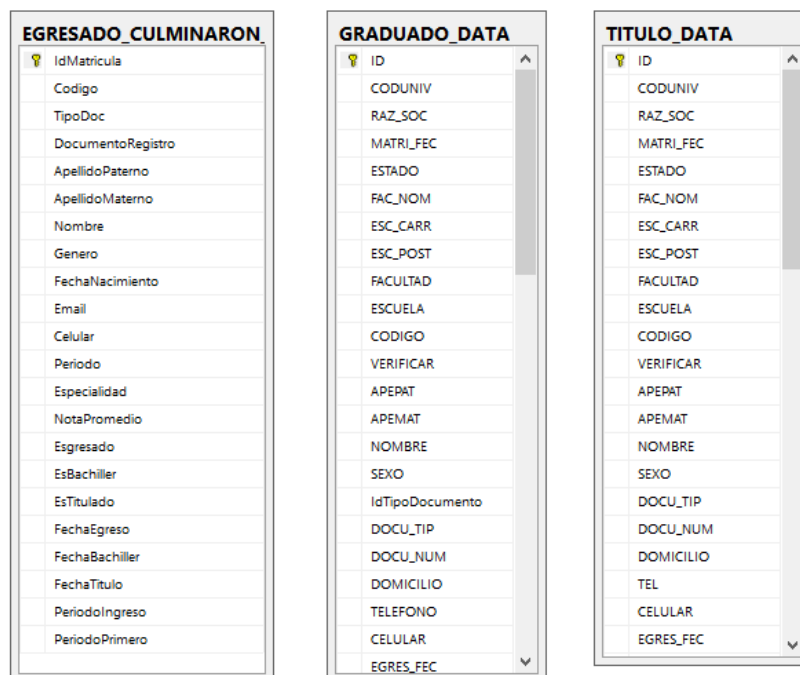
En la UNJBG, existen varias fuentes de origen de datos. En los cuales tenemos al base de datos de Registro académico, Bienestar Universitario y Grados y Títulos.



**Figura 51.** Diagrama de la base de datos de registro académico.  
Fuente: Propia.



**Figura 52.** Diagrama de la base de datos de bienestar universitario.  
Fuente: Propia.



**Figura 53.** Diagrama de la base de datos de grados y títulos.  
Fuente: Propia.

#### 5.2.2.4. Diccionario de datos de la base de datos transaccional

**Tabla 13** Descripción de la tabla OPE\_Entidad

| Nombre de Columna   | Tipo de Dato | Permitir | Descripción             |
|---------------------|--------------|----------|-------------------------|
| IdEntidad           | int          | Not Null | <b>Clave Primaria</b>   |
| ApellidoPaterno     | nvarchar(30) | Null     | Apellido Paterno        |
| ApellidoMaterno     | nvarchar(30) | Null     | Apellido Materno        |
| Nombre              | nvarchar(75) | Null     | Nombres                 |
| DocumentoRegistro   | nvarchar(15) | Null     | Documento Registro      |
| FechaNacimiento     | datetime     | Null     | Fecha de Nacimiento     |
| Telefono            | nvarchar(15) | Null     | Teléfono                |
| TelefonoProcedencia | nvarchar(15) | Null     | Teléfono de Procedencia |
| CelularProcedencia  | nvarchar(15) | Null     | Celular de Procedencia  |

| Nombre de Columna  | Tipo de Dato  | Permitir | Descripción             |
|--------------------|---------------|----------|-------------------------|
| Celular            | nvarchar(15)  | Null     | Celular                 |
| Celular2           | nvarchar(15)  | Null     | Celular2                |
| Email              | nvarchar(75)  | Null     | Correo Electrónico      |
| PaginaWeb          | nvarchar(75)  | Null     | Pagina Web              |
| Zona               | nvarchar(125) | Null     | Nombre de la Zona       |
| Via                | nvarchar(125) | Null     | Nombre de la Vía        |
| Numero             | nvarchar(75)  | Null     | Número                  |
| Referencia         | nvarchar(250) | Null     | Referencia              |
| IdSexo             | int           | Null     | Sexo                    |
| IdEstadoCivil      | int           | Null     | Estado Civil            |
| IdTipoDocumentoReg | int           | Null     | Tipo Documento Registro |
| IdLugarResidencia  | nvarchar(8)   | Null     | Lugar de Residencia     |
| IdLugarProcedencia | nvarchar(8)   | Null     | Lugar de Procedencia    |
| IdLugarNacimiento  | nvarchar(8)   | Null     | Lugar de Nacimiento     |
| IdTipoZona         | int           | Null     | Tipo de Zona            |
| IdTipoVia          | int           | Null     | Tipo de Vía             |

Fuente: Elaboración Propia.

**Tabla 14** Descripción de la tabla MAT\_Matricula

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdMatricula       | int          | Not Null | <b>Clave Primaria</b> |
| Codigo            | nvarchar(12) | Null     | Código de Ingreso     |
| Fecha             | datetime     | Null     | Fecha                 |
| PuntajeIngreso    | float        | Null     | Puntaje de Ingreso    |
| AnioResidencia    | int          | Null     | Año de Residencia     |

| Nombre de Columna   | Tipo de Dato   | Permitir | Descripción                    |
|---------------------|----------------|----------|--------------------------------|
| Observacion         | nvarchar(2000) | Null     | Observación                    |
| IdEntidad           | int            | Null     | Entidad                        |
| IdPlanCurricular    | int            | Null     | Plan de Estudio                |
| IdTipoIngreso       | int            | Null     | Tipo de Ingreso                |
| IdSituacionLaboral  | int            | Null     | Situación Laboral              |
| IdOcupacion         | int            | Null     | Ocupación                      |
| IdCentroEstudio     | int            | Null     | Centro de Estudio              |
| IdGradoInstruccion  | int            | Null     | Grado de Instrucción           |
| IdPreparacionUniv   | int            | Null     | Preparación Universitaria      |
| IdEstadoMatricula   | int            | Null     | Estado de Matrícula            |
| IdCategoriaAlumno   | int            | Null     | Categoría de Alumno            |
| IdModalidadEstudio  | int            | Null     | Modalidad de Estudio           |
| IdSede              | int            | Null     | Sede                           |
| IdSistemaEvaluacion | int            | Null     | Sistema de Evaluación de Rango |
| IdPeriodo           | int            | Null     | Periodo                        |

Fuente: Elaboración Propia.

**Tabla 15** Descripción de la tabla MAT\_Ficha

| Nombre de Columna | Tipo de Dato   | Permitir | Descripción           |
|-------------------|----------------|----------|-----------------------|
| IdFicha           | int            | Not Null | <b>Clave Primaria</b> |
| Denominacion      | nvarchar(20)   | Null     | Código de Numeración  |
| Fecha             | datetime       | Null     | Fecha                 |
| Firmado           | bit            | Null     | Firmado               |
| Sellado           | bit            | Null     | Sellado               |
| Observacion       | nvarchar(2000) | Null     | Observación           |

| Nombre de Columna    | Tipo de Dato | Permitir | Descripción            |
|----------------------|--------------|----------|------------------------|
| IdMatricula          | int          | Null     | Matrícula              |
| IdGradoEducativo     | int          | Null     | Grado Educativo        |
| IdClase              | int          | Null     | Clase                  |
| IdTurno              | int          | Null     | Turno                  |
| IdSituacionMatricula | int          | Null     | Situación de Matrícula |
| IdResolucion         | int          | Null     | Resolución             |
| IdEstadoFicha        | int          | Null     | Estado de Ficha        |
| IdPlanCurricular     | int          | Null     | Plan de Estudio        |
| IdMedioMatricula     | int          | Null     | Medio de Matrícula     |
| IdTipoMatricula      | int          | Null     | Tipo de Matrícula      |
| IdModalidadEstudio   | int          | Null     | Modalidad de Estudio   |
| IdSede               | int          | Null     | Sede                   |
| IdTabla              | int          | Null     | Tabla                  |
| IdPeriodo            | int          | Null     | Periodo                |

Fuente: Elaboración Propia.

**Tabla 16** Descripción de la tabla ACT\_FichaDetalle

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdFichaDetalle    | int          | Not Null | <b>Clave Primaria</b> |
| NroMatricula      | nvarchar(20) | Null     | Código de Numeración  |
| Nota              | int          | Null     | Nota                  |
| Fecha             | datetime     | Null     | Fecha                 |
| IdFicha           | int          | Null     | Ficha de Matrícula    |
| IdActa            | int          | Null     | Acta                  |
| IdResolucion      | int          | Null     | Resolución            |

| Nombre de Columna | Tipo de Dato | Permitir | Descripción       |
|-------------------|--------------|----------|-------------------|
| IdEstadoFicha     | int          | Null     | Estado de Ficha   |
| IdTipoMatricula   | int          | Null     | Tipo de Matrícula |

Fuente: Elaboración Propia.

**Tabla 17** Descripción de la tabla ACT\_Acta

| Nombre de Columna | Tipo de Dato   | Permitir | Descripción                |
|-------------------|----------------|----------|----------------------------|
| IdActa            | int            | Not Null | <b>Clave Primaria</b>      |
| Denominacion      | nvarchar(20)   | Null     | Código de Numeración       |
| Fecha             | datetime       | Null     | Fecha                      |
| Firmado           | bit            | Null     | Firmado                    |
| Sellado           | bit            | Null     | Sellado                    |
| Observacion       | nvarchar(2000) | Null     | Observación                |
| Resolucion        | nvarchar(250)  | Null     | Resolución                 |
| Referencia        | nvarchar(250)  | Null     | Referencia                 |
| NroOrdenTipoActa  | int            | Null     | Nro de Orden del Tipo Acta |
| IdAula            | int            | Null     | Aula                       |
| IdTipoActa        | int            | Null     | Tipo de Acta               |

Fuente: Elaboración Propia.

**Tabla 18** Descripción de la tabla ACT\_Aula

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdAula            | int          | Not Null | <b>Clave Primaria</b> |
| Denominacion      | nvarchar(20) | Null     | Código de Numeración  |
| Fecha             | datetime     | Null     | Fecha                 |
| Capacidad         | int          | Null     | Capacidad             |

| Nombre de Columna  | Tipo de Dato   | Permitir | Descripción          |
|--------------------|----------------|----------|----------------------|
| IdMateria          | int            | Null     | Materia              |
| IdClase            | nvarchar(2000) | Null     | Clase                |
| IdTurno            | nvarchar(250)  | Null     | Turno                |
| IdUbicacion        | nvarchar(250)  | Null     | Ubicación            |
| IdSede             | int            | Null     | Sede                 |
| IdModalidadEstudio | int            | Null     | Modalidad de Estudio |
| IdPeriodo          | int            | Null     | Periodo              |

Fuente: Elaboración Propia.

**Tabla 19** Descripción de la tabla ADS\_Materia

| Nombre de Columna  | Tipo de Dato  | Permitir | Descripción           |
|--------------------|---------------|----------|-----------------------|
| IdMateria          | int           | Not Null | <b>Clave Primaria</b> |
| Denominacion       | nvarchar(250) | Null     | Denominación          |
| Codigo             | nvarchar(15)  | Null     | Código Generado       |
| HoraTeorica        | int           | Null     | Hora Teórica          |
| HoraPractica       | int           | Null     | Hora Práctica         |
| HoraLaboratorio    | int           | Null     | Hora Laboratorio      |
| Credito            | float         | Null     | Crédito               |
| Obligatorio        | bit           | Null     | Obligatorio           |
| IdPadre            | int           | Null     | Padre                 |
| IdModalidadEstudio | int           | Null     | Modalidad de Estudio  |
| IdPlanCurricular   | int           | Null     | Plan de Estudio       |
| IdGradoEducativo   | int           | Null     | Grado Educativo       |
| IdSubArea          | int           | Null     | Sub Área              |
| IdArea             | int           | Null     | Área                  |

| Nombre de Columna | Tipo de Dato | Permitir | Descripción     |
|-------------------|--------------|----------|-----------------|
| IdRegimen         | int          | Null     | Régimen         |
| IdTipoMateria     | int          | Null     | Tipo de Materia |
| IdEtiqueta        | int          | Null     | Etiqueta        |

Fuente: Elaboración Propia.

**Tabla 20** Descripción de la tabla ADS\_Plancurricular

| Nombre de Columna    | Tipo de Dato | Permitir | Descripción                |
|----------------------|--------------|----------|----------------------------|
| IdPlanCurricular     | int          | Not Null | <b>Clave Primaria</b>      |
| Denominacion         | nvarchar(30) | Null     | Denominación               |
| Numero               | int          | Null     | Número                     |
| Credito              | int          | Null     | Total de Creditaje         |
| Ciclo                | int          | Null     | Cantidad de Ciclo          |
| FechaInicio          | datetime     | Null     | Fecha de Inicio            |
| FechaFinalizacion    | datetime     | Null     | Fecha de Finalización      |
| IdResolucionInicio   | int          | Null     | Resolución de Inicio       |
| IdResolucionFinaliza | int          | Null     | Resolución de Finalización |
| IdPadre              | int          | Null     | Padre                      |
| IdEspecialidad       | int          | Null     | Especialidad               |
| IdTipoRegimen        | int          | Null     | Tipo de Régimen            |
| IdTipoCiclo          | int          | Null     | Tipo de Ciclo              |
| IdEtiquetaCiclo      | int          | Null     | Etiqueta de Ciclo          |

Fuente: Elaboración Propia.

**Tabla 21** Descripción de la tabla ADS\_Especialidad

| Nombre de Columna   | Tipo de Dato  | Permitir | Descripción                        |
|---------------------|---------------|----------|------------------------------------|
| IdEspecialidad      | int           | Not Null | <b>Clave Primaria</b>              |
| NroRegistro         | int           | Null     | Nro de Registro                    |
| Denominacion        | nvarchar(250) | Null     | Denominación                       |
| Abreviatura         | nvarchar(30)  | Null     | Abreviatura                        |
| FechaInicio         | datetime      | Null     | Fecha de Inicio                    |
| FechaFinalizacion   | datetime      | Null     | Fecha de Finalización              |
| FechaFuncionamiento | datetime      | Null     | Fecha de Funcionamiento            |
| IdResolucion        | int           | Null     | Resolución de Finalización         |
| IdTipoEspecialidad  | int           | Null     | Tipo de Especialidad               |
| IdClaseEspecialidad | int           | Null     | Clase de Especialidad              |
| IdDuracionAcademica | int           | Null     | Duración Académica de Especialidad |
| IdEtiqueta          | int           | Null     | Etiqueta                           |
| IdEscuela           | int           | Null     | Escuela                            |

Fuente: Elaboración Propia.

**Tabla 22** Descripción de la tabla MAT\_Periodo

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdPeriodo         | int          | Not Null | <b>Clave Primaria</b> |
| Denominacion      | nvarchar(50) | Null     | Denominación          |
| FechaInicio       | datetime     | Null     | Fecha de Inicio       |
| FechaFinalizacion | datetime     | Null     | Fecha de Finalización |
| Anio              | int          | Null     | Año del Periodo       |
| Ciclo             | int          | Null     | Ciclo del Periodo     |

| Nombre de Columna    | Tipo de Dato | Permitir | Descripción                |
|----------------------|--------------|----------|----------------------------|
| IdResolucion         | int          | Null     | Resolución                 |
| IdTipoPeriodo        | int          | Null     | Tipo de Periodo            |
| IdTipoRegimenEstudio | int          | Null     | Tipo de Régimen de Estudio |

Fuente: Elaboración Propia.

**Tabla 23** Descripción de la tabla Alumno

| Nombre de Columna | Tipo de Dato | Permitir | Descripción             |
|-------------------|--------------|----------|-------------------------|
| Idalumno          | Int          | Not Null | <b>Clave Primaria</b>   |
| Codigo            | varchar(45)  | Null     | Código de Universitario |
| Dni               | varchar(8)   | Null     | Documento de Identidad  |
| Nombre            | varchar(80)  | Null     | Nombres                 |
| Apellido          | varchar(80)  | Null     | Apellidos               |
| Pass              | varchar(45)  | Null     | Password                |
| Anio_ingreso      | int          | Null     | Año de Ingreso          |
| Sexo              | varchar(1)   | Null     | Sexo                    |
| Fnacimiento       | date         | Null     | Fecha de Nacimiento     |
| Edad              | int          | Null     | Edad                    |
| Celular           | varchar(15)  | Null     | Celular                 |
| Email             | nvarchar(80) | Null     | Correo Electrónico      |
| Paisnac           | varchar(45)  | Null     | País                    |
| Dptonac           | varchar(45)  | Null     | Departamento            |
| Provnac           | varchar(45)  | Null     | Provincia               |
| Distnac           | varchar(45)  | Null     | Distrito                |
| Idtipocolegio     | int          | Null     | Tipo Colegio            |
| Idpreparacion     | int          | Null     | Preparación             |

| Nombre de Columna | Tipo de Dato | Permitir | Descripción  |
|-------------------|--------------|----------|--------------|
| Idmodalidad       | int          | Null     | Modalidad    |
| Idespecialidad    | int          | Null     | Especialidad |

Fuente: Elaboración Propia.

**Tabla 24** Descripción de la tabla FichaSocio

| Nombre de Columna | Tipo de Dato | Permitir | Descripción                |
|-------------------|--------------|----------|----------------------------|
| Idfichasocio      | Int          | Not Null | <b>Clave Primaria</b>      |
| Estado            | varchar(45)  | Null     | Estado                     |
| Fecha_envio       | datetime     | Null     | Fecha de Envío             |
| Fecha_evaluado    | datetime     | Null     | Fecha de Evaluación        |
| Anio              | int          | Null     | Año                        |
| Puntaje           | real         | Null     | Puntaje de Evaluación      |
| Categoria         | varchar(1)   | Null     | Categoría de la Evaluación |
| Foraneo           | int          | Null     | Foráneo                    |
| Tutor             | int          | Null     | Tutor                      |
| Ingresomensual    | int          | Null     | Ingreso Mensual            |
| Estudiosfam       | varchar(45)  | Null     | Estudios de Familiares     |
| Familiarenfermo   | varchar(80)  | Null     | Familiar Emfermo           |
| Observacion       | varchar(MAX) | Null     | Observación                |
| Idalumno          | int          | Null     | Alumno                     |
| Idusuario         | int          | Null     | Usuario                    |
| Idsituacionpadres | int          | Null     | Situación del Padre        |
| Idsostiene        | int          | Null     | Sostiene                   |
| Idcargafamiliar   | int          | Null     | Carga Familiar             |
| Idreligion        | int          | Null     | Religión                   |

| Nombre de Columna | Tipo de Dato | Permitir | Descripción              |
|-------------------|--------------|----------|--------------------------|
| Idriesgosalud_fam | int          | Null     | Riesgo de Salud Familiar |

Fuente: Elaboración Propia.

**Tabla 25** Descripción de la tabla Especialidad

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdEspecialidad    | int          | Not Null | <b>Clave Primaria</b> |
| Especialidad      | varchar(120) | Null     | Especialidad          |
| Siglaespecialidad | varchar(6)   | Null     | Sigla de Especialidad |
| Idescuela         | int          | Null     | Escuela               |

Fuente: Elaboración Propia.

**Tabla 26** Descripción de la tabla Especialidad

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdEspecialidad    | int          | Not Null | <b>Clave Primaria</b> |
| Especialidad      | varchar(120) | Null     | Especialidad          |
| Siglaespecialidad | varchar(6)   | Null     | Sigla de Especialidad |
| Idescuela         | int          | Null     | Escuela               |

Fuente: Elaboración Propia.

**Tabla 27** Descripción de la tabla CargaFamiliar

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| Idcargafamiliar   | int          | Not Null | <b>Clave Primaria</b> |
| item              | varchar(45)  | Null     | Denominación          |
| puntaje           | int          | Null     | Puntaje               |

Fuente: Elaboración Propia.

**Tabla 28** Descripción de la tabla Comunicacion

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| Idcomunicacion    | int          | Not Null | <b>Clave Primaria</b> |
| item              | varchar(45)  | Null     | Denominación          |
| puntaje           | int          | Null     | Puntaje               |

Fuente: Elaboración Propia.

**Tabla 29** Descripción de la tabla RiesgoSalud\_Familiar

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| Idriesgosalud_fam | int          | Not Null | <b>Clave Primaria</b> |
| Pregunta          | varchar(45)  | Null     | Pregunta              |
| item              | varchar(45)  | Null     | Denominación          |
| puntaje           | int          | Null     | Puntaje               |

Fuente: Elaboración Propia.

**Tabla 30** Descripción de la tabla SituacionPadres

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| Idsituacionpadres | int          | Not Null | <b>Clave Primaria</b> |
| item              | varchar(45)  | Null     | Denominación          |
| puntaje           | int          | Null     | Puntaje               |

Fuente: Elaboración Propia.

**Tabla 31** Descripción de la tabla Sostiene

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| Idsostiene        | int          | Not Null | <b>Clave Primaria</b> |

| Nombre de Columna | Tipo de Dato | Permitir | Descripción  |
|-------------------|--------------|----------|--------------|
| item              | varchar(45)  | Null     | Denominación |

Fuente: Elaboración Propia.

**Tabla 32** Descripción de la tabla Familiar

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| idfamiliar        | Int          | Not Null | <b>Clave Primaria</b> |
| Idfichasocio      | int          | Null     | Ficha Socio           |
| Nombre            | varchar(80)  | Null     | Nombres               |
| Apellido          | varchar(80)  | Null     | Apellidos             |
| Edad              | int          | Null     | Edad                  |
| Sexo              | varchar(1)   | Null     | Sexo                  |
| Ingreso           | int          | Null     | Ingreso               |
| Centrolaboral     | varchar(100) | Null     | Centro Laboral        |
| Miemb_hogar       | smallint     | Null     | Miembro del Hogar     |
| Idparentesco      | int          | Null     | Parentesco            |
| Idestadocivil     | int          | Null     | Estado Civil          |
| Idniveleducativo  | int          | Null     | Nivel Educativo       |
| Idtrabajo         | int          | Null     | Trabajo               |
| Idsegurosalud     | int          | Null     | Seguro de Salud       |

Fuente: Elaboración Propia.

**Tabla 33** Descripción de la tabla EgresadoCulminaron

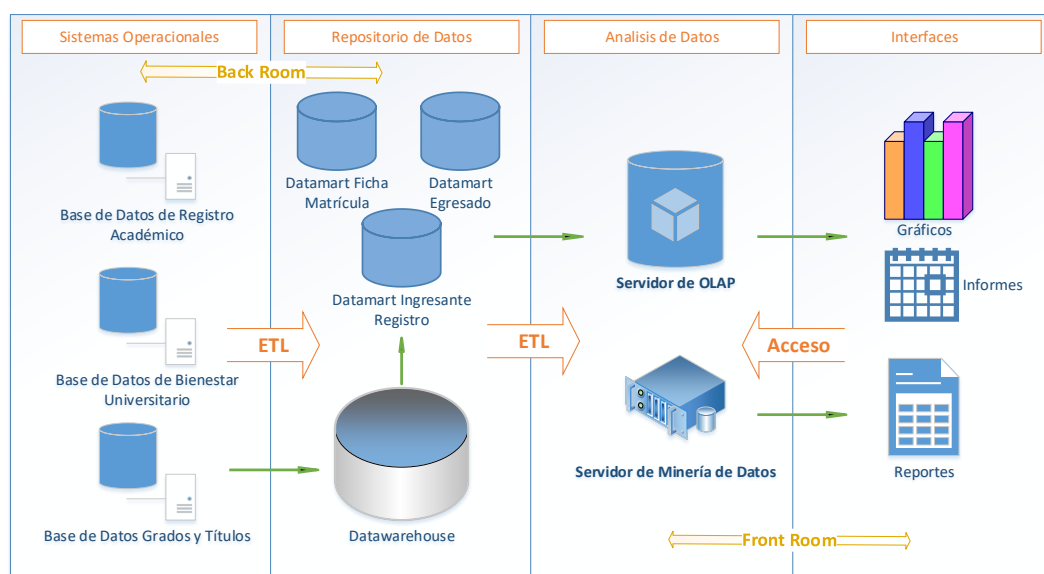
| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdMatricula       | Int          | Not Null | <b>Clave Primaria</b> |
| Codigo            | nvarchar(12) | Null     | Código de Ingreso     |

| <b>Nombre de Columna</b> | <b>Tipo de Dato</b> | <b>Permitir</b> | <b>Descripción</b>              |
|--------------------------|---------------------|-----------------|---------------------------------|
| TipoDoc                  | int                 | Null            | Tipo de Documento               |
| DocumentoRegistro        | nvarchar(15)        | Null            | Documento Registro              |
| ApellidoPaterno          | nvarchar(30)        | Null            | Apellido Paterno                |
| ApellidoMaterno          | nvarchar(30)        | Null            | Apellido Materno                |
| Nombre                   | nvarchar(75)        | Null            | Nombres                         |
| Genero                   | int                 | Null            | Foráneo                         |
| FechaNacimiento          | datetime            | Null            | Fecha de Nacimiento             |
| Email                    | nvarchar(75)        | Null            | Correo Electrónico              |
| Celular                  | nvarchar(15)        | Null            | Celular                         |
| Periodo                  | nvarchar(30)        | Null            | Periodo Lectivo                 |
| Especialidad             | nvarchar(250)       | Null            | Especialidad                    |
| NotaPromedio             | float               | Null            | Nota de Promedio Ponderado      |
| Egresado                 | int                 | Null            | Es Egresado                     |
| EsBachiller              | int                 | Null            | Es Bachiller                    |
| EsTitulado               | int                 | Null            | Es Titulado                     |
| FechaEgreso              | datetime            | Null            | Fecha de Egreso                 |
| FechaBachiller           | datetime            | Null            | Fecha de Bachiller              |
| FechaTitulo              | datetime            | Null            | Fecha de Título                 |
| PeriodoIngreso           | nvarchar(30)        | Null            | Periodo de Ingreso              |
| PeriodoPrimero           | nvarchar(30)        | Null            | Periodo de la Primera Matrícula |

Fuente: Elaboración Propia.

### 5.2.3. Diseño de la arquitectura

Es la base de todo proyecto, idea o plan que con lleva a ser útil o funcional, construible y estéticamente bello, que nos permita tener una visión general de todo el panorama del diseño arquitectónico y que sea una guía para la selección de las bases de datos operacionales, construcción de un DW, análisis de datos y finalmente diseños de interfaces para los usuarios finales. Como se puede visualizar en la figura 51.



**Figura 54.** Diseño de la arquitectura.

Fuente: Propia.

De acuerdo con Ralph Kimball plantea tres columnas fundamentales de la arquitectura del marco de trabajo: La columna de datos, técnica e infraestructura.

**Tabla 34** Columnas fundamentales de la arquitectura.

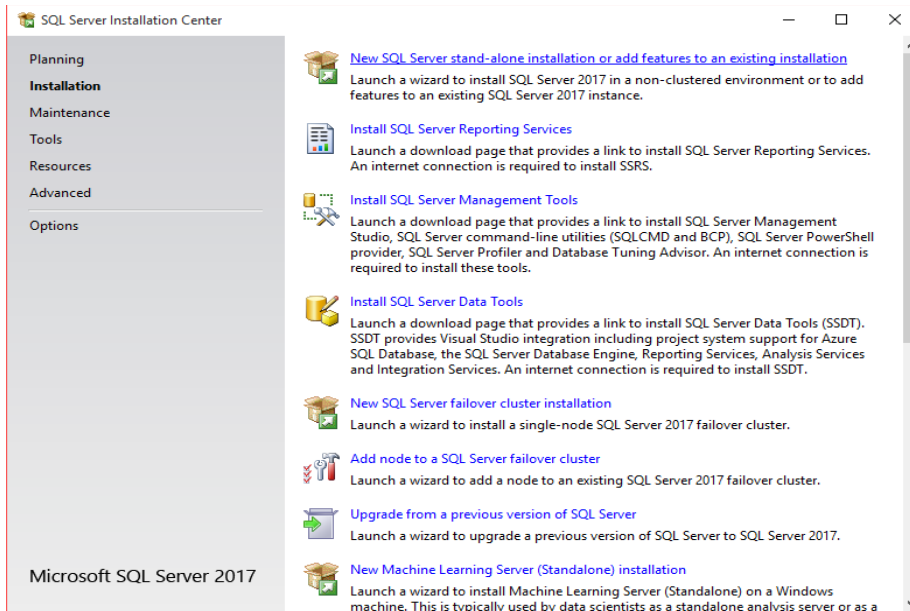
| <b>Dato</b>  | <b>Técnica Back Room</b>  | <b>Técnica Front Room</b>   | <b>Infraestructura</b>   |
|--|---|---|--|
| ¿el qué?   | ¿el cómo?   | ¿el cómo?   | ¿el dónde?   |
| Se analizará las Base de Datos Registros Académicos, Bienestar Universitario, Grados y Títulos.<br>Se creará los Datamarts IngresanteRegistro, FichaMatricula y Egresado bajo el esquema de Estrella, orientado al modelo MOLAP. | Se realizará el proceso de Extracción, transformación y carga de las base de datos de Registros Académicos, Bienestar Universitario, Grados y Títulos al Datawarehouse. | Se encargará de acceder a los datos de los modelos dimensionales: CUB_IngresanteRegistro, CUB_FichaMatricula y CUB_Egresado para mostrarlos a los usuarios finales. | Se instalará en el Servidor de la Dirección Académica de Actividades y Servicios Académicos, bajo la plataforma de Windows Server 2008 Standard. |

Fuente: Elaboración Propia.

#### **5.2.4. Instalación y selección del producto**

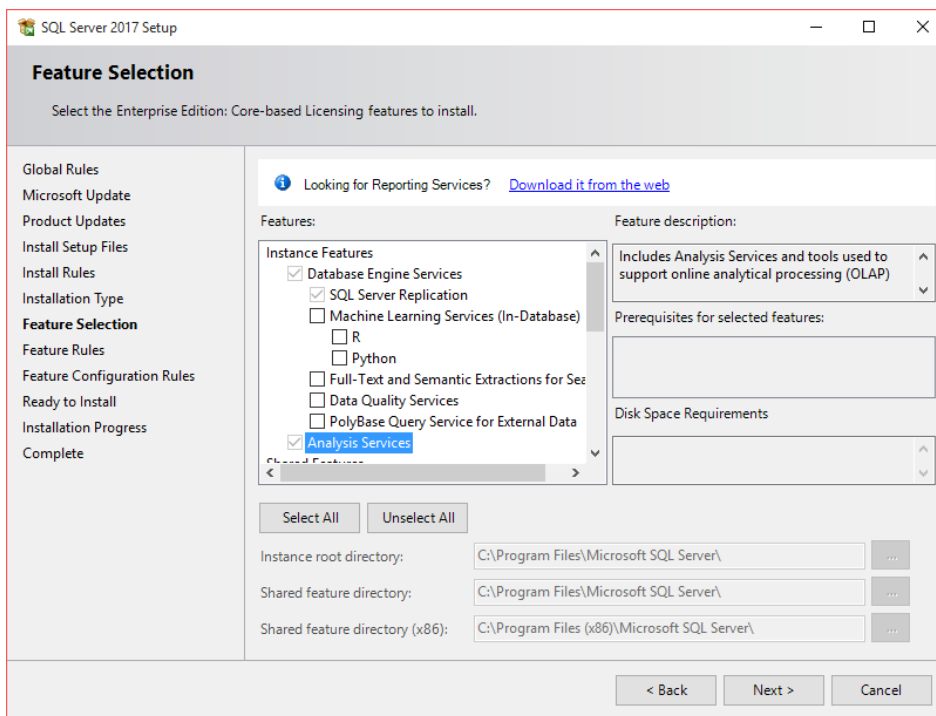
De acuerdo con el manejo del software que tiene a disposición la UNJBG y alineaciones con este proyecto se establece el uso de los siguientes productos:

- CA Erwin Data Modeler r7.3
- Microsoft SQL Server 2017 (Database, Analysis Services)
- Microsoft SQL Server Management Studio 17.9
- Microsoft SQL Server Data Tools



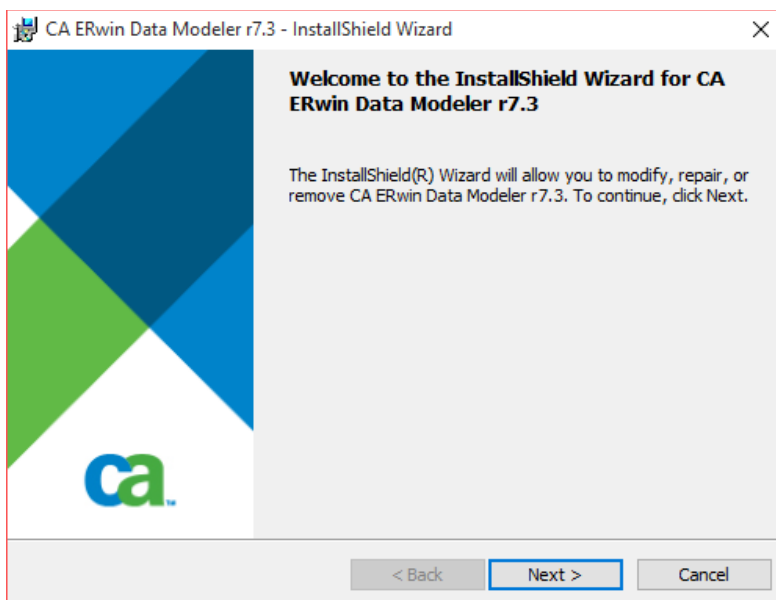
**Figura 55.** Instalación de Microsoft SQL Server 2017.

Fuente: Propia.



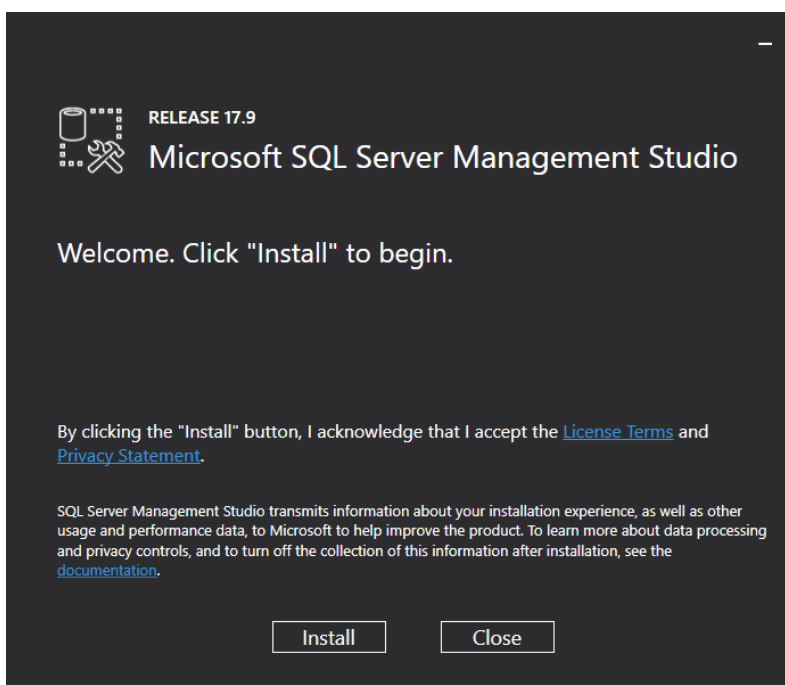
**Figura 56.** Selección de las características de Microsoft SQL Server 2017.

Fuente: Propia.



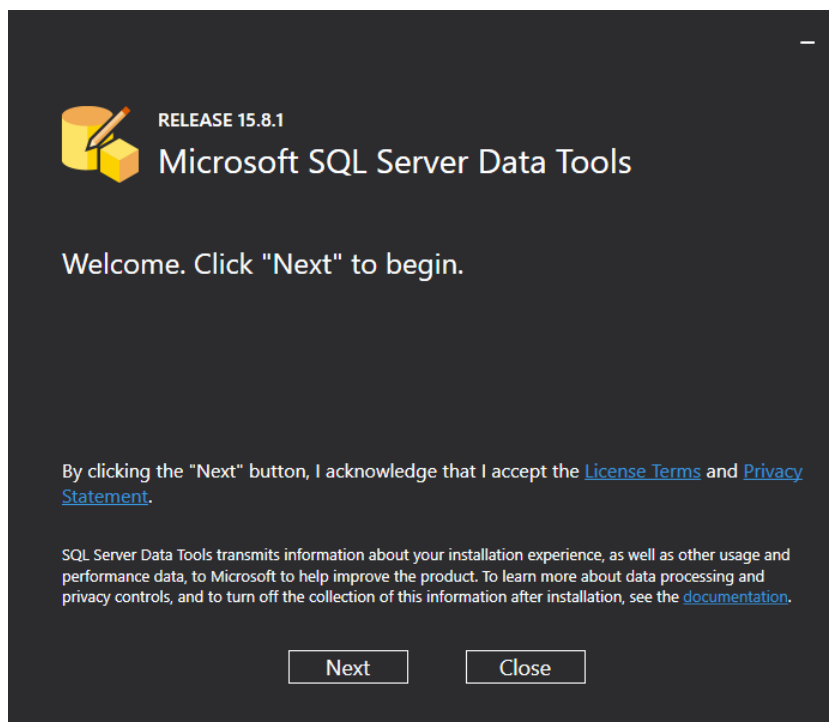
**Figura 57.** Instalación de CA Erwin Data Modeler r7.3.

Fuente: Propia.



**Figura 58.** Instalación de Microsoft SQL Server Management Studio 17.9.

Fuente: Propia.



**Figura 59.** Instalación de Microsoft SQL Server Data Tools 15.8.

Fuente: Propia

## **5.2.5. Modelamiento dimensional**

Es una técnica para nos ha permitido diseñar los modelos lógicos de las bases de datos de Registros Académicos, Bienestar Universitario y Grados y Títulos bajo el modelo estrella donde nos permita definir nuestras dimensiones, tablas de hechos y relaciones. A continuación, detallaremos las siguientes fases del modelamiento dimensional.

### **5.2.5.1. Elegir el proceso de negocio**

En esta fase elegiremos tres procesos de negocios que detallemos a continuación:

- Analizar que ingresantes están desertando, en que situación individual, académica, socioeconómica, sanitaria y en que periodos académicos.
- Analizar que estudiantes están desertando, en que situación académica, socioeconómica, individual, medidas académicas institucionales y en que periodos académicos.
- Analizar que egresados se están demorando en tramitar sus documentos académicos, en que situación individual, y en qué meses, años.

#### **5.2.5.2. Establecer el nivel de granularidad**

En esta fase el nivel de granularidad será más bajo para los tres procesos de negocios.

- Un registro de ingresante desertor representa el grano encontrado por cada matricula realizada en los periodos académicos lectivos, esto quiere decir que cada registro contiene situaciones individuales (Estudiante, Lugar Residencia), académicas (Situación de Matricula, Grado Educativo, Tipo de Ingreso, Programa Académico, Periodo, Centro de Estudio, Preparación de Estudio), socioeconómicas (Nivel Educativos de los Padres, Ámbito Vivienda, Carga Familiar, Uso de Medio de Comunicación) y sanitaria(Riesgo de Salud Familiar).

- Un registro de estudiante desertor representa el grano encontrado, esto quiere decir que cada registro contiene situaciones individuales (Estudiante, Lugar Residencia), académicas (Grado Educativo, Tipo de Ingreso, Programa Académico, Periodo, Centro de Estudio), socioeconómicas (Ámbito Vivienda, Servicio de Vivienda, Material de Construcción, Uso de Medio de Comunicación).
- Un registro de egresado que demora sus tramite de documentos académicos representa el grano a encontrado, esto quiere decir que cada registro contiene situaciones individuales (Estudiante, Lugar Residencia), académicas (Tipo de Ingreso, Programa Académico, Periodo, Centro de Estudio, Fechas de Egresados).

### 5.2.5.3. Elegir las dimensiones

En esta fase identificamos las dimensiones que nos describen los datos que resulten de los tres procesos de negocios.

**Tabla 35** Dimensiones encontradas en los procesos de negocios

| Nombre de la Dimensión  | Descripción  |
|-------------------------|--|
| DIM_Estudiante          | Registro de Estudiantes                              |
| DIM_Ubicaciongeografica | Registros de las Ubicaciones Geográficas             |
| DIM_CentroEstudio       | Registro de los Centros de Estudios                  |
| DIM_PreparacionEstudio  | Registro de Preparación de Estudios (Cepu, Academia) |
| DIM_NivelEducativoPadre | Registro del Nivel Educativo del Padre               |

| <b>Nombre de la Dimensión</b> | <b>Descripción</b>  |
|-------------------------------|---|
| DIM_AmbitoVivienda            | Registro del Ámbito de Vivienda (Urbano, Rural)                 |
| DIM_TenenciaVivienda          | Registro de Tenencia de Vivienda (Alquilada, Propia)            |
| DIM_MaterialConstruccion      | Registro de Material de Construcción de la Casa                 |
| DIM_ServicioVivienda          | Registro de Servicio de Vivienda (Luz, agua y desagüe)          |
| DIM_UsoMedioComunicacion      | Registro de Medios de Comunicación (Cable, Internet, Celular)   |
| DIM_Sostiene                  | Registro del Sostén de la Familia (Padre, Madre)                |
| DIM_SituacionPadre            | Registro de la Situación del Padre ( Casado, Divorciado, Viudo) |
| DIM_RiesgoSaludFamiliar       | Registro de Salud Familiar (Sano, Enfermo)                      |
| DIM_CargaFamiliar             | Registro de Cargar Familiar (01, 02, 03 personas)               |
| DIM_GradoEducativo            | Registro de Grados Educativos                                   |
| DIM_SituacionMatricula        | Registro de Situación de Matricula(Retirado, Amnistía)          |
| DIM_TipoIngreso               | Registro de Tipos de Ingresos (CEPU, Fase de Admisión)          |
| DIM_ProgramaAcademico         | Registro de Programas Académicos.                               |
| DIM_Fecha                     | Registro de Fechas (Días, Semanas, Meses, Años)                 |
| DIM_Periodo                   | Registro de Periodos Lectivos (2012-2018)                       |

Fuente: Elaboración Propia.

**Tabla 36** Descripción de la dimensión de estudiante

| <b>Nombre de Columna</b> | <b>Tipo de Dato</b> | <b>Permitir</b> | <b>Descripción</b>      |
|--------------------------|---------------------|-----------------|-------------------------|
| IdEstudiante             | int                 | Not Null        | <b>Clave Primaria</b>   |
| Codigo                   | nvarchar(15)        | Null            | Código de Ingreso       |
| ApellidosyNombres        | nvarchar(250)       | Null            | Apellidos y Nombres     |
| Sexo                     | nvarchar(50)        | Null            | Sexo                    |
| SexoAbreviatura          | nvarchar(30)        | Null            | Sexo Abreviatura (M, F) |
| EstadoCivil              | nvarchar(50)        | Null            | Estado Civil            |

| Nombre de Columna        | Tipo de Dato  | Permitir | Descripción                     |
|--------------------------|---------------|----------|---------------------------------|
| EstadoCivilAbreviatura   | nvarchar(30)  | Null     | Estado Civil Abreviatura (S, C) |
| NivelCategorizacion      | nvarchar(50)  | Null     | Nivel de Categorización         |
| NivelCategorizacionAbrev | nvarchar(30)  | Null     | Abreviatura de Categorización   |
| NivelCategorizacionRango | nvarchar(30)  | Null     | Rango de Categorización         |
| TipoNivelCategorizacion  | nvarchar(50)  | Null     | Tipo de Nivel Categorización    |
| DistritoNacimiento       | nvarchar(250) | Null     | Distrito de Nacimiento          |
| ProvinciaNacimiento      | nvarchar(250) | Null     | Provincia de Nacimiento         |
| DepartamentoNacimiento   | nvarchar(250) | Null     | Departamento de Nacimiento      |
| PeriodoIngreso           | nvarchar(50)  | Null     | Periodo de Ingreso              |
| PeriodoIngresoAnio       | nvarchar(30)  | Null     | Año de Periodo de Ingreso       |

Fuente: Elaboración Propia.

**Tabla 37** Descripción de la dimensión de ubicación geográfica

| Nombre de Columna     | Tipo de Dato  | Permitir | Descripción           |
|-----------------------|---------------|----------|-----------------------|
| IdUbicacionGeografica | nvarchar(8)   | Not Null | <b>Clave Primaria</b> |
| Distrito              | nvarchar(250) | Null     | Distrito              |
| Provincia             | nvarchar(250) | Null     | Provincia             |
| Departamento          | nvarchar(250) | Null     | Departamento          |

Fuente: Elaboración Propia.

**Tabla 38** Descripción de la dimensión de centro de estudios

| Nombre de Columna | Tipo de Dato  | Permitir | Descripción           |
|-------------------|---------------|----------|-----------------------|
| IdCentroEstudio   | int           | Not Null | <b>Clave Primaria</b> |
| CentroEstudio     | nvarchar(250) | Null     | Centro de Estudios    |
| Codigo            | nvarchar(50)  | Null     | Código de Registro    |

| Nombre de Columna | Tipo de Dato  | Permitir | Descripción       |
|-------------------|---------------|----------|-------------------|
| Turno             | nvarchar(50)  | Null     | Turno             |
| TipoGenero        | nvarchar(125) | Null     | Tipo de Genero    |
| GestionEducativa  | nvarchar(125) | Null     | Gestión Educativa |
| Distrito          | nvarchar(250) | Null     | Distrito          |
| Provincia         | nvarchar(250) | Null     | Provincia         |
| Departamento      | nvarchar(250) | Null     | Departamento      |

Fuente: Elaboración Propia.

**Tabla 39** Descripción de la dimensión de preparación de estudios

| Nombre de Columna       | Tipo de Dato | Permitir | Descripción             |
|-------------------------|--------------|----------|-------------------------|
| IdPreparacionEstudio    | int          | Not Null | <b>Clave Primaria</b>   |
| PreparacionEstudio      | nvarchar(50) | Null     | Preparación de Estudios |
| PreparacionEstudioAbrev | nvarchar(30) | Null     | Abreviatura             |

Fuente: Elaboración Propia.

**Tabla 40** Descripción de la dimensión de nivel educativo

| Nombre de Columna   | Tipo de Dato  | Permitir | Descripción           |
|---------------------|---------------|----------|-----------------------|
| IdNivelEducativo    | int           | Not Null | <b>Clave Primaria</b> |
| NivelEducativo      | nvarchar(250) | Null     | Nivel Educativo       |
| NivelEducativoAbrev | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 41** Descripción de la dimensión de ámbito de vivienda

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| IdAmbitoVivienda  | int          | Not Null | <b>Clave Primaria</b> |

| Nombre de Columna   | Tipo de Dato  | Permitir | Descripción        |
|---------------------|---------------|----------|--------------------|
| AmbitoVivienda      | nvarchar(250) | Null     | Ámbito de Vivienda |
| AmbitoViviendaAbrev | nvarchar(30)  | Null     | Abreviatura        |

Fuente: Elaboración Propia.

**Tabla 42** Descripción de la dimensión de tenencia de vivienda

| Nombre de Columna   | Tipo de Dato  | Permitir | Descripción           |
|---------------------|---------------|----------|-----------------------|
| IdTenenciaVivienda  | int           | Not Null | <b>Clave Primaria</b> |
| TenenciaVivienda    | nvarchar(250) | Null     | Tenencia de Vivienda  |
| AmbitoViviendaAbrev | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 43** Descripción de la dimensión de material de construcción

| Nombre de Columna       | Tipo de Dato  | Permitir | Descripción              |
|-------------------------|---------------|----------|--------------------------|
| IdMaterialConstruccion  | int           | Not Null | <b>Clave Primaria</b>    |
| MaterialConstruccion    | nvarchar(250) | Null     | Material de Construcción |
| MaterialConstruccionAbr | nvarchar(30)  | Null     | Abreviatura              |

Fuente: Elaboración Propia.

**Tabla 44** Descripción de la dimensión de servicio de vivienda

| Nombre de Columna     | Tipo de Dato  | Permitir | Descripción           |
|-----------------------|---------------|----------|-----------------------|
| IdServicioVivienda    | int           | Not Null | <b>Clave Primaria</b> |
| ServicioVivienda      | nvarchar(250) | Null     | Servicio de Vivienda  |
| ServicioViviendaAbrev | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 45** Descripción de la dimensión de uso de medio de comunicación

| Nombre de Columna      | Tipo de Dato  | Permitir | Descripción                  |
|------------------------|---------------|----------|------------------------------|
| IdUsoMedioComunicacion | nvarchar(4)   | Not Null | <b>Clave Primaria</b>        |
| UsoMedioComunicacion   | nvarchar(250) | Null     | Uso de Medio de Comunicación |
| UsoMedioComunicacionA  | nvarchar(30)  | Null     | Abreviatura                  |

Fuente: Elaboración Propia.

**Tabla 46** Descripción de la dimensión de sostiene

| Nombre de Columna  | Tipo de Dato  | Permitir | Descripción             |
|--------------------|---------------|----------|-------------------------|
| IdSotiene          | int           | Not Null | <b>Clave Primaria</b>   |
| Sotiene            | nvarchar(250) | Null     | Sostiene (Padre, Madre) |
| SotieneAbreviatura | nvarchar(30)  | Null     | Abreviatura             |

Fuente: Elaboración Propia.

**Tabla 47** Descripción de la dimensión de situación del padre

| Nombre de Columna         | Tipo de Dato  | Permitir | Descripción           |
|---------------------------|---------------|----------|-----------------------|
| IdSituacionPadre          | int           | Not Null | <b>Clave Primaria</b> |
| SituacionPadre            | nvarchar(250) | Null     | Situación del Padre   |
| SituacionPadreAbreviatura | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 48** Descripción de la dimensión de riesgo de salud familiar

| Nombre de Columna     | Tipo de Dato  | Permitir | Descripción              |
|-----------------------|---------------|----------|--------------------------|
| IdRiesgoSaludFamiliar | int           | Not Null | <b>Clave Primaria</b>    |
| RiesgoSaludFamiliar   | nvarchar(250) | Null     | Riesgo de Salud Familiar |

| Nombre de Columna        | Tipo de Dato | Permitir | Descripción |
|--------------------------|--------------|----------|-------------|
| RiesgoSaludFamiliarAbrev | nvarchar(30) | Null     | Abreviatura |

Fuente: Elaboración Propia.

**Tabla 49** Descripción de la dimensión de carga familiar

| Nombre de Columna        | Tipo de Dato  | Permitir | Descripción           |
|--------------------------|---------------|----------|-----------------------|
| IdCargaFamiliar          | int           | Not Null | <b>Clave Primaria</b> |
| CargaFamiliar            | nvarchar(250) | Null     | Carga Familiar        |
| CargaFamiliarAbreviatura | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 50** Descripción de la dimensión de grado educativo

| Nombre de Columna   | Tipo de Dato  | Permitir | Descripción           |
|---------------------|---------------|----------|-----------------------|
| IdGradoEducativo    | int           | Not Null | <b>Clave Primaria</b> |
| GradoEducativo      | nvarchar(250) | Null     | Grado Educativo       |
| GradoEducativoAbrev | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 51** Descripción de la dimensión de situación de matrícula

| Nombre de Columna       | Tipo de Dato  | Permitir | Descripción            |
|-------------------------|---------------|----------|------------------------|
| IdSituacionMatricula    | int           | Not Null | <b>Clave Primaria</b>  |
| SituacionMatricula      | nvarchar(250) | Null     | Situación de Matrícula |
| SituacionMatriculaAbrev | nvarchar(30)  | Null     | Abreviatura            |

Fuente: Elaboración Propia.

**Tabla 52** Descripción de la dimensión de tipo de ingreso

| Nombre de Columna      | Tipo de Dato  | Permitir | Descripción           |
|------------------------|---------------|----------|-----------------------|
| IdTipoIngreso          | int           | Not Null | <b>Clave Primaria</b> |
| TipoIngreso            | nvarchar(250) | Null     | Tipo de Ingreso       |
| TipoIngresoAbreviatura | nvarchar(30)  | Null     | Abreviatura           |

Fuente: Elaboración Propia.

**Tabla 53** Descripción de la dimensión de programa académicos

| Nombre de Columna      | Tipo de Dato  | Permitir | Descripción                |
|------------------------|---------------|----------|----------------------------|
| IdProgramaAcademico    | int           | Not Null | <b>Clave Primaria</b>      |
| ProgramaAcademico      | nvarchar(250) | Null     | Tipo de Ingreso            |
| ProgramaAcademicoAbrev | nvarchar(50)  | Null     | Abreviatura                |
| CarreraProfesional     | nvarchar(250) | Null     | Carrera Profesional        |
| Facultad               | nvarchar(250) | Null     | Facultad                   |
| TipoProgramaAcademico  | nvarchar(250) | Null     | Tipo de Programa Académico |

Fuente: Elaboración Propia.

**Tabla 54** Descripción de la dimensión de periodo

| Nombre de Columna  | Tipo de Dato  | Permitir | Descripción                 |
|--------------------|---------------|----------|-----------------------------|
| IdPeriodo          | int           | Not Null | <b>Clave Primaria</b>       |
| Periodo            | nvarchar(250) | Null     | Periodo Académico           |
| Ciclo              | int           | Null     | Ciclo Académico             |
| Anio               | int           | Null     | Año del Periodo Académico   |
| TipoRegimenEstudio | nvarchar(30)  | Null     | Tipo de Régimen de Estudios |

Fuente: Elaboración Propia.

**Tabla 55** Descripción de la dimensión de fechas

| Nombre de Columna | Tipo de Dato | Permitir | Descripción           |
|-------------------|--------------|----------|-----------------------|
| DateKey           | Int          | Not Null | <b>Clave Primaria</b> |
| Date              | Datetime     | Null     | Fecha                 |
| EnglishMonthName  | nvarchar(50) | Null     | Nombre del Mes        |
| CalendarYear      | nvarchar(4)  | Null     | Año Calendario        |

Fuente: Elaboración Propia.

#### 5.2.5.4. Identificar medidas y las tablas de hechos

En esta fase identificamos los valores numéricos de las medidas para las tablas de hechos, que representan los valores de negocios.

**Tabla 56** Tablas de hechos encontradas en los procesos de negocios

| Nombre de la Tabla Hecho | Descripción                     |
|--------------------------|---------------------------------|
| THE_IngresanteRegistro   | Registro de Ingresante Registro |
| THE_FichaMatricula       | Registros de Ficha de Matrícula |
| THE_Egresado             | Registro de Egresados           |

Fuente: Elaboración Propia.

**Tabla 57** Medidas de las tablas de hechos encontradas en los procesos de negocios

| Nombre de la Tabla Hecho | Medidas                 | Descripción   |
|--------------------------|-------------------------|---|
| THE_IngresanteRegistro   | MatriculaNoRegistradaPI | Matrícula no registrada al periodo inicial.                     |
|                          | DesaprobadoMas50PC      | Desaprobado más del 50 por ciento en la primer periodo lectivo. |
|                          | DesertorPortencial      | Desertor Potencial  |
|                          | Desertor                | Desertor  |
|                          | MatriculaNoRegistradaSR | Matrícula no registrada en aplicación de los reglamentos.       |

| Nombre de la Tabla Hecho | Medidas              | Descripción  |
|--------------------------|----------------------|--|
| THE_FichaMatricula       | DesaprobadoMas70PC4S | Desaprobado más del 70 por ciento dentro de los 4 semestres. |
|                          | DesaprobadoMas90PC2S | Desaprobado más del 90 por ciento dentro de los 2 semestres. |
|                          | DesaprobadoMas50PC   | Desaprobado más del 50 por ciento del programa académico.    |
| THE_Egresado             | DesertorPortencial   | Desertor Potencial   |
|                          | Desertor             | Desertor   |
|                          | EgresadoNoTramitaron | Egresado que no tramitaron su documentos académicos.         |
|                          | Egresado             | Estudiante que tramitan su constancia de egresado.           |
|                          | Bachiller            | Egresado que tramitan su grado académico.                    |
|                          | Titulado             | Bachilleres que tramitan su título profesional.              |

Fuente: Elaboración Propia.

**Tabla 58** Descripción de la tabla de hecho ingresante registro

| Nombre de Columna      | Tipo de Dato | Permitir | Descripción                   |
|------------------------|--------------|----------|-------------------------------|
| IdIngresanteRegistro   | int          | Not Null | <b>Clave Primaria</b>         |
| IdEstudiante           | int          | Null     | Estudiante                    |
| IdLugarResidencia      | nvarchar(8)  | Null     | Ubicación Geográficas         |
| IdCentroEstudio        | int          | Null     | Centro de Estudios            |
| IdPreparacionEstudio   | int          | Null     | Preparación de Estudios       |
| IdCantidadPostulacion  | int          | Null     | Cantidad de Postulación       |
| IdNivelEducativoPadre  | int          | Null     | Nivel Educativo               |
| IdAmbitoVivienda       | int          | Null     | Ámbito de Vivienda            |
| IdTenenciaVivienda     | int          | Null     | Tenencia de Vivienda          |
| IdMaterialConstruccion | int          | Null     | Material de Construcción      |
| IdServicioVivienda     | int          | Null     | Servicio de Vivienda          |
| IdUsoMedioComunicacion | nvarchar(4)  | Null     | Uso del Medio de Comunicación |
| IdSostiene             | int          | Null     | Sostiene                      |
| IdSituacionPadre       | int          | Null     | Situación del Padre           |

| Nombre de Columna       | Tipo de Dato | Permitir | Descripción                       |
|-------------------------|--------------|----------|-----------------------------------|
| IdRiesgoSaludFamiliar   | int          | Null     | Riesgo de Salud Familiar          |
| IdCargaFamiliar         | int          | Null     | Carga Familiar                    |
| IdGradoEducativo        | int          | Null     | Grado Educativo                   |
| IdSituacionMatricula    | int          | Null     | Situación de Matrícula            |
| IdTipoIngreso           | int          | Null     | Tipo de Ingreso                   |
| IdProgramaAcademico     | int          | Null     | Programa Académicos               |
| IdPeriodo               | int          | Null     | Periodo                           |
| MatriculaNoRegistradaPI | int          | Null     | Matrícula No Registrada al Inicio |
| DesaprobadoMas50PC      | int          | Null     | Desaprobado más del 50%           |
| DesertorPortencial      | int          | Null     | Desertor Potencial                |
| Desertor                | int          | Null     | Desertor                          |

Fuente: Elaboración Propia.

**Tabla 59** Descripción de la tabla de hecho ficha de matrícula

| Nombre de Columna      | Tipo de Dato | Permitir | Descripción                   |
|------------------------|--------------|----------|-------------------------------|
| IdFichaMatricula       | int          | Not Null | <b>Clave Primaria</b>         |
| IdEstudiante           | int          | Null     | Estudiante                    |
| IdAmbitoVivienda       | int          | Null     | Ámbito de Vivienda            |
| IdMaterialConstruccion | int          | Null     | Material de Construcción      |
| IdServicioVivienda     | int          | Null     | Servicio de Vivienda          |
| IdUsoMedioComunicacion | nvarchar(4)  | Null     | Uso del Medio de Comunicación |
| IdSostiene             | int          | Null     | Sostiene                      |
| IdLugarResidencia      | nvarchar(8)  | Null     | Ubicación Geográficas         |
| IdCentroEstudio        | int          | Null     | Centro de Estudios            |
| IdTipoIngreso          | int          | Null     | Tipo de Ingreso               |

| Nombre de Columna       | Tipo de Dato | Permitir | Descripción                       |
|-------------------------|--------------|----------|-----------------------------------|
| IdGradoEducativo        | int          | Null     | Grado Educativo                   |
| IdProgramaAcademico     | int          | Null     | Programa Académicos               |
| IdPeriodoUltimo         | int          | Null     | Periodo                           |
| MatriculaNoRegistradaPI | int          | Null     | Matrícula No Registrada al Inicio |
| DesaprobadoMas70PC4S    | int          | Null     | Desaprobado más del 70%           |
| DesaprobadoMas90PC2S    | int          | Null     | Desaprobado más del 90%           |
| DesaprobadoMas50PC      | int          | Null     | Desaprobado más del 50%           |
| DesertorPortencial      | int          | Null     | Desertor Potencial                |
| Desertor                | int          | Null     | Desertor                          |

Fuente: Elaboración Propia.

**Tabla 60** Descripción de la tabla de hecho egresado

| Nombre de Columna    | Tipo de Dato | Permitir | Descripción                |
|----------------------|--------------|----------|----------------------------|
| IdEgresado           | Int          | Not Null | <b>Clave Primaria</b>      |
| IdEstudiante         | Int          | Null     | Estudiante                 |
| IdFecha              | Datetime     | Null     | Fecha                      |
| IdLugarResidencia    | nvarchar(8)  | Null     | Ubicación Geográficas      |
| IdCentroEstudio      | Int          | Null     | Centro de Estudios         |
| IdTipoIngreso        | Int          | Null     | Tipo de Ingreso            |
| IdProgramaAcademico  | Int          | Null     | Programa Académicos        |
| IdPeriodoFinalizo    | Int          | Null     | Periodo de Finalización    |
| EgresadoNoTramitaron | Int          | Null     | Egresado que no tramitaron |
| Egresado             | Int          | Null     | Egresado                   |
| Bachiller            | Int          | Null     | Bachiller                  |
| Titulado             | Int          | Null     | Titulado                   |

Fuente: Elaboración Propia.

### 5.2.5.5. Elegimos el esquema del modelo dimensional

En esta fase elegimos el esquema estrella, para la representación de los tres procesos de negocios.

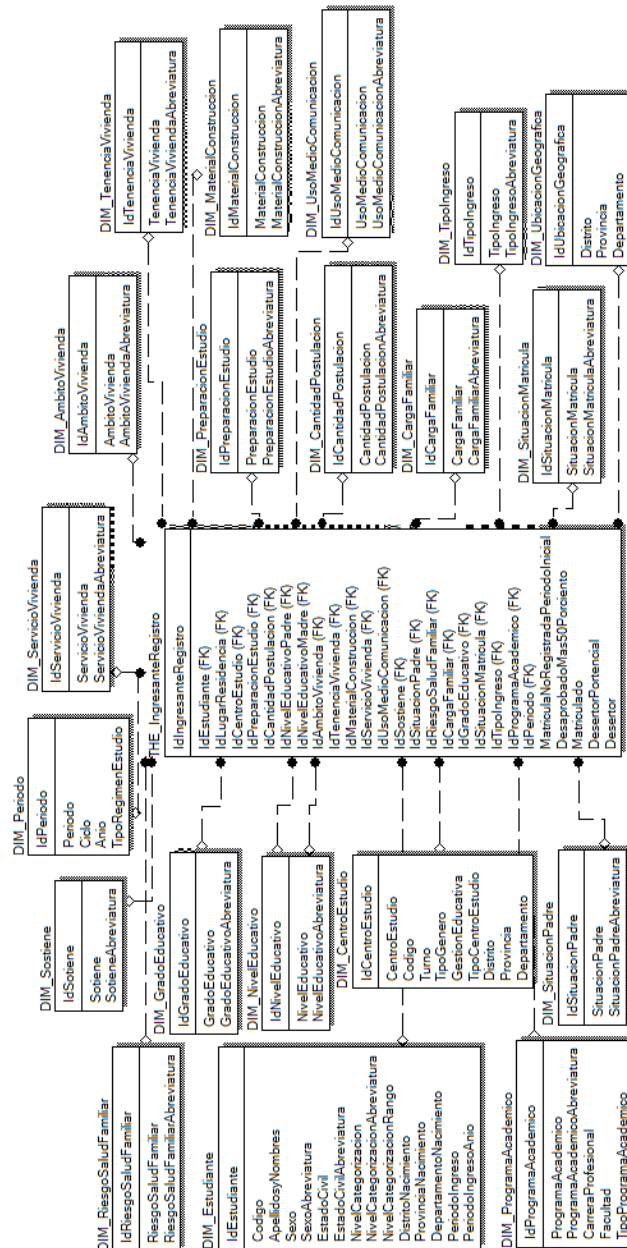
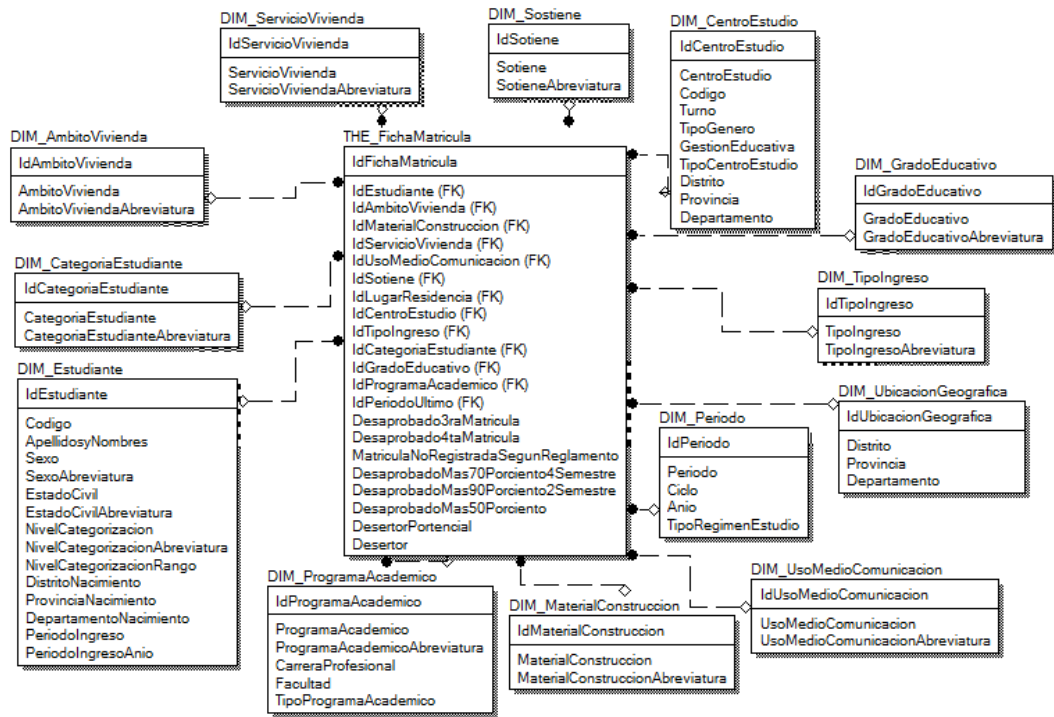


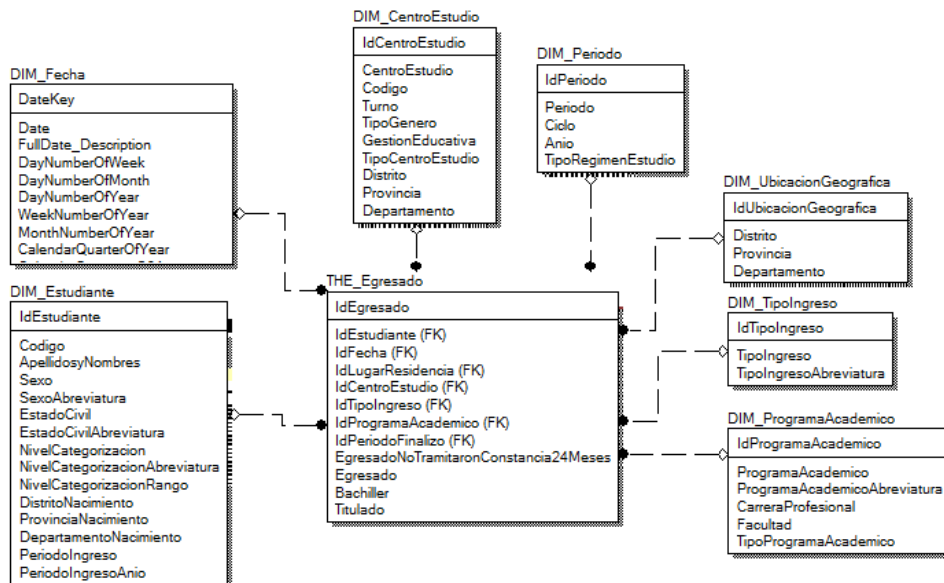
Figura 60. Diseño lógico de los ingresantes desertores.

Fuente: Propia



**Figura 61.** Diseño lógico de los estudiantes desertores.

Fuente: Propia



**Figura 62.** Diseño lógico de los egresados sin tramites de documentos académicos.

Fuente: Propia

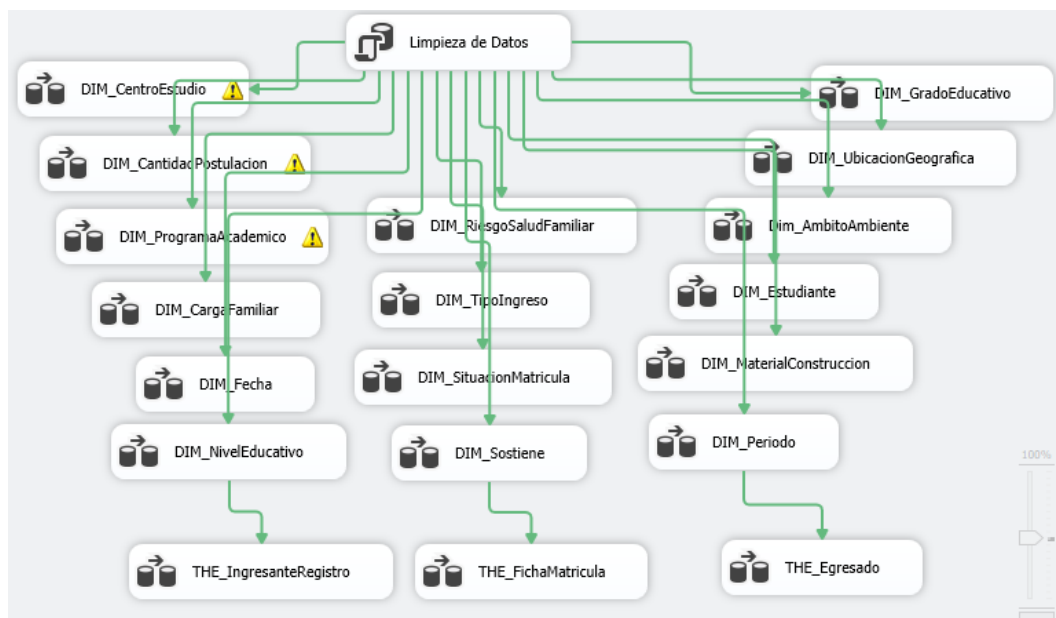
### **5.2.6. Diseño físico**

En esta fase se procede a la creación física de los diseños lógicos de los procesos de negocios, que adelante lo conoceremos como Datamart, como se detalla en el anexo 4, para esto seguimos los pasos siguientes:

- Creación de la base de Datos: SYSA\_DAWH\_R4.
- Creación de las estructuras físicas de los Data marts:  
Ingresante Registro, Ficha Matrícula, Egresado
- Identificación los campos de indexación y particionamiento de las tablas de hechos que permitan optimizar las consultas.
- Establecemos un rango de 07 años de datos históricos.
- Creación de los Scripts para poblar los objetos de los Data Marts.
- La instalación de Microsoft SQL Server 2017 (Analysis Services), no permite realizar análisis de datos en la etapa desarrollo, prueba y producción.

### **5.2.7. Diseño e implementación del subsistema de extracción, transformación y carga (ETL)**

En esta fase se describe las tareas de flujo de datos, conexiones de distintos orígenes de datos, la preparación del proceso ETL y esto sea reflejado en la carga de los modelos dimensionales. Como se detalla en el anexo 5.



**Figura 63.** Diseño del subsistema de ETL.

Fuente: Propia.

## 5.2.8. Especificación de aplicaciones de business intelligence (BI)

### 5.2.8.1. Informes estándar

En esta sección la información se expresó en forma de gráficos, reportes estadísticos de los desertores de ingresantes, matriculados por los distintos factores individuales, académicos y socioeconómicos.

### 5.2.8.2. Aplicaciones analíticas

En esta sección se aplicó modelos de minería de datos de árbol de decisión, redes bayesianas y regresión logística para la predicción de los desertores de los estudiantes basadrinos.

### **5.2.8.3. Usuarios finales**

Sobre los roles y accesos a la solución BI, se está clasificando por el cargo que ocupan en la UNJBG:

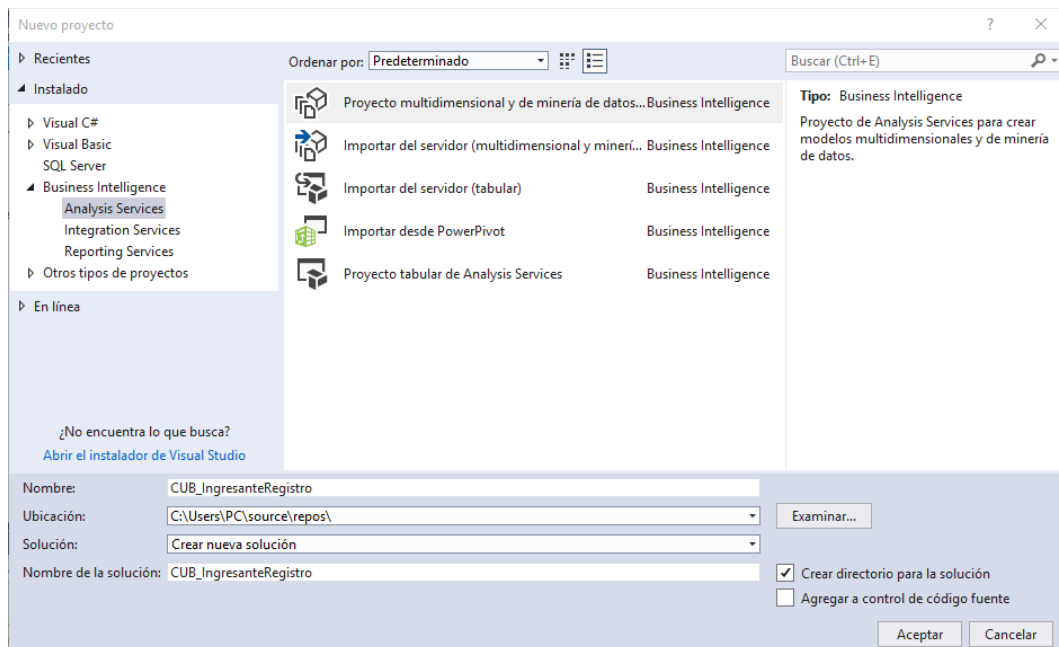
- Vicerrector Académico: Podrá visualizar los reportes de deserción estudiantil basadrina.
- Directores de Escuelas: Podrán visualizar los reportes y crearlos para una mejor toma de decisión sobre sus estudiantes desertores.
- Personal de Sistemas: Tendrán acceso total a la solución, estos podrán crear o modificar las consultas previas de los reportes de la solución BI.

### **5.2.9. Desarrollo de aplicación BI**

Para el desarrollo de las aplicaciones BI se utilizó Microsoft Visual Studio 2017, donde nos permite crear conexiones con distintos orígenes de datos, establecer vista de orígenes de datos, crear de cubos y diversos reportes que nos permiten realizar análisis de datos, para sí tomar buenas decisiones y oportunas.

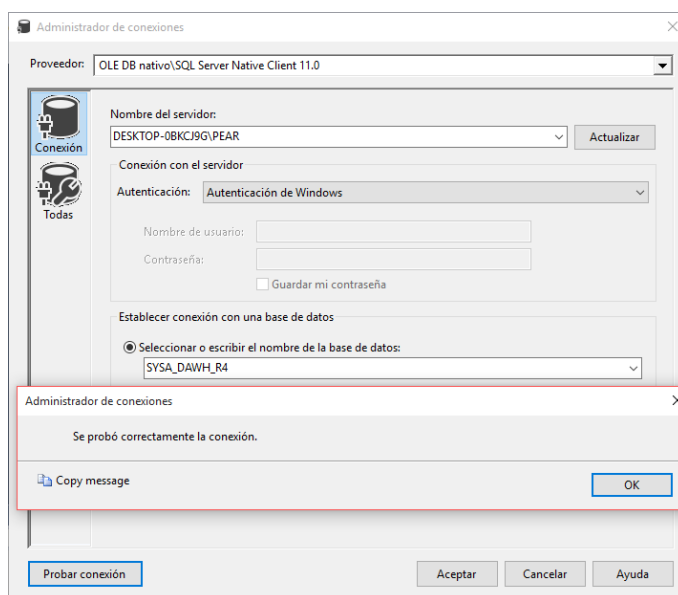
#### **5.2.9.1. Elegir el proyecto multidimensional**

En esta sección elegimos el tipo de proyecto dimensional para el desarrollo de las aplicaciones BI.



**Figura 64.** Proyecto multidimensional.  
Fuente: Propia.

### 5.2.9.2. Elegir el origen de datos



**Figura 65.** Conexión con el origen de datos.  
Fuente: Propia.

### 5.2.9.3. Crear vista de origen de datos

En esta sección seleccionamos las dimensiones y tabla de hecho para la vista de datos.

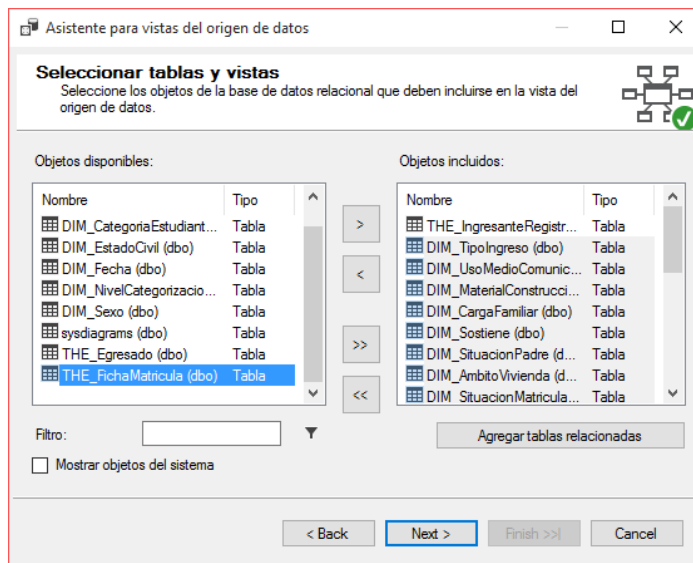


Figura 66. Selección de tablas.

Fuente: Propia.

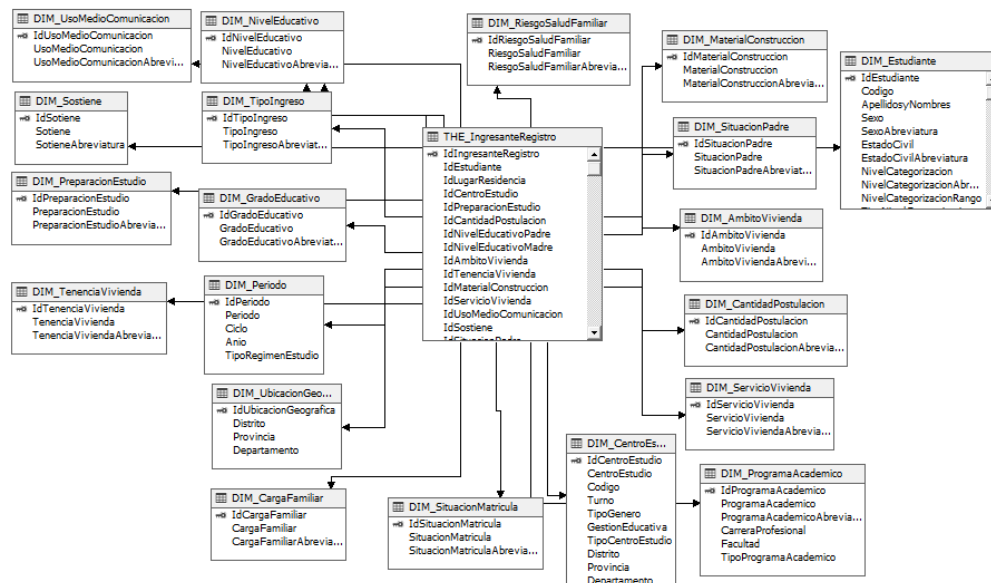
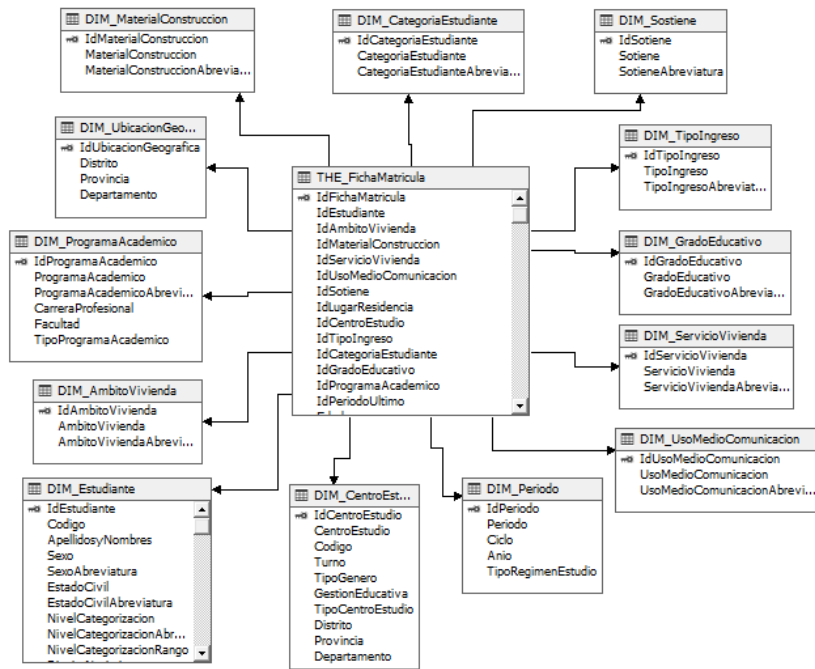


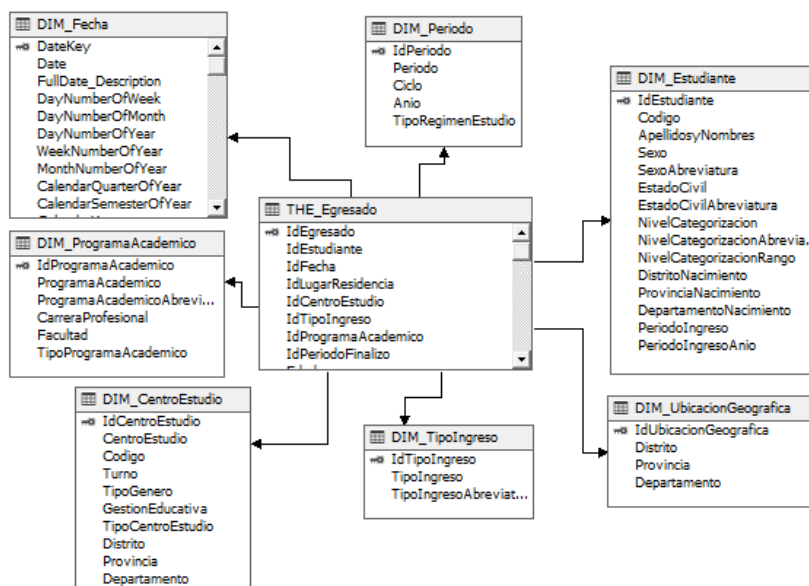
Figura 67. Vista de datos de ingresante registro.

Fuente: Propia.



**Figura 68.** Vista de datos de ficha de matrícula.

Fuente: Propia.



**Figura 69.** Vista de datos de egresado.

Fuente: Propia.

#### 5.2.9.4. Crear cubos

En esta sección seleccionamos el método de creación, las dimensiones, tabla de hechos y sus medidas para la modelo dimensional.

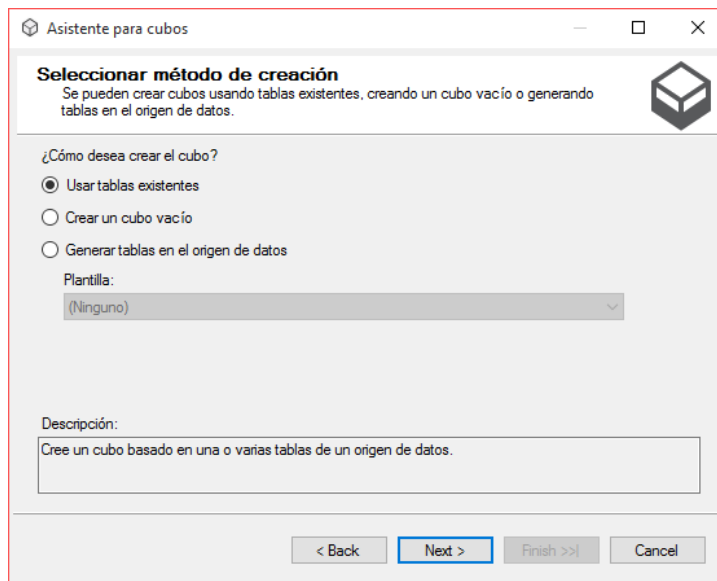


Figura 70. Selección de método de creación.

Fuente: Propia.

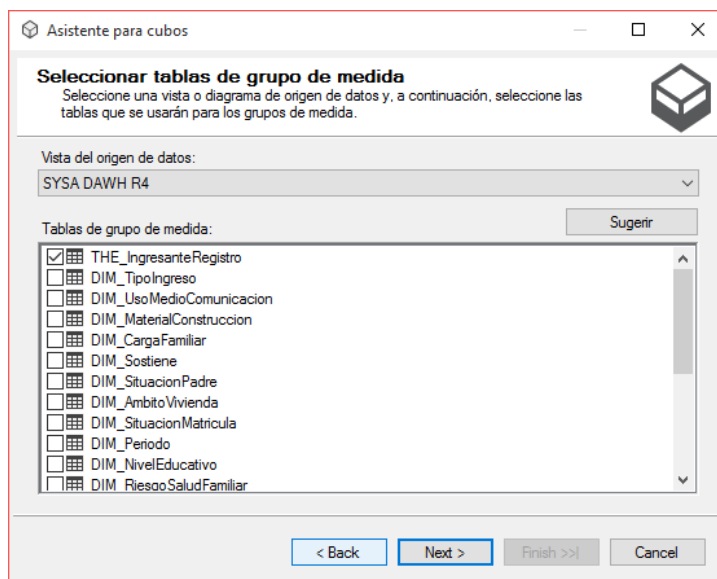
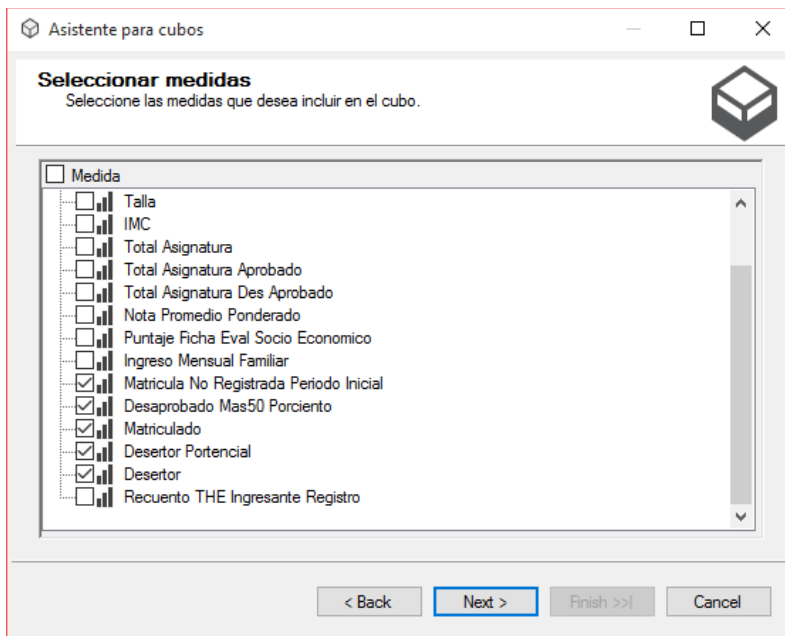
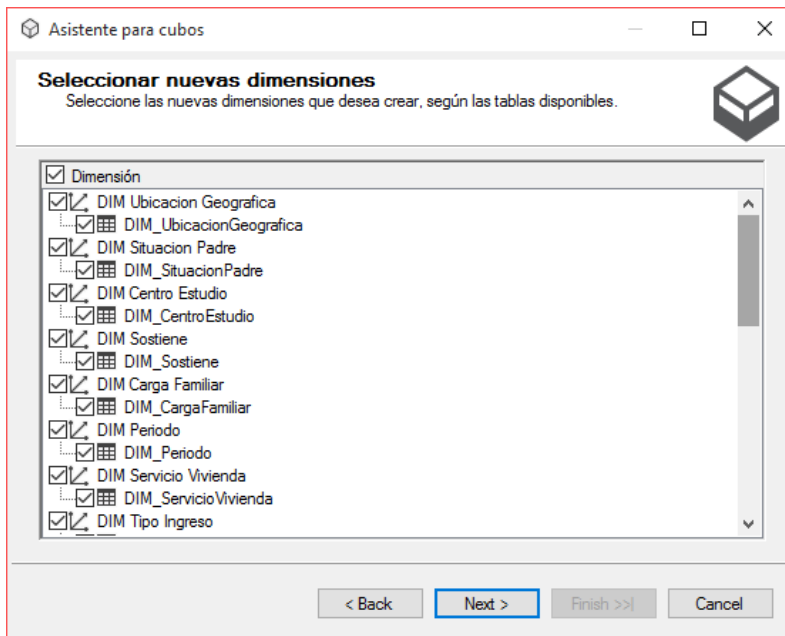


Figura 71. Selección de la tabla de hecho ingresante registro.

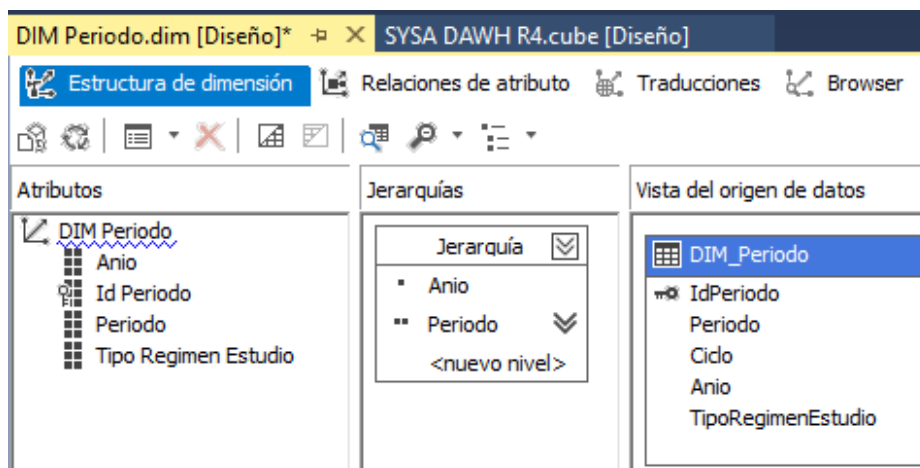
Fuente: Propia.



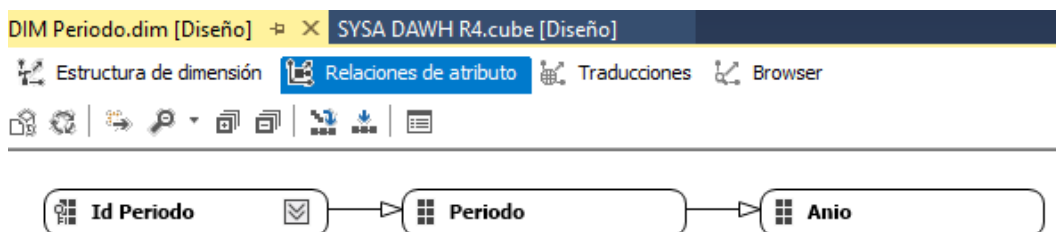
**Figura 72.** Selección de medidas de la tabla de hecho ingresante registro.  
Fuente: Propia.



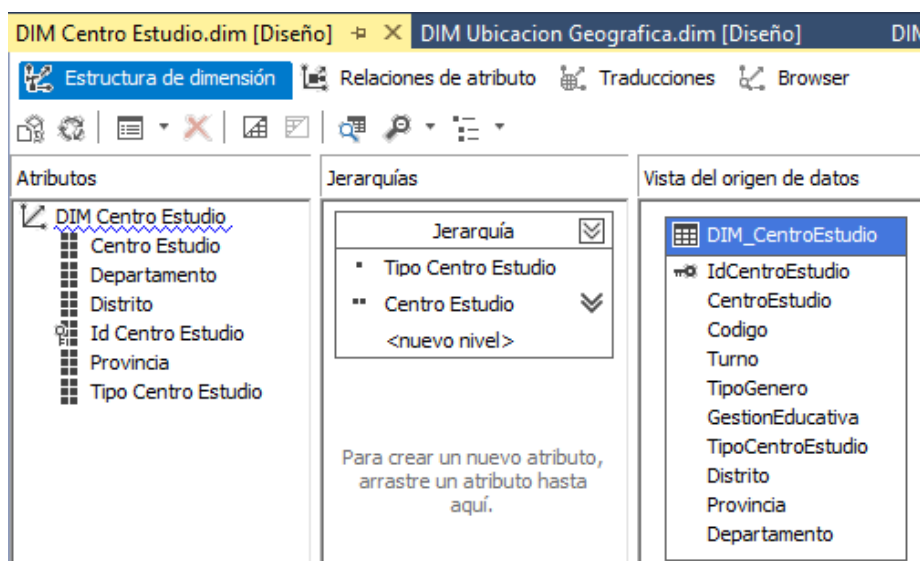
**Figura 73.** Selección de dimensiones.  
Fuente: Propia.



**Figura 74.** Selección de atributos y jerarquías de la dimensión periodo.  
Fuente: Propia.



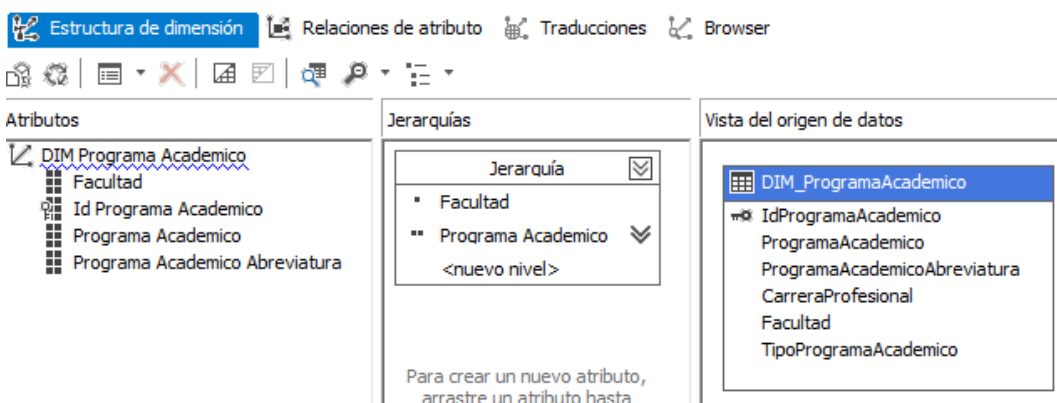
**Figura 75.** Selección de relaciones de atributos de la dimensión periodo.  
Fuente: Propia.



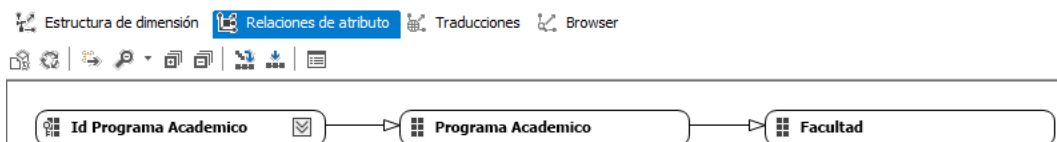
**Figura 76.** Selección de atributos y jerarquías de la dimensión centro de estudios.  
Fuente: Propia.



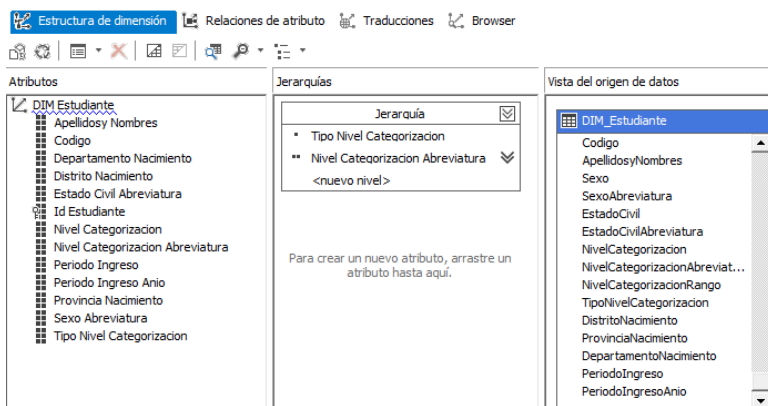
**Figura 77.** Selección de relaciones de atributos de la dimensión centro de estudios.  
Fuente: Propia.



**Figura 78.** Selección de atributos y jerarquías de la dimensión programa académicos.  
Fuente: Propia.



**Figura 79.** Selección de relaciones de atributos de la dimensión programa académicos.  
Fuente: Propia.



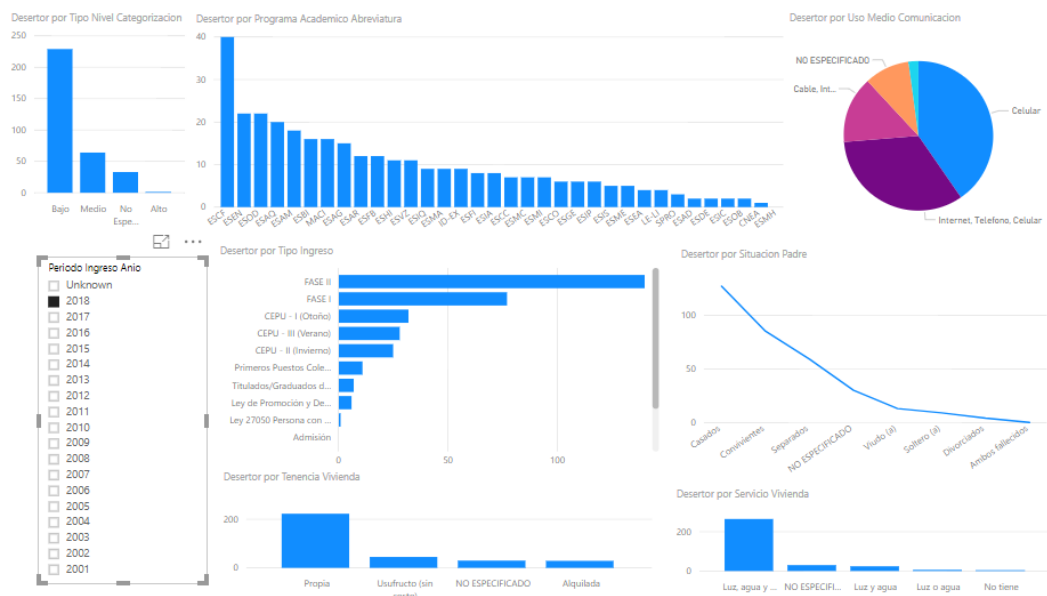
**Figura 80.** Selección de atributos y jerarquías de la dimensión estudiante.  
Fuente: Propia.



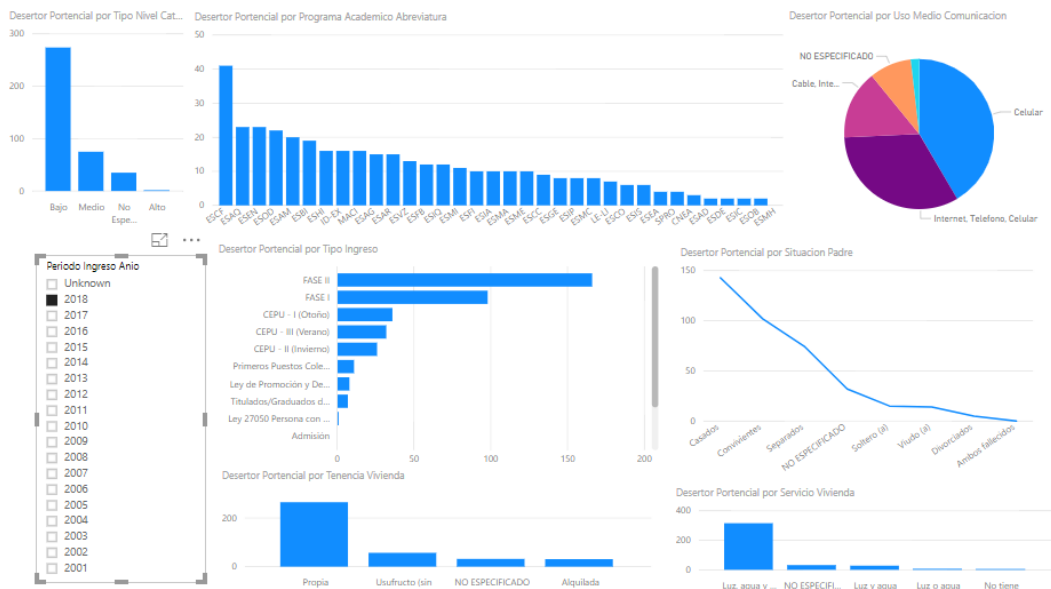
**Figura 81.** Selección de relaciones de atributos de la dimensión estudiante.  
Fuente: Propia.

### 5.2.9.5. Explorar Resultados

En esta sección se exploró los resultados de ingresantes, estudiantes desertores y desertores potenciales por distintos criterios, seguidamente se realizó también con los egresados que no tramitaron sus documentos académicos.



**Figura 82.** Desertores de los ingresantes.  
Fuente: Propia.



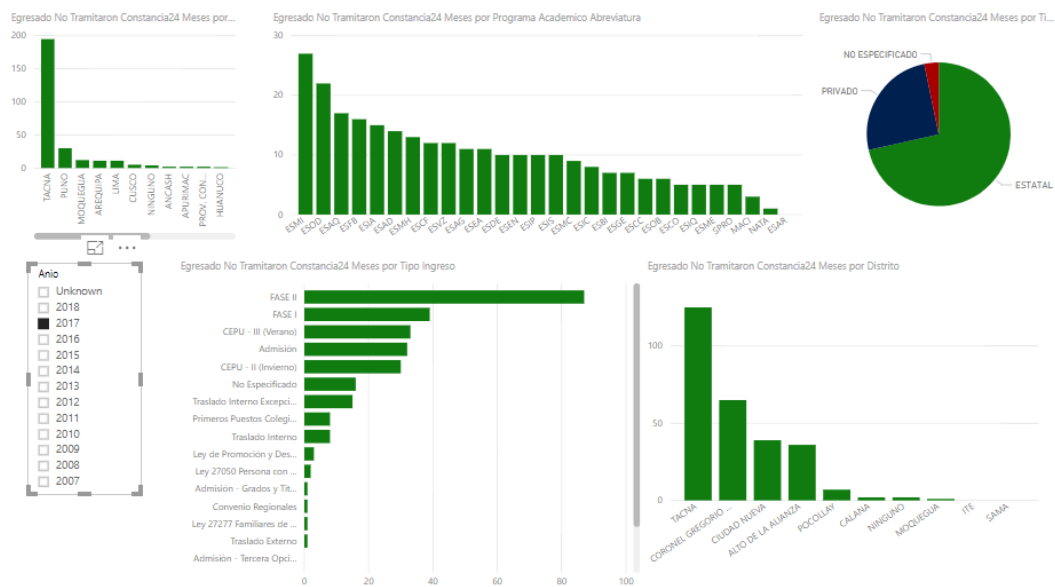
**Figura 83.** Desertores potenciales de los ingresantes.  
Fuente: Propia.



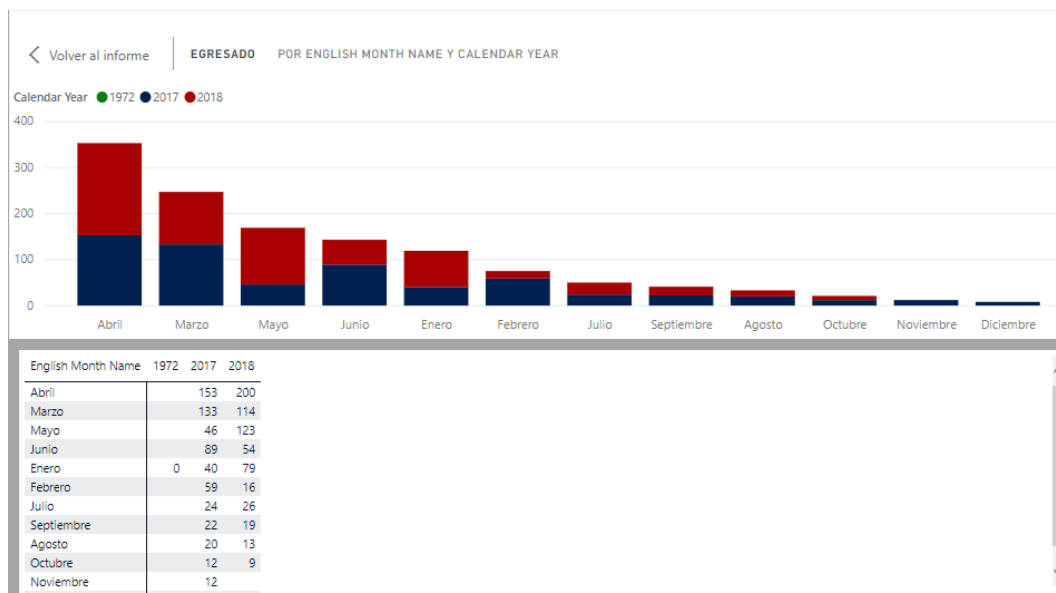
**Figura 84.** Desertores de los matriculados.  
Fuente: Propia.



**Figura 85.** Desertores potenciales de los matriculados.  
Fuente: Propia.



**Figura 86.** Egresados que no tramitaron sus documentos académicos.  
Fuente: Propia.



**Figura 87.** Egresados que tramitaron sus documentos académicos.

Fuente: Propia.

### 5.2.9.6. Exploración de estructuras de minería de datos

En esta sección se trabajó con una base de datos relacional, seguidamente se utilizó las técnicas de minería de datos para minar la zona de Ingresante Registro y serie temporal para establecer los campos de entradas y de predicciones. Como se puede ver en el anexo 6 como mucho más detalle.

## 5.3. Selección de una metodología para el desarrollo del proyecto de Minería de Datos

Existen varias metodologías para el desarrollo de proyectos de minería de datos como lo describen el autor Moine (2013) en su investigación denominada "Metodologías para el descubrimiento de conocimiento en

bases de datos: un estudio comparativo” (p.1). Moine realiza un análisis descriptivo de las metodologías mencionadas a continuación:

Cross Industry Standard Process for Data Mining (CRISP-DM) es una metodología más usadas por los investigadores y la comunidad académica por el nivel detalle con el que describen las tareas en cada fase del proceso, y además también porque incorporan actividades de gestión proyecto.

La metodología SEMMA cuyo acrónimo proviene de las cinco fases de sus procesos: Sample, Explore, Modify, Model, Assess. De acuerdo con Moine (2013), esta metodología está enfocada al modelado de la minería de datos en otros términos en sus procesos de selección, exploración y modelado de grandes volúmenes de datos para descubrir patrones de negocio desconocidos.

La metodología Catalyst, conocida como P3TQ (Product, Place, Price, Time, Quantity), fue propuesta por Dorian Pyle en el año 2003. “Esta metodología está formada por dos partes: Metodología para el Modelado del Negocio y Metodología para la Minería de Datos” (Moine, 2013, p.24).

Para este trabajo de investigación se está escogiendo la metodología de CRISP-DM, por el nivel detalle con el que describen las tareas en cada fase del proceso.

## **5.4. Metodología CRISP-DM**

### **5.4.1. Fase de comprensión del negocio**

#### **5.4.1.1. Determinar los objetivos del negocio**

El objetivo es realizar las predicciones de los estudiantes ingresantes, matriculados que tiene la probabilidad de desertar a partir de la información de las bases de datos de Registro Académico, Bienestar Universitario de la UNJBG.

Contexto: Considerando la situación del negocio el presente trabajo tiene lugar en la Dirección Académica de Actividades y Servicios Académicos y la Dirección de Bienestar Universitaria de la UNJBG, donde la información a gestionar parte de las bases de datos de REAC y BUN.

Objetivos del negocio: El objetivo es realizar las predicciones de datos fiables de los nuevos estudiantes ingresantes y estudiantes matriculados a partir de la información de los estudiantes antiguos registradas en la base de datos de REAC y BUN. Resultado de esta nueva información nos permita tomar medidas y buenas decisiones con respecto a los estudiantes desertores.

Criterios de éxito: Desde el punto de vista del negocio se considera la posibilidad de éxito el predecir los estudiantes desertores, igualmente, minimizar el riesgo de la deserción estudiantil.

#### **5.4.1.2. Evaluación de la situación**

La plataforma de trabajo es Windows, además cuenta con un gestor de base de datos SQL Server para la administración de las bases de datos de REAC y BUN, para esta investigación se tomó la información de los periodos 2012 hasta 2018 donde incluye datos: académicos, socioeconómicos y individuales.

Recursos de software y hardware: En cuanto a los recursos de software, se dispone del software Microsoft SQL Server, para realizar las tareas de minería de datos. Con respecto a los recursos de hardware la UNJBG cuenta con un servidor HPE ProLiant DL380 Gen10 donde se almacena toda la información vinculada a la parte académica y socioeconómica.

Requisitos, supuestos y restricciones: Con respecto de los requisitos se necesita la autorización del director de DASA y con respecto a los supuestos restricciones no existe ningún tipo de restricción frente a la obtención de la información de los estudiantes vinculada a la parte académica y socioeconómica.

Costos y beneficios: Esta investigación no genera ningún costo adicional a la UNJBG ya que las informaciones obtenidas son de las bases de datos de REAC y BUN. Además, los activos fijos son proporcionados por la UNJBG, frente a los beneficios, se podrá tener la información oportuna de los estudiantes desertores y esto contribuirá a tomar buenas decisiones o medidas correctivas.

#### **5.4.1.3. Determinar los objetivos de la minería de datos**

Entre los objetivos de minería de datos se tiene:

- Predecir los estudiantes desertores teniendo en cuenta los criterios individuales, académicos y socioeconómicos.
- Predecir las tendencias de los estudiantes desertores durante los próximos cinco años.

#### **5.4.1.4. Generación del plan del proyecto**

El proyecto se organizó en las siguientes etapas:

- Etapa 1: Análisis de requerimientos del proyecto.
- Etapa 2: Análisis de información de la estructura de los datos y de tablas de la base de datos REAC, BUN.
- Etapa 3: Preparación de los datos que comprende selección, limpieza, conversión y formateo, esto para facilitar la minería de datos sobre la información.

- Etapa 4: Elección de las técnicas de modelado apropiadas y ejecución de las mismas.
- Etapa 5: Análisis de los datos de las técnicas aplicadas.
- Etapa 6: Generar informes con los resultados obtenidos en función a los objetivos del negocio y los criterios de éxito establecidos.
- Etapa 7: Preparación de los resultados finales.

#### **5.4.1.5. Evaluación inicial de herramientas y técnicas**

La herramienta que se usó para llevar a cabo ésta minería de datos es el software Microsoft SQL Server Data Tools y en cuanto a las técnicas que se van a emplear para la extracción de conocimiento es:

##### Predictivas

- Árboles de decisión
- Regresión Logística
- Redes Neuronales
- Redes Bayesianas
- Series temporales

##### Descriptivas

- Agrupamiento

## 5.4.2. Fase de comprensión de los datos

### 5.4.2.1. Recopilación de datos iniciales

Los datos han sido recolectados de las bases de datos REAC, BUN, en los pasos descritos a continuación:

- Para la gestión y administración de los datos se trabajó con el software Microsoft SQL Server Management Studio.
- Se crearon distinta consulta SQL apropiadas de los módulos de matrículas y bienestar universitario.
- Se extrajo información de los modelos dimensionales construidos con anterioridad.
- Los datos obtenidos de las bases de datos comprenden del periodo 2012 al 2018.

### 5.4.2.2. Descripción de los datos

Después de describir, diseñar y construir modelos dimensionales de los estudiantes desertores. Nos toca ahora describir las nuevas tablas que nos permitirá predecir y estimar las tendencias de los nuevos desertores, como se describe a continuación:

**Tabla 61** Descripción de la tabla registro de ingresante

| Nombre de Columna    | Tipo de Dato | Permitir | Descripción           |
|----------------------|--------------|----------|-----------------------|
| IdIngresanteRegistro | int          | Not Null | <b>Clave Primaria</b> |
| Codigo               | nvarchar(15) | Null     | Código de Ingreso     |

| <b>Nombre de Columna</b> | <b>Tipo de Dato</b> | <b>Permitir</b> | <b>Descripción</b>            |
|--------------------------|---------------------|-----------------|-------------------------------|
| Estudiante               | nvarchar(250)       | Null            | Apellidos y Nombres           |
| Sexo                     | nvarchar(30)        | Null            | Sexo                          |
| EstadoCivil              | nvarchar(30)        | Null            | Estado Civil                  |
| NivelCategorizacion      | nvarchar(50)        | Null            | Nivel de Categorización       |
| NivelCategorizacionAbrev | nvarchar(30)        | Null            | Abreviatura de Categorización |
| TipoNivelCategorizacion  | nvarchar(30)        | Null            | Tipo de Nivel Categorización  |
| DistritoNacimiento       | nvarchar(250)       | Null            | Distrito de Nacimiento        |
| ProvinciaNacimiento      | nvarchar(250)       | Null            | Provincia de Nacimiento       |
| DepartamentoNacimiento   | nvarchar(250)       | Null            | Departamento de Nacimiento    |
| PeriodoIngreso           | nvarchar(50)        | Null            | Periodo de Ingreso            |
| PeriodoIngresoAnio       | nvarchar(30)        | Null            | Año de Periodo de Ingreso     |
| LugarResidencia          | nvarchar(250)       | Null            | Lugar de Residencia           |
| CentroEstudio            | nvarchar(250)       | Null            | Centro de Estudio             |
| PreparacionEstudio       | nvarchar(250)       | Null            | Preparación de Estudio        |
| CantidadPostulacion      | nvarchar(250)       | Null            | Cantidad de Postulación       |
| NivelEducativoPadre      | nvarchar(250)       | Null            | Nivel Educativo del Padre     |
| NivelEducativoMadre      | nvarchar(250)       | Null            | Nivel Educativo de la Madre   |
| AmbitoVivienda           | nvarchar(250)       | Null            | Ambito de Vivienda            |
| TenenciaVivienda         | nvarchar(250)       | Null            | Tenencia de Vivienda          |
| MaterialConstruccion     | nvarchar(250)       | Null            | Material de Construcción      |
| ServicioVivienda         | nvarchar(250)       | Null            | Servicio de Vivienda          |
| UsoMedioComunicacion     | nvarchar(250)       | Null            | Uso de Medio Comunicación     |
| Sostiene                 | nvarchar(250)       | Null            | Familiar que Sostiene         |
| SituacionPadre           | nvarchar(250)       | Null            | Situación del Padre           |
| RiesgoSaludFamiliar      | nvarchar(250)       | Null            | Riesgo Salud Familiar         |
| CargaFamiliar            | nvarchar(250)       | Null            | Carga Familiar                |

| Nombre de Columna        | Tipo de Dato  | Permitir | Descripción                 |
|--------------------------|---------------|----------|-----------------------------|
| GradoEducativo           | nvarchar(250) | Null     | Grado Educativo             |
| TipoIngreso              | nvarchar(250) | Null     | Tipo de Ingreso             |
| ProgramaAcademico        | nvarchar(250) | Null     | Programa Académico          |
| Edad                     | float         | Null     | Edad                        |
| Peso                     | float         | Null     | Peso                        |
| Talla                    | float         | Null     | Talla                       |
| IMC                      | float         | Null     | Indice de Masa Corporal     |
| TotalAsignatura          | int           | Null     | Total de Asignatura         |
| TotalAsignaturaAprobado  | int           | Null     | Total Asignatura Aprobada   |
| TotalAsignaturaDesAprob  | int           | Null     | Total Asignatura Desprobada |
| NotaPromedioPonderado    | float         | Null     | Nota de Promedio Ponderado  |
| PuntajeFichaEvalSocioEco | float         | Null     | Puntaje Socioeconómico      |
| IngresoMensualFamiliar   | float         | Null     | Ingreso Mensual Familiar    |
| MatriculaNoRegistradaPI  | int           | Null     | Matrícula No Registrada     |
| DesaprobadoMas50PC       | int           | Null     | Desaprobado Mas 50%         |
| Matriculado              | int           | Null     | Matriculado                 |
| DesertorPortencial       | int           | Null     | Desertor Potencial          |
| Desertor                 | int           | Null     | Desertor                    |

Fuente: Elaboración Propia.

**Tabla 62** Descripción de la tabla registro de serie temporal

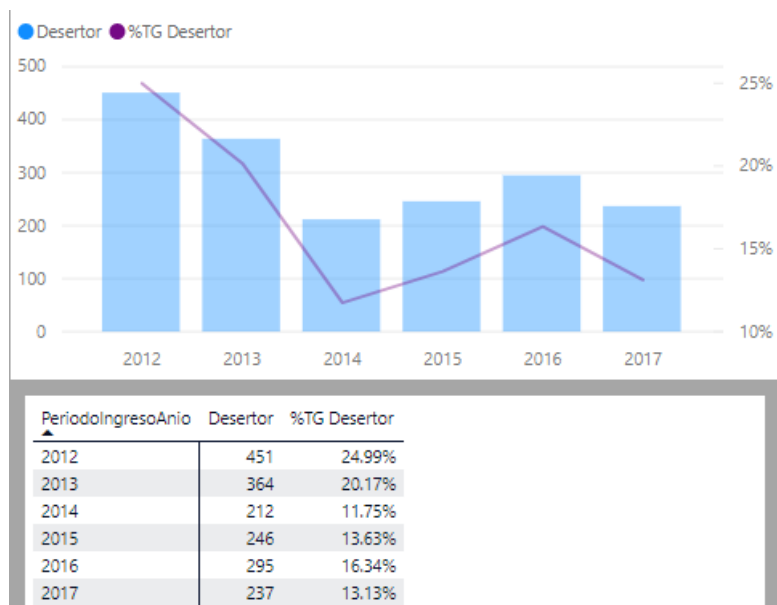
| Nombre de Columna       | Tipo de Dato  | Permitir | Descripción           |
|-------------------------|---------------|----------|-----------------------|
| IdRegistroSerieTemporal | int           | Not Null | <b>Clave Primaria</b> |
| FechaPeriodo            | datetime      | Null     | Fecha del Periodo     |
| ProgramaAcademico       | nvarchar(250) | Null     | Programa Académico    |

| Nombre de Columna        | Tipo de Dato | Permitir | Descripción                |
|--------------------------|--------------|----------|----------------------------|
| Edad                     | float        | Null     | Edad                       |
| NotaPromedioPonderado    | float        | Null     | Nota de Promedio Ponderado |
| PuntajeFichaEvalSocioEco | float        | Null     | Puntaje Socioeconómico     |
| IngresoMensualFamiliar   | float        | Null     | Ingreso Mensual Familiar   |
| Matriculado              | int          | Null     | Matriculado                |
| DesertorPortencial       | int          | Null     | Desertor Potencial         |
| Desertor                 | int          | Null     | Desertor                   |

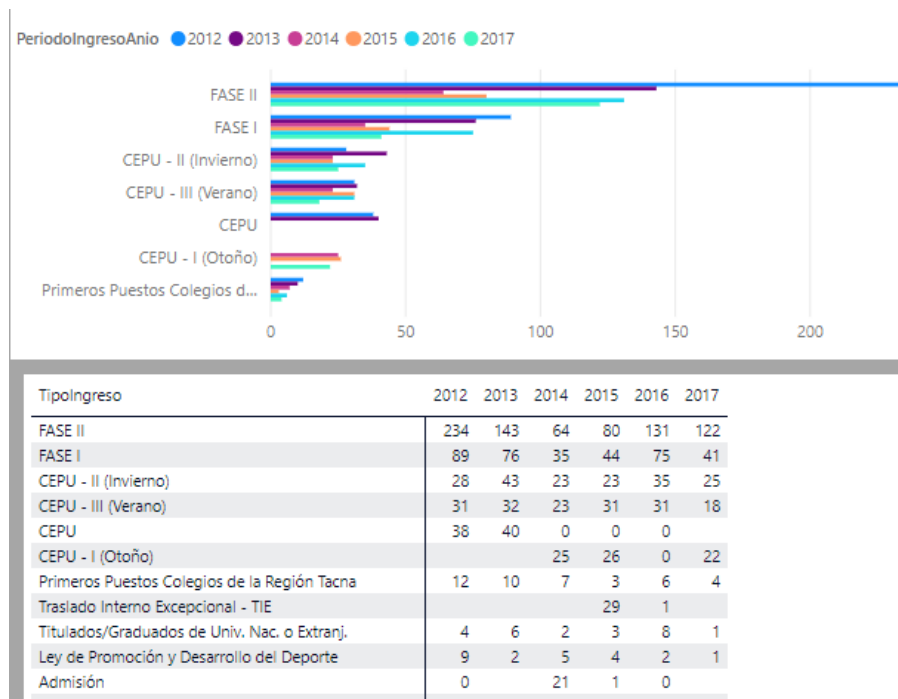
Fuente: Elaboración Propia.

#### 5.4.2.3. Exploración de los datos

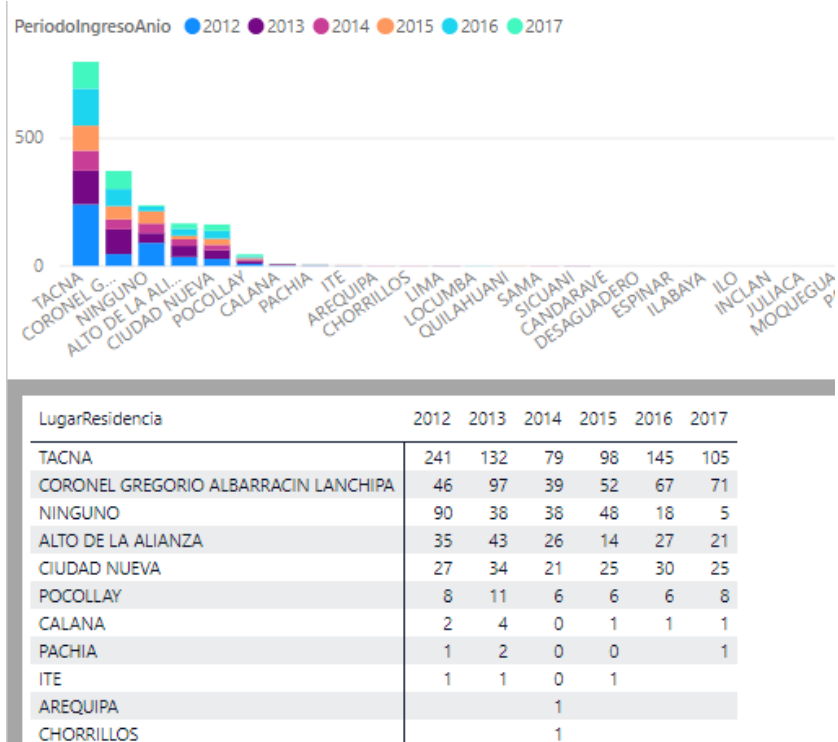
En esta etapa se realiza la exploración de datos sobre la información extraída de los periodos lectivos 2012 hasta el 2017 bajo el software empleado Power BI. Los resultados obtenidos van a mostrar información estadística de los estudiantes desertores por centro de estudios, lugar de residencia, lugar de nacimiento, año de ingreso, tipo de ingresos, programa de académicos, tipo de nivel de categorización, uso de medios de comunicación, nivel educativo del padre.



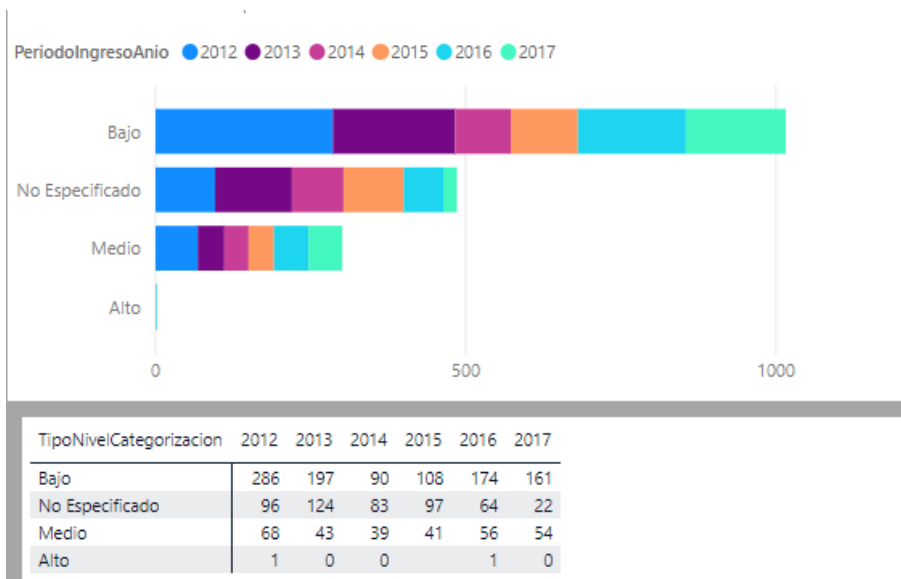
**Figura 88.** Estudiantes desertores por año de ingreso.  
Fuente: Propia



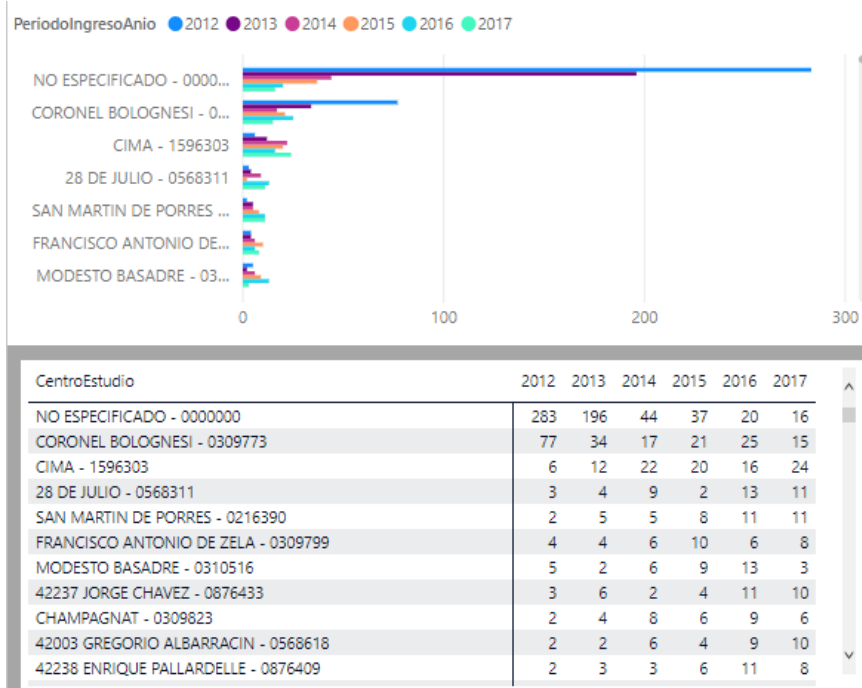
**Figura 89.** Estudiantes desertores por tipo de ingreso y año de ingreso.  
Fuente: Propia



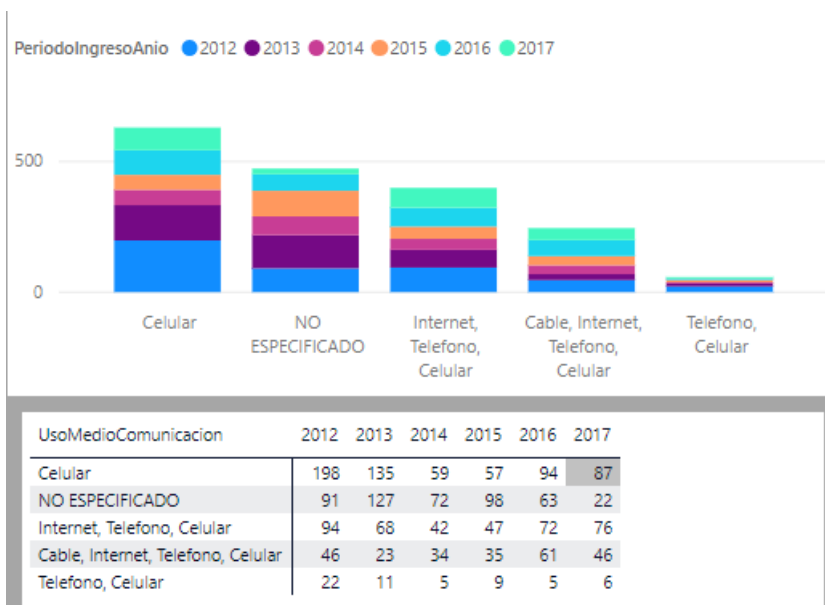
**Figura 90.** Estudiantes desertores por lugar de residencia y año de ingreso.  
Fuente: Propia



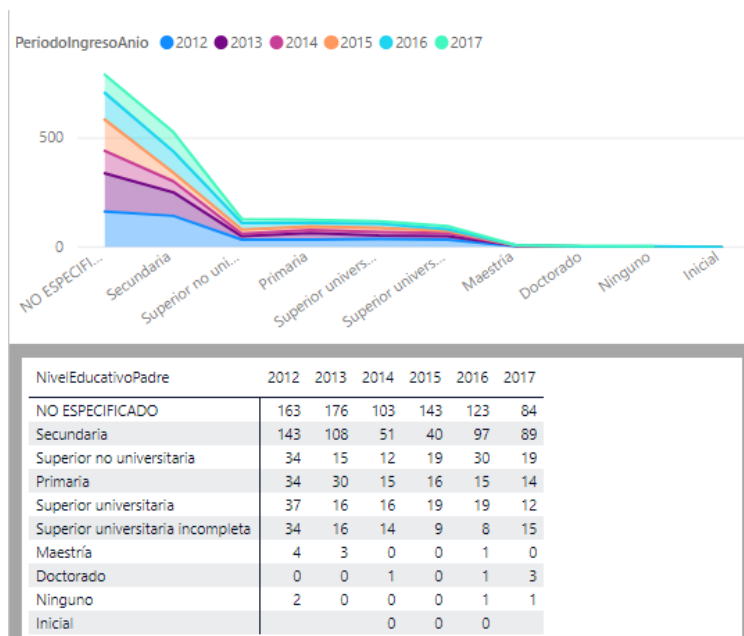
**Figura 91.** Estudiantes desertores por nivel de categorización y año de ingreso.  
Fuente: Propia



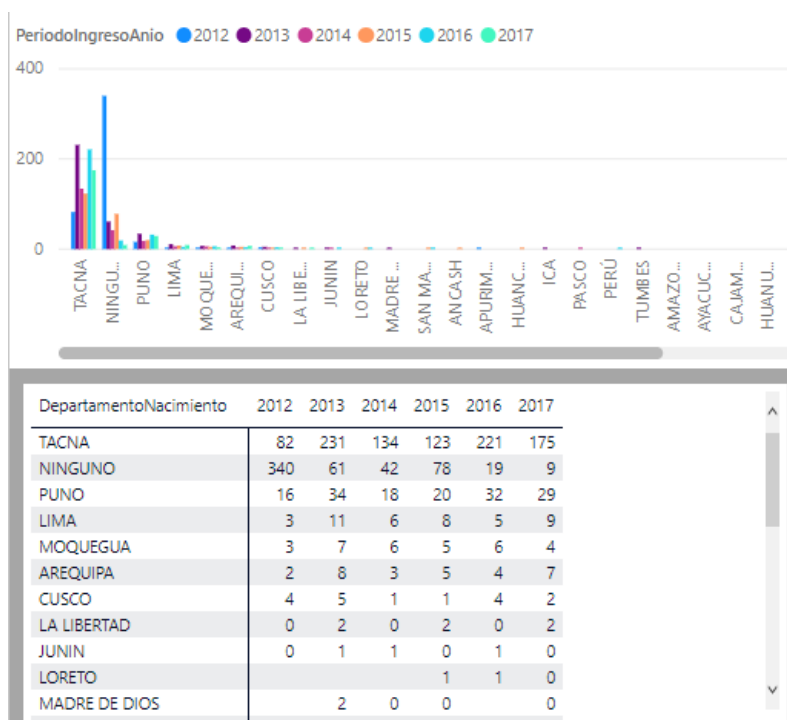
**Figura 92.** Estudiantes desertores por centro de estudios y año de ingreso.  
Fuente: Propia



**Figura 93.** Estudiantes desertores por uso de medios de comunicación y año de ingreso.  
Fuente: Propia



**Figura 94.** Estudiantes desertores por nivel educativo del padre y año de ingreso.  
Fuente: Propia



**Figura 95.** Estudiantes desertores por lugar de nacimiento y año de ingreso.  
Fuente: Propia

| ProgramaAcademico  | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|--|------|------|------|------|------|------|
| Carrera Profesional de Física Aplicada                       | 34   | 30   | 9    | 6    | 11   | 5    |
| Carrera Profesional de Ingeniería en Informática y Sistemas  | 16   | 20   | 11   | 14   | 19   | 15   |
| Carrera Profesional de Ingeniería en Industrias Alimentarias | 30   | 10   | 12   | 13   | 12   | 17   |
| Carrera Profesional de Medicina Veterinaria y Zootecnia      | 17   | 19   | 8    | 15   | 23   | 10   |
| Carrera Profesional de Biología - Microbiología              | 24   | 20   | 8    | 9    | 14   | 14   |
| Carrera Profesional de Ingeniería Química                    | 17   | 21   | 10   | 9    | 19   | 13   |
| Carrera Profesional de Ingeniería Metalúrgica                | 25   | 14   | 11   | 11   | 13   | 8    |
| Carrera Profesional de Ingeniería Geológica - Geotecnia      | 22   | 12   | 16   | 9    | 7    | 12   |
| Carrera Profesional de Ingeniería de Minas                   | 21   | 11   | 13   | 10   | 13   | 8    |
| Carrera Profesional de Ingeniería Pesquera                   | 24   | 15   | 4    | 8    | 11   | 6    |
| Carrera Profesional de Arquitectura                          | 15   | 15   | 8    | 4    | 10   | 10   |

**Figura 96.** Estudiantes desertores por carrera profesional y año de ingreso.

Fuente: Propia

#### 5.4.2.4. Validación de los datos

En esta etapa no a merita la validación de datos por que la información proviene de los modelos dimensionales que dentro de sus etapas se realizó el proceso ETL donde nos garantiza la consistencia de los datos.

#### 5.4.3. Fase de preparación de los datos

A partir de los datos almacenados en los modelos dimensionales se ha seleccionado, limpiado, integrado, formateado y finalmente se ha construido los nuevos datos, de tal manera esta organización de los datos facilitará una buena aplicación de las técnicas de minería de datos.

##### 5.4.3.1. Selección de datos

Se ha seleccionado los atributos, medidas de las dimensiones y de la tabla de hecho del cubo ingresante de registro para esta etapa, como se detalla a continuación:

**Tabla 63** Selección de atributos y medidas de la tabla registro ingresante

| <b>Nombre del Objeto</b> | <b>Descripción</b>  |
|--------------------------|---|
| DIM_Estudiante           | Codigo, ApellidosyNombres,Sexo, SexoAbreviatura, EstadoCivil, EstadoCivilAbreviatura, NivelCategorizacion, NivelCategorizacionAbreviatura, NivelCategorizacionRango, TipoNivelCategorizacion, DistritoNacimiento, ProvinciaNacimiento, DepartamentoNacimiento, PeriodoIngreso, PeriodoIngresoAnio |
| DIM_UbicacionGeografica  | Distrito  |
| DIM_CentroEstudio        | CentroEstudio   |
| DIM_PreparacionEstudio   | PreparacionEstudio  |
| DIM_NivelEducativo       | NivelEducativo (Padre, Madre)   |
| DIM_AmbitoVivienda       | AmbitoVivienda  |
| DIM_TenenciaVivienda     | TenenciaVivienda  |
| DIM_MaterialConstruccion | MaterialConstruccion  |
| DIM_ServicioVivienda     | ServicioVivienda  |
| DIM_UsoMedioComunicacion | UsoMedioComunicacion  |
| DIM_Sostiene             | Sostiene  |
| DIM_SituacionPadre       | SituacionPadre  |
| DIM_RiesgoSaludFamiliar  | RiesgoSaludFamiliar   |
| DIM_GradoEducativo       | GradoEducativo  |
| DIM_TipoIngreso          | TipoIngreso   |
| DIM_ProgramaAcademico    | ProgramaAcademico   |
| THE_IngresanteRegistro   | Edad  |
| THE_IngresanteRegistro   | TotalAsignaturaDesAprobado  |
| THE_IngresanteRegistro   | NotaPromedioPonderado   |
| THE_IngresanteRegistro   | PuntajeFichaEvalSocioEconomico  |
| THE_IngresanteRegistro   | IngresoMensualFamiliar  |

| Nombre del Objeto      | Descripción                         |
|------------------------|-------------------------------------|
| THE_IngresanteRegistro | MatriculaNoRegistradaPeriodoInicial |
| THE_IngresanteRegistro | DesaprobadoMas50Porciento           |
| THE_IngresanteRegistro | Matriculado                         |
| THE_IngresanteRegistro | DesertorPortencial                  |
| THE_IngresanteRegistro | Desertor                            |

Fuente: Elaboración Propia.

**Tabla 64** Selección de atributos y medidas de la tabla registro serie temporal

| Nombre del Objeto      | Descripción                    |
|------------------------|--------------------------------|
| DIM_Estudiante         | PeriodoIngresoAnio             |
| DIM_ProgramaAcademico  | ProgramaAcademico              |
| THE_IngresanteRegistro | Edad                           |
| THE_IngresanteRegistro | NotaPromedioPonderado          |
| THE_IngresanteRegistro | PuntajeFichaEvalSocioEconomico |
| THE_IngresanteRegistro | IngresoMensualFamiliar         |
| THE_IngresanteRegistro | Matriculado                    |
| THE_IngresanteRegistro | DesertorPortencial             |
| THE_IngresanteRegistro | Desertor                       |

Fuente: Elaboración Propia.

#### 5.4.3.2. Limpieza de datos

En esta etapa no a merita la limpia de los datos de las tablas registro de ingresante y serie temporal por que la información proviene de los modelos dimensionales dado que dentro de su etapa se realizó un proceso ETL.

### 5.4.3.3. Construcción de datos

En esta etapa se ha construido nuevos datos para las tablas de registro ingresante y serie temporal como mostramos a continuación:

**Tabla 65** Nuevos datos de la tabla registro de ingresante

| Nombre de la Columna       | Descripción  |
|----------------------------|--|
| TotalAsignaturaAprobado    | Cálculo de Ida suma de la asignatura aprobado              |
| TotalAsignaturaDesAprobado | Cálculo de la suma de la asignatura desaprobado            |
| TotalAsignatura            | Cálculo de la suma de la asignatura aprobado y desaprobado |
| NotaPromedioPonderado      | Cálculo del promedio arimético de la nota final.           |

Fuente: Elaboración Propia.

**Tabla 66** Nuevos datos de la tabla registro de serie temporal

| Nombre de la Columna           | Descripción  |
|--------------------------------|--|
| FechaPeriodo                   | Asignación del día 1 del mes de abril y del año del periodo de ingreso |
| Edad                           | Cálculo del promedio arimético de la edad                              |
| NotaPromedioPonderado          | Cálculo del promedio ponderado de la nota final.                       |
| PuntajeFichaEvalSocioEconomico | Cálculo del promedio arimético del puntaje de la ficha socioeconómico  |
| IngresoMensualFamiliar         | Cálculo del promedio arimético del ingreso mensual familiar            |
| Matriculado                    | Cálculo de la suma de los matriculados                                 |
| DesertorPortencial             | Cálculo de la suma de los desertores potenciales                       |
| Desertor                       | Cálculo de la suma de los desertores                                   |

Fuente: Elaboración Propia.

#### 5.4.3.4. Integración de datos

En la tabla de registro de serie de temporal se ha integrado el año del periodo de ingreso de la dimensión de estudiante para la generación del nuevo campo fecha del periodo.

#### 5.4.4. Fase de modelado

##### 5.4.4.1. Selección de la técnica de modelado

En este trabajo de investigación se ha seleccionado estas técnicas de minerías de datos para la predicción de los estudiantes desertores de la UNJBG.

- Árboles de decisión
- Regresión Logística
- Redes Bayesianas
- Series temporales

##### 5.4.4.2. Diseño de la evaluación

En esta etapa se establece los criterios de evaluación de las técnicas de minería de datos. Objetivamente el parámetro más importante es la precisión y subjetivamente es la interpretación de los resultados. A continuación, mostraremos el diseño de evaluación.

**Tabla 67** Diseño de evaluación de las técnicas de DM - Ingresante

| Técnica de DM       | Precisión | Tasa Error | Curva ROC |
|---------------------|-----------|------------|-----------|
| Árboles de decisión | a         | A          | X         |

| Técnica de DM       | Precisión | Tasa Error | Curva ROC |
|---------------------|-----------|------------|-----------|
| Regresión Logística | b         | B          | Y         |
| Redes Bayesianas    | c         | C          | Z         |

Fuente: Elaboración Propia.

**Tabla 68** Diseño de evaluación de las técnicas de DM – serie temporal

| Técnica de DM     | Precisión | Tasa Error | Curva ROC |
|-------------------|-----------|------------|-----------|
| Series temporales | b         | B          | Y         |

Fuente: Elaboración Propia.

#### 5.4.4.3. Construcción del modelo

En esta etapa se ha considerado los mismos escenarios de aprendizaje para todas las técnicas de minería de datos.

Escenario de entrada

- Ámbito vivienda
- Carga familiar
- Centro de Estudio
- Departamento de nacimiento
- Lugar de residencia
- Estado civil
- Sexo
- Nivel educativo del padre
- Riesgo salud familiar

- Servicio de vivienda
- Tipo de nivel de categorización
- Tipo de ingreso
- Programa académico
- Edad
- Ingreso mensual familiar
- Puntaje de la ficha socioeconómico
- Nota promedio final
- Desaprobado más del 50 por ciento
- Total, de asignaturas desaprobadas
- Desertor potencial

#### Escenario de predicción

- Desertor

En el anexo 6 se detalla el desarrollo y la aplicación de las técnicas de minerías de datos a las tablas registro de ingresante y serie temporal, para esta etapa rescataremos algunos datos importantes que sirva para las técnicas de DM que veremos a continuación.

Para el esquema de registro de ingresante se trabajó con tres técnicas de mineras de datos

La primera es el Árbol de decisión donde se aplicó las formulas de la matriz de confusión y la curva ROC.

**Tabla 69** Matriz de confusión de árbol de decisión

| Arboles de decisión | Predicho negativo | Predicho positivo |
|---------------------|-------------------|-------------------|
| Real negativo       | 1566              | 40                |
| Real positivo       | 77                | 317               |

Fuente: Elaboración Propia.

La precisión global se aplica de la fórmula (2).

$$AC = \frac{1566 + 317}{1566 + 40 + 77 + 317} = 0,94$$

La tasa de error global se aplica de la fórmula (3).

$$TE = \frac{40 + 77}{1566 + 40 + 77 + 317} = 0,06$$

La tasa positiva verdadera se aplica de la fórmula (4).

$$TP = \frac{317}{394} = 0,80$$

La tasa negativa verdadera se aplica de la fórmula (5).

$$TN = \frac{1566}{1606} = 0,98$$

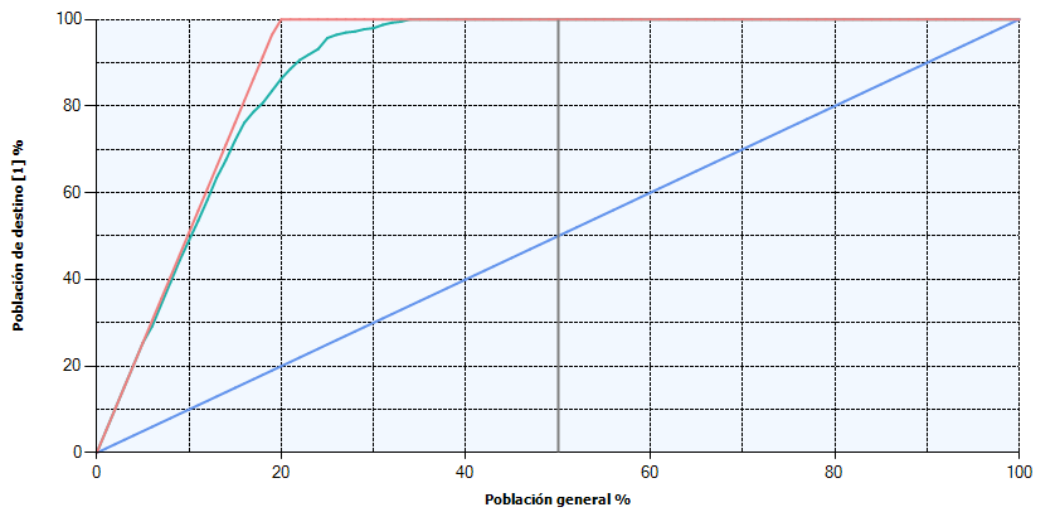
La tasa de falsos positivos se aplica de la fórmula (6).

$$PF = \frac{40}{1606} = 0,02$$

La tasa de falsos negativos se aplica de la fórmula (7).

$$FN = \frac{77}{394} = 0,20$$

Realizamos el análisis de la técnica mediante el uso de la Curva ROC ver la figura 83, en la que podemos apreciar que el modelo tiene un grado de certeza del 99%.



**Figura 97.** Cueva ROC de la técnica del árbol de decisión.

Fuente: Propia

La segunda es la regresión logística donde se aplicó las formulas de la matriz de confusión y la curva ROC.

**Tabla 70** Matriz de confusión de la regresión logística

| Regresión logística | Predicho negativo | Predicho positivo |
|---------------------|-------------------|-------------------|
| Real negativo       | 1497              | 109               |
| Real positivo       | 58                | 336               |

Fuente: Elaboración Propia.

La precisión global se aplica de la fórmula (2).

$$AC = \frac{1497 + 336}{1497 + 109 + 58 + 336} = 0,92$$

La tasa de error global se aplica de la fórmula (3).

$$TE = \frac{109 + 58}{1566 + 40 + 77 + 317} = 0,08$$

La tasa positiva verdadera se aplica de la fórmula (4).

$$TP = \frac{336}{394} = 0,85$$

La tasa negativa verdadera se aplica de la fórmula (5).

$$TN = \frac{1497}{1606} = 0,93$$

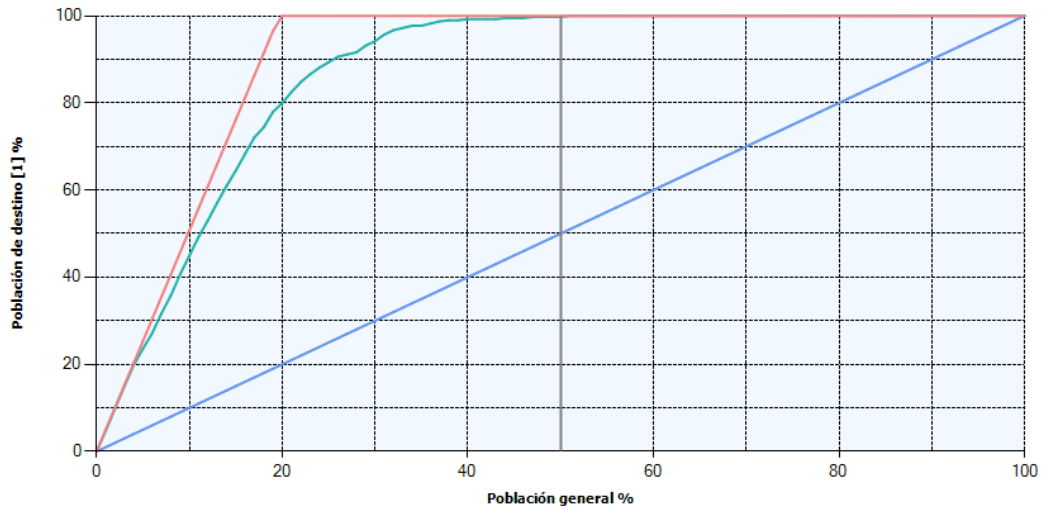
La tasa de falsos positivos se aplica de la fórmula (6).

$$PF = \frac{109}{1606} = 0,07$$

La tasa de falsos negativos se aplica de la fórmula (7).

$$FN = \frac{58}{394} = 0,15$$

Realizamos el análisis de la técnica mediante el uso de la Curva ROC ver la figura 84, en la que podemos apreciar que el modelo tiene un grado de certeza del 97%.



**Figura 98.** Cueva ROC de la técnica de la regresión logística.

Fuente: Propia

La tercera son las Redes Bayesianas donde se aplicó las fórmulas de la matriz de confusión y la curva ROC.

**Tabla 71** Matriz de confusión de las redes bayesianas

| Redes Bayesianas | Predicho negativo | Predicho positivo |
|------------------|-------------------|-------------------|
| Real negativo    | 1549              | 57                |
| Real positivo    | 179               | 215               |

Fuente: Elaboración Propia.

La precisión global se aplica de la fórmula (2).

$$AC = \frac{1549 + 215}{1549 + 57 + 179 + 215} = 0,88$$

La tasa de error global se aplica de la fórmula (3).

$$TE = \frac{57 + 179}{1549 + 57 + 179 + 215} = 0,12$$

La tasa positiva verdadera se aplica de la fórmula (4).

$$TP = \frac{215}{394} = 0,55$$

La tasa negativa verdadera se aplica de la fórmula (5).

$$TN = \frac{1549}{1606} = 0,96$$

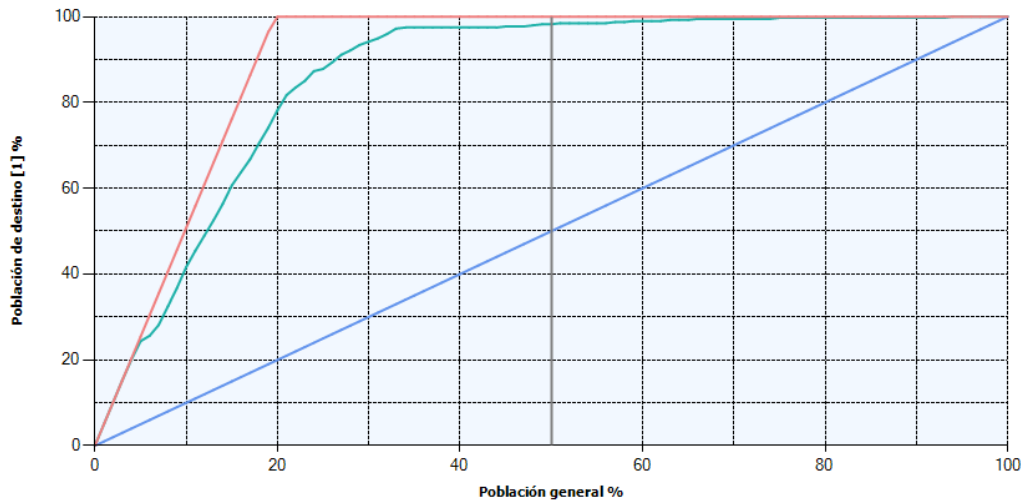
La tasa de falsos positivos se aplica de la fórmula (6).

$$PF = \frac{57}{1606} = 0,04$$

La tasa de falsos negativos se aplica de la fórmula (7).

$$FN = \frac{179}{394} = 0,45$$

Realizamos el análisis de la técnica mediante el uso de la Curva ROC ver la figura 85, en la que podemos apreciar que el modelo tiene un grado de certeza del 96%.



**Figura 99.** Cueva ROC de la técnica de redes bayesianas

Fuente: Propia

Para el esquema de registro de serie temporales solamente se trabajó con la técnica de minería de datos de series temporales para la estimación en el tiempo. Visto de otra manera se podrá estimar cuantos estudiantes desertarán dentro de de 5 años.

#### 5.4.4.4. Evaluación del modelo

En esta etapa se analiza que la técnica arboles de decisión tiene una mayor precisión de 0,94 y una tasa de error mínima de 0,06.

**Tabla 72** Resultados de evaluación de las técnicas de DM - ingresante

| Técnica de DM       | Precisión | Tasa Error | Curva ROC |
|---------------------|-----------|------------|-----------|
| Arboles de decisión | 0,94      | 0,06       | 0,99      |
| Regresión Logística | 0,92      | 0,08       | 0,97      |
| Redes Bayesianas    | 0,88      | 0,12       | 0,96      |

Fuente: Elaboración Propia.

Es apropiado realizar la validación de cruzada para todas las técnicas de minería de datos. A continuación, mostramos los datos recogidos del anexo 6.

**Tabla 73** Validación cruzada de los errores de las técnicas de DM

| Iteración | Arboles de decisión | Regresión Logística | Redes Bayesianas |
|-----------|---------------------|---------------------|------------------|
| 1         | 10                  | 59                  | 23               |
| 2         | 11                  | 46                  | 29               |
| 3         | 13                  | 44                  | 17               |
| 4         | 14                  | 48                  | 26               |
| 5         | 17                  | 56                  | 28               |
| 6         | 12                  | 43                  | 22               |
| 7         | 10                  | 48                  | 26               |
| 8         | 16                  | 44                  | 28               |
| 9         | 14                  | 48                  | 20               |
| 10        | 12                  | 46                  | 18               |

Fuente: Elaboración Propia.

Error de validación cruzada para la técnica de DM arboles de decisión en aplicación de la fórmula (1).

$$E = \frac{10 + 11 + 13 + 14 + 17 + 12 + 10 + 16 + 14 + 12}{10} = 12,9$$

Error de validación cruzada para la técnica de DM regresión logística en aplicación de la fórmula (1).

$$E = \frac{59 + 46 + 44 + 48 + 56 + 43 + 48 + 44 + 48 + 46}{10} = 48,2$$

Error de validación cruzada para la técnica de DM redes bayesianas en aplicación de la fórmula (1).

$$E = \frac{23 + 29 + 17 + 26 + 28 + 22 + 26 + 28 + 20 + 18}{10} = 23,7$$

La técnica de minería de DM de árboles de decisión tiene el mínimo menor error de validación cruzada con respecto a las otras técnicas de DM.

Concluimos que la técnica de DM de árboles de decisión es la más apropiada para la predicción de los estudiantes desertores de la UNJBG.

#### 5.4.5. Fase de evaluación

En esta etapa se realizó la predicción de los datos para los ingresantes del 2018 dando como resultado que 348 estudiantes tienen la probabilidad de desertar, como se muestra en la tabla 74.

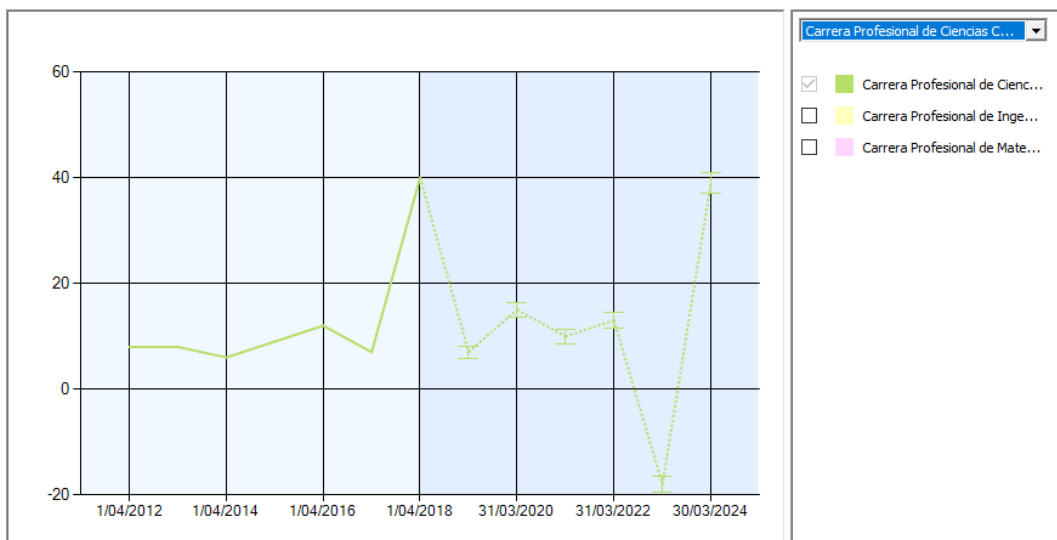
**Tabla 74** Ingresantes desertores del 2018 por programas académicos

| N° | Programa Académico  | Cantidad |
|----|---|----------|
| 1  | Carrera Profesional de Ciencias Contables y Financieras     | 38       |
| 2  | Carrera Profesional de Arquitectura                         | 23       |
| 3  | Carrera Profesional de Enfermería                           | 22       |
| 4  | Carrera Profesional de Odontología                          | 22       |
| 5  | Carrera Profesional de Ingeniería Ambiental                 | 19       |
| 6  | Carrera Profesional de Biología – Microbiología             | 19       |
| 7  | CAR.PROF. Educación - Matemática, Computación e Informática | 16       |
| 8  | Carrera Profesional de Agronomía                            | 14       |
| 9  | Carrera Profesional de Artes                                | 14       |

| <b>N°</b> | <b>Programa Académico</b>   | <b>Cantidad</b> |
|-----------|---|-----------------|
| 10        | CAR.PROF. Educación - Idioma Extranjero   | 13              |
| 11        | Carrera Profesional de Farmacia y Bioquímica                                    | 12              |
| 12        | Carrera Profesional de Farmacia y Bioquímica                                    | 12              |
| 13        | Carrera Profesional de Medicina Veterinaria y Zootecnia                         | 12              |
| 14        | Carrera Profesional de Ingeniería de Minas                                      | 10              |
| 15        | Carrera Profesional de Ingeniería en Industrias Alimentarias                    | 9               |
| 16        | Carrera Profesional de Ingeniería Metalúrgica                                   | 9               |
| 17        | Carrera Profesional de Ingeniería Química                                       | 9               |
| 18        | Carrera Profesional de Matemática   | 9               |
| 19        | Carrera Profesional de Ingeniería Pesquera                                      | 7               |
| 20        | Carrera Profesional de Física Aplicada  | 7               |
| 21        | Carrera Profesional de Ciencias de la Comunicación                              | 7               |
| 22        | CAR.PROF. Educación - Lengua, Literatura  | 7               |
| 23        | Carrera Profesional de Ingeniería Geológica - Geotecnia                         | 6               |
| 24        | Carrera Profesional de Ingeniería Mecánica                                      | 6               |
| 25        | Carrera Profesional de Ingeniería Comercial                                     | 6               |
| 26        | Carrera Profesional de Ingeniería en Informática y Sistemas                     | 4               |
| 27        | Carrera Profesional de Economía Agraria   | 4               |
| 28        | CAR.PROF. Educación - Ciencias Sociales y Promoción Socio Cultural              | 3               |
| 29        | Carrera Profesional de Derecho y Ciencias Políticas                             | 2               |
| 30        | Carrera Profesional de Ciencias Administrativas                                 | 2               |
| 31        | Carrera Profesional de Ingeniería Civil   | 2               |
| 32        | Carrera Profesional de Obstetricia  | 2               |
| 33        | CAR.PROF. Educación - Ciencias de la Naturaleza y Promoción Educativa Ambiental | 1               |

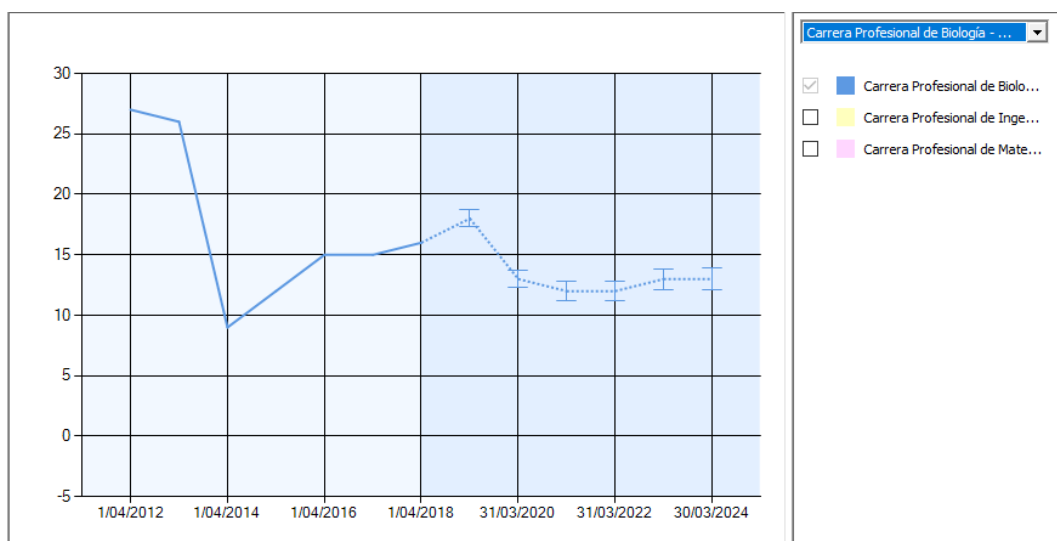
Fuente: Elaboración Propia.

Se ha estimado la cantidad de desertores por programas académicos dentro de los seis años próximos dentro de esos tenemos a las carreras profesionales de Ciencias Contables y Financieras y Biología, Microbiología y Agronomía.



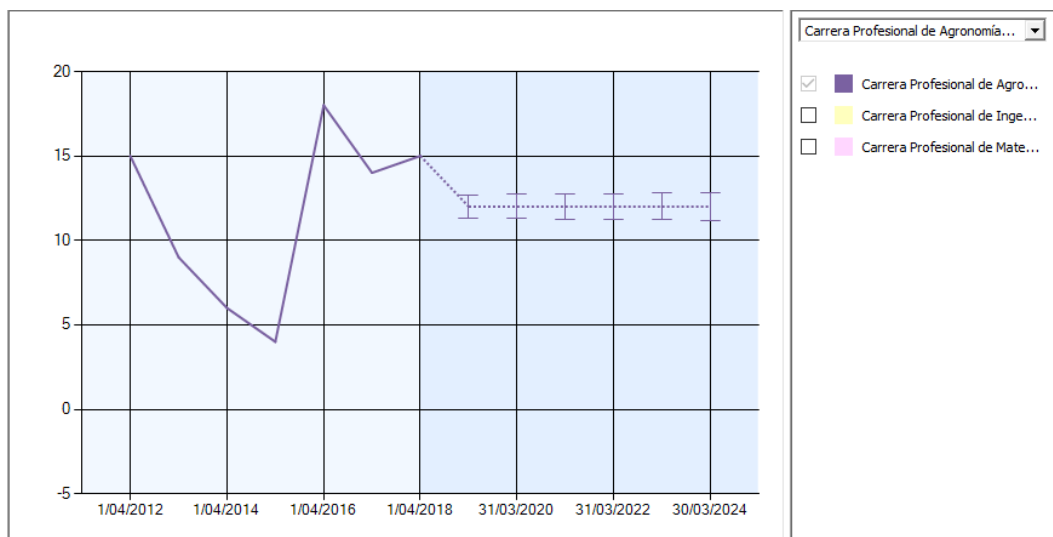
**Figura 100.** Desertores de la carrera profesional ciencias contables financieras

Fuente: Propia



**Figura 101.** Desertores de la carrera profesional biología - microbiología

Fuente: Propia



**Figura 102.** Desertores de la carrera profesional agronomía  
Fuente: Propia

Finalmente damos respuesta a los criterios de negocio planteando inicialmente de predecir los estudiantes que tiene alta probabilidad de desertar. Queda establecido los algoritmos o técnicas de minería más apropiada para este proyecto de investigación para los dos escenarios son: arboles de decisión y series temporales.

#### **5.4.6. Fase de despliegue**

En esta etapa se realizó el escenario despliegue donde los usuarios finales utilizarán los reportes finales de esta manera podrán tomar buenas decisiones de manera oportuna. Ver anexo 7.

## ANEXO 4: CÓDIGO AUTOGENERADO DEL ERWIN DATA MODELE r7

El código autogenerado es el siguiente:

```
USE [master]
GO

CREATE DATABASE [SYSA_DAWH_R4]
    CONTAINMENT = NONE
    ON PRIMARY
    ( NAME = N'SYSA_DAWH_R4', FILENAME =
N'F:\DATA\SQL2017\SYSA_DAWH\SYSA_DAWH_R4.mdf' , SIZE = 33792KB , MAXSIZE =
UNLIMITED, FILEGROWTH = 1024KB )
    LOG ON
    ( NAME = N'SYSA_DAWH_R4_log', FILENAME =
N'F:\DATA\SQL2017\SYSA_DAWH\SYSA_DAWH_R4_log.ldf' , SIZE = 69760KB , MAXSIZE =
2048GB , FILEGROWTH = 10%)
GO

USE [SYSA_DAWH_R4]
GO

CREATE TABLE [dbo].[DIM_Estudiante](
    [IdEstudiante] [int] NOT NULL,
    [Codigo] [nvarchar](15) NULL,
    [ApellidosyNombres] [nvarchar](250) NULL,
    [Sexo] [nvarchar](50) NULL,
    [SexoAbreviatura] [nvarchar](30) NULL,
    [EstadoCivil] [nvarchar](50) NULL,
    [EstadoCivilAbreviatura] [nvarchar](30) NULL,
    [NivelCategorizacion] [nvarchar](50) NULL,
    [NivelCategorizacionAbreviatura] [nvarchar](30) NULL,
    [NivelCategorizacionRango] [nvarchar](30) NULL,
    [TipoNivelCategorizacion] [nvarchar](30) NULL,
    [DistritoNacimiento] [nvarchar](250) NULL,
    [ProvinciaNacimiento] [nvarchar](250) NULL,
    [DepartamentoNacimiento] [nvarchar](250) NULL,
    [PeriodoIngreso] [nvarchar](50) NULL,
    [PeriodoIngresoAnio] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_Estudiante] PRIMARY KEY CLUSTERED
    (
        [IdEstudiante] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_AmbitoVivienda', 'U') IS NOT NULL
    DROP TABLE DIM_AmbitoVivienda
GO

CREATE TABLE [dbo].[DIM_AmbitoVivienda](
    [IdAmbitoVivienda] [int] NOT NULL,
```

```

        [AmbitoVivienda] [nvarchar](250) NULL,
        [AmbitoViviendaAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_AmbitoVivienda] PRIMARY KEY CLUSTERED
    (
        [IdAmbitoVivienda] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_Asignatura', 'U') IS NOT NULL
    DROP TABLE DIM_Asignatura
GO

CREATE TABLE [dbo].[DIM_Asignatura](
    [IdAsignatura] [int] NOT NULL,
    [AsignaturaCodigo] [nvarchar](50) NULL,
    [Asignatura] [nvarchar](250) NULL,
    [AsignaturaHoraTotal] [float] NULL,
    [GradoEducativoNumero] [int] NULL,
    [GradoEducativo] [nvarchar](50) NULL,
    [TipoMateria] [nvarchar](250) NULL,
    [SubArea] [nvarchar](250) NULL,
    [Area] [nvarchar](250) NULL,
    [Regimen] [nvarchar](250) NULL,
    [PlanEstudio] [nvarchar](50) NULL,
    [PlanEstudioNroCiclo] [int] NULL,
    [PlanEstudioTipoCiclo] [nvarchar](50) NULL,
    [PlanEstudioTipoRegimen] [nvarchar](50) NULL,
    CONSTRAINT [PK_DIM_Asignatura] PRIMARY KEY CLUSTERED
    (
        [IdAsignatura] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_CantidadPostulacion', 'U') IS NOT NULL
    DROP TABLE DIM_CantidadPostulacion
GO

CREATE TABLE [dbo].[DIM_CantidadPostulacion](
    [IdCantidadPostulacion] [int] NOT NULL,
    [CantidadPostulacion] [nvarchar](50) NULL,
    [CantidadPostulacionAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_CantidadPostulacion] PRIMARY KEY CLUSTERED
    (
        [IdCantidadPostulacion] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

```

```

IF OBJECT_ID ('DIM_CargaFamiliar', 'U') IS NOT NULL
    DROP TABLE DIM_CargaFamiliar
GO

CREATE TABLE [dbo].[DIM_CargaFamiliar](
    [IdCargaFamiliar] [int] NOT NULL,
    [CargaFamiliar] [nvarchar](250) NULL,
    [CargaFamiliarAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_CargaFamiliar] PRIMARY KEY CLUSTERED
    (
        [IdCargaFamiliar] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_CategoriaEstudiante', 'U') IS NOT NULL
    DROP TABLE DIM_CategoriaEstudiante
GO

CREATE TABLE [dbo].[DIM_CategoriaEstudiante](
    [IdCategoriaEstudiante] [int] NOT NULL,
    [CategoriaEstudiante] [nvarchar](50) NULL,
    [CategoriaEstudianteAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_CategoriaEstudiante] PRIMARY KEY CLUSTERED
    (
        [IdCategoriaEstudiante] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_CentroEstudio', 'U') IS NOT NULL
    DROP TABLE DIM_CentroEstudio
GO

CREATE TABLE [dbo].[DIM_CentroEstudio](
    [IdCentroEstudio] [int] NOT NULL,
    [CentroEstudio] [nvarchar](250) NULL,
    [Codigo] [nvarchar](50) NULL,
    [Turno] [nvarchar](50) NULL,
    [TipoGenero] [nvarchar](125) NULL,
    [GestionEducativa] [nvarchar](125) NULL,
    [TipoCentroEstudio] [nvarchar](125) NULL,
    [Distrito] [nvarchar](250) NULL,
    [Provincia] [nvarchar](250) NULL,
    [Departamento] [nvarchar](250) NULL,
    CONSTRAINT [PK_DIM_CentroEstudio] PRIMARY KEY CLUSTERED
    (
        [IdCentroEstudio] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]

```

```

GO

IF OBJECT_ID ('DIM_EstadoCivil', 'U') IS NOT NULL
    DROP TABLE DIM_EstadoCivil
GO

CREATE TABLE [dbo].[DIM_EstadoCivil](
    [IdEstadoCivil] [int] NOT NULL,
    [EstadoCivil] [nvarchar](50) NULL,
    [EstadoCivilAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_EstadoCivil] PRIMARY KEY CLUSTERED
    (
        [IdEstadoCivil] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_GradoEducativo', 'U') IS NOT NULL
    DROP TABLE DIM_GradoEducativo
GO

CREATE TABLE [dbo].[DIM_GradoEducativo](
    [IdGradoEducativo] [int] NOT NULL,
    [GradoEducativo] [nvarchar](250) NULL,
    [GradoEducativoAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_GradoEducativo] PRIMARY KEY CLUSTERED
    (
        [IdGradoEducativo] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_MaterialConstruccion', 'U') IS NOT NULL
    DROP TABLE DIM_MaterialConstruccion
GO

CREATE TABLE [dbo].[DIM_MaterialConstruccion](
    [IdMaterialConstruccion] [int] NOT NULL,
    [MaterialConstruccion] [nvarchar](250) NULL,
    [MaterialConstruccionAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_MaterialConstruccion] PRIMARY KEY CLUSTERED
    (
        [IdMaterialConstruccion] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_NivelCategorizacion', 'U') IS NOT NULL
    DROP TABLE DIM_NivelCategorizacion
GO

CREATE TABLE [dbo].[DIM_NivelCategorizacion](

```

```

        [IdNivelCategorizacion] [int] NOT NULL,
        [NivelCategorizacion] [nvarchar](250) NULL,
        [NivelCategorizacionAbreviatura] [nvarchar](30) NULL,
        CONSTRAINT [PK_DIM_NivelCategorizacion] PRIMARY KEY CLUSTERED
    (
        [IdNivelCategorizacion] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

```

```

IF OBJECT_ID ('DIM_NivelEducativo', 'U') IS NOT NULL
    DROP TABLE DIM_NivelEducativo
GO

```

```

CREATE TABLE [dbo].[DIM_NivelEducativo](
    [IdNivelEducativo] [int] NOT NULL,
    [NivelEducativo] [nvarchar](250) NULL,
    [NivelEducativoAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_NivelEducativo] PRIMARY KEY CLUSTERED
    (
        [IdNivelEducativo] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

```

```

IF OBJECT_ID ('DIM_Periodo', 'U') IS NOT NULL
    DROP TABLE DIM_Periodo
GO

```

```

CREATE TABLE [dbo].[DIM_Periodo](
    [IdPeriodo] [int] NOT NULL,
    [Periodo] [nvarchar](50) NULL,
    [Ciclo] [int] NULL,
    [Anio] [int] NULL,
    [TipoRegimenEstudio] [nvarchar](50) NULL,
    CONSTRAINT [PK_DIM_Periodo] PRIMARY KEY CLUSTERED
    (
        [IdPeriodo] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

```

```

IF OBJECT_ID ('DIM_PreparacionEstudio', 'U') IS NOT NULL
    DROP TABLE DIM_PreparacionEstudio
GO

```

```

CREATE TABLE [dbo].[DIM_PreparacionEstudio](
    [IdPreparacionEstudio] [int] NOT NULL,
    [PreparacionEstudio] [nvarchar](50) NULL,
    [PreparacionEstudioAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_PreparacionEstudio] PRIMARY KEY CLUSTERED
    (

```

```

        [IdPreparacionEstudio] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_ProgramaAcademico', 'U') IS NOT NULL
    DROP TABLE DIM_ProgramaAcademico
GO

CREATE TABLE [dbo].[DIM_ProgramaAcademico](
    [IdProgramaAcademico] [int] NOT NULL,
    [ProgramaAcademico] [nvarchar](250) NULL,
    [ProgramaAcademicoAbreviatura] [nvarchar](50) NULL,
    [CarreraProfesional] [nvarchar](250) NULL,
    [Facultad] [nvarchar](250) NULL,
    [TipoProgramaAcademico] [nvarchar](250) NULL,
    CONSTRAINT [PK_DIM_ProgramaAcademico] PRIMARY KEY CLUSTERED
    (
        [IdProgramaAcademico] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_RiesgoSaludFamiliar', 'U') IS NOT NULL
    DROP TABLE DIM_RiesgoSaludFamiliar
GO

CREATE TABLE [dbo].[DIM_RiesgoSaludFamiliar](
    [IdRiesgoSaludFamiliar] [int] NOT NULL,
    [RiesgoSaludFamiliar] [nvarchar](250) NULL,
    [RiesgoSaludFamiliarAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_RiesgoSaludFamiliar] PRIMARY KEY CLUSTERED
    (
        [IdRiesgoSaludFamiliar] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_ServicioVivienda', 'U') IS NOT NULL
    DROP TABLE DIM_ServicioVivienda
GO

CREATE TABLE [dbo].[DIM_ServicioVivienda](
    [IdServicioVivienda] [int] NOT NULL,
    [ServicioVivienda] [nvarchar](250) NULL,
    [ServicioViviendaAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_ServicioVivienda] PRIMARY KEY CLUSTERED
    (
        [IdServicioVivienda] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

```

```

IF OBJECT_ID ('DIM_Sexo', 'U') IS NOT NULL
    DROP TABLE DIM_Sexo
GO

CREATE TABLE [dbo].[DIM_Sexo](
    [IdSexo] [int] NOT NULL,
    [Sexo] [nvarchar](50) NULL,
    [SexoAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_Sexo] PRIMARY KEY CLUSTERED
    (
        [IdSexo] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_SituacionPadre', 'U') IS NOT NULL
    DROP TABLE DIM_SituacionPadre
GO

CREATE TABLE [dbo].[DIM_SituacionPadre](
    [IdSituacionPadre] [int] NOT NULL,
    [SituacionPadre] [nvarchar](250) NULL,
    [SituacionPadreAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_SituacionPadre] PRIMARY KEY CLUSTERED
    (
        [IdSituacionPadre] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_SituacionMatricula', 'U') IS NOT NULL
    DROP TABLE DIM_SituacionMatricula
GO

CREATE TABLE [dbo].[DIM_SituacionMatricula](
    [IdSituacionMatricula] [int] NOT NULL,
    [SituacionMatricula] [nvarchar](250) NULL,
    [SituacionMatriculaAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_SituacionMatricula] PRIMARY KEY CLUSTERED
    (
        [IdSituacionMatricula] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_Sostiene', 'U') IS NOT NULL
    DROP TABLE DIM_Sostiene
GO

CREATE TABLE [dbo].[DIM_Sostiene](
    [IdSotiene] [int] NOT NULL,

```

```

        [Sotiene] [nvarchar](250) NULL,
        [SotieneAbreviatura] [nvarchar](30) NULL,
        CONSTRAINT [PK_DIM_Sostiene] PRIMARY KEY CLUSTERED
    (
        [IdSotiene] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_TenenciaVivienda', 'U') IS NOT NULL
    DROP TABLE DIM_TenenciaVivienda
GO

CREATE TABLE [dbo].[DIM_TenenciaVivienda](
    [IdTenenciaVivienda] [int] NOT NULL,
    [TenenciaVivienda] [nvarchar](250) NULL,
    [TenenciaViviendaAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_TenenciaVivienda] PRIMARY KEY CLUSTERED
    (
        [IdTenenciaVivienda] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_TipoIngreso', 'U') IS NOT NULL
    DROP TABLE DIM_TipoIngreso
GO

CREATE TABLE [dbo].[DIM_TipoIngreso](
    [IdTipoIngreso] [int] NOT NULL,
    [TipoIngreso] [nvarchar](250) NULL,
    [TipoIngresoAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_TipoIngreso] PRIMARY KEY CLUSTERED
    (
        [IdTipoIngreso] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

IF OBJECT_ID ('DIM_UbicacionGeografica', 'U') IS NOT NULL
    DROP TABLE DIM_UbicacionGeografica
GO

CREATE TABLE [dbo].[DIM_UbicacionGeografica](
    [IdUbicacionGeografica] [nvarchar](8) NOT NULL,
    [Distrito] [nvarchar](250) NULL,
    [Provincia] [nvarchar](250) NULL,
    [Departamento] [nvarchar](250) NULL,
    CONSTRAINT [PK_DIM_UbicacionGeografica] PRIMARY KEY CLUSTERED
    (
        [IdUbicacionGeografica] ASC
    )

```

```
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

```
IF OBJECT_ID ('DIM_UsoMedioComunicacion', 'U') IS NOT NULL
    DROP TABLE DIM_UsoMedioComunicacion
GO
```

```
CREATE TABLE [dbo].[DIM_UsoMedioComunicacion](
    [IdUsoMedioComunicacion] [nvarchar](4) NOT NULL,
    [UsoMedioComunicacion] [nvarchar](250) NULL,
    [UsoMedioComunicacionAbreviatura] [nvarchar](30) NULL,
    CONSTRAINT [PK_DIM_UsoMedioComunicacion] PRIMARY KEY CLUSTERED
(
    [IdUsoMedioComunicacion] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

```
IF OBJECT_ID ('DIM_Fecha', 'U') IS NOT NULL
    DROP TABLE DIM_Fecha
GO
```

```
--create schema Dimension
CREATE TABLE DIM_Fecha(
    [DateKey] [int] NOT NULL,
    [Date] [datetime] NOT NULL,
    [FullDate_Description] [nvarchar](100) NULL,
    [DayNumberOfWeek] [tinyint] NULL,
    [DayNumberOfMonth] [tinyint] NULL,
    [DayNumberOfYear] [smallint] NULL,
    [WeekNumberOfYear] [tinyint] NULL,
    [MonthNumberOfYear] [tinyint] NULL,
    [CalendarQuarterOfYear] [tinyint] NULL,
    [CalendarSemesterOfYear] [tinyint] NULL,
    [CalendarYear] [char](4) NULL,
    [CalendarYearWeek] [nvarchar](25) NULL,
    [CalendarYearMonth] [nvarchar](25) NULL,
    [CalendarYearQuarter] [nvarchar](25) NULL,
    [CalendarYearSemester] [nvarchar](25) NULL,
    [CalendarYearWeek_Description] [nvarchar](25) NULL,
    [CalendarYearMonth_Description] [nvarchar](25) NULL,
    [CalendarYearQuarter_Description] [nvarchar](25) NULL,
    [CalendarYearSemester_Description] [nvarchar](25) NULL,
    [CalendarYear_Description] [nvarchar](25) NULL,
    [EnglishDayNameOfWeek] [nvarchar](10) NULL,
    [SpanishDayNameOfWeek] [nvarchar](10) NULL,
    [CatalanDayNameOfWeek] [nvarchar](10) NULL,
    [EnglishMonthName] [nvarchar](10) NULL,
    [SpanishMonthName] [nvarchar](10) NULL,
    [CatalanMonthName] [nvarchar](10) NULL
CONSTRAINT [PK_DimDate_DateKey] PRIMARY KEY CLUSTERED
([DateKey] ASC)WITH (PAD_INDEX = OFF, IGNORE_DUP_KEY = OFF) ON [PRIMARY],
CONSTRAINT [AK_DimDate_Date] UNIQUE NONCLUSTERED
```

```

    ([Date] ASC)WITH (PAD_INDEX = OFF, IGNORE_DUP_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]

```

```
GO
```

```
----- CREACIÓN DE TABLAS DE HECHOS INGRESANTE REGISTRO -----
```

```
IF OBJECT_ID ('THE_IngresanteRegistro', 'U') IS NOT NULL
    DROP TABLE THE_IngresanteRegistro
```

```
GO
```

```
CREATE TABLE [dbo].[THE_IngresanteRegistro](
    [IdIngresanteRegistro] [int] NOT NULL,
    [IdEstudiante] [int] NULL,
    [IdLugarResidencia] [nvarchar](8) NULL,
    [IdCentroEstudio] [int] NULL,
    [IdPreparacionEstudio] [int] NULL,
    [IdCantidadPostulacion] [int] NULL,
    [IdNivelEducativoPadre] [int] NULL,
    [IdNivelEducativoMadre] [int] NULL,
    [IdAmbitoVivienda] [int] NULL,
    [IdTenenciaVivienda] [int] NULL,
    [IdMaterialConstruccion] [int] NULL,
    [IdServicioVivienda] [int] NULL,
    [IdUsoMedioComunicacion] [nvarchar](4) NULL,
    [IdSostiene] [int] NULL,
    [IdSituacionPadre] [int] NULL,
    [IdRiesgoSaludFamiliar] [int] NULL,
    [IdCargaFamiliar] [int] NULL,
    [IdGradoEducativo] [int] NULL,
    [IdSituacionMatricula] [int] NULL,
    [IdTipoIngreso] [int] NULL,
    [IdProgramaAcademico] [int] NULL,
    [IdPeriodo] [int] NULL,
    [Edad] [float] NULL,
    [Peso] [float] NULL,
    [Talla] [float] NULL,
    [IMC] [float] NULL,
    [TotalAsignatura] [int] NULL,
    [TotalAsignaturaAprobado] [int] NULL,
    [TotalAsignaturaDesAprobado] [int] NULL,
    [NotaPromedioPonderado] [float] NULL,
    [PuntajeFichaEvalSocioEconomico] [float] NULL,
    [IngresoMensualFamiliar] [float] NULL,
    [MatriculaNoRegistradaPeriodoInicial] [int] NULL,
    [DesaprobadoMas50Porcentaje] [int] NULL,
    [Matriculado] [int] NULL,
    [DesertorPortencial] [int] NULL,
    [Desertor] [int] NULL,
    CONSTRAINT [PK_THE_Ingresante] PRIMARY KEY CLUSTERED
(
    [IdIngresanteRegistro] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]

```

```

GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_AmbitoVivienda] FOREIGN KEY([IdAmbitoVivienda])
REFERENCES [dbo].[DIM_AmbitoVivienda] ([IdAmbitoVivienda])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_AmbitoVivienda]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CantidadPostulacion] FOREIGN
KEY([IdCantidadPostulacion])
REFERENCES [dbo].[DIM_CantidadPostulacion] ([IdCantidadPostulacion])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CantidadPostulacion]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CargaFamiliar] FOREIGN KEY([IdCargaFamiliar])
REFERENCES [dbo].[DIM_CargaFamiliar] ([IdCargaFamiliar])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CargaFamiliar]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CentroEstudio] FOREIGN KEY([IdCentroEstudio])
REFERENCES [dbo].[DIM_CentroEstudio] ([IdCentroEstudio])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_CentroEstudio]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Estudiante] FOREIGN KEY([IdEstudiante])
REFERENCES [dbo].[DIM_Estudiante] ([IdEstudiante])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Estudiante]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_GradoEducativo] FOREIGN KEY([IdGradoEducativo])
REFERENCES [dbo].[DIM_GradoEducativo] ([IdGradoEducativo])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_GradoEducativo]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_MaterialConstruccion] FOREIGN
KEY([IdMaterialConstruccion])
REFERENCES [dbo].[DIM_MaterialConstruccion] ([IdMaterialConstruccion])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_MaterialConstruccion]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_NivelEducativo] FOREIGN
KEY([IdNivelEducativoPadre])

```

```

REFERENCES [dbo].[DIM_NivelEducativo] ([IdNivelEducativo])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_NivelEducativo]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_NivelEducativo1] FOREIGN
KEY([IdNivelEducativoMadre])
REFERENCES [dbo].[DIM_NivelEducativo] ([IdNivelEducativo])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_NivelEducativo1]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Periodo] FOREIGN KEY([IdPeriodo])
REFERENCES [dbo].[DIM_Periodo] ([IdPeriodo])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Periodo]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_PreparacionEstudio] FOREIGN
KEY([IdPreparacionEstudio])
REFERENCES [dbo].[DIM_PreparacionEstudio] ([IdPreparacionEstudio])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_PreparacionEstudio]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_ProgramaAcademico] FOREIGN
KEY([IdProgramaAcademico])
REFERENCES [dbo].[DIM_ProgramaAcademico] ([IdProgramaAcademico])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_ProgramaAcademico]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_RiesgoSaludFamiliar] FOREIGN
KEY([IdRiesgoSaludFamiliar])
REFERENCES [dbo].[DIM_RiesgoSaludFamiliar] ([IdRiesgoSaludFamiliar])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_RiesgoSaludFamiliar]
GO

ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_ServicioVivienda] FOREIGN
KEY([IdServicioVivienda])
REFERENCES [dbo].[DIM_ServicioVivienda] ([IdServicioVivienda])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_ServicioVivienda]
GO

```

```

ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_SituacionMatricula] FOREIGN
KEY([IdSituacionMatricula])
REFERENCES [dbo].[DIM_SituacionMatricula] ([IdSituacionMatricula])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_SituacionMatricula]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_SituacionPadre] FOREIGN KEY([IdSituacionPadre])
REFERENCES [dbo].[DIM_SituacionPadre] ([IdSituacionPadre])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_SituacionPadre]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Sostiene] FOREIGN KEY([IdSostiene])
REFERENCES [dbo].[DIM_Sostiene] ([IdSostiene])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_Sostiene]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_TenenciaVivienda] FOREIGN
KEY([IdTenenciaVivienda])
REFERENCES [dbo].[DIM_TenenciaVivienda] ([IdTenenciaVivienda])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_TenenciaVivienda]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_TipoIngreso] FOREIGN KEY([IdTipoIngreso])
REFERENCES [dbo].[DIM_TipoIngreso] ([IdTipoIngreso])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_TipoIngreso]
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_UbicacionGeografica] FOREIGN
KEY([IdLugarResidencia])
REFERENCES [dbo].[DIM_UbicacionGeografica] ([IdUbicacionGeografica])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_UbicacionGeografica]
GO

ALTER TABLE [dbo].[THE_IngresanteRegistro] WITH CHECK ADD CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_UsoMedioComunicacion] FOREIGN
KEY([IdUsoMedioComunicacion])
REFERENCES [dbo].[DIM_UsoMedioComunicacion] ([IdUsoMedioComunicacion])
GO
ALTER TABLE [dbo].[THE_IngresanteRegistro] CHECK CONSTRAINT
[FK_THE_IngresanteRegistro_DIM_UsoMedioComunicacion]
GO

```

```

----- CREACIÓN DE LA TABLA HECHO FICHA DE MATRÍCULA -----
IF OBJECT_ID ('THE_FichaMatricula', 'U') IS NOT NULL
    DROP TABLE THE_FichaMatricula
GO

CREATE TABLE [dbo].[THE_FichaMatricula](
    [IdFichaMatricula] [int] NOT NULL,
    [IdEstudiante] [int] NULL,
    [IdAmbitoVivienda] [int] NULL,
    [IdMaterialConstruccion] [int] NULL,
    [IdServicioVivienda] [int] NULL,
    [IdUsoMedioComunicacion] [nvarchar](4) NULL,
    [IdSotiene] [int] NULL,
    [IdLugarResidencia] [nvarchar](8) NULL,
    [IdCentroEstudio] [int] NULL,
    [IdTipoIngreso] [int] NULL,
    [IdCategoriaEstudiante] [int] NULL,
    [IdGradoEducativo] [int] NULL,
    [IdProgramaAcademico] [int] NULL,
    [IdPeriodoUltimo] [int] NULL,
    [Edad] [int] NULL,
    [TotalAsignatura] [int] NULL,
    [TotalAsignaturaAprobado] [int] NULL,
    [TotalAsignaturaDesaprobado] [int] NULL,
    [CantidadAsignaturaMatriculado] [int] NULL,
    [NotaPromedioPonderado] [float] NULL,
    [Desaprobado3raMatricula] [int] NULL,
    [Desaprobado4taMatricula] [int] NULL,
    [MatriculaNoRegistradaSegunReglamento] [int] NULL,
    [DesaprobadoMas50Porciento4Semestre] [int] NULL,
    [DesaprobadoMas70Porciento4Semestre] [int] NULL,
    [DesaprobadoMas70Porciento2Semestre] [int] NULL,
    [DesaprobadoMas90Porciento2Semestre] [int] NULL,
    [DesaprobadoMas50Porciento] [int] NULL,
    [DesaprobadoMas70Porciento] [int] NULL,
    [CambioCurricular] [int] NULL,
    [DesertorPortencial] [int] NULL,
    [Desertor] [int] NULL,
    CONSTRAINT [PK_THE_FichaMatricula] PRIMARY KEY CLUSTERED
    (
        [IdFichaMatricula] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO

ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_AmbitoVivienda] FOREIGN KEY([IdAmbitoVivienda])
REFERENCES [dbo].[DIM_AmbitoVivienda] ([IdAmbitoVivienda])
GO

ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_AmbitoVivienda]
GO

```

```

ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_CategoriaEstudiante] FOREIGN
KEY([IdCategoriaEstudiante])
REFERENCES [dbo].[DIM_CategoriaEstudiante] ([IdCategoriaEstudiante])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_CategoriaEstudiante]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_CentroEstudio] FOREIGN KEY([IdCentroEstudio])
REFERENCES [dbo].[DIM_CentroEstudio] ([IdCentroEstudio])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_CentroEstudio]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_Estudiante] FOREIGN KEY([IdEstudiante])
REFERENCES [dbo].[DIM_Estudiante] ([IdEstudiante])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_Estudiante]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_GradoEducativo] FOREIGN KEY([IdGradoEducativo])
REFERENCES [dbo].[DIM_GradoEducativo] ([IdGradoEducativo])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_GradoEducativo]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_MaterialConstruccion] FOREIGN
KEY([IdMaterialConstruccion])
REFERENCES [dbo].[DIM_MaterialConstruccion] ([IdMaterialConstruccion])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_MaterialConstruccion]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_Periodo] FOREIGN KEY([IdPeriodoUltimo])
REFERENCES [dbo].[DIM_Periodo] ([IdPeriodo])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_Periodo]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_ProgramaAcademico] FOREIGN KEY([IdProgramaAcademico])
REFERENCES [dbo].[DIM_ProgramaAcademico] ([IdProgramaAcademico])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_ProgramaAcademico]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_ServicioVivienda] FOREIGN KEY([IdServicioVivienda])
REFERENCES [dbo].[DIM_ServicioVivienda] ([IdServicioVivienda])
GO

```

```

ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_ServicioVivienda]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_Sostiene] FOREIGN KEY([IdSotiene])
REFERENCES [dbo].[DIM_Sostiene] ([IdSotiene])
GO

ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_Sostiene]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_TipoIngreso] FOREIGN KEY([IdTipoIngreso])
REFERENCES [dbo].[DIM_TipoIngreso] ([IdTipoIngreso])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_TipoIngreso]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_UbicacionGeografica] FOREIGN KEY([IdLugarResidencia])
REFERENCES [dbo].[DIM_UbicacionGeografica] ([IdUbicacionGeografica])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_UbicacionGeografica]
GO
ALTER TABLE [dbo].[THE_FichaMatricula] WITH CHECK ADD CONSTRAINT
[FK_THE_FichaMatricula_DIM_UsoMedioComunicacion] FOREIGN
KEY([IdUsoMedioComunicacion])
REFERENCES [dbo].[DIM_UsoMedioComunicacion] ([IdUsoMedioComunicacion])
GO
ALTER TABLE [dbo].[THE_FichaMatricula] CHECK CONSTRAINT
[FK_THE_FichaMatricula_DIM_UsoMedioComunicacion]
GO
----- CREACIÓN DE LA TABLA HECHO EGRESADO -----
IF OBJECT_ID ('THE_Egresado', 'U') IS NOT NULL
    DROP TABLE THE_Egresado
GO
CREATE TABLE [dbo].[THE_Egresado](
    [IdEgresado] [int] NOT NULL,
    [IdEstudiante] [int] NULL,
    [IdFecha] [int] NULL,
    [IdLugarResidencia] [nvarchar](8) NULL,
    [IdCentroEstudio] [int] NULL,
    [IdTipoIngreso] [int] NULL,
    [IdProgramaAcademico] [int] NULL,
    [IdPeriodoFinalizo] [int] NULL,
    [Edad] [int] NULL,
    [NotaPromedioPonderado] [float] NULL,
    [EgresadoNoTramitaronConstancia24Meses] [int] NULL,
    [EgresadoNoTramitaronBachiller36Meses] [int] NULL,
    [BachillerNoTramitaronTitulo60Meses] [int] NULL,
    [IngresanteVinculado] [int] NULL,
    [Egresado] [int] NULL,
    [Bachiller] [int] NULL,
    [Titulado] [int] NULL,

```

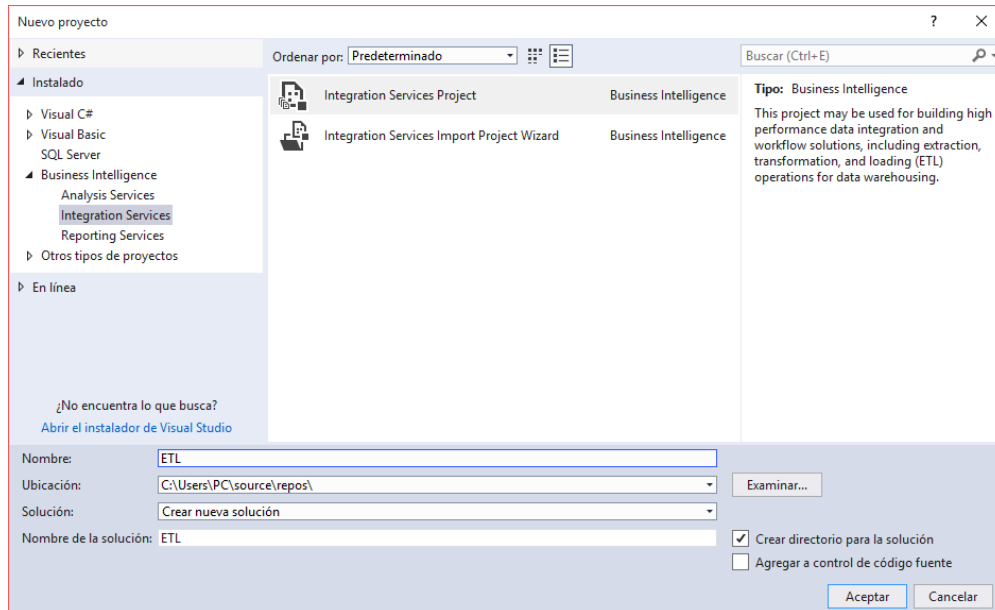
```

        [DesertorPortencial] [int] NULL,
        [Desertor] [int] NULL,
    CONSTRAINT [PK_THE_Egresado] PRIMARY KEY CLUSTERED
    (
        [IdEgresado] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
    ) ON [PRIMARY]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_CentroEstudio] FOREIGN KEY([IdCentroEstudio])
REFERENCES [dbo].[DIM_CentroEstudio] ([IdCentroEstudio])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT
[FK_THE_Egresado_DIM_CentroEstudio]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_Estudiante] FOREIGN KEY([IdEstudiante])
REFERENCES [dbo].[DIM_Estudiante] ([IdEstudiante])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT
[FK_THE_Egresado_DIM_Estudiante]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_Fecha] FOREIGN KEY([IdFecha])
REFERENCES [dbo].[DIM_Fecha] ([DateKey])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT [FK_THE_Egresado_DIM_Fecha]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_Periodo] FOREIGN KEY([IdPeriodoFinalizo])
REFERENCES [dbo].[DIM_Periodo] ([IdPeriodo])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT [FK_THE_Egresado_DIM_Periodo]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_ProgramaAcademico] FOREIGN KEY([IdProgramaAcademico])
REFERENCES [dbo].[DIM_ProgramaAcademico] ([IdProgramaAcademico])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT
[FK_THE_Egresado_DIM_ProgramaAcademico]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_TipoIngreso] FOREIGN KEY([IdTipoIngreso])
REFERENCES [dbo].[DIM_TipoIngreso] ([IdTipoIngreso])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT
[FK_THE_Egresado_DIM_TipoIngreso]
GO
ALTER TABLE [dbo].[THE_Egresado] WITH CHECK ADD CONSTRAINT
[FK_THE_Egresado_DIM_UbicacionGeografica] FOREIGN KEY([IdLugarResidencia])
REFERENCES [dbo].[DIM_UbicacionGeografica] ([IdUbicacionGeografica])
GO
ALTER TABLE [dbo].[THE_Egresado] CHECK CONSTRAINT
[FK_THE_Egresado_DIM_UbicacionGeografica]

```

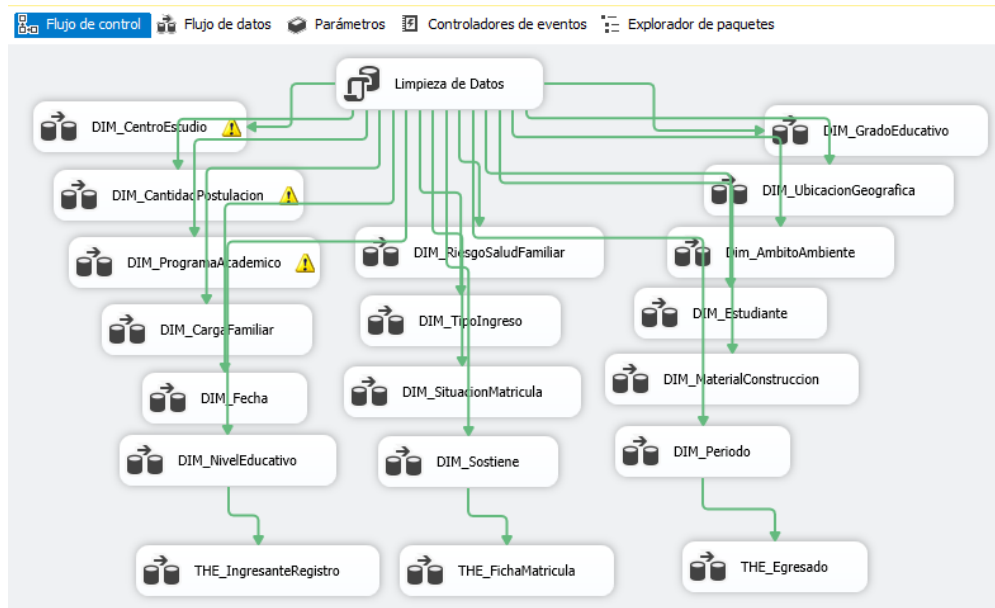
## ANEXO 5: ETL PARA EL MODELO DIMENSIONAL

Seleccionamos el proyecto de Servicio de integración para el proceso ETL.

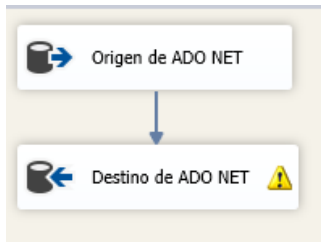


Nota: Nuevo Proyecto de Servicio de Integración.

Diseño del flujo de tareas del proceso ETL de información de las dimensiones y tablas de hechos.



Nota: Diseño del flujo de tareas.



Nota: Conexiones de ADO NET de origen y destino.

### Tabla: Proceso de Limpieza

| Nombre de Proceso | Descripción   |
|-------------------|---|
| Conexión          | SYSA_DAWH_R4  |
| Sentencia SQL     | DELETE FROM DIM_AmbitoVivienda<br>DELETE FROM DIM_CantidadPostulacion<br>DELETE FROM DIM_CargaFamiliar<br>DELETE FROM DIM_CentroEstudio<br>DELETE FROM DIM_Estudiante<br>DELETE FROM DIM_Fecha<br>DELETE FROM DIM_GradoEducativo<br>DELETE FROM DIM_MaterialConstruccion<br>DELETE FROM DIM_NivelEducativo<br>DELETE FROM DIM_Periodo<br>DELETE FROM DIM_PreparacionEstudio<br>DELETE FROM DIM_ProgramaAcademico<br>DELETE FROM DIM_RiesgoSaludFamiliar<br>DELETE FROM DIM_ServicioVivienda<br>DELETE FROM DIM_SituacionMatricula<br>DELETE FROM DIM_SituacionPadre<br>DELETE FROM DIM_Sostiene<br>DELETE FROM DIM_TenenciaVivienda<br>DELETE FROM DIM_TipoIngreso<br>DELETE FROM DIM_UbicacionGeografica<br>DELETE FROM DIM_UsoMedioComunicacion<br>DELETE FROM THE_IngresanteRegistro<br>DELETE FROM THE_FichaMatricula<br>DELETE FROM THE_Egresado |

Fuente: Elaboración Propia.

### Tabla: Carga de la Dimensión Ambito de Vivienda

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | SELECT idambito, ítem, puntaje<br>FROM ambito<br>UNION ALL<br>SELECT Id=0, Denominacion='NO ESPECIFICADO', Abreviatura='0' |
| Tabla Destino | DIM_AmbitoVivienda   |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Carga Familiar**

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <code>SELECT idcargafamiliar, ítem, puntaje<br/>FROM cargafamiliar<br/>UNION ALL<br/>SELECT Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0'</code> |
| Tabla Destino | DIM_CargaFamiliar   |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Centro de Estudio**

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <code>SELECT IdCentroEstudio=CE.IdCentroEstudio<br/>,CentroEstudio=CE.Denominacion+'-' +CE.Codigo<br/>,Codigo=CE.Codigo<br/>,Turno=TU.Denominacion<br/>,TipoGenero=TS.Denominacion<br/>,GestionEducativa=GE.Denominacion<br/>,TipoCentroEstudio=TC.Denominacion<br/>,Distrito=DT.Denominacion<br/>,Provincia=PR.Denominacion<br/>,Departamento=DP.Denominacion<br/>FROM ADS_CentroEstudio CE<br/>LEFT JOIN ADS_Catalogo TU ON TU.IdCatalogo = CE.IdTurno<br/>LEFT JOIN ADS_Catalogo TS ON TS.IdCatalogo = CE.IdTipoSexo<br/>LEFT JOIN ADS_Catalogo TC ON TC.IdCatalogo = CE.IdTipoCentroEstudio<br/>LEFT JOIN ADS_Catalogo GE ON GE.IdCatalogo = CE.IdGestionEducativa<br/>LEFT JOIN ADS_UbicacionGeografica DT ON DT.IdUbicacionGeografica =<br/>CE.IdUbicacionGeografica<br/>LEFT JOIN ADS_UbicacionGeografica PR ON PR.IdUbicacionGeografica = DT.IdPadre<br/>LEFT JOIN ADS_UbicacionGeografica DP ON DP.IdUbicacionGeografica = PR.IdPadre<br/>WHERE CE.Eliminado = 0</code> |
| Tabla Destino | DIM_CentroEstudio  |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Estudiante**

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;                           |
| Sentencia SQL | <code>SELECT<br/>IdEstudiante<br/>,Codigo<br/>,ApellidosyNombres</code> |

```

.Sexo
.SexoAbreviatura
.EstadoCivil
.EstadoCivilAbreviatura
.NivelCategorizacion
.NivelCategorizacionAbreviatura
.NivelCategorizacionRango
.TipoNivelCategorizacion=
CASE
WHEN NivelCategorizacionAbreviatura='A' THEN 'Alto'
WHEN NivelCategorizacionAbreviatura='B' OR
NivelCategorizacionAbreviatura='C' THEN 'Medio'
WHEN NivelCategorizacionAbreviatura='D' OR
NivelCategorizacionAbreviatura='E' OR NivelCategorizacionAbreviatura='F' THEN 'Bajo'
END
.DistritoNacimiento
.ProvinciaNacimiento
.DepartamentoNacimiento
.PeriodoIngreso
.PeriodoIngresoAnio
FROM #DATA

```

Tabla Destino DIM\_Estudiente

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Fecha

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre> Declare @StartDate datetime, @EndDate datetime set @StartDate = '19720101' set @EndDate = cast(cast(year(getdate())) + 5 as char(4)) + '1231' as datetime) while @StartDate &lt;= @EndDate begin exec DIM_I_InsertarFecha @StartDate set @StartDate = dateadd(day, 1, @StartDate) end </pre> |
| Tabla Destino | DIM_Fecha  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Grado Educativo

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;                              |
| Sentencia SQL | SELECT IdGradoEducativo, Denominacion, Abreviatura FROM ADS_GradoEducativo |
| Tabla Destino | DIM_GradoEducativo   |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Material de Construcción**

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT idambito, ítem, puntaje FROM material UNION ALL SELECT Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0'</pre> |
| Tabla Destino | DIM_MaterialConstruccion  |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Nivel Educativo**

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <pre>SELECT idambito, ítem, puntaje FROM niveleducativo UNION ALL SELECT Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0'</pre> |
| Tabla Destino | DIM_NivelEducativo  |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Periodo**

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT IdPeriodo=PE.IdPeriodo ,Periodo=PE.Denominacion ,Ciclo=PE.Ciclo ,Anio=PE.Anio ,TipoRegimenEstudio=TR.Denominacion FROM MAT_Periodo PE LEFT JOIN ADS_Catalogo TR ON TR.IdCatalogo = PE.IdTipoRegimenEstudio WHERE PE.Denominacion IS NOT NULL</pre> |
| Tabla Destino | DIM_Periodo  |

Fuente: Elaboración Propia.

**Tabla: Carga de la Dimensión Preparación Estudio**

| Proceso | Descripción |
|---------|-------------|
|---------|-------------|

|               |   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <pre>SELECT IdPreparacionEstudio=IdCatalogo ,PreparacionEstudio=Denominacion ,PreparacionEstudioAbreviatura=Abreviatura FROM SYSA_UNJBG_R4.dbo.ADS_Catalogo WHERE Tipo='0190'</pre> |
| Tabla Destino | DIM_PreparacionEstudio  |

Fuente: Elaboración Propia.

### Tabla: Carga de la Dimensión Programa Académico

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT IdProgramaAcademico=EP.IdEspecialidad, ProgramaAcademico=EP.DenominacionLargo, ProgramaAcademicoAbreviatura=EP.Abreviatura, CarreraProfesional=ES.DenominacionLargo, Facultad=FA.DenominacionLargo, TipoProgramaAcademico=TP.Denominacion FROM ADS_VW_Especialidad EP LEFT JOIN ADS_VW_Escuela ES ON ES.IdEscuela = EP.IdEscuela LEFT JOIN ADS_VW_Facultad FA ON FA.IdFacultad = ES.IdFacultad LEFT JOIN ADS_Catalogo TP ON TP.IdCatalogo = EP.IdTipoEspecialidad WHERE EP.Eliminado = 0</pre> |
| Tabla Destino | DIM_ProgramaAcademico  |

Fuente: Elaboración Propia.

### Tabla: Carga de la Dimensión Riesgo Salud Familiar

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT idriesgosalud_fam, ítem, puntaje FROM riesgosalud_fam UNION ALL SELECT Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0'</pre> |
| Tabla Destino | DIM_RiesgoSaludFamiliar   |

Fuente: Elaboración Propia.

### Tabla: Carga de la Dimensión Servicio de Vivienda

| Proceso | Descripción |
|---------|-------------|
|---------|-------------|

|               |   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <code>SELECT</code> idservicio, ítem, puntaje<br><code>FROM</code> servicio<br><code>UNION ALL</code><br><code>SELECT</code> Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0' |
| Tabla Destino | DIM_ServicioVivienda  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Situación Matrícula

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <code>SELECT</code> IdSituacionMatricula=IdCatalogo<br>,SituacionMatricula=Denominacion<br>,SituacionMatriculaAbreviatura=Abreviatura<br><code>FROM</code> SYSA_UNJBG_R4.dbo.ADS_Catalogo<br><code>WHERE</code> Tipo LIKE '0340%' |
| Tabla Destino | DIM_SituacionMatricula  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Situación del Padre

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <code>SELECT</code> idsituacionpadres, ítem, puntaje<br><code>FROM</code> situacionpadres<br><code>UNION ALL</code><br><code>SELECT</code> Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0' |
| Tabla Destino | DIM_SituacionPadre  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Sostiene

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <code>SELECT</code> idsostiene, ítem, puntaje<br><code>FROM</code> sostiene<br><code>UNION ALL</code><br><code>SELECT</code> Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0' |
| Tabla Destino | DIM_Sostiene  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Tenencia de Vivienda

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT idtenencia, ítem, puntaje FROM tenencia UNION ALL SELECT Id=0, Denominacion='NO ESPECIFICADO',Abreviatura='0'</pre> |
| Tabla Destino | DIM_TenenciaVivienda  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Tipo de Ingreso

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre>SELECT IdTipoIngreso=IdCatalogo ,TipoIngreso=Denominacion ,TipoIngresoAbreviatura=Abreviatura FROM SYSA_UNJBG_R4.dbo.ADS_Catalogo WHERE Tipo LIKE '0240%'</pre> |
| Tabla Destino | DIM_TipoIngreso  |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Ubicación Geografica

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <pre>SELECT DT.IdUbicacionGeografica, Distrito=DT.Denominacion, Provincia=PR.Denominacion, Departamento=DP.Denominacion FROM ADS_UbicacionGeografica DT LEFT JOIN ADS_UbicacionGeografica PR ON PR.IdUbicacionGeografica = DT.IdPadre LEFT JOIN ADS_UbicacionGeografica DP ON DP.IdUbicacionGeografica = PR.IdPadre WHERE DT.IdTipoUbicacionGeografica=1533</pre> |
| Tabla Destino | DIM_UbicacionGeografica   |

Fuente: Elaboración Propia.

#### Tabla: Carga de la Dimensión Uso de Medio de Comunicación

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre> SELECT IdUsoMedioComunicacion='0000', UsoMedioComunicacion='NO ESPECIFICADO', UsoMedioComunicacionAbreviatura='--' UNION ALL SELECT '0001','Celular','D' UNION ALL SELECT '0010','Telefono','C' UNION ALL SELECT '0011','Telefono, Celular','C,D' UNION ALL SELECT '0100','Internet','B' UNION ALL SELECT '0101','Internet, Celular','B,D' UNION ALL SELECT '0110','Internet, Telefono','B,C' UNION ALL SELECT '0111','Internet, Telefono, Celular','B,C,D' UNION ALL SELECT '1000','Cable','A' UNION ALL SELECT '1001','Cable, Celular','A,D' UNION ALL SELECT '1010','Cable, Telefono','A,C' UNION ALL SELECT '1011','Cable, Telefono, Celular','A,C,D' UNION ALL SELECT '1100','Cable, Internet','A,B,' UNION ALL SELECT '1101','Cable, Internet, Celular','A,B,D' UNION ALL SELECT '1110','Cable, Internet, Telefono','A,B,C' UNION ALL SELECT '1111','Cable, Internet, Telefono, Celular','A,B,C,D' </pre> |
| Tabla Destino | DIM_UsoMedioComunicacion  |

Fuente: Elaboración Propia.

### Tabla: Carga de la Tabla de Hecho Ingresante Registro

| Proceso       | Descripción  |
|---------------|--|
| Conexión      | Origen: SYSA_UNJBG_R4, TEST_DBUN_R4; Destino: SYSA_DAWH_R4;  |
| Sentencia SQL | <pre> SELECT Id ,IdEstudiante ,IdLugarResidencia ,IdCentroEstudio ,IdPreparacionEstudio ,IdCantidadPostulacion ,IdNivelEducativoPadre ,IdNivelEducativoMadre ,IdAmbitoVivienda ,IdTenenciaVivienda ,IdMaterialConstruccion ,IdServicioVivienda ,IdUsoMedioComunicacion ,IdSostiene ,IdSituacionPadre ,IdRiesgoSaludFamiliar ,IdCargaFamiliar ,IdGradoEducativo ,IdSituacionMatricula ,IdTipoIngreso ,IdProgramaAcademico ,IdPeriodo ,Edad ,Peso ,Talla ,IMC ,TotalAsignatura ,TotalAsignaturaAprobado </pre> |

```

,TotalAsignaturaDesAprobado
,NotaPromedioPonderado
,PuntajeFichaEvalSocioEconomico
,IngresoMensualFamiliar
,MatriculaNoRegistradaPeriodoInicial
,DesaprobadoMas50Porciento
,Matriculado
,DesertorPortencial
,Desertor
FROM #THE_IngresanteRegistro

```

Tabla Destino THE\_IngresanteRegistro

Fuente: Elaboración Propia.

### Tabla: Carga de la Tabla de Hecho Ficha de Matrícula

| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4, TEST_DBUN_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <pre> SELECT ,Id=ROW_NUMBER() OVER (ORDER BY IdEstudiante) ,IdEstudiante ,IdAmbitoVivienda ,IdMaterialConstruccion ,IdServicioVivienda ,IdUsoMedioComunicacion ,IdSostiene ,IdLugarResidencia ,IdCentroEstudio ,IdTipoIngreso ,IdCategoriaEstudiante=610 ,IdGradoEducativo ,IdProgramaAcademico ,IdPeriodo ,Edad ,TotalAsignatura ,TotalAsignaturaAprobado ,TotalAsignaturaDesAprobado ,CantidadAsignaturaMatriculado ,NotaPromedioPonderado = CASE WHEN Tipo = 0 THEN CASE WHEN SumaCreditos = 0 THEN 0.0 ELSE SumaNotas/SumaCreditos END WHEN Tipo = 1 THEN Promedio ELSE 0.0 END ,Desaprobado3raMatricula ,Desaprobado4taMatricula ,MatriculaNoRegistradaSegunReglamento= CASE WHEN CambioCurricular=0 THEN CASE WHEN Tipo=0 AND CantidadPeriodoSinMatricular&gt;2 THEN 1 WHEN Tipo=1 AND CantidadPeriodoSinMatricular&gt;4 THEN 1 ELSE 0 END ELSE 0 END ,DesaprobadoMas50Porciento4Semestre=CASE WHEN TotalAsignatura4Semestre &gt; 0 THEN CASE WHEN (TotalAsignaturaDesaprobado4Semestre /TotalAsignatura4Semestre)&gt;=0.5 THEN 1 ELSE 0 END ELSE 0 END ,DesaprobadoMas70Porciento4Semestre= </pre> |

```

CASE WHEN TotalAsignatura4Semestre > 0 THEN CASE WHEN
(TotalAsignaturaDesaprobado4Semestre /TotalAsignatura4Semestre)>=0.7 THEN 1
ELSE 0 END ELSE 0 END
.DesaprobadoMas70Porciento2Semestre=
CASE WHEN TotalAsignatura2Semestre > 0 THEN CASE WHEN
(TotalAsignaturaDesaprobado2Semestre /TotalAsignatura2Semestre)>=0.7 THEN 1
ELSE 0 END ELSE 0 END
.DesaprobadoMas90Porciento2Semestre=
CASE WHEN TotalAsignatura2Semestre > 0 THEN CASE WHEN
(TotalAsignaturaDesaprobado2Semestre /TotalAsignatura2Semestre)>=0.9 THEN 1
ELSE 0 END ELSE 0 END
.DesaprobadoMas50Porciento=
CASE WHEN TotalAsignatura > 0 THEN CASE WHEN (TotalAsignaturaDesAprobado
/TotalAsignatura)>=0.5 THEN 1 ELSE 0 END ELSE 0 END
.DesaprobadoMas70Porciento=
CASE WHEN TotalAsignatura > 0 THEN CASE WHEN (TotalAsignaturaDesAprobado
/TotalAsignatura)>=0.7 THEN 1 ELSE 0 END ELSE 0 END
.CambioCurricular
.DesertorPortencial=
CASE
  WHEN CambioCurricular=0 THEN
    CASE
      WHEN Tipo=0 AND CantidadPeriodoSinMatricular>2 THEN 1
      WHEN Tipo=1 AND CantidadPeriodoSinMatricular>4 THEN 1
      ELSE
        CASE
          WHEN TotalAsignatura2Semestre > 0 THEN
            CASE
              WHEN (TotalAsignaturaDesaprobado2Semestre
/TotalAsignatura2Semestre)>=0.7 THEN 1
            ELSE
              CASE
                WHEN TotalAsignatura4Semestre > 0 THEN
                  CASE WHEN (TotalAsignaturaDesaprobado4Semestre
/TotalAsignatura4Semestre)>=0.5 THEN 1
                ELSE DesertorPortencial END
              ELSE DesertorPortencial END
            END
          ELSE DesertorPortencial END
        END
      ELSE DesertorPortencial END
    END
  ELSE DesertorPortencial END
.Desertor=
CASE
  WHEN CambioCurricular=0 THEN
    CASE
      WHEN Tipo=0 AND CantidadPeriodoSinMatricular>2 THEN 1
      WHEN Tipo=1 AND CantidadPeriodoSinMatricular>4 THEN 1
      ELSE
        CASE
          WHEN TotalAsignatura2Semestre > 0 THEN
            CASE
              WHEN (TotalAsignaturaDesaprobado2Semestre
/TotalAsignatura2Semestre)>=0.9 THEN 1
            ELSE
              CASE
                WHEN TotalAsignatura4Semestre > 0 THEN
                  CASE WHEN (TotalAsignaturaDesaprobado4Semestre
/TotalAsignatura4Semestre)>=0.7 THEN 1
                ELSE Desertor END
              ELSE Desertor END
            END
          ELSE Desertor END
        END
      ELSE Desertor END
    END
  END
END

```

```

ELSE Desertor END
FROM #FichaMatricula

```

Tabla Destino THE\_FichaMatricula

---

Fuente: Elaboración Propia.

**Tabla: Carga de la Tabla de Hecho Egresado**

---

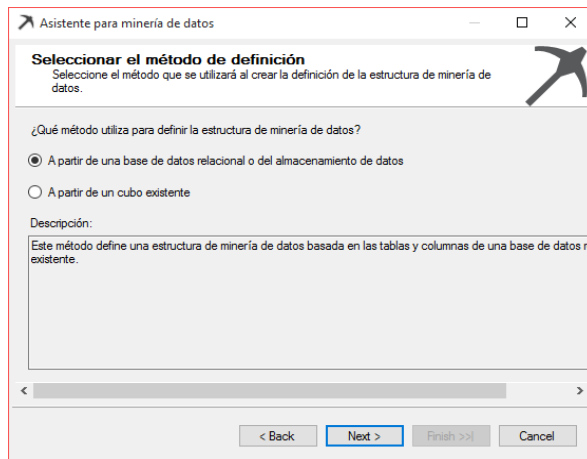
| Proceso       | Descripción   |
|---------------|---|
| Conexión      | Origen: SYSA_UNJBG_R4; Destino: SYSA_DAWH_R4;   |
| Sentencia SQL | <pre> SELECT IdEgresado ,IdEstudiante ,IdFecha ,IdLugarResidencia ,IdCentroEstudio ,IdTipoIngreso ,IdProgramaAcademico ,IdPeriodoFinalizo ,Edad ,NotaPromedioPonderado ,EgresadoNoTramitaronConstancia24Meses ,EgresadoNoTramitaronBachiller36Meses ,BachillerNoTramitaronTitulo60Meses ,IngresanteVinculado ,Egresado ,Bachiller ,Titulado ,DesertorPortencial ,Desertor FROM #EGRESADO </pre> |
| Tabla Destino | THE_Egresado  |

---

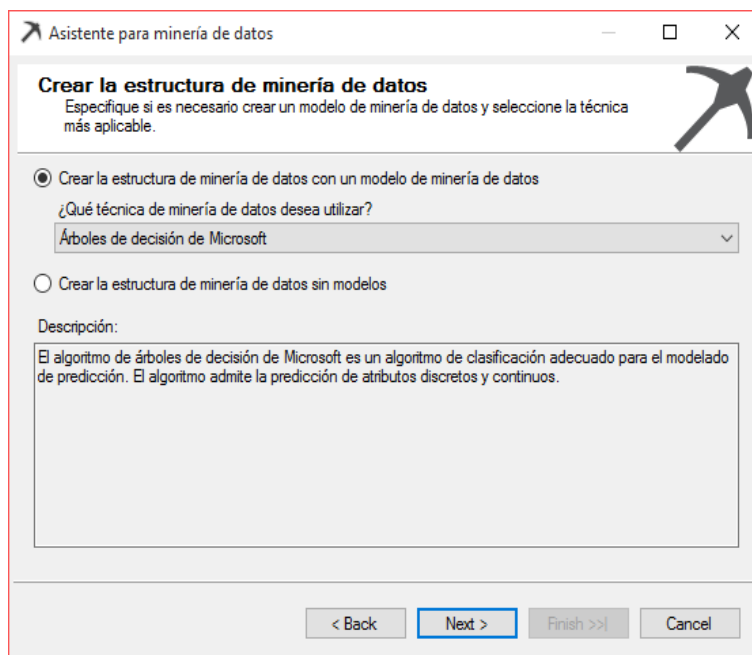
Fuente: Elaboración Propia.

## ANEXO 6: CONSTRUCCIÓN DEL MODELO DE MINERÍA DE DATOS

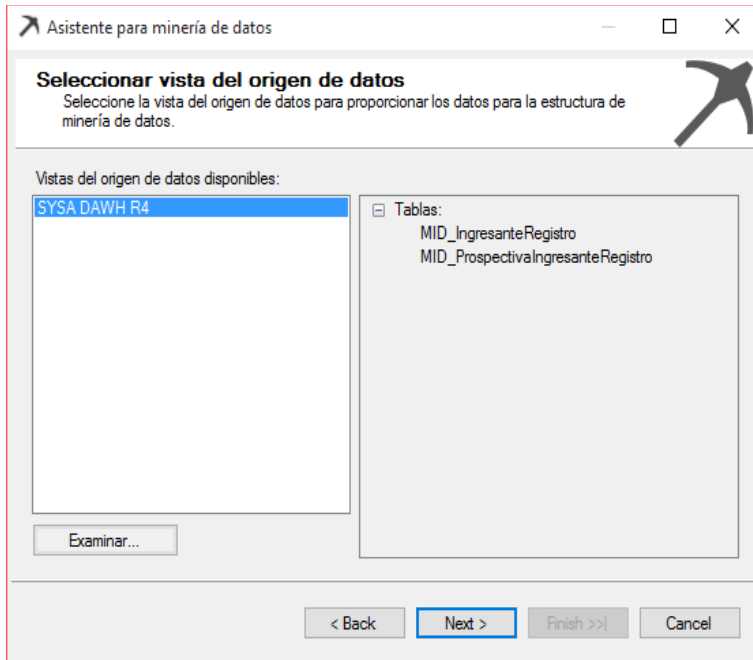
En esta sección se trabajó con una base de datos multidimensional, seguidamente se utilizó las técnicas de minería de datos para minar la zona de Ingresante Registro para establecer los campos de entradas y de predicciones.



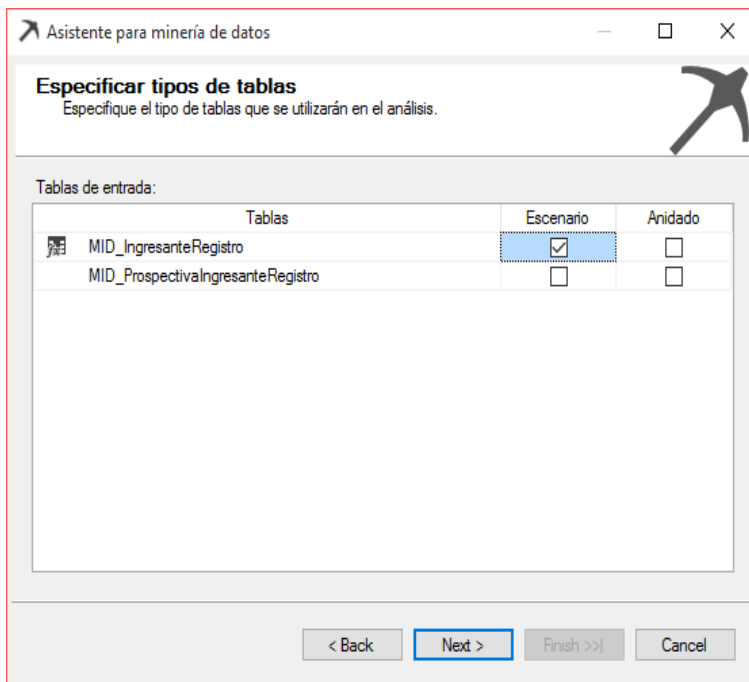
Nota: Selección del método de definición de la estructura de minería de datos.



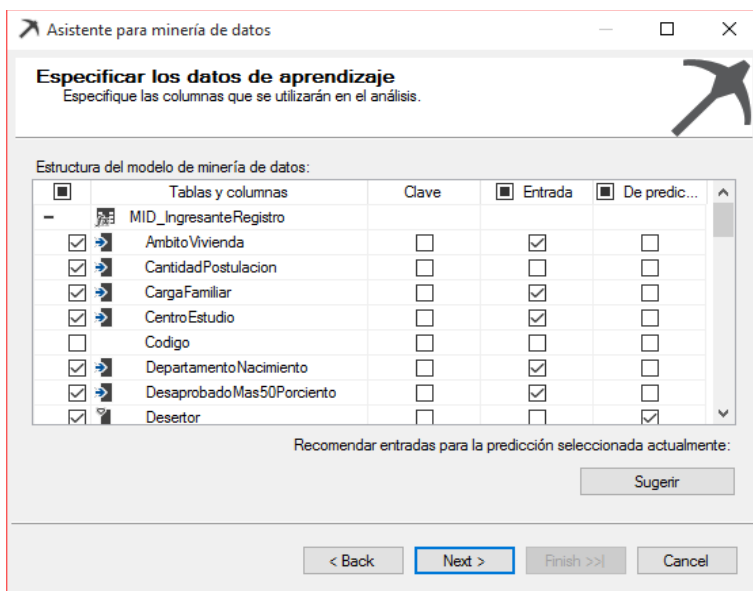
Nota: Selección de la técnica de minería de datos.



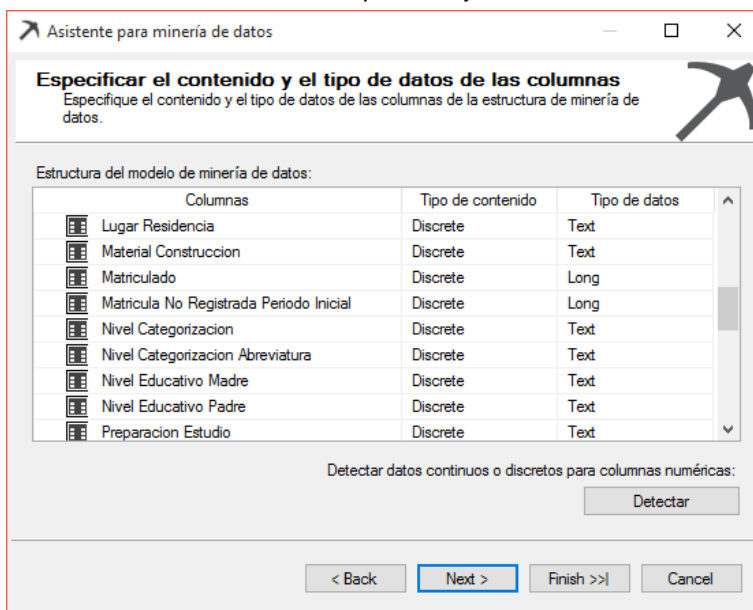
Nota: Selección de la vista del origen de datos.



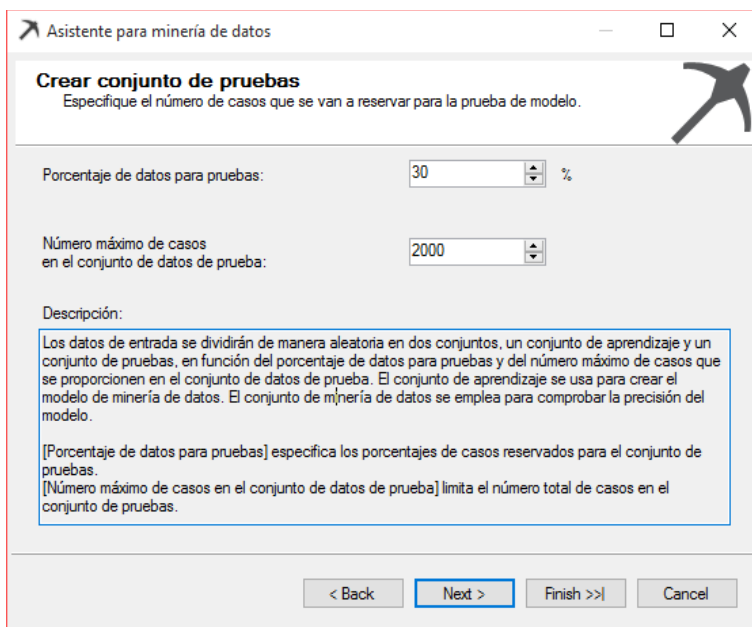
Nota: Selección de la tabla para el escenario.



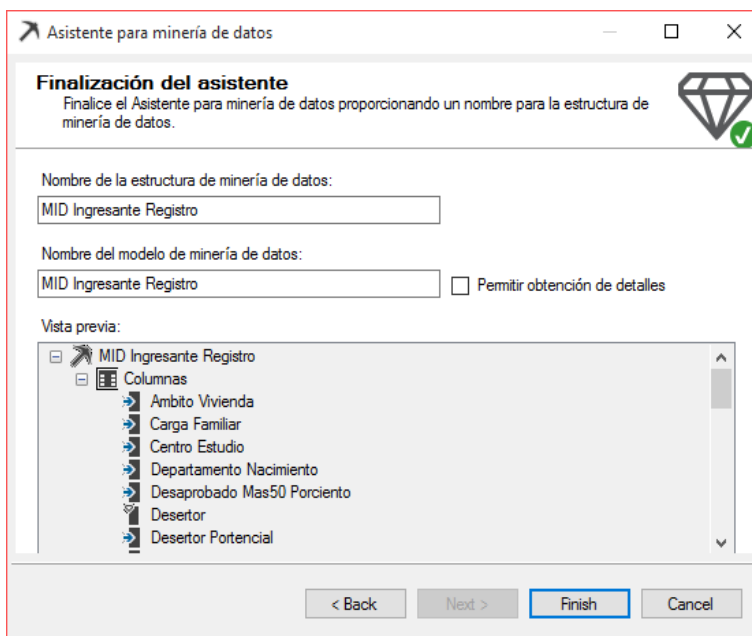
Nota: Selección de datos de aprendizaje.



Nota: Especificación de contenido y tipo de datos de las columnas.



Nota: Selección de datos de aprendizaje.



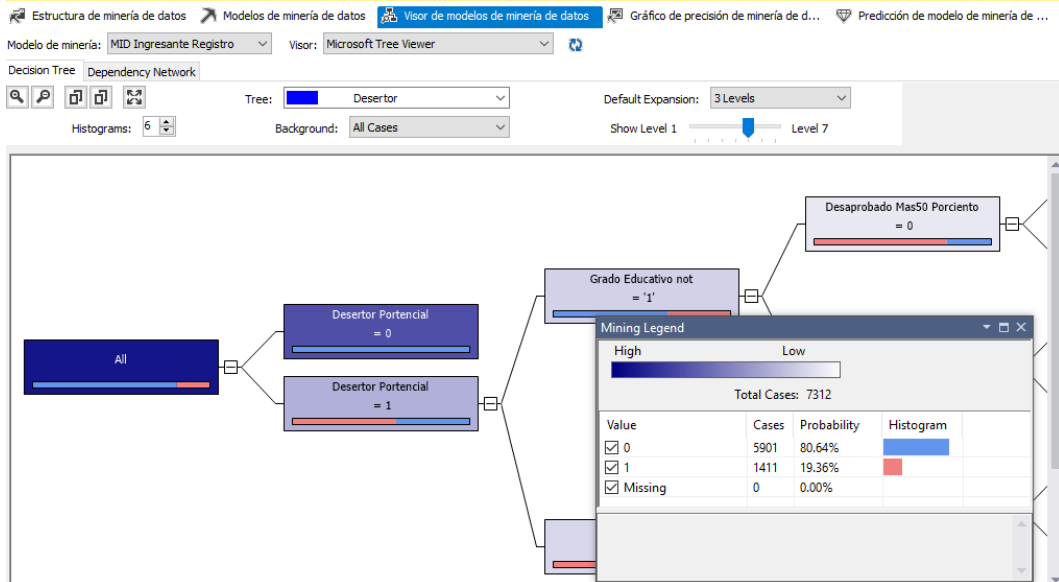
Nota: Selección del nombre del estructura y modelo de minería de datos.

MID Ingresante Registro.dmm [Diseño] X

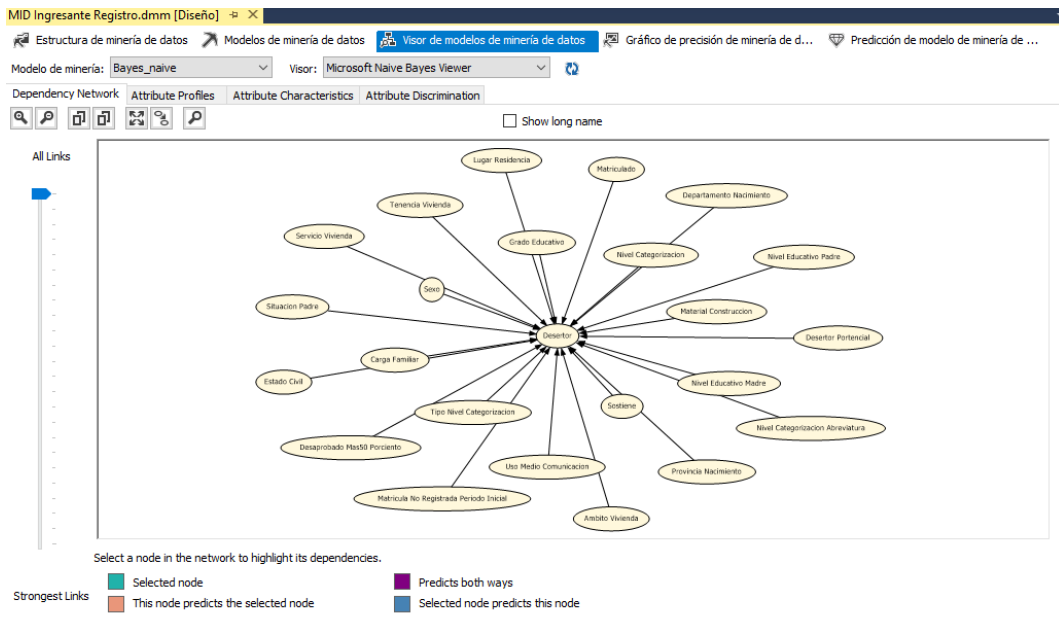
Estructura de minería de datos Modelos de minería de datos Visor de modelos de minería de datos Gráfico de precisión de minería de d... Predicción de modelo de

| Estructura                   | MID Ingresante Registro  | Bayes_naive           | Cluster              | Red_Neural               |
|------------------------------|--------------------------|-----------------------|----------------------|--------------------------|
|                              | Microsoft_Decision_Trees | Microsoft_Naive_Bayes | Microsoft_Clustering | Microsoft_Neural_Network |
| Ambito Vivienda              | Input                    | Input                 | Input                | Input                    |
| Cantidad Postulacion         | Omitir                   | Omitir                | Omitir               | Omitir                   |
| Carga Familiar               | Input                    | Input                 | Input                | Input                    |
| Centro Estudio               | Input                    | Input                 | Input                | Input                    |
| Departamento Nacimiento      | Input                    | Input                 | Input                | Input                    |
| Desaprobado Mas50 Por ciento | Input                    | Input                 | Input                | Input                    |
| Desertor Portencial          | Input                    | Input                 | Input                | Input                    |
| Desertor                     | PredictOnly              | PredictOnly           | PredictOnly          | PredictOnly              |
| Distrito Nacimiento          | Input                    | Input                 | Input                | Input                    |
| Edad                         | Input                    | Omitir                | Input                | Input                    |
| Estado Civil                 | Input                    | Input                 | Input                | Input                    |
| Grado Educativo              | Input                    | Input                 | Input                | Input                    |
| Id Ingresante Registro       | Key                      | Key                   | Key                  | Key                      |
| Ingreso Mensual Familiar     | Omitir                   | Omitir                | Omitir               | Omitir                   |
| Lugar Residencia             | Input                    | Input                 | Input                | Input                    |
| Material Construccion        | Input                    | Input                 | Input                | Input                    |

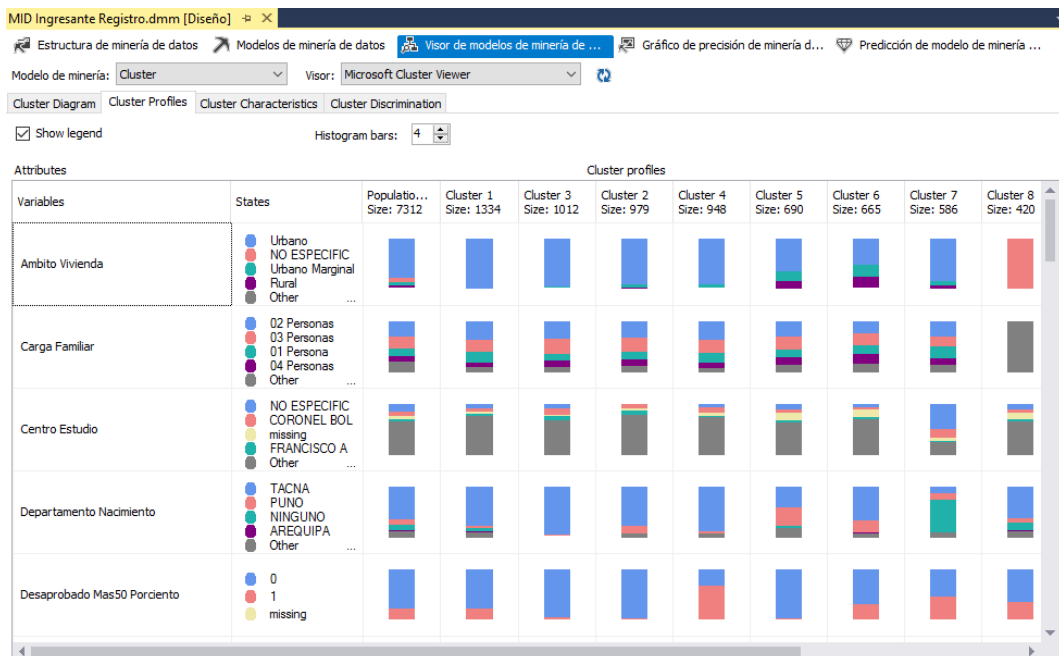
Nota: Creación de varios modelos de minería de datos.



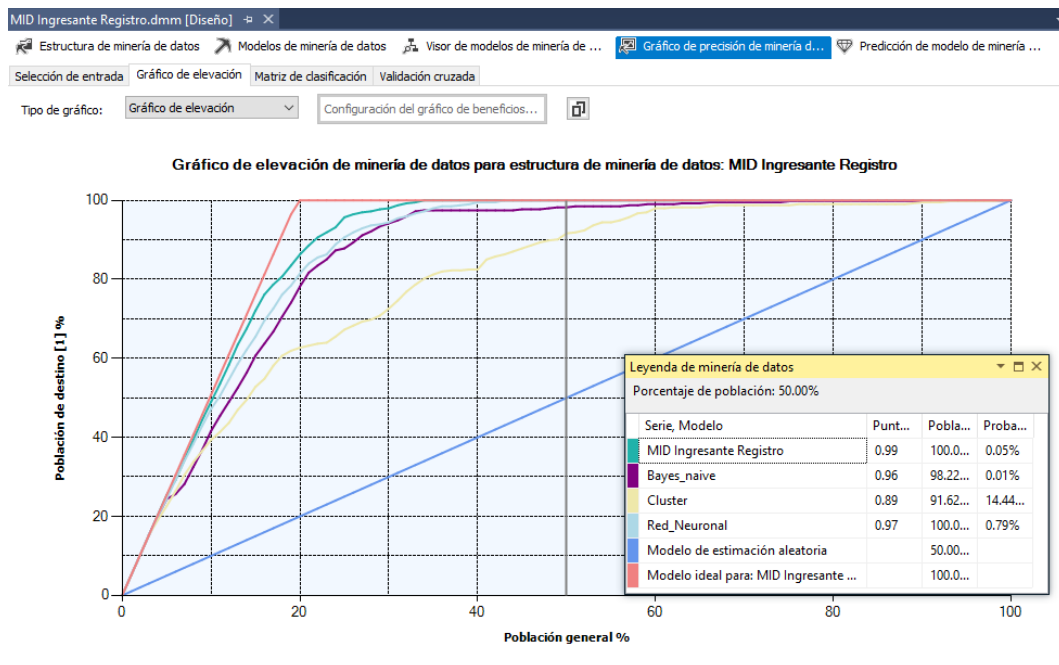
Nota: Visor de modelos de minería de datos de árbol de decisión.



Nota: Visor de modelos de minería de datos de naive bayes.



Nota: Visor de modelos de minería de datos de clúster



Nota: Gráfico de precisión de minería de datos.

Las columnas de las matrices de clasificación corresponden a valores reales; las filas corresponden a valores de predicción

Recuentos para MID Ingresante Registro en Desertor:

| Previsto | 0 (Real) | 1 (Real) |
|----------|----------|----------|
| 0        | 1566     | 77       |
| 1        | 40       | 317      |

Recuentos para RegresionLogistica en Desertor:

| Previsto | 0 (Real) | 1 (Real) |
|----------|----------|----------|
| 0        | 1497     | 58       |
| 1        | 109      | 336      |

Recuentos para NaiveBayes en Desertor:

| Previsto | 0 (Real) | 1 (Real) |
|----------|----------|----------|
| 0        | 1549     | 179      |
| 1        | 57       | 215      |

Nota: Gráfico de la matriz de confusión.

|    |     |                |      |                     |        |
|----|-----|----------------|------|---------------------|--------|
| 1  | 200 | Classification | Fail | 10                  |        |
| 2  | 200 | Classification | Fail | 11                  |        |
| 3  | 200 | Classification | Fail | 13                  |        |
| 4  | 200 | Classification | Fail | 14                  |        |
| 5  | 200 | Classification | Fail | 17                  |        |
| 6  | 200 | Classification | Fail | 12                  |        |
| 7  | 200 | Classification | Fail | 10                  |        |
| 8  | 200 | Classification | Fail | 16                  |        |
| 9  | 200 | Classification | Fail | 14                  |        |
| 10 | 200 | Classification | Fail | 12                  |        |
|    |     |                |      | Promedio            | 12.9   |
|    |     |                |      | Desviación estándar | 2.2561 |

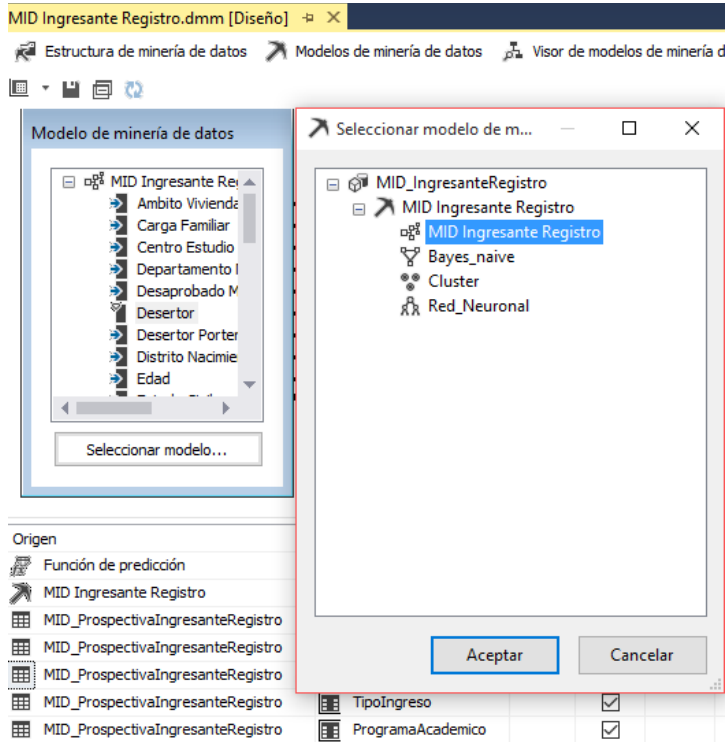
Nota: Valores fallidos de modelo de árbol decisión.

|    |     |                |      |                     |       |
|----|-----|----------------|------|---------------------|-------|
| 1  | 200 | Classification | Fail | 59                  |       |
| 2  | 200 | Classification | Fail | 46                  |       |
| 3  | 200 | Classification | Fail | 44                  |       |
| 4  | 200 | Classification | Fail | 48                  |       |
| 5  | 200 | Classification | Fail | 56                  |       |
| 6  | 200 | Classification | Fail | 43                  |       |
| 7  | 200 | Classification | Fail | 48                  |       |
| 8  | 200 | Classification | Fail | 44                  |       |
| 9  | 200 | Classification | Fail | 48                  |       |
| 10 | 200 | Classification | Fail | 46                  |       |
|    |     |                |      | Promedio            | 48.2  |
|    |     |                |      | Desviación estándar | 4.996 |

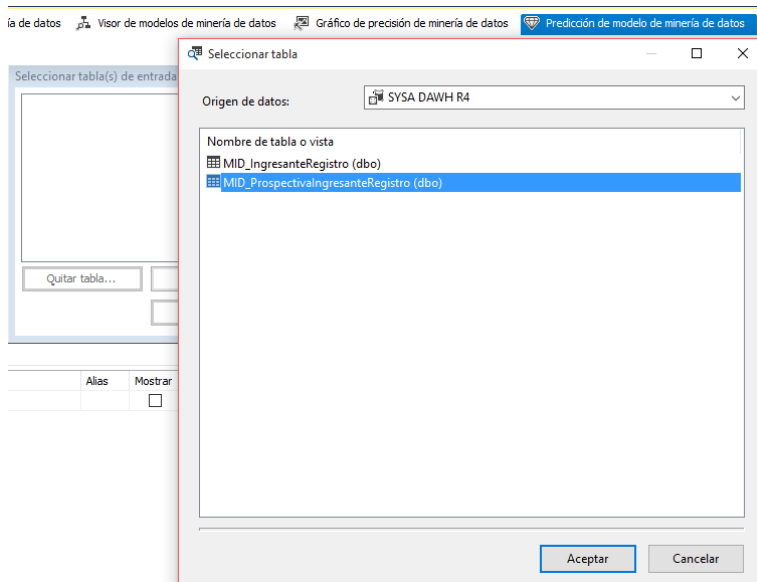
Nota: Valores fallidos de modelo de regresión logística.

|    |     |                |      |                     |        |
|----|-----|----------------|------|---------------------|--------|
| 1  | 200 | Classification | Fail | 23                  |        |
| 2  | 200 | Classification | Fail | 29                  |        |
| 3  | 200 | Classification | Fail | 17                  |        |
| 4  | 200 | Classification | Fail | 26                  |        |
| 5  | 200 | Classification | Fail | 28                  |        |
| 6  | 200 | Classification | Fail | 22                  |        |
| 7  | 200 | Classification | Fail | 26                  |        |
| 8  | 200 | Classification | Fail | 28                  |        |
| 9  | 200 | Classification | Fail | 20                  |        |
| 10 | 200 | Classification | Fail | 18                  |        |
|    |     |                |      | Promedio            | 23.7   |
|    |     |                |      | Desviación estándar | 4.1243 |

Nota: Valores fallidos de modelo de redes bayesianas.



**Nota:** Selección del modelo de minería de datos.



**Nota:** Selección de la tabla de entrada.

| Origen                            | Campo              | Alias | Mostrar                             | Grupo | Y/O | Criterios o argumento                |
|-----------------------------------|--------------------|-------|-------------------------------------|-------|-----|--------------------------------------|
| Función de predicción             | PredictProbability |       | <input checked="" type="checkbox"/> |       |     | [MID Ingresante Registro].[Desertor] |
| MID Ingresante Registro           | Desertor           |       | <input checked="" type="checkbox"/> |       |     | =1                                   |
| MID_ProspectivaIngresanteRegistro | Codigo             |       | <input checked="" type="checkbox"/> |       |     |                                      |
| MID_ProspectivaIngresanteRegistro | Estudiante         |       | <input checked="" type="checkbox"/> |       |     |                                      |
| MID_ProspectivaIngresanteRegistro | LugarResidencia    |       | <input checked="" type="checkbox"/> |       |     |                                      |
| MID_ProspectivaIngresanteRegistro | TipoIngreso        |       | <input checked="" type="checkbox"/> |       |     |                                      |
| MID_ProspectivaIngresanteRegistro | ProgramaAcademico  |       | <input checked="" type="checkbox"/> |       |     |                                      |
|                                   |                    |       | <input type="checkbox"/>            |       |     |                                      |

Nota: Selección de campos, funciones y criterios de la consulta de predicción.

| Expression      | Desertor | Codigo      | Estudiante                      | LugarResidencia  | TipoIngreso               | ProgramaAcademico                           |
|-----------------|----------|-------------|---------------------------------|------------------|---------------------------|---|
| 0.8740808153... | 1        | 2018-178020 | LAURA ESCOBAR, LUIS FERNANDO    | CORONEL GRE...   | FASE I                    | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178026 | FARFAN MAMANI, JESUS RODRIGO    | TACNA            | CEPU - III (Verano)       | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178029 | MAMANI MAMANI, NESTOR HUGO      | CORONEL GRE...   | CEPU - III (Verano)       | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178031 | FERNANDEZ DÁVILA FIGUEROA, S... | CORONEL GRE...   | Ley de Promoción y De...  | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178033 | NAVARRO BABOJA, DIEGO           | TACNA            | Primeros Puestos Coleg... | Carrera Profesional de Ingeniería Ambiental |
| 0.8991068489... | 1        | 2018-178040 | CHAÑE ALANOCA, MILTHON WILBER   | CALANA           | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8991068489... | 1        | 2018-178041 | FIGUEROA LAURA, ALEXIS VALENTIN | CORONEL GRE...   | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178045 | QUENTA LLANO, REBECA MERCEDES   | PACHIA           | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178046 | QUISPE CORI, RONALD DAVID       | CIUDAD NUEVA     | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178047 | QUISPE SINTICALA, TOMAS EDISON  | CORONEL GRE...   | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178048 | ROQUE TAPIA, DAIRA MILENA       | TACNA            | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.8740808153... | 1        | 2018-178050 | YNADO ZEBALLOS, JEAN FRANCO     | TACNA            | FASE II                   | Carrera Profesional de Ingeniería Ambiental |
| 0.9999693811... | 1        | 2018-179001 | RAMOS CONTRERAS, FIORELLA LU... | CORONEL GRE...   | CEPU - I (Otoño)          | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179006 | CALISAYA CALDERON, HELEN LISETH | CORONEL GRE...   | FASE I                    | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179010 | QUISPE CHAMBILLA, MILAGROS Y... | CIUDAD NUEVA     | FASE I                    | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179013 | ALAVE YUCRA, SHEYLA YOSELYN     | CIUDAD NUEVA     | CEPU - III (Verano)       | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179014 | HUAYHUA HUAYHUA, MARLENY MA...  | TACNA            | CEPU - III (Verano)       | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179016 | BENITEZ CALIZAYA, LUCÍA BELÉN   | TACNA            | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179018 | CALIZAYA POMA, IMMER JHAIR      | CORONEL GRE...   | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179022 | ESPEJO GOMEZ, EDU               | CORONEL GRE...   | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179026 | LAYME GOMEZ, JUAN ALEXANDER     | CIUDAD NUEVA     | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179028 | MAMANI CONDORI, MASI LAUTARO    | ALTO DE LA AL... | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179029 | MAMANI GUILLERMO, GABRIELA F... | TACNA            | FASE II                   | Carrera Profesional de Historia             |
| 0.8740808153... | 1        | 2018-179033 | TORRES ILLACHURA, JEROIN        | CIUDAD NUEVA     | FASE II                   | Carrera Profesional de Historia             |

Ejecución de consulta finalizada con 348 filas recuperadas

isto Lin 78 Col 50 Car 50 INS

Nota: Predicción de modelo de minería de datos de árboles de decisión.

| Expresion      | Dese... | Codigo      | Estudiante       | ProgramaAcademico                                  | TipoIngreso                                  |
|----------------|---------|-------------|------------------|--|--|
| 0.860813757... | 1       | 2018-101015 | MAMANI ARMU...   | Carrera Profesional de Ingeniería de Minas         | FASE I                                       |
| 0.556099464... | 1       | 2018-101018 | SILVA VALDIVI... | Carrera Profesional de Ingeniería de Minas         | FASE I                                       |
| 0.689663299... | 1       | 2018-101020 | CANDIA MAMA...   | Carrera Profesional de Ingeniería de Minas         | CEPU - III (Verano)                          |
| 0.715430332... | 1       | 2018-101025 | CHAMBILLA AD...  | Carrera Profesional de Ingeniería de Minas         | Ley de Promoción y Desarrollo del Deporte    |
| 0.858473661... | 1       | 2018-101026 | CHAMBILLA PA...  | Carrera Profesional de Ingeniería de Minas         | Primeros Puestos Colegios de la Región Tacna |
| 0.999798876... | 1       | 2018-101030 | CACALLACA H...   | Carrera Profesional de Ingeniería de Minas         | FASE II                                      |
| 0.978013343... | 1       | 2018-101039 | MAMANI MAMA...   | Carrera Profesional de Ingeniería de Minas         | FASE II                                      |
| 0.624400732... | 1       | 2018-101045 | YUPANQUI PEÑ...  | Carrera Profesional de Ingeniería de Minas         | FASE II                                      |
| 0.999797923... | 1       | 2018-101046 | VILCA CALACH...  | Carrera Profesional de Ingeniería de Minas         | FASE II                                      |
| 0.999802954... | 1       | 2018-102001 | LOPEZ MAMAN...   | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.530832920... | 1       | 2018-102003 | ALANOCA MAM...   | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.691307857... | 1       | 2018-102005 | VELARDE GAR...   | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.699781767... | 1       | 2018-102006 | HUALPA MAMA...   | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.619687411... | 1       | 2018-102008 | ARRATIA PERC...  | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.802857292... | 1       | 2018-102009 | MAMANI YAPU...   | Carrera Profesional de Ciencias Contables y Fin... | CEPU - I (Otoño)                             |
| 0.543547539... | 1       | 2018-102017 | CALAHUILLE S...  | Carrera Profesional de Ciencias Contables y Fin... | CEPU - II (Invierno)                         |
| 0.564112808... | 1       | 2018-102022 | CASTILLO CALI... | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.559829304... | 1       | 2018-102025 | CONDORI CHU...   | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.732212321... | 1       | 2018-102026 | CONTRERAS R...   | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.732346111... | 1       | 2018-102027 | COTRADO NIN...   | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.637617217... | 1       | 2018-102035 | LUPACA LUPAC...  | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.582614927... | 1       | 2018-102037 | MAMANI QUIJ...   | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |
| 0.650131234... | 1       | 2018-102041 | RAJANI CHALC...  | Carrera Profesional de Ciencias Contables y Fin... | FASE I                                       |

Ejecución de consulta finalizada con 329 filas recuperadas

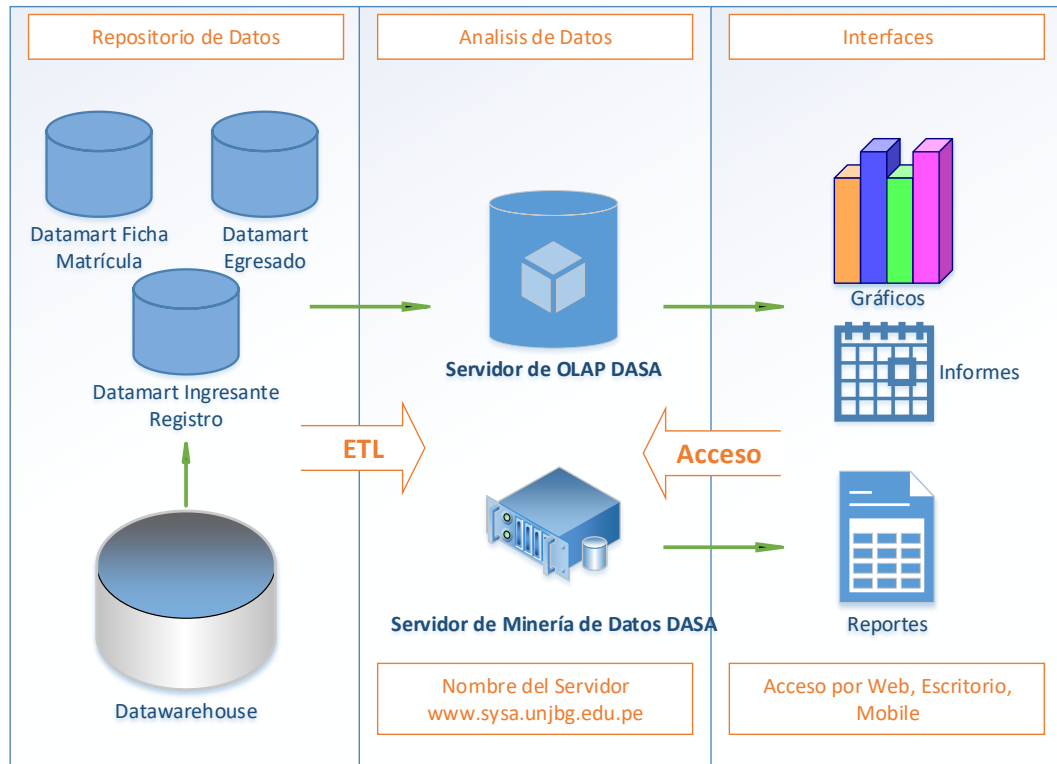
Nota: Predicción de modelo de minería de datos de regresión logística.

| Expresion     | De... | Codigo      | Estudiante             | ProgramaAcademico                                     | TipoIngreso                           |
|---------------|-------|-------------|------------------------|---|---------------------------------------|
| 0.99972658... | 1     | 2018-101030 | CACALLACA HOLGUIN...   | Carrera Profesional de Ingeniería de Minas            | FASE II                               |
| 0.99972658... | 1     | 2018-101046 | VILCA CALACHUA, BR...  | Carrera Profesional de Ingeniería de Minas            | FASE II                               |
| 0.99972658... | 1     | 2018-102001 | LOPEZ MAMANI, MARC...  | Carrera Profesional de Ciencias Contables y Financ... | CEPU - I (Otoño)                      |
| 0.60851206... | 1     | 2018-102085 | JALIRI CENTENO, BLA... | Carrera Profesional de Ciencias Contables y Financ... | FASE II                               |
| 0.99972658... | 1     | 2018-103020 | CCALLATA CARBAJAL, ... | Carrera Profesional de Ingeniería Metalúrgica         | FASE I                                |
| 0.99972658... | 1     | 2018-103029 | TRAVEZAÑO BELISARI...  | Carrera Profesional de Ingeniería Metalúrgica         | FASE I                                |
| 0.66974540... | 1     | 2018-104002 | PARI CANSAYA, JOSU...  | Carrera Profesional de Ingeniería Mecánica            | CEPU - I (Otoño)                      |
| 0.96184180... | 1     | 2018-104010 | CONDORI HUARINGA, ...  | Carrera Profesional de Ingeniería Mecánica            | CEPU - II (Invierno)                  |
| 0.53369286... | 1     | 2018-104018 | LOZA ARONI, MANUEL ... | Carrera Profesional de Ingeniería Mecánica            | FASE I                                |
| 0.67214113... | 1     | 2018-105033 | MAMANI SARMIENTO, ...  | Carrera Profesional de Ciencias Administrativas       | Titulados/Graduados de Univ. Nac. ... |
| 0.99972658... | 1     | 2018-106036 | MAMANI ARRAZOLA, ...   | Carrera Profesional de Ingeniería Comercial           | CEPU - III (Verano)                   |
| 0.68525667... | 1     | 2018-107028 | CHOQUE MAQUERA, G...   | Carrera Profesional de Ingeniería Pesquera            | FASE II                               |
| 0.99972658... | 1     | 2018-107035 | GAUNA CHOQUEMAMA...    | Carrera Profesional de Ingeniería Pesquera            | FASE II                               |
| 0.65380375... | 1     | 2018-108009 | MORALES GUTIERREZ,...  | Carrera Profesional de Agronomía                      | CEPU - II (Invierno)                  |
| 0.72593196... | 1     | 2018-108014 | COHAILA MAMANI, PIE... | Carrera Profesional de Agronomía                      | FASE I                                |
| 0.88228877... | 1     | 2018-108031 | TUEROS GUTIERREZ, C... | Carrera Profesional de Agronomía                      | CEPU - III (Verano)                   |
| 0.67026454... | 1     | 2018-108037 | QUEQUE VILDOSO, RE...  | Carrera Profesional de Agronomía                      | Titulados/Graduados de Univ. Nac. ... |
| 0.99972650... | 1     | 2018-110004 | PARICAHUA MAMANI, ...  | Carrera Profesional de Medicina Veterinaria y Zoot... | CEPU - I (Otoño)                      |
| 0.99972658... | 1     | 2018-110016 | ESCALANTE GODOY, A...  | Carrera Profesional de Medicina Veterinaria y Zoot... | FASE I                                |
| 0.99972658... | 1     | 2018-110017 | HEREDIA ZAVALA, PA...  | Carrera Profesional de Medicina Veterinaria y Zoot... | FASE I                                |
| 0.99972658... | 1     | 2018-110018 | IZQUIERDO URBANO, ...  | Carrera Profesional de Medicina Veterinaria y Zoot... | FASE I                                |
| 0.99972658... | 1     | 2018-110019 | MACHACCA CAÑI, WE...   | Carrera Profesional de Medicina Veterinaria y Zoot... | FASE I                                |
| 0.99972658... | 1     | 2018-110023 | ALANOCA RODRIGUEZ...   | Carrera Profesional de Medicina Veterinaria y Zoot... | CEPU - III (Verano)                   |

Ejecución de consulta finalizada con 74 filas recuperadas

Nota: Predicción del modelo de minería de datos redes bayesianas.

## ANEXO 7: IMPLEMENTACIÓN DE DW Y DM



## ANEXO 8: VALIDEZ DEL INSTRUMENTO

Se ha sometido a 3 jueces para que valoren la validez del instrumento, según formato adjunto:

### I. DATOS GENERALES

1.1. Apellidos y nombres del juez:

1.2. Cargo e institución donde labora:

1.3. Nombre del instrumento evaluado: **Ficha de datos de indicadores**

1.4. Autor del Instrumento: Ing. Nelson Abrahan Pablo Mollo Condorí

### II. ASPECTOS DE EVALUACIÓN

| INDICADORES     | CRITERIOS  | Deficiente<br>00-20% | Regular<br>21-40% | Buena<br>41-60% | Muy Buena<br>61-80% | Excelente<br>81-100% |
|-----------------|--|----------------------|-------------------|-----------------|---------------------|----------------------|
| 1. CLARIDAD     | Está formulado con lenguaje apropiado y comprensible.          |                      |                   |                 |                     | X                    |
| 2. OBJETIVIDAD  | Está expresado en conductas observables.                       |                      |                   |                 | X                   |                      |
| 3. ACTUALIDAD   | Adecuado al avance de la ciencia y tecnología.                 |                      |                   |                 |                     | X                    |
| 4. ORGANIZACIÓN | Existe una organización lógica.                                |                      |                   |                 |                     | X                    |
| 5. SUFICIENCIA  | Comprende los aspectos en cantidad y calidad suficiente.       |                      |                   |                 |                     | X                    |
| 6. PERTINENCIA  | Permite conseguir datos de acuerdo a los objetivos planteados. |                      |                   |                 | X                   |                      |
| 7. CONSISTENCIA | Basado en aspecto teórico-científico.                          |                      |                   |                 |                     | X                    |
| 8. COHERENCIA   | Entre los índices, indicadores y las dimensiones.              |                      |                   |                 | X                   |                      |
| 9. METODOLOGÍA  | La estrategia responde al propósito de la investigación.       |                      |                   |                 |                     | X                    |
| 10. APLICACIÓN  | Los datos permiten un tratamiento estadístico pertinente.      |                      |                   |                 |                     | X                    |

2. OPINIÓN DE APLICABILIDAD: Es factible la aplicabilidad

3. PROMEDIO DE LA VALORACIÓN: 93,3%

Diciembre, 2018.

Firma del Experto Informante

## RELACIÓN DE EXPERTOS

| <b>N°</b> | <b>Apellidos y Nombres</b>        | <b>Cargo</b>                             | <b>Experto</b>            |
|-----------|-----------------------------------|--|---------------------------|
| 1         | Dr. Oscar Mamani Aguilar          | Director de DASA                         | Validador del Instrumento |
| 2         | Ing. Olga Huaman Guzman           | Auditor                                  | Validador del Instrumento |
| 3         | SAP. Lourdes Mendiá Morrón        | Tecnico de registro académico            | Validador del Instrumento |
| 4         | SAP. Alicia Angélica Montes Roque | Jefa de Oficina de registro central      | Encuestado                |
| 5         | SAP. Elba Virginia Plata Ticona   | Asesora del Vicerector académico         | Encuestado                |
| 6         | SAP. Eliana yatto Becerra         | Especialista en registros académico      | Encuestado                |
| 7         | SAP. Sulema Valle Aduvire         | Especialista en registros académico      | Encuestado                |
| 8         | SAP. Isabel Pastor Herrera        | Especialista en registros académico DASA | Encuestado                |

## CUESTIONARIO DE LA FICHA DE DATOS DE INDICADORES DE DESERCIÓN ESTUDIANTIL

| Factores   | Dimensiones   | Indicadores   | Escala                     |
|--|---|---|----------------------------|
| Predictivos  | Individuales  | ¿Es la edad un indicador que influye en la deserción estudiantil?                                   | No -<br>Mediamente<br>- Si |
|  |   | ¿Es el estado civil un indicador que influye en la deserción estudiantil?                           |                            |
|  |   | ¿Es el género un indicador que influye en la deserción estudiantil?                                 |                            |
|  |   | ¿Es el peso un indicador que influye en la deserción estudiantil?                                   |                            |
|  |   | ¿Es la talla un indicador que influye en la deserción estudiantil?                                  |                            |
|  |   | ¿Es el indicador de masa corporal un indicador que influye en la deserción estudiantil?             |                            |
|  |   | ¿Es el lugar de residencia un indicador que influye en la deserción estudiantil?                    |                            |
|  |   | ¿Es el lugar de nacimiento un indicador que influye en la deserción estudiantil?                    |                            |
|  | Académicos  | ¿Es el tipo de ingreso un indicador que influye en la deserción estudiantil?                        | No -<br>Mediamente<br>- Si |
|  |   | ¿Es el centro de estudio un indicador que influye en la deserción estudiantil?                      |                            |
|  |   | ¿Considera Ud. el indicador tipo de centro estudios que influye en la deserción?                    |                            |
|  |   | ¿Ud cree que el programa académico es un indicador que influye en la deserción?                     |                            |
|  |   | ¿El promedio de nota final es un indicador de deserción estudiantil?                                |                            |
|  |   | ¿Considera ud. Que el total de asignaturas desaprobadas influye en la deserción?                    |                            |
|  |   | ¿Considera ud. Que las asignaturas desaprobadas mas del 50 % del semestres influye en la deserción? |                            |
|  |   | ¿Considera ud. Que la preparación de estudio es un indicador que influye en la deserción?           |                            |
|  |   | ¿Considera ud. Que el grado educativo es un indicador relevante en la deserción?                    |                            |
|  |   | ¿Considera ud. Que el periodo lectivo de matrícula es un indicador relevante en la deserción?       |                            |
|  | ¿Considera ud. Que el periodo lectivo de ingreso es un indicador relevante en la deserción? |   |                            |
|  | Socioeconómicos   | ¿Es el nivel educativo del padre un indicador que influye en la deserción estudiantil?              | No -<br>Mediamente<br>- Si |
| ¿Es el nivel educativo del madre un indicador que influye en la deserción estudiantil? |   |   |                            |
| ¿Es el ámbito de vivienda un indicador que influye en la deserción estudiantil?        |   |   |                            |
| ¿Es el tenencia de vivienda un indicador que influye en la deserción estudiantil?      |   |   |                            |

| Factores | Dimensiones | Indicadores  | Escala |
|----------|-------------|--|--------|
|          |             | ¿Es el material de construcción de vivienda es un indicador que influye en la deserción estudiantil? |        |
|          |             | ¿Es el servicio de vivienda un indicador que influye en la deserción estudiantil?                    |        |
|          |             | ¿Es el uso del medio de comunicación un indicador que influye en la deserción estudiantil?           |        |
|          |             | ¿Es la sostención familiar un indicador que influye en la deserción estudiantil?                     |        |
|          |             | ¿Es la situación del padre un indicador que influye en la deserción estudiantil?                     |        |
|          |             | ¿Es el riesgo de salud familiar un indicador que influye en la deserción estudiantil?                |        |
|          |             | ¿Es el carga familiar un indicador que influye en la deserción estudiantil?                          |        |
|          |             | ¿Es el ingreso familiar un indicador que influye en la deserción estudiantil?                        |        |
|          |             | ¿Es el nivel de categorización un indicador que influye en la deserción estudiantil?                 |        |
|          |             | ¿Es el tipo de categorización un indicador que influye en la deserción estudiantil?                  |        |
|          |             | ¿Es la puntuación socioeconómico un indicador que influye en la deserción estudiantil?               |        |

### RESULTADOS DE LA FICHA DE DATOS DE INDICADORES DE DESERCIÓN ESTUDIANTIL

| Factores    | Dimensiones  | N° | EXPERTO 01 | EXPERTO 02 | EXPERTO 03 | EXPERTO 04 | EXPERTO 05 |
|-------------|--------------|----|------------|------------|------------|------------|------------|
| Predictivos | Individuales | 1  | 2          | 1          | 2          | 3          | 1          |
|             |              | 2  | 1          | 2          | 2          | 2          | 2          |
|             |              | 3  | 1          | 1          | 1          | 2          | 1          |
|             |              | 4  | 1          | 1          | 1          | 1          | 1          |
|             |              | 5  | 1          | 1          | 1          | 1          | 1          |
|             |              | 6  | 3          | 3          | 3          | 3          | 3          |
|             |              | 7  | 2          | 2          | 2          | 3          | 2          |
|             |              | 8  | 2          | 2          | 2          | 2          | 2          |
|             | Académicos   | 9  | 3          | 3          | 3          | 2          | 3          |
|             |              | 10 | 3          | 2          | 3          | 2          | 2          |
|             |              | 11 | 3          | 2          | 2          | 2          | 2          |
|             |              | 12 | 2          | 2          | 2          | 2          | 2          |

| Factores | Dimensiones     | N° | EXPERTO 01 | EXPERTO 02 | EXPERTO 03 | EXPERTO 04 | EXPERTO 05 |
|----------|-----------------|----|------------|------------|------------|------------|------------|
|          |                 | 13 | 3          | 3          | 3          | 3          | 3          |
|          |                 | 14 | 3          | 3          | 3          | 3          | 3          |
|          |                 | 15 | 3          | 2          | 3          | 2          | 3          |
|          |                 | 16 | 2          | 2          | 2          | 2          | 2          |
|          |                 | 17 | 1          | 2          | 2          | 1          | 2          |
|          |                 | 18 | 1          | 1          | 1          | 1          | 1          |
|          |                 | 19 | 2          | 2          | 3          | 3          | 3          |
|          | Socioeconómicos | 20 | 3          | 2          | 2          | 3          | 2          |
|          |                 | 21 | 3          | 2          | 2          | 3          | 2          |
|          |                 | 22 | 2          | 2          | 3          | 2          | 2          |
|          |                 | 23 | 2          | 3          | 2          | 2          | 2          |
|          |                 | 24 | 3          | 2          | 2          | 2          | 2          |
|          |                 | 25 | 3          | 2          | 3          | 2          | 2          |
|          |                 | 26 | 2          | 2          | 2          | 2          | 2          |
|          |                 | 27 | 3          | 2          | 3          | 2          | 3          |
|          |                 | 28 | 3          | 3          | 3          | 2          | 3          |
|          |                 | 29 | 3          | 3          | 3          | 3          | 2          |
|          |                 | 30 | 2          | 3          | 2          | 2          | 2          |
|          |                 | 31 | 3          | 3          | 3          | 3          | 2          |
|          |                 | 32 | 2          | 2          | 2          | 2          | 2          |
| 33       | 2               | 2  | 2          | 2          | 2          |            |            |
| 34       | 2               | 1  | 2          | 2          | 2          |            |            |

| Estadísticas de fiabilidad |                |
|----------------------------|----------------|
| Alfa de Cronbach           | N de elementos |
| 0,895                      | 34             |