

UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO

CARRERA:
INGENIERÍA DE SISTEMAS

Trabajo de titulación previo a la obtención del título de:
INGENIERO DE SISTEMAS

TEMA:
ANÁLISIS, DISEÑO Y CONSTRUCCIÓN DE UN ALMACÉN DE DATOS DE
COLOCACIONES DE CRÉDITO Y CAPTACIONES DE LOS BANCOS
PRIVADOS PARA APLICAR ALGORITMOS DE MINERÍA DE DATOS.

AUTOR:
JUAN CARLOS PESANTEZ ANGULO

TUTOR:
WALTER FERNANDO GAIBOR NARANJO

Quito, julio del 2016

Declaratoria de coautoría del docente tutor

Yo, declaro que bajo mi dirección y asesoría fue desarrollado el trabajo de titulación ANÁLISIS, DISEÑO Y CONSTRUCCIÓN DE UN ALMACÉN DE DATOS DE COLOCACIONES DE CRÉDITO Y CAPTACIONES DE LOS BANCOS PRIVADOS PARA APLICAR ALGORITMOS DE MINERÍA DE DATOS realizado por Juan Carlos Pesantez Angulo, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana para ser considerados como trabajo final de titulación,

Quito, julio del 2016,



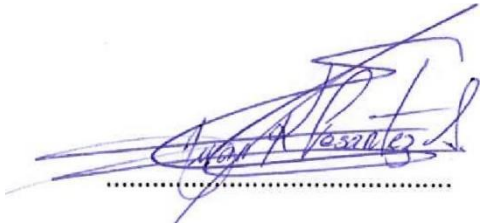
.....

Walter Fernando Gaibor Naranjo

Cédula de identidad: 1713106647

Cesión de derechos de autor

Yo Juan Carlos Pesantez Angulo, con documento de identificación N° 172033739-1, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación intitulado: ANÁLISIS, DISEÑO Y CONSTRUCCIÓN DE UN ALMACÉN DE DATOS DE COLOCACIONES DE CRÉDITO Y CAPTACIONES DE LOS BANCOS PRIVADOS PARA APLICAR ALGORITMOS DE MINERÍA DE DATOS, mismo que ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada, En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.



Juan Carlos Pesantez Angulo

172033739-1

Julio del 2016

DEDICATORIA

Quiero con mucho afecto dedicar éste proyecto a mis padres Víctor Aurelio Pesantez Córdova y Marena Semiramis Angulo, quienes con su paciencia, amor y sabiduría me enseñaron el valor del trabajo, la perseverancia, la honestidad y quienes son mi guía, mi inspiración y mi razón de vivir.

AGRADECIMIENTO

A la Universidad Politécnica Salesiana, donde pude vivir grandes experiencias que me enriquecieron como persona, aquella institución que forjo mediante sus docentes e instalaciones a la persona que ahora pretende cursar un nuevo rumbo de vida.

Al Ingeniero Walter Gaibor, quien desde las aulas impartiendo conocimiento y posterior dirección hizo posible la ejecución de este proyecto.

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO 1	5
ESTADO DEL ARTE.....	5
1.1. Fundamento teórico.....	5
1.1.1. Sistema financiero.....	5
1.1.2. Instituciones financieras.	5
1.1.3. Captaciones.....	5
1.1.4. Colocación o crédito.	5
1.2. Inteligencia de negocios (BI).....	6
1.2.1. Estructura de los sistemas de BI	6
1.3. Data Warehouse.....	7
1.3.1. Data Warehousing.....	7
1.3.2. Arquitectura de un almacén de datos	8
1.3.3. Esquemas para la estructura de datos	9
1.3.4. Elementos de una arquitectura Data Warehouse	10
1.3.5. Función de un Data Warehouse	11
1.3.6. Elementos de un Data Warehouse	12
1.3.6.1. Tablas de hechos.	12
1.3.6.2. Tablas de dimensiones.	13
1.4. Minería de datos.	14
1.4.1. Características de la minería de datos.....	15
1.4.2. Diferencias entre el análisis estadístico y la minería de datos.....	16
CAPÍTULO 2	18
ANÁLISIS Y DISEÑO	18
2.1. Artefactos – Diagramas	18
2.1.1. Herramientas de Hardware.	18
2.1.2. Herramientas de Software.....	18
2.1.2.1 Herramientas de ETL.	18
2.1.2.2. Herramientas de Data Warehouse y consultas OLAP.....	18
2.1.2.3. Herramientas para la gestión de datos.....	19
2.2. Diagrama conceptual de la base de datos.	19

2.3. Arquitectura.....	21
2.3.1. Arquitectura del bus del Data Warehouse.	21
2.3.2. Arquitectura de las tablas de dimensiones.....	21
2.3.2.1. Dimensión fechas.....	22
2.3.2.2. Dimensión entidades.....	22
2.3.2.3. Dimensión tipo_captación.....	22
2.3.2.4. Dimensión tipo_colocación.....	23
2.3.2.5. Dimensión geografía.....	23
2.3.3. Arquitectura de la tabla de hechos.....	24
CAPÍTULO 3.....	26
CONSTRUCCIÓN Y PRUEBAS.....	26
3.1. Descripción de clases.	26
a) Módulo entidad.....	26
b) Módulo región.	27
c) Módulo provincias.....	27
d) Módulo cantón.....	28
e) Módulo tipo depósito.....	28
f) Módulo número de clientes.	28
g) Módulo número de cuentas.	28
h) Módulo saldo.....	28
3.2. Diagrama físico de la base de datos.	29
3.3. Proceso ETL.	31
3.4. Diseño multidimensional.....	32
3.5. Almacén de datos (Cubo).	35
3.6. Minería de datos.	40
3.6.1. Algoritmo de clústeres de Microsoft.	40
3.6.1.1. Funcionamiento del algoritmo de clústeres de Microsoft.....	40
3.6.1.2. Ejecución del algoritmo clústeres de Microsoft para captaciones de crédito.....	41
3.6.1.3. Ejecución del algoritmo clústeres de Microsoft para colocaciones. ..	49
3.6.2. Algoritmo Serie Temporal de Microsoft.	53
3.6.2.1. ¿Cómo funciona el algoritmo de serie temporal de Microsoft?.....	54
3.6.2.2. Ejecución del algoritmo serie temporal de Microsoft.....	54
3.6.3. Algoritmo Bayes naive de Microsoft.....	60
3.6.3.1. ¿Cómo funciona el algoritmo Bayes naive de Microsoft?.....	60

3.6.3.2. Ejecución del algoritmo Bayes naive de Microsoft.	60
3.6.4. Algoritmo Reglas de asociación de Microsoft.....	66
3.6.4.1. ¿Cómo funciona el algoritmo Reglas de asociación de Microsoft?... ..	66
3.6.4.2 Ejecución del algoritmo Reglas de asociación de Microsoft.	66
3.7. Reportes.....	72
3.7.1. Reporte para captaciones de bancos privados del Ecuador.	72
3.7.1.1. Construcción de reportes.....	72
CONCLUSIONES.	80
RECOMENDACIONES.	80
REFERENCIAS.....	81
ANEXOS.....	82

ÍNDICE DE TABLAS

Tabla 1. Arreglo del Data Warehouse.....	21
Tabla 2. Dimensión fechas.....	22
Tabla 3. Dimensión entidades.....	22
Tabla 4. Dimensión tipo_captación.....	22
Tabla 5. Dimensión tipo_colocación.....	23
Tabla 6. Dimensión geografía.....	23
Tabla 7. Tabla de hechos (“Captaciones”).....	24
Tabla 8. Tabla de hechos (“Colocaciones”).....	24

ÍNDICE DE FIGURAS

Figura 1. Etapas Hefesto.	3
Figura 2. Arquitectura del Sistema BI.....	7
Figura 3. Arquitectura de un Almacén de Datos.....	8
Figura 4. Esquema tipo Estrella y tabla de Hechos.....	9
Figura 5. Esquema Copo de Nieve / Tabla de Hechos.....	10
Figura 6. Diagrama Conceptual Captaciones.....	20
Figura 7. Diagrama Conceptual Colocaciones.....	20
Figura 8 Código Módulo Entidad.	27
Figura 9. Ejecución Resultante de Módulos.	29
Figura 10. Diagrama Físico Captaciones.	30
Figura 11. Diagrama Físico Colocaciones.	31
Figura 12. Proyecto Integration Services.	31
Figura 13. Flujo de Datos.....	32
Figura 14. Orígenes de Datos.....	33
Figura 15. Asistente-Vista de Origen de Datos.....	34
Figura 16. Vista de Orígenes de Datos.....	34
Figura 17. Selección del método de creación.....	35
Figura 18. Selección de tablas de grupo de medida.....	36
Figura 19. Selección de medidas.....	37
Figura 20. Selección de nuevas dimensiones.....	37
Figura 21. Finalización del asistente.....	38
Figura 22. Data Warehouse Captaciones.	38
Figura 23. Estructura del Cubo Colocaciones.....	39
Figura 24. Examinador del Cubo Colocaciones.....	39
Figura 25. Gráfico de Dispersión.....	40
Figura 26. Selección del Método de Definición.....	41
Figura 27. Crear la Estructura de Minería de Datos.....	42
Figura 28. Selección de la Dimensión de Cubo de Origen.	43
Figura 29. Dimensión geografía.....	43
Figura 30. Selección de la Clave de Caso.	44

Figura 31. Selección de Columnas de Nivel de Caso.	44
Figura 32. Campos pertenecientes a la tabla de hechos.	45
Figura 33. Especificar el Uso de las Columnas del Modelo.	45
Figura 34. Segmentación del Cubo de Origen.	46
Figura 35. Casilla obtención de detalles.	46
Figura 36. Ingreso de Nombres a la Minería de Datos.	47
Figura 37. Visor de Modelos de Minería de Datos.	48
Figura 38. Detalles de Clústeres 1 y 2.	48
Figura 39. Selección de Algoritmo de Clústeres de Microsoft para Colocaciones... 49	49
Figura 40. Selección de Dimensiones y Columnas de nivel de Clústeres de Microsoft.	50
Figura 41. Campos de Predicción y Tipos de Datos.	50
Figura 42. Segmentación del Cubo y Finalización.	51
Figura 43. Clústeres de Microsoft para Colocaciones.	52
Figura 44. Clústeres con mayor densidad.	53
Figura 45. Crear la Estructura de Minería de Datos.	55
Figura 46. Selección de la Dimensión del Cubo de Origen.	55
Figura 47. Selección de la Clave de Caso.	56
Figura 48. Selección de Columnas de Nivel de Caso.	56
Figura 49. Especificar el uso de las Columnas del Modelo.	57
Figura 50. Segmentación del Cubo de Origen.	57
Figura 51. Finalización del Asistente.	58
Figura 52. Leyenda para Series Temporales.	58
Figura 53. Visor del Modelo de Minería de Datos Serie Temporal.	59
Figura 54. Predicción del Modelo de Minería de Datos.	59
Figura 55. Creación de la Estructura de Minería de Datos "Bayes naive"	61
Figura 56. Dimensión y clave de Caso para "Bayes naive"	62
Figura 57. Columnas de Nivel de Caso y Columnas de Predicción para "Bayes naive"	62
Figura 58. Segmentación del Cubo de Origen y Finalización del asistente para "Bayes naive"	63
Figura 59. Red De Dependencias de "Bayes naive"	64
Figura 60. Leyenda de Morosidad para "Bayes naive"	64
Figura 61. Leyenda de Cartera Vencida para "Bayes naive"	65

Figura 62. Predicciones del Algoritmo "Bayes naive".....	65
Figura 63. Creación de Estructura para "Series Temporales".....	67
Figura 64. Dimensión de origen y clave de caso para "Series Temporales".....	68
Figura 65. Columnas de Nivel de Caso y Columnas de Predicción para "Series Temporales".....	68
Figura 66. Segmentación del Cubo de Origen para "Reglas de Asociación".	69
Figura 67 Finalización del Asistente para "Reglas de Asociación de Microsoft". ...	69
Figura 68. Premisas del Algoritmo "Reglas de Asociación de Microsoft".....	70
Figura 69. Soporte y Tamaño de conjuntos de elementos para “Reglas de Asociación de Microsoft”.	70
Figura 70. Red de Dependencias para "Reglas de Asociación de Microsoft".....	71
Figura 71. Selección para Generar Reportes.....	72
Figura 72. Origen de Datos Compartidos para Reportes.	73
Figura 73. Propiedades de la Conexión del Servidor de Reportes.....	73
Figura 74. Origen de Datos para Informes.....	74
Figura 75. Generador de Consultas para Informes.....	75
Figura 76. Consultas para Informes sobre Captaciones.....	75
Figura 77. Consultas para Informes sobre Colocaciones.....	76
Figura 78. Consulta en Lenguaje SQL.....	76
Figura 79. Selección del Tipo de Informe.....	77
Figura 80. Diseño de la Matriz para Informes.	77
Figura 81. Finalización del Asistente para Informes.	78
Figura 82. Informe Captaciones.....	78
Figura 83. Presentación de Informe Colocaciones.....	79

ÍNDICE DE ANEXOS

Anexo 1. Módulo Región.....	82
Anexo 2. Módulo Provincias.....	82
Anexo 3. Módulo Cantón.....	83
Anexo 4. Módulo Tipo Depósito.	84
Anexo 5. Módulo Número de Clientes.	85
Anexo 6. Módulo Número de Cuentas.....	86
Anexo 7. Módulo Saldo.	87

RESUMEN

El presente proyecto técnico pretende de dar a conocer la minería de datos y su utilidad aplicada a entidades privadas, con el propósito de descubrir información que existe dentro de la data, esta información es manejada mediante archivos planos que son de acceso público, haciendo que con los resultados obtenidos se generen posibilidades de negocio en sitios aun no abarrotados, incrementando ingresos y reduciendo egresos para las organizaciones.

La minería de datos, necesita de estudio y de indagación acerca de su uso e interpretación, así mismo utiliza herramientas algorítmicas para la búsqueda de información oculta.

El presente documento consta de tres capítulos, en donde se detalla a profundidad la relación y utilidad propuesta para otorgar adecuado uso y manipulación a los algoritmos de minería de datos.

El capítulo uno, describe los instrumentos y materiales proporcionados para comenzar la elaboración y análisis de la información para posteriormente agilizar el proceso de toma de decisiones.

El capítulo dos, describe toda la sustentación científica y teórica que cubre con los artefactos y terminología utilizada en la ejecución del proyecto.

El capítulo tres, describe la implementación de la data obtenida e implantada en distintos Data Warehouses, así como los resultados obtenidos del estudio y aplicación de algoritmos de minería de datos para predecir futuros comportamientos dentro del negocio de cada entidad bancaria.

Por último, se generan reportes para colocaciones y captaciones bancarias para que estos sean manipuladas por personas interesadas y que estén dentro de la toma de decisiones de las entidades.

ABSTRACT

This technical project aims to publicize data mining and applied to private entities for the purpose of discovering information that exists within the data useful, this information is handled through flat files that are publicly accessible, making with the results generated business opportunities in sites not yet overcrowded, increasing revenue and reducing expenditures for organizations.

Data mining, you need to study and inquiry into their use and interpretation, also uses algorithmic tools for finding hidden information.

This document consists of three chapters, where depth is detailed the relationship and utility proposed to grant proper use and handling of the data mining algorithms.

Chapter one describes the tools and materials provided to begin the processing and analysis of information to further streamline the decision-making process.

Chapter two describes the scientific and theoretical basis covered with artifacts and terminology used in project implementation.

Chapter three describes the implementation of the data obtained and implemented in different data warehouses as well as the results of the study and application of data mining algorithms to predict future behavior in the business of each bank.

Finally, they report for bank loans and deposits for them to be manipulated by people interested and are within the decision-making entities are generated.

INTRODUCCIÓN

Antecedentes.

La extracción de la data oculta y predecible de grandes bases de datos ayuda a compañías a concentrarse en información útil y relevante, refiriendo a la calidad de la información para las personas. La minería de datos junto con sus herramientas hace que los mismos adquieran valor agregado desde el momento del uso y aplicación en las distintas plataformas, por lo que estos datos se convierten en información y consecutivamente en conocimiento.

La Superintendencia de Bancos del Ecuador (“SBE”), genera boletines financieros con información remitida por las mismas instituciones financieras bajo responsabilidad propia con la finalidad de proporcionar a usuarios en general una visión de la situación financiera de la entidad o grupo de entidades.

Los datos que son publicados por la Superintendencia de Bancos del Ecuador son en formato Excel y al ser consolidados con múltiples entidades financieras se obtiene una data extensa lo cual hace prácticamente inmanejable en el formato anteriormente descrito.

La elaboración de reportes estáticos de forma manual conlleva mucho tiempo al personal de distintas áreas de sistemas que se dedican a la extracción de la data de manera manual para presentarlos en informes según las necesidades del auspiciante o personas interesadas, además en el proceso de carga de datos el usuario interesado tiene que depurar, clasificar y organizar la información.

Dentro del boletín informativo se puede encontrar diferente información relacionada a los depósitos de ahorro, a plazo, en garantía, restringidos y por confirmar con sus respectivos números de clientes, de igual manera se incluye los reportes por línea de negocio clasificada por cartera por vencer, improductiva, total cartera, morosidad, provisiones e indicadores de cobertura.

Justificación.

La Superintendencia de Bancos del Ecuador publica datos históricos de las colocaciones de crédito y captaciones de la banca privada en formato Excel, lo cual hace que sea difícil la administración de la información, por lo que al crear un almacén de datos se facilita la toma de decisiones y posteriormente se aplica

algoritmos de minería de datos para descubrir información de las captaciones y colocaciones en los que respecta a plazas, provincias, regiones correspondientes a captaciones y de asignaciones de créditos correspondientes a colocaciones.

De esta manera se puede presentar reportes para la toma de decisiones en lo que compete a la concesión de créditos y con el uso algoritmos de minería de datos se busca obtener información relevante o que se encuentra escondida dentro de la data.

Según Microsoft, la minería de datos busca ahorrar grandes cantidades de dinero a una empresa para lo cual proporciona un sentido automatizado para identificar información clave desde grandes volúmenes de datos, la aplicación de Data Mining a los almacenes de datos busca dividir los problemas en distintos sectores y esto provoca que exista distintos grupos de trabajo especializados en el ámbito de ese problema, para así optimizar el tiempo y recursos dentro de las instituciones financieras.

Objetivos:

Objetivo general.

Analizar, diseñar y construir un almacén de datos de colocaciones de crédito y captaciones de los bancos privados para aplicar algoritmos de minería de datos.

Objetivos específicos.

- Estudiar y aplicar los diferentes algoritmos de minería de datos para descubrir información que exista dentro de las colocaciones y captaciones.
- Diseñar y construir un almacén de datos para las colocaciones de créditos y captaciones de la banca privada.
- Construir un dashboard de reportería para la toma de decisiones.

Marco metodológico.

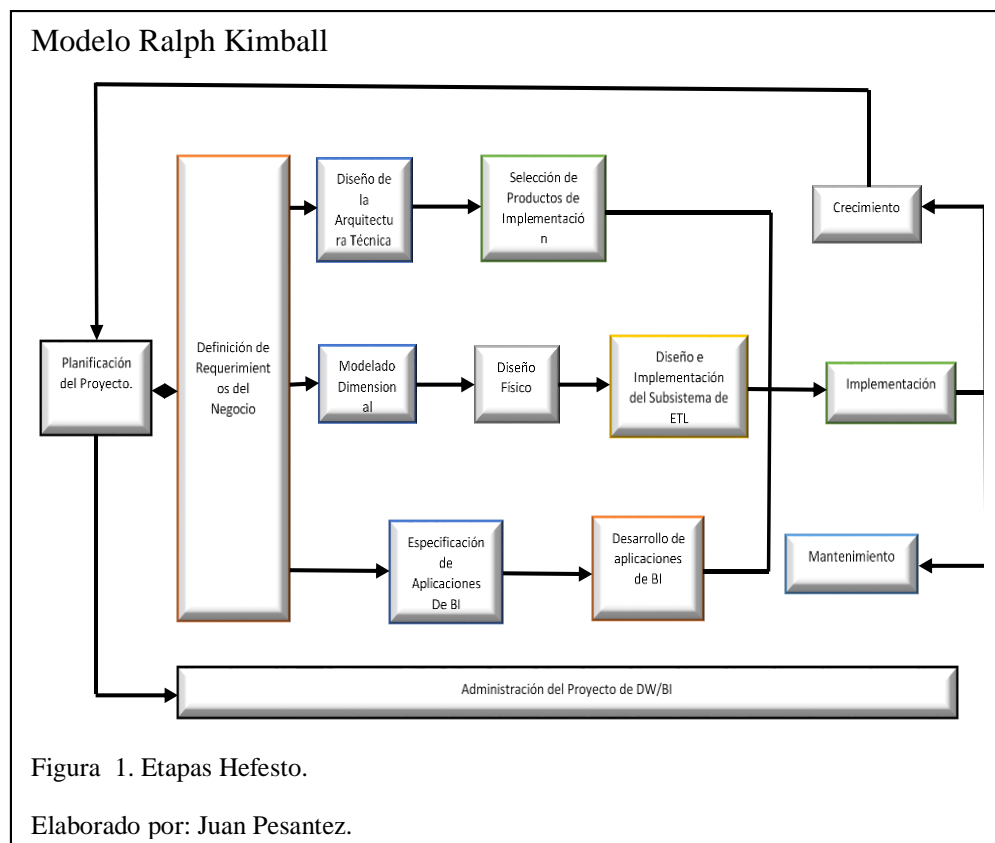
Metodología de Ralph Kimball.

La metodología de Kimball, llamada Modelo Dimensional (Dimensional Modeling), se basa en lo que se denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Esta metodología es considerada una de las técnicas favoritas a la hora de construir un Data Warehouse. (Ralph Kimball, Margy Ross, 2013)

En el Modelo Dimensional se conforman modelos de tablas y relaciones con el propósito de perfeccionar la toma de decisiones, con base en las consultas hechas en una base de datos relacional que están ligadas con un cálculo o un conjunto de mediciones de los resultados de los procesos de negocio.

El Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo exhibir los datos dentro de un cuadro de compromiso estándar e instintivo, para permitir su acceso con un alto rendimiento. (Ralph Kimball, Margy Ross, 2013)

El modelo presentado por Ralph Kimball deja de lado la suficiencia de un data warehouse agrupado debido a que la mayoría de interesados desea obtener información más detallada. Además, impugna que la mejor práctica es almacenar los datos en data marts independientes y utilizando dimensiones lógicamente conectadas. Kimball formula el modelo de datos tipo estrella, lo cual admite la optimización de consultas y facilidad de uso de los data marts (Ralph Kimball, Margy Ross, 2013) (ver Figura1).



Este modelo también es distinguido como “Bottom-up” (“de abajo hacia arriba”), y se fundamenta en la edificación de modelos que son acreditados como data marts autónomos, conformados de una manera usual traspasando la estructura de bus, las uniones de los mismos diseñan el DWH asociado. Con esta característica se hace más maleable y sencilla de implementar, ya que se puede cimentar un data mart como primer elemento del sistema de inteligencia de negocios, y luego ir añadiendo seguidamente otros de acuerdo a los exigencias de cada área. (Ralph Kimball, Margy Ross, 2013)

En este sistema, los procesos ETL (“Extracción, Transformación y Carga”), extraen información de los sistemas operacionales y los procesan, cargando los data mart de

una forma individual, teniendo en cuenta la estandarización de las dimensiones. (Ralph Kimball, Margy Ross, 2013)

Kimball presenta 4 fases para la construcción del DWH que son:

- Elegir el proceso de negocio.

El primer paso es elegir el área a modelar. Esta es una decisión del patrocinador del proyecto, y depende fundamentalmente del análisis de requerimientos y de los problemas analíticos obtenidos de los usuarios.

- Establecer el nivel de granularidad.

La granularidad se trata de definir el nivel más bajo de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que pueda ser útil de los datos actuales.

- Elegir las dimensiones.

Las dimensiones surgen de los requerimientos del negocio discutidas por el equipo de trabajo, también son facilitadas por el nivel de granularidad previamente establecido.

- Identificar medidas y las tablas de hechos.

Este paso consiste en identificar las medidas que nacen de los flujos del negocio. Una medida usa criterios definidos en las dimensiones, es un atributo de una tabla que se desea analizar, sumalizando o agrupando sus datos. (Ralph Kimball, Margy Ross, 2013)

CAPÍTULO 1

ESTADO DEL ARTE

1.1. Fundamento teórico

1.1.1. Sistema financiero.

En el Ecuador el sistema financiero está compuesto principalmente por bancos privados, banca pública, cooperativas, mutualistas, entre otras, las cuales prestan servicios de captación y colocación a sus clientes. (Superintendencia de Bancos del Ecuador, 2016)

Las organizaciones financieras para forjar competencia entre los medios similares del mercado realizan estudios sobre: “Captaciones y Colocaciones” (Superintendencia de Bancos del Ecuador, 2016), en base a la información provista por el organismo de control.

1.1.2. Instituciones financieras.

“Sociedad que interviene en los mercados financieros, cuya actividad consiste en captar fondos del público e invertirlos en activos financieros” (Superintendencia de Bancos del Ecuador, 2016). Con respecto a la banca privada los bancos, las sociedades financieras o corporaciones de inversión y desarrollo, las asociaciones mutualistas de ahorro y crédito para la vivienda y las cooperativas de ahorro y crédito que son intermediarios financieros del público (Superintendencia de Bancos del Ecuador, 2016).

1.1.3. Captaciones.

Según la superintendencia de bancos del Ecuador, las captaciones componen todos los recursos del público que han sido situados en un establecimiento financiero a través de depósitos a la vista, depósitos a plazo u otros mecanismos (Superintendencia de Bancos del Ecuador, 2016).

1.1.4. Colocación o crédito.

En la página web de la superintendencia de bancos del Ecuador, se ha apreciado a la colocación o crédito como un préstamo de dinero que una institución concede a sus interesados, con la responsabilidad de que, en un tiempo expectante el cliente devolverá dicha prestación en forma paulatina, mediante el pago de cuotas, o en un solo pago y con un interés agregado que

compensa al fiador por el período que no tuvo ese capital (Superintendencia de Bancos del Ecuador, 2016).

1.2. Inteligencia de negocios (BI)

“BI es la combinación de prácticas, capacidades y tecnologías usadas por las compañías para recopilar e integrar la información, aplicar reglas del negocio y asegurar la visibilidad de la información en función de una mejor comprensión del mismo y, en última instancia, para mejorar el desempeño” (Rodríguez Parrilla, 2014).

Curto en el libro *Introducción al Business Intelligence* menciona que la inteligencia de negocios es una evolución de los sistemas de soporte a la toma de decisiones, integrando un conjunto de inventivas, aplicaciones, destrezas y capacidades enfocadas a la creación y gobierno de información que permite tomar mejores decisiones a los usuarios de una organización (Curto, 2010).

En definitiva, los sistemas de inteligencia de negocio buscan responder a las preguntas:

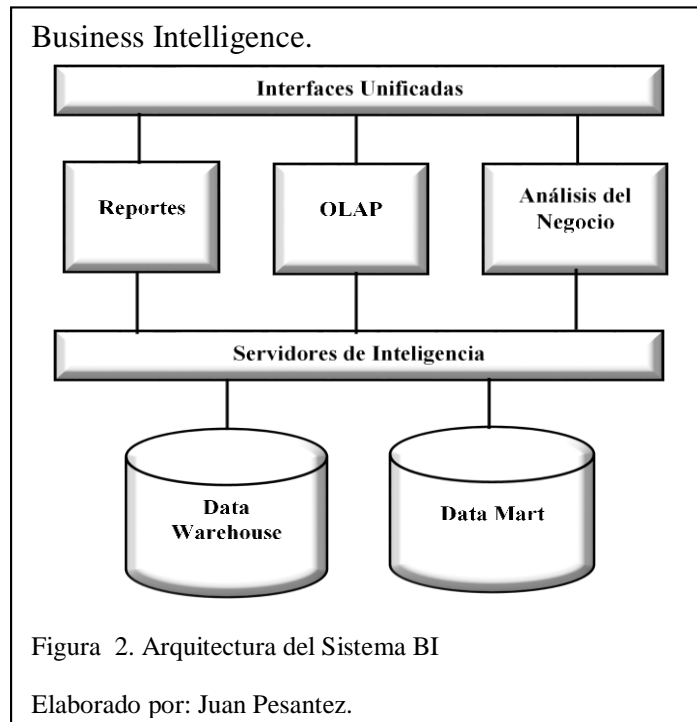
- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿Por qué pasó?
- ¿Qué pasará?

1.2.1. Estructura de los sistemas de BI

La gestión del conocimiento para el apoyo a la toma de decisiones se realiza mediante lo que se conoce como minería de datos, que es el procesamiento y manipulación de la información que se encuentra en el data warehouse para encontrar patrones de desempeño dentro de una organización, como pueden ser tendencias, predicciones entre otros. Las funciones de BI (“Business Intelligence”) en las grandes empresas consisten en crear y manipular los cuadros de mando, crear reportes, realizar análisis en línea (OLAP), así como implementar alarmas y predicciones asociadas a los indicadores clave del desempeño empresarial.

Como lugares de almacenamiento de información, existen dos tipos de bases de datos: las operacionales, que aseguran el sistema ERP (“Enterprise Resource Planning”) o, la planificación de todo tipo de recursos y las bases de datos para el apoyo a la toma de decisiones, entre las que están el data warehouse y los data mart.

En la Figura 2 se puede observar la arquitectura y la relación que tiene con las interfaces unificadas como pueden ser interfaces web o escritorio para usuarios finales.



1.3. Data Warehouse

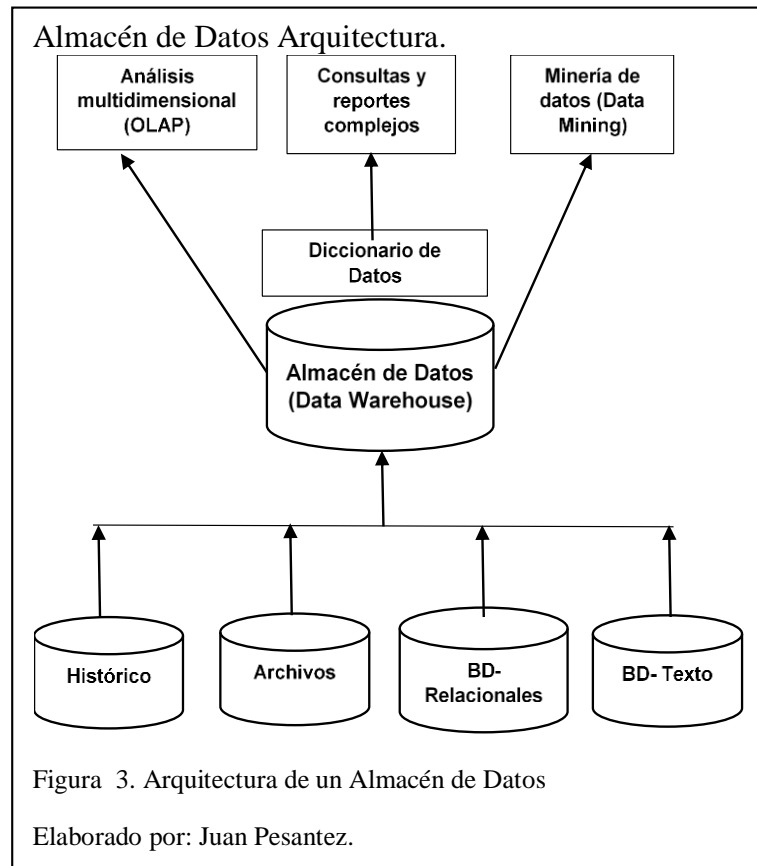
Un almacén de datos o data warehouse es una base de datos corporativa cuya característica principal es la integración y el filtrado de información de una o varias fuentes y posteriormente se procesará para su análisis desde distintos puntos de vista y con una gran velocidad de respuesta (Reinosa, E. J., Maldonado, C. A., & Muñoz, R, 2012).

1.3.1. Data Warehousing

Es el proceso de extraer y filtrar datos asociados, los cuales proceden de diferentes sistemas de información internos o externos, con el objetivo de transformarlos, integrarlos y conservarlos en un almacén de datos o Data Warehouse (Rodríguez Parrilla, 2014).

1.3.2. Arquitectura de un almacén de datos

Un sistema de almacenamiento de datos consta de tres niveles: el primero conformado por las bases de datos, fuentes (de producción e históricos); el segundo, es una base de datos concisa extraída de las bases de datos de producción y, el tercero, compuesto por las interfaces orientadas a consumidores que extraen información para la toma de decisiones (ver Figura3).



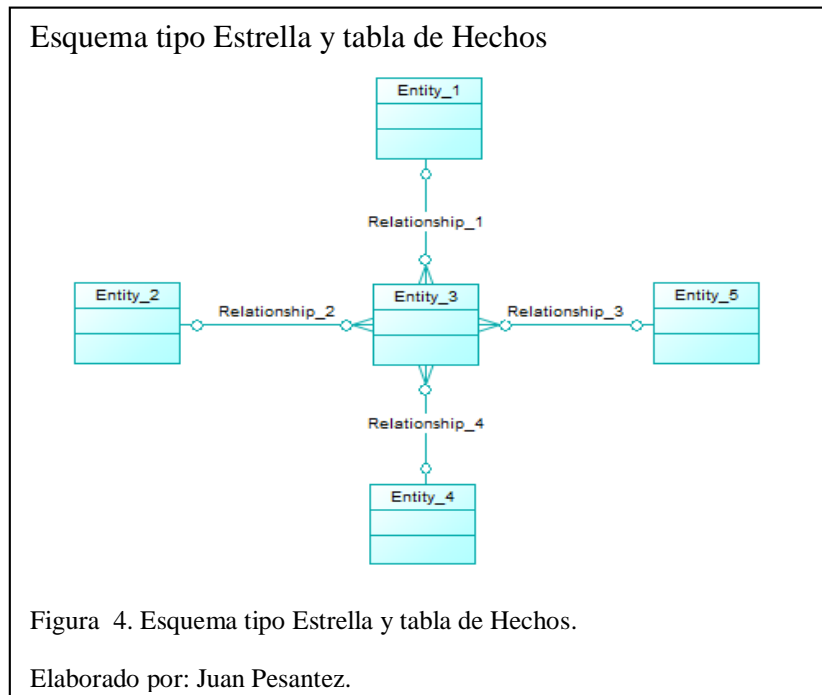
Esta arquitectura se estructura en cinco partes principales:

- Fuentes de datos externas y los procesos responsables de realizar la extracción de los datos de 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 o front-end.
- Servidores OLAP que atienden a consultas de las herramientas de los usuarios finales y los lanza al servidor del almacén de datos.

Los datos históricos no se encuentran generalmente en las bases de datos transaccionales que recogen datos diarios producidos por la empresa, además implica un alto costo por su gran volumen y almacenamiento por lo que se necesita un repositorio específico para los datos históricos.

1.3.3. Esquemas para la estructura de datos

- Esquema en estrella: Se caracteriza por tener una tabla de hechos en el centro de su estructura, las tablas de dimensiones se relacionan a esta mediante claves primarias simples pertenecientes a cada dimensión. La tabla de hechos tiene múltiples relaciones con otras tablas mediante (foreign keys), las tablas de dimensión se encuentran fuera de las formas normales para las bases de datos o desnormalizadas. A continuación, se muestra en la Figura 4 un diagrama tipo estrella.



- Esquema copo de nieve: Es un esquema basado en el tipo estrella, las tablas de dimensiones se normalizan en varias tablas, permitiendo tener varios caminos para llegar a los datos influyendo directamente sobre el rendimiento de las consultas. Por lo tanto, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con varias tablas, donde se crea una tabla por cada nivel de la jerarquía. (Curto, 2010), en la Figura 5 se muestra un ejemplo de esquema tipo copo de nieve.

Esquema tipo Copo de Nieve y tabla de Hechos.

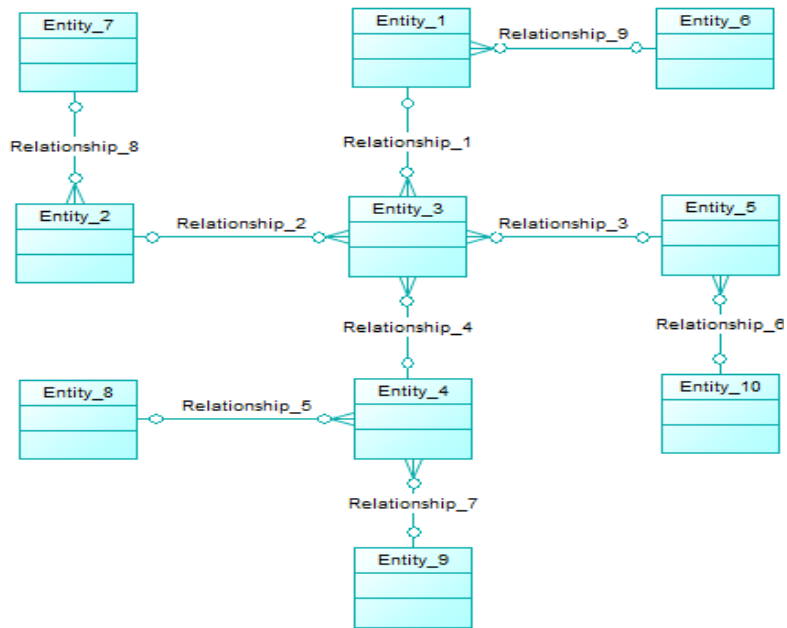


Figura 5. Esquema Copo de Nieve / Tabla de Hechos.

Elaborado por: Juan Pesantez.

1.3.4. Elementos de una arquitectura Data Warehouse

Estos elementos representan la estructura total de datos, y tienen una relación con los siguientes componentes: procesamiento, comunicación y presentación, que interactúan con los usuarios finales. A continuación, se describe cada uno de los elementos:

- Base de datos operacional: Tienen información histórica que proveen una estructura de procesamiento eficiente, donde se almacenan todas las transacciones de la organización.
- Nivel de acceso a la información: Es el nivel donde incluye el hardware y software que permite al usuario final interactuar con las herramientas y aplicaciones desarrolladas para emitir reportes en hojas de cálculos y gráficos que normalmente usa para sus labores diarias. Por ejemplo: Excel, power pivot, crystal reports, etc.

- Nivel de acceso a datos: El nivel de acceso a los datos se encarga de definir una interface entre las herramientas de acceso a la información y las bases de datos operacionales de forma universal. De tal manera que permite a los usuarios finales acceder a cualquier información en la empresa para su trabajo diario.
- Nivel de directorio de datos (metadata): Permite almacenar todas las definiciones de los metadatos, es decir contiene información cuya finalidad es ayudar a la comprensión y navegación de los usuarios a través del DWH.
- Nivel de gestión de procesos: Permite realizar diversas tareas para construir y mantener el DWH actualizado.
- Nivel de mensajes de aplicación: El nivel de mensajes de aplicación permite circular la información a toda la red de la institución.
- Nivel data warehouse (físico): Es un repositorio central de información donde se almacenan de forma que sea fácil de acceder.
- Nivel de organización de datos: Es el componente final de un DWH, incluye los procesos replica necesarios para seleccionar, editar y cargar datos en un DWH para acceder a los datos desde una base de datos operacional.

1.3.5. Función de un Data Warehouse

Los almacenes de datos no solo brindan investigación renovada, sino que también proporciona la obtención de conocimiento a los responsables de la toma de decisiones. Inclusive contiene la capacidad de modelar y remodelar los antecedentes provenientes de la data.

La función del DWH es conseguir datos desde varias bases de información provenientes de fuentes internas y externas, que almacenan recursos necesarios para grandes establecimientos, suministran un ambiente para que las sociedades realicen una mejor apreciación de la información que está siendo gobernada por diversas aplicaciones estratégicas.

Los sistemas para el soporte de decisiones gerenciales basadas en un DWH, pueden ser más prácticas con la explotación de datos, con el fin de ayudar en el estudio del empuje del negocio.

Las principales características que tiene un data warehouse (“DWH”) son:

- Orientado al tema: Aprueba clasificar la información en base a los objetivos estratégicos planteados por la asociación.
- Integrado: En un ambiente data warehouse la investigación encontrada al interior está siempre integrada y se muestra de muchas maneras.
- Tiempo variante: Registra los cambios que se producen a lo largo del tiempo, reflejando los datos originales en cualquier momento.
- No volátil: La información es solo de lectura no se puede eliminar ni modificar una vez acumulada, siempre y cuando se encuentre sólida.

1.3.6. Elementos de un Data Warehouse

1.3.6.1. Tablas de hechos.

Una tabla de hechos es la representación de un proceso de negocio, se constituyen para crear indicadores y son usados por los analistas de negocios para el apoyo a la toma de decisiones, estos permiten almacenar dos tipos de atributos caracterizados:

- Medidas de causa
- Claves foráneas hacia búsquedas de una tabla de dimensión.

Existen diferentes tipos de tablas de hechos:

- Básicos: Son los que se encuentran representados por un campo de una tabla de hechos, campos como precio, unidad, campos que comúnmente no se necesita realizar ningún tipo de operación para poder definir como operación.
- Derivados: Estos se constituyen al adoptar uno o más hechos con alguna acción matemática, poseen la ventaja de que pueden almacenarse previamente calculadas, sus tablas son como: total ventas, rentabilidad, promedio.

- Agregadas: Se forjan luego que se resolvió la consulta proporcionada a la tabla de hechos, se genera activamente a través de instrucciones.
- Pre-Agregadas: Se generan antes de que se realice la consulta correspondiente, de esta manera la consulta se realiza contra una tabla que ya fue previamente sumariada.
- Indicadores: Son valores numéricos generados a través de sumariaciones (suma, promedios, mínimos, máximos, totales, porcentajes, formulas definidas, etc.), dependiendo del requerimiento de información requerida por el negocio.

1.3.6.2. Tablas de dimensiones.

Las tablas de dimensiones delimitan como se encuentra la información organizada en sentido lógico y proporcionan el medio para examinar el argumento del negocio. En todo data warehouse la tabla de dimensión Tiempo es obligatoria, la dimensión tiempo mantiene niveles jerárquicos.

Estas tablas de dimensiones acopian los puntos de análisis de un hecho, estos se pueden catalogar como, SCD que son siglas de Slowly Changing Dimension, y se cuenta como una política de reajuste de datos de una dimensión.

Existen varias formaciones de las tablas de dimensiones:

- SCD 0.

En ningún tiempo se cambia la información, ni se reescribe.

- SCD 1.

Estrictamente sustituye el valor de una propiedad dentro de un campo en una de las dimensiones, supliéndolo con el valor actual.

Asimismo, la particularidad siempre irradia la asignación más nueva.

- SCD 2.

Es la técnica primordial para el alcance de exactitud intercalando lentamente atributos provenientes de una tabla de dimensión, se pide tener claves primarias que cumplan el papel de un identificador único que se fija a cada registro de una tabla de dimensión.

- SCD 3.

Las permutaciones en los caracteres consolidados indican que el valor de la columna no debe pasar de un extremo a otro, agrega una columna dimensión.

Este cambio no adiciona una nueva fila a la dimensión, sino que se amplía una nueva columna para sujetar el cambio de atributo.

Admite ver nuevos e históricos datos de hechos por cualquiera de los valores ignorados de los atributos o antes.

- SCD 4.

A este punto de análisis se le conoce regularmente como tablas históricas

1.4. Minería de datos.

La minería de datos es un conjunto de técnicas que se utilizan para obtener información implícita de las grandes bases de datos.

Según (Trujillo, 2013) 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.

En si la minería de datos la conforman otras disciplinas como la de bases de datos, la estadística, la de visualización de información o machine learning, entre otras.

Dentro de la minería de datos se puede encontrar algunos de sus usos como:

- Descubrir hechos y relaciones de datos.
- Requiere poca intervención humana.
- Encuentra patrones dentro de los datos.
- Determina y establece reglas.
- Almacena y reutiliza reglas.
- Presenta información útil a los usuarios.
- El usuario final debe ser capaz de analizar resultados.

La minería de datos se encarga a través de un conjunto de herramientas y técnicas algorítmicas de buscar los patrones de interés ocultos, que son los que permiten la anticipación de futuros acontecimientos gracias a la predicción de acontecimientos o al pronóstico de situación con cierto grado de probabilidad.

1.4.1. Características de la minería de datos.

Para que la minería de datos se pueda ejecutar y cumplir con su objetivo, debe tener las siguientes características:

- **Recolección de datos a gran escala:** unifica el contenido de la información de todas las bases de datos disponibles, internas o externas. De esta manera se disminuyen los errores y desvíos si la información disponible contiene amplitud y profundidad.
- **Alta tecnología y gran almacenamiento:** como la minería de datos procesa un volumen de información considerable y realiza un importante número de combinaciones, necesita múltiples y veloces procesadores, así como una gran capacidad de memoria RAM, debido a los procesos intermedios de recolección y combinación de datos e información.
- **Algoritmos de minería de datos:** estos funcionan a través de la aplicación de diversas herramientas algorítmicas que son las que permiten la búsqueda de información oculta.

Dentro de la minería de datos se tiene técnicas comúnmente usadas entre las cuales están:

- **Redes neuronales artificiales:** modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.

- Árboles de decisión: estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Algunos de los métodos de árboles de decisión incluyen Árboles de Clasificación y Regresión, Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection).
- Algoritmos Genéticos: técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
- Método del vecino más cercano: una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases de los k registros más similares a él en un conjunto de datos históricos.
- Regla de inducción: la extracción de las reglas if-then de datos basados en significados estadísticos.

La tecnología de la minería de datos, con bases de datos de suficiente tamaño y cantidad, genera nuevas oportunidades de negocios que proveen las siguientes capacidades:

- Predicción automatizada de tendencias y comportamientos. La minería de datos, al automatizar la búsqueda de información predecible en grandes bases de datos, puede inferir, ante una nueva situación o estímulo determinado, cuál sería el comportamiento y la tendencia de respuesta probable de los clientes, proveedores y otros sujetos de la organización.
- Obtención automatizada de modelos previamente desconocidos. Las organizaciones cuentan con parámetros conocidos que permiten la evaluación del comportamiento de sus clientes y productos, obtenidos del histórico de sus bases de datos con respecto al comportamiento de la organización y así proceder a la identificación de los nuevos patrones de tendencia, lo cual hace necesaria el uso de la minería de datos para que barra estos datos de un solo paso, a través de sus herramientas algorítmicas de búsqueda de información oculta.

1.4.2. Diferencias entre el análisis estadístico y la minería de datos.

En los dos casos lo que se desea es mejorar la toma de decisiones a través de un mejor conocimiento del entorno, que se puede estudiar gracias a los datos

que se encuentran almacenados en la organización y también mediante información propia o del medio en que ésta se desenvuelve.

Las técnicas estadísticas se centran generalmente en técnicas confirmatorias, esto es que, a partir de patrones conocidos, demuestran la ocurrencia de los hechos en el paso del tiempo, pero en cambio, las técnicas de la minería de datos son exploratorias, es decir, rastrean el universo de posibilidades para encontrar nuevos patrones desconocidos.

La minería de datos provee mejores soluciones cuando la dimensión del problema es mayor. En un problema con gran cantidad de variables, resulta muy complicado encontrar hipótesis de partida interesantes. En este caso, las técnicas de minería de datos encuentran nuevas relaciones para luego investigar sobre las variables más interesantes. Las técnicas de minería de datos tienen menos limitaciones que las estadísticas, pero la minería de datos al ser menos restrictiva que la estadística, se puede utilizar con supuestos mínimos.

Las técnicas de la minería de datos son muy útiles si los datos de la organización son muy dinámicos. En los casos en los que el almacén de datos es más estático, se justifica una inversión en análisis estadístico. En cambio, en uno dinámico, se necesita que las técnicas de la minería de datos exploren los cambios y determinen si una regla de negocio ha cambiado para abordar ciertas cuestiones relacionadas con el corto y mediano plazo. De esta manera, la minería de datos incidirá en la inversión y en la actualización del negocio. (Reinosa, E. J., Maldonado, C. A., & Muñoz, R, 2012)

CAPÍTULO 2

ANÁLISIS Y DISEÑO

2.1. Artefactos – Diagramas

2.1.1. Herramientas de Hardware.

La herramienta de hardware que se encuentra en uso para la elaboración del proyecto técnico consiente en:

- **Marca:** Sony Vaio
- **CPU:** Intel Core i3, M380, 2.53Ghz
- **Memoria:** 5 GB
- **Sistema Operativo:** Microsoft Windows 7 Ultimate Service Pack 1
- **Modelo:** VPCEA45FL

2.1.2. Herramientas de Software.

2.1.2.1 Herramientas de ETL.

Para la obtención del proyecto técnico, desde sus inicios se propuso usar como instrumento todo el paquete de software SQL Server 2012, ya que en su versión Enterprise con service pack 3, se pueden instalar otras utilidades tales como:

- Analysis Services
- Integration Services
- SQL Server Data Tools

2.1.2.2. Herramientas de Data Warehouse y consultas OLAP.

Para la construcción del cubo o data warehouse se usa la herramienta SQL Server Data Tools, en la que al ingresar se crea un nuevo proyecto llamado Proyecto Multidimensional y de Minería de Datos, propio para Business Intelligence.

Dentro de este proyecto y una vez ya construido el cubo se pueden hacer consultas ya sean estas por medio de lenguaje SQL o en su caso por medio de un entorno visual propio de la herramienta.

2.1.2.3. Herramientas para la gestión de datos.

Para la gestión de los datos se usa el motor de base de datos llamado SQL Server 2012 dentro del cual se alojan todos los datos extraídos mediante procesos, consultas, entre otros, resultantes de los archivos planos proporcionados por la Superintendencia de Bancos del Ecuador. (Superintendencia de Bancos del Ecuador, 2016)

2.2. Diagrama conceptual de la base de datos.

El diagrama conceptual es una descripción de alto nivel del contenido de una base de datos, independiente del sistema gestor que se manipula, esta primera etapa es de percepción, identificación y descripción de los fenómenos tanto internos como externos que afectan al proyecto, usando el lenguaje conocido como modelado conceptual.

Para la construcción del modelo conceptual, inevitablemente se parte de los requerimientos del cliente o auspiciante (“stakeholder”), necesarios para la ejecución y puesta a producción del proyecto, haciendo de esto una vía para alinear los procesos del negocio a las herramientas de T.I. de esta manera se plantean varias preguntas tales como:

- ¿Quién?
- ¿Qué?
- ¿Cuándo?
- ¿Dónde?
- ¿Por qué?

Todas estas preguntas dadas en una vista general de la solución del proyecto puesto a producción.

Para la construcción del data mart a partir de los datos proporcionados es necesario levantar un modelo conceptual que sirve como eje para la base de datos de la cual se va a partir, este modelo conceptual es la plataforma inicial donde se ubican los datos que corresponden a captaciones de bancos privados del Ecuador, como se muestra en la Figura 6.

Diagrama Conceptual Captaciones.

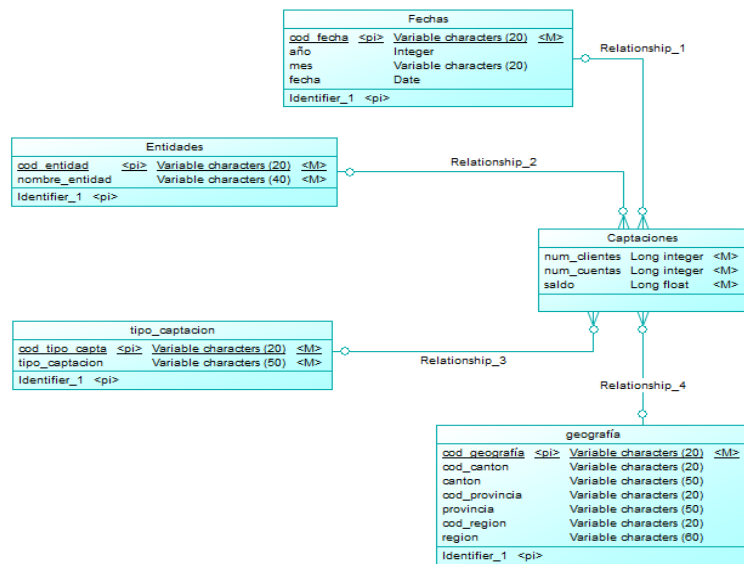


Figura 6. Diagrama Conceptual Captaciones.

Elaborado por: Juan Pesantez.

El modelo conceptual que pertenece a las colocaciones bancarias se construye a base del estándar tipo estrella, así como se vio en el modelo anterior mencionado el cual facilita las relaciones o vínculos que se tiene entre tablas de hechos como de dimensiones, así como se puede observar en la Figura 7.

Diagrama Conceptual Colocaciones

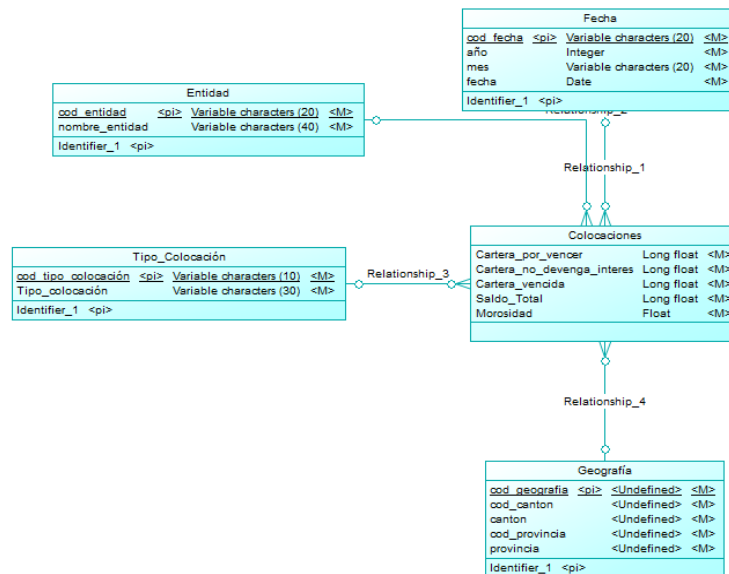


Figura 7. Diagrama Conceptual Colocaciones.

Elaborado por: Juan Pesantez.

2.3. Arquitectura.

2.3.1. Arquitectura del bus del Data Warehouse.

Para esta arquitectura se observan las tablas resultantes de dimensiones y de hechos conformadas previamente, esto permite la integración entre diferentes data mart, lo cual examina que no queden aislados entre sí y pueden ser usados durante todo el proyecto, esto se le conoce como el bus del data warehouse.

La siguiente tabla detalla la arquitectura del bus del data warehouse, con sus tablas de dimensiones y la conformación de las mismas con la unificación de los datos proporcionados, la cual permite la unión de los data marts tanto de captaciones así como de colocaciones bancarias.

Tabla 1. Arreglo del Data Warehouse

Información sobre Captaciones	Entidades	Tipo_captacion	Fecha
Información sobre Entidades	Entidades	Geografía	Fecha
Información Geográfica	Cantón	Provincia	Región
Información sobre Fechas	Fecha	Mes	Año

Nota: La tabla muestra las tablas de dimensiones del DataWarehouse.

Elaborado por: Juan Pesantez.

2.3.2. Arquitectura de las tablas de dimensiones.

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. (Trujillo, 2013).

Según Curto en su libro Introducción al Business Intelligence explica que las dimensiones son la representación en el data warehouse de una vista para un cierto proceso de negocio (Curto, 2010).

2.3.2.1. Dimensión fechas.

Esta dimensión es una de las más importantes para su uso futuro en cuanto a la distribución de los datos, la misma que otorga una jerarquía organizacional ya sea esta por años, meses o por una fecha exacta de donde se extrajo un determinado dato.

Tabla 2. Dimensión fechas.

Fechas	
cod_fecha <pi>	Variable characters (20) <M>
año	Integer
mes	Variable characters (20)
fecha	Date

Nota: La tabla muestra el contenido de la dimensión fechas.

Elaborado por: Juan Pesantez.

2.3.2.2. Dimensión entidades.

Esta dimensión contiene los nombres de las entidades bancarias privadas del Ecuador, así como una clave primaria la cual tendría como función de relacionar las entidades con las otras tablas de dimensiones y con la tabla de hechos.

Tabla 3. Dimensión entidades.

Entidades	
cod_entidad <pi>	Variable characters (20) <M>
nombre_entidad	Variable characters (40) <M>

Nota: La tabla muestra el contenido de la dimensión entidades.

Elaborado por: Juan Pesantez.

2.3.2.3. Dimensión tipo_captación.

En esta dimensión se colocan los 14 ejemplos de tipos de depósitos que se tiene para cada una de las entidades pertenecientes a un cantón específico, así como una clave primaria para la asociación entre las distintas tablas.

Tabla 4. Dimensión tipo_captación

Tipo_Captación	
cod_tipo_capta <pi>	Variable characters (20) <M>

Tipo_Captación	Variable characters (50) <M>
----------------	------------------------------

Nota: La tabla muestra el contenido de la dimensión tipo_captacion.

Elaborado por: Juan Pesantez.

2.3.2.4. Dimensión tipo_colocación.

Esta dimensión es exclusiva para el uso dentro del data mart de colocaciones bancarias, la cual facilita la relación entre los tipos de colocaciones que existen para estas entidades, ya sean de tipo comercial, de consumo, vivienda, entre otros.

Tabla 5. Dimensión tipo_colocación

Tipo_Colocación	
cod_tipo_colocación <pi>	Variable characters (10) <M>
Tipo_colocación	Variable characters (30) <M>

Nota: La tabla muestra el contenido de la dimensión tipo_colocación.

Elaborado por: Juan Pesantez.

2.3.2.5. Dimensión geografía.

Esta dimensión es específica para la localización geográfica de cada una de las entidades dentro del territorio ecuatoriano, contiene campos como región, provincia y cantón, así como un código para cada uno de estos campos.

Tabla 6. Dimensión geografía

Geografía	
cod_geografía <pi>	Variable characters (20) <M>
cod_canton	Variable characters (20)
canton	Variable characters (50)
cod_provincia	Variable characters (20)
provincia	Variable characters (50)
cod_region	Variable characters (20)
region	Variable characters (60)

Nota: La tabla muestra el contenido de la dimensión geografía.

Elaborado por: Juan Pesantez.

2.3.3. Arquitectura de la tabla de hechos.

Los hechos contienen atributos o medidas a analizar. Estos representan normalmente relaciones muchos a mucho con todas las dimensiones y muchos a uno con cada dimensión en particular (Trujillo, 2013).

Una tabla de hecho es una representación de un proceso de negocio. A nivel de diseño una tabla de hechos es aquella que permite almacenar dos tipos de propiedades especiales:

- Medidas del proceso/actividad/flujo de trabajo/evento que se pretende modelizar.
- Claves foráneas hacia registros en una tabla de dimensión. (Curto, 2010)

Tabla 7. Tabla de hechos (“Captaciones”)

Tabla de Hechos. ("Captaciones")			
cod_tipo_capta	pk tabla tipo_captacion	fk1 en tabla de hechos	tipo varchar(20)
cod_entidad	pk tabla Entidades	fk2 en tabla de hechos	tipo varchar(20)
cod_fecha	pk tabla Fechas	fk3 en tabla de hechos	tipo varchar(20)
cod_geografía	pk tabla Geografía	fk4 en tabla de hechos	tipo varchar(20)
num_clientes			tipo bigint
num_cuentas			tipo bigint
saldo			double precision

Nota: La tabla muestra los aspectos principales de la tabla de hechos para captaciones.

Elaborado por: Juan Pesantez.

Tabla 8. Tabla de hechos (“Colocaciones”)

Tabla de Hechos. ("Colocaciones")			
cod_fecha	pk tabla Fecha	fk1 en tabla de hechos	tipo varchar(20)
cod_entidad	pk tabla Entidades	fk2 en tabla de hechos	tipo varchar(20)
cod_geografia	pk tabla Geografía	fk3 en tabla de	tipo

		hechos	varchar(20)
cod_tipo_colocación	pk tabla Tipo_Colocación	fk4 en tabla de hechos	tipo varchar(10)
Cartera_por_vencer			tipo double precisión
Cartera_no_devenga_interes			tipo double precisión
Cartea_vencida			tipo double precisión
Saldo_Total			tipo double precisión
Morosidad			tipo float

Nota: La tabla muestra los aspectos principales de la tabla de hechos para colocaciones.

Elaborado por: Juan Pesantez.

CAPÍTULO 3

CONSTRUCCIÓN Y PRUEBAS.

3.1. Descripción de clases.

Para acelerar el proceso de ETL usando la herramienta de Integration Services de SQL Server Data Tools, se utiliza Visual Basic para Aplicaciones, dentro de los archivos en formato Excel proporcionados por la Superintendencia de Bancos del Ecuador, Visual Basic proporciona una ayuda para la organización de los datos y hacer que se pueda manipular de una mejor manera los campos y celdas dentro del utilitario Excel, se le llama lenguaje de macros para ser utilizado dentro del paquete Microsoft Office, con la sola limitación que el producto final no se puede compilar separadamente del documento, hoja o base de datos en que fue creado, es decir, se convierte en una macro.

Su beneficio principal es automatizar tareas frecuentes, así como crear ejecutables y transacciones de bases de datos (“MS Access”), para uso de plenamente de escritorio ya sea de forma local o remota en el caso de Ms Windows Server. Esto permite acceder a las funcionalidades de un lenguaje orientado a eventos dentro del paquete office con lenguaje basado en Basic 4.0 y 5.0.

El código creado para cada uno de los archivos proporcionados se divide en módulos cada uno con una forma de ejecución individual dentro de una o varias hojas de cálculo.

a) Módulo entidad.

Su función es asignar en las celdas los nombres de las entidades pertenecientes para cada región, provincia, cantón, tipo de captación, número de cuentas, número de clientes y saldo (ver Figura 8).

Código para Módulo Entidad.

```
Sub macro_entidad() 'inicio, nombre del módulo
Dim valor1 As String
Dim valor2 As String
Dim valor3 As String
Dim entidad As String
Dim inicio As Long '-----> asignación de variables
Dim i As Long
Dim j As Long
Dim k As Integer
Dim cuenta As Integer
Dim acumulador As Long
Dim aumento As Long
aumento = 1 'inicializa la variable aumento en 1
acumulador = 156 'asigna el valor de 156 en la variable acumulador
cuenta = 0 'inicializa la variable cuenta
inicio = 4 'donde se encuentra el primer valor
valor1 = Cells(inicio, "A").Value 'toma el valor de la celda asignada
valor2 = Cells(5, "A").Value 'toma el valor de la celda asignada
For i = 1 To 73631 'inicio, asignacion del ciclo for
    valor3 = "Total " & Right(valor1, cuenta) 'asignación a la variable con el número de caracteres de cuenta
    If valor1 = valor2 Or valor1 = valor3 Then 'condición para verificar si es o no una entidad en la celda
        inicio = inicio + 1 'aumenta la cuenta para la variable inicio
        valor1 = Cells(inicio, "A").Value 'toma el valor de la celda asignada
    Else 'caso contrario
        entidad = valor1 'asigno valor de variable a una auxiliar
        For j = aumento To acumulador 'inicio ciclo for
            Cells(j, "AQ") = entidad 'Celdas donde se colocaran los datos
        Next j 'aumento de la variable j del ciclo for
        aumento = aumento + 156 'crecimiento de la variable aumento
        acumulador = acumulador + 156 'crecimiento de la variable acumulador
        inicio = inicio + 1 'aumenta la cuenta para la variable inicio
        cuenta = Len(valor1) 'asigno el número de caracteres correspondientes al nombre de la entidad
        valor1 = Cells(inicio, "A").Value 'toma el valor de la celda asignada
    End If 'fin del ciclo if
Next i 'aumento de la variable i del ciclo for
End Sub 'fin del modulo
```

Figura 8 Código Módulo Entidad.

Elaborado: Juan Pesantez.

b) Módulo región.

Este módulo utiliza dos hojas de cálculo del archivo de Excel, analiza si el campo provincia se encuentra en una de ellas, compara los campos entre las hojas de cálculo y procede a colocar la región a la que pertenece determinada provincia (ver Anexo 1).

c) Módulo provincias.

Para este módulo se utilizó el concepto base del módulo región, el cambio principal radica en que compara el cantón previamente colocado en la hoja de cálculo, con una lista obtenida de todos los cantones del Ecuador en otra hoja de cálculo, para así colocar el nombre de la provincia a quien pertenece dicho cantón (ver Anexo 2).

d) Módulo cantón.

Este módulo coloca el cantón donde pertenece determinada entidad, analizando la celda donde se encuentra el nombre de la entidad y la celda donde se localiza el nombre del cantón a donde pertenece (ver Anexo 3).

e) Módulo tipo depósito.

Para este módulo, su función principal es la de colocar los tipos de depósito pertenecientes a los valores de: número de clientes, número de cuentas y saldo (ver Anexo 4).

f) Módulo número de clientes.

El código para este módulo cumple la función de colocar los valores de números de clientes en forma horizontal junto a su respectivo tipo de depósito, cantón, provincia, región y entidad bancaria. (ver Anexo 5).

g) Módulo número de cuentas.

El módulo número de cuentas es muy similar al módulo número de clientes, su función principal es colocar a nivel de vista vertical a los datos originales a una vista horizontal, alineando dichos valores con los campos predecesores a este módulo (ver Anexo 6).

h) Módulo saldo.

Este módulo no tiene mucha diferencia frente a los anteriores dos módulos, pero sus cambios radican en la creación de variables tipo long, ya que la longitud de datos hace que sea necesario obtener más espacios en cada bloque, lo cual implica mayor espacio de almacenamiento en memoria (ver Anexo 7).

Una vez concluida la ejecución de todos los módulos con la herramienta de Visual Basic para aplicaciones, da como resultante una hoja de cálculo con los datos, valores, nombres y tipos de captaciones (ver Figura 9), organizados de forma que se facilita el proceso de ETL, esto trae como consecuencias la mejora en tiempos de ejecución, optimización de recursos para el procesador, memoria RAM y mayor comprensión en la carga de información al motor de base de datos entre otros.

Hoja de Cálculo Resultante.

	A	B	C	D	E	F	G	H
1	Entidad	Region	Provincia	Canton	Tipo Deposito	num_clientes	num_cuentas	saldo
170	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	520	525	2529162,6
171	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	532	544	2660347,4
172	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	528	541	2815256,5
173	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	531	548	2794077,1
174	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	507	519	2685910,6
175	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	526	546	2881473,7
176	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	532	549	2725621,6
177	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	539	555	2978733,7
178	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	543	607	3116090,9
179	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	516	569	2738352,7
180	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	534	558	2993767,8
181	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS QUE NO GENERAN INTERESES	538	559	2418896,2
182	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	861419,8
183	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	912812,32
184	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	1279700,4
185	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	1459623,3
186	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	115302,55
187	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	282716,55
188	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	25749,39
189	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	24517,1
190	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	22762,33
191	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	3	3	24431,89
192	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	2	2	21403,97
193	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS MONETARIOS DE INSTITUCIONES FINANCIERAS	2	2	27908,87
194	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS DE AHORRO	1356	1384	4888037,8
195	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS DE AHORRO	1369	1397	5093544
196	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS DE AHORRO	1354	1382	5023296,5
197	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS DE AHORRO	1371	1398	5191577,1
198	BP AMAZONAS	Sierra	PICHINCHA	QUITO	DEPOSITOS DE AHORRO	1374	1402	4527993,8

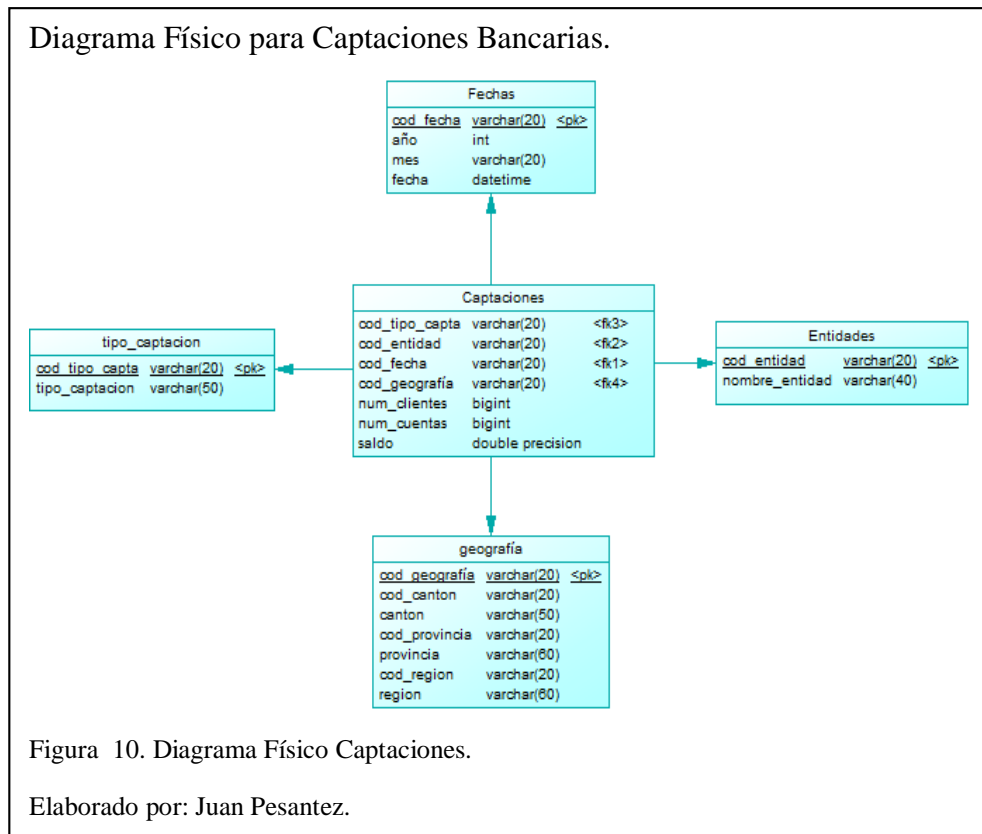
Figura 9. Ejecución Resultante de Módulos.
Elaborado por: Juan Pesantez.

3.2. Diagrama físico de la base de datos.

En la metodología según Kimball dice, “las tablas de dimensión deben tener un único índice sobre las claves primarias y sería recomendable que el índice estuviera compuesto de un único atributo”. También recomienda el uso de índices tipo árbol-B en atributos de alta cardinalidad y aplicar índices de mapas de bits en atributos de cardinalidad media o baja.

El nivel físico es el más bajo de los distintos niveles de donde se encuentran almacenados los datos. No es nada fácil realizar el proceso de crear una base de datos, puesto que se comprende desde el análisis del problema hasta la implementación física en un DMS (DataBase Manager/Management System). El manejador de bases de datos (DMS) consiste en un conjunto de datos interrelacionados y en todos los programas de acceso (Jiménez Capel, María Yolanda, 2014).

En la Figura 10, se muestra el diagrama físico de la base de datos para Captaciones, partiendo del diagrama conceptual mostrado previamente.



Este diagrama permite tener una visión más amplia de cómo es la base de datos resultante, muestra un esquema en estrella donde las tablas que se encuentran por la parte exterior son las tablas de dimensiones y la tabla interior es la resultante tabla de hechos, que por medio de las relaciones de las tablas de uno a muchos obtiene como claves foráneas (fk) varias entidades dentro de esta misma tabla, que no se podía visualizar por medio del diagrama conceptual.

Para el diagrama físico de colocaciones bancarias mostrado en la Figura 11, se sigue la metodología Kimball aplicada a todo el proyecto, colocando una única clave primaria a las tablas de dimensiones la cual pasara por la relación entre tablas hasta la tabla de hechos.

Diagrama Físico para Colocaciones Bancarias.

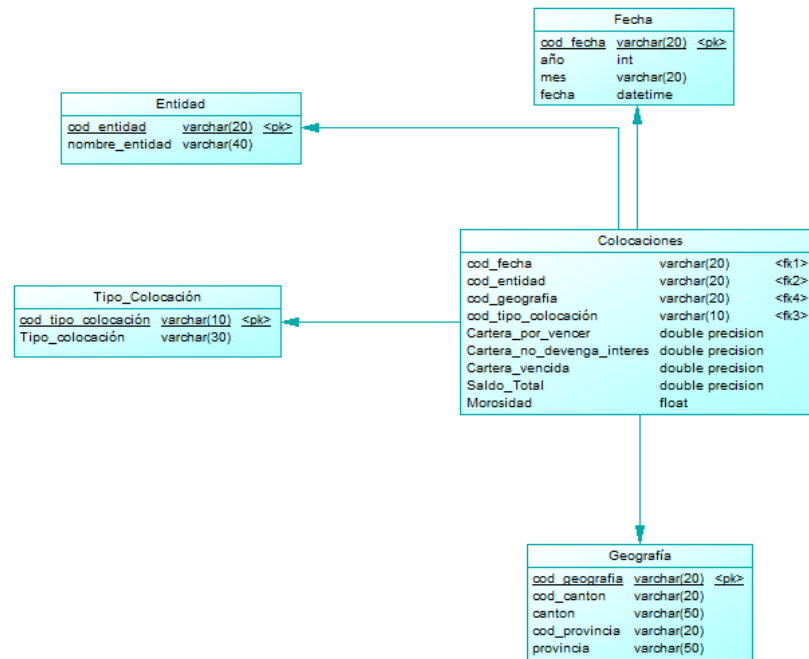


Figura 11. Diagrama Físico Colocaciones.

Elaborado por: Juan Pesantez.

3.3. Proceso ETL.

El proceso de Extracción, Transformación y Carga (ETL), tiene como función principal introducir los datos que se encuentran distribuidos en archivos planos (“EXCEL”), hasta una base de datos ya creada en determinado gestor.

Este proceso se lo realiza mediante la herramienta SQL Server Data Tools en donde se crea un proyecto de Integration Services como lo muestra la Figura 12:

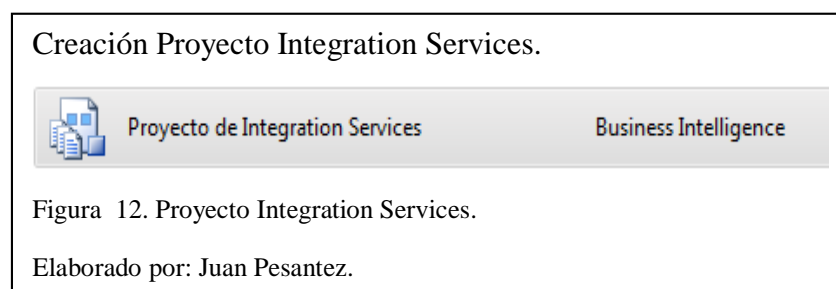
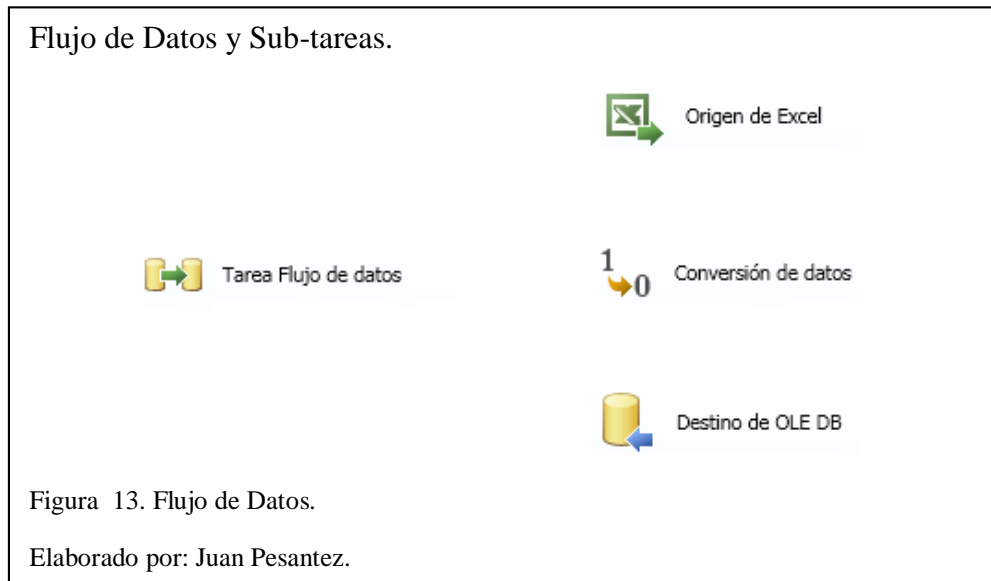


Figura 12. Proyecto Integration Services.

Elaborado por: Juan Pesantez.

Una vez creado el proyecto, se debe establecer un flujo de datos el cual toma como tarea o conjunto de actividades primordiales el almacenamiento de sub procesos a ser ejecutados dentro de la vida del proyecto, estas tareas son las

encargadas de realizar el ETL de manera automática, confiable y rápida, haciendo que los tiempos de ejecución sean mucho menores, comparando con un ingreso manual mediante código SQL dentro del motor de base de datos, en la Figura 13, se muestra el flujo de datos y las tareas a utilizar:



Por la gran cantidad de datos lo recomendable es realizar un ETL por archivo en este caso por año de captación, el proceso se lo puede realizar por medio de un solo ETL con todos los datos unificados, pero demanda gran uso del procesador y esto puede provocar varios errores en su ejecución al contener grandes cantidades de datos para ser procesados.

3.4. Diseño multidimensional.

El diseño multidimensional en contraste del modelo relacional, permite estudiar datos mediante cubos realizados OLAP (On-Line Analytical Processing). Además, de su gran escalabilidad ya que se pueden agregar tablas de hechos y tablas de dimensiones acorde se requiera y de acuerdo a las exigencias del negocio. Se estableció dos data marts que a su vez conforman el data warehouse del cual se consume la información necesaria para la evaluación, construcción y ejecución de las consultas necesarias para la elaboración de la minería de datos.

Para poder crear el diseño multidimensional usando la herramienta SQL Server Data Tools, se procede a llenar la base de datos con la información proveniente de captaciones de los bancos privados del Ecuador desde el año 2011 hasta el año 2015 según el modelo físico con las tablas de dimensiones y las tablas de hechos.

Una vez llena la base de datos, se crea un origen de datos entre la herramienta SQL Server Data Tools y el motor de base de datos SQL Server 2012.

En la Figura 14, se muestra como se hace la conexión entre las 2 herramientas a una base de datos relacional.

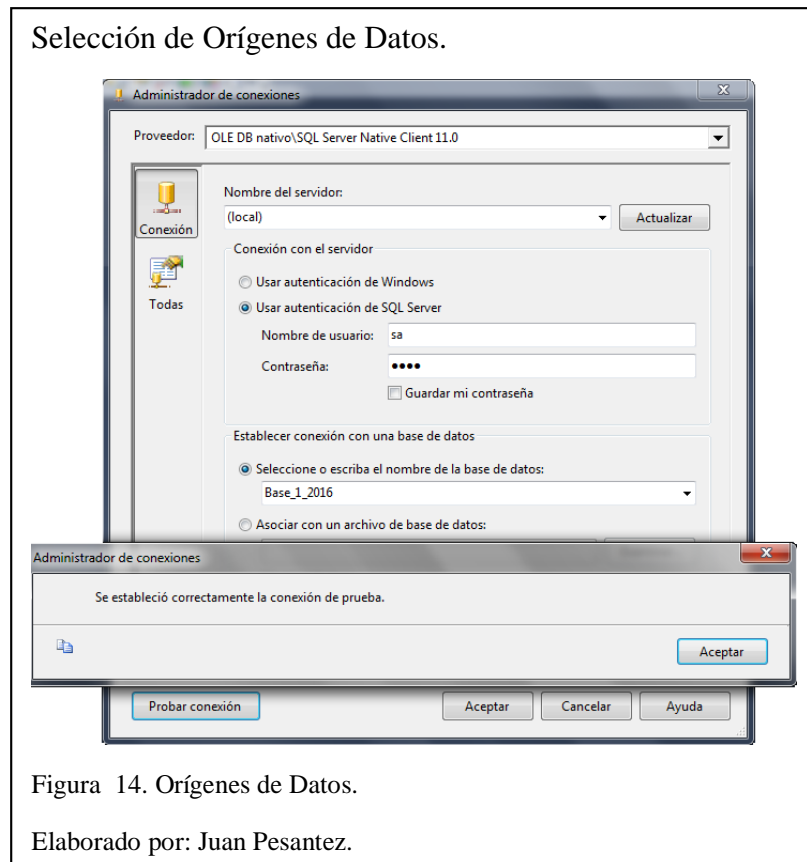


Figura 14. Orígenes de Datos.

Elaborado por: Juan Pesantez.

Una vez realizada la conexión, se crea una vista del origen de datos la cual muestra un diagrama físico de la base de datos relacional, indicando al usuario todas las características tales como los metadatos, relaciones entre tablas, creación de cálculos y configuración de claves lógicas entre una o más tablas (ver Figura 15), que por medio de la herramienta no se tiene impacto en el motor de base de datos o en los metadatos almacenados.

Asistente para Vista de Origen de Datos.

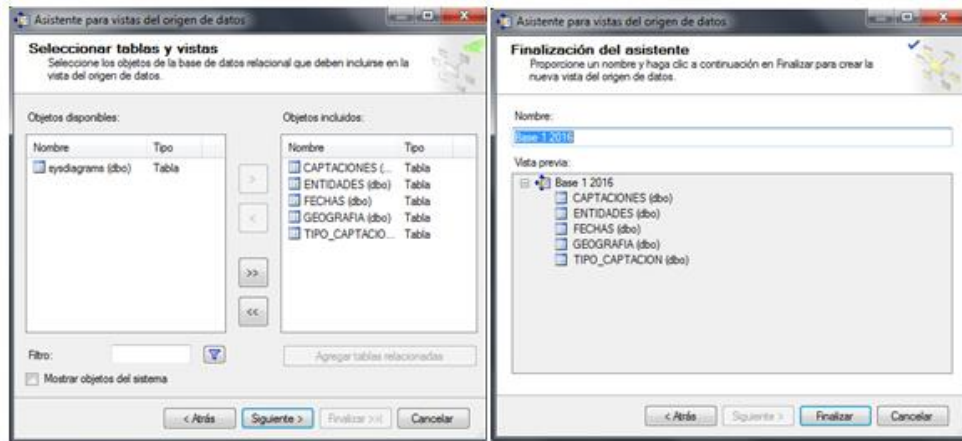


Figura 15. Asistente-Vista de Origen de Datos.

Elaborado por: Juan Pesantez.

Ya que se han escogido las tablas que se va a utilizar para el proyecto, se procede a finalizar el asistente, con lo cual aparece la Figura 16:

Vista Previa de Orígenes de Datos.

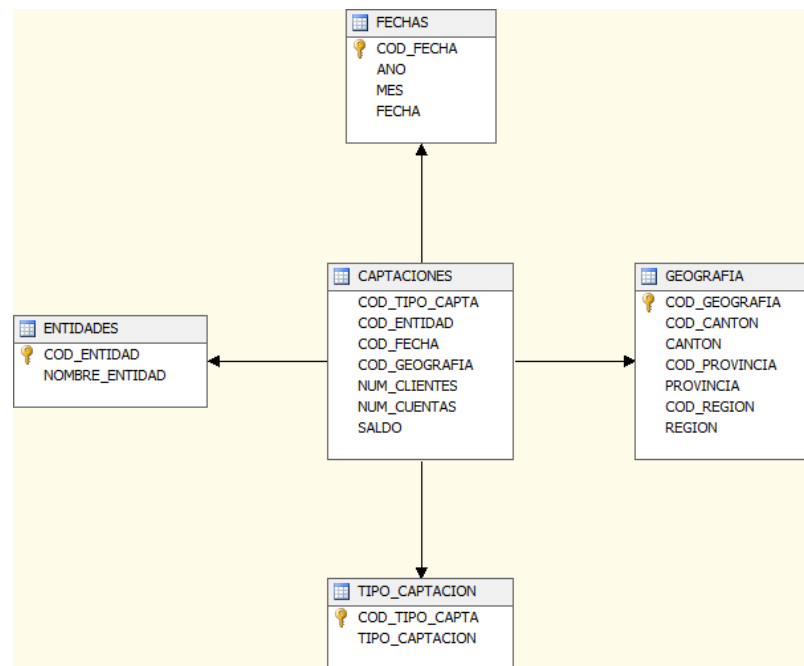


Figura 16. Vista de Orígenes de Datos.

Elaborado por: Juan Pesantez.

3.5. Almacén de datos (Cubo).

“Un almacén de datos, es una recopilación de distintas fuentes de información orientados por tema, compuestos, versátiles en el tiempo y no volátiles que se emplea como soporte a la adquisición de decisiones estratégicas” (Trujillo, 2013).

Los almacenes de datos son como un conjunto de procedimientos que recopilan datos históricos para ser utilizados posteriormente para la toma de decisiones estratégicas. Por ello un almacén de datos no es una base de datos en el sentido tradicional, donde cualquier aplicación de usuario final puede realizar inserciones, actualizaciones y borrados sobre la base de datos (Trujillo, 2013).

Para la construcción del almacén de datos previamente se debe tener realizada una conexión a una base de datos y tener formada una vista de origen de datos para proporcionar un panorama único del modelado relacional de la base de datos explicados en numerales anteriores.

Como primer paso se toma como guía el asistente de creación de cubos, como se muestra en la Figura 17.

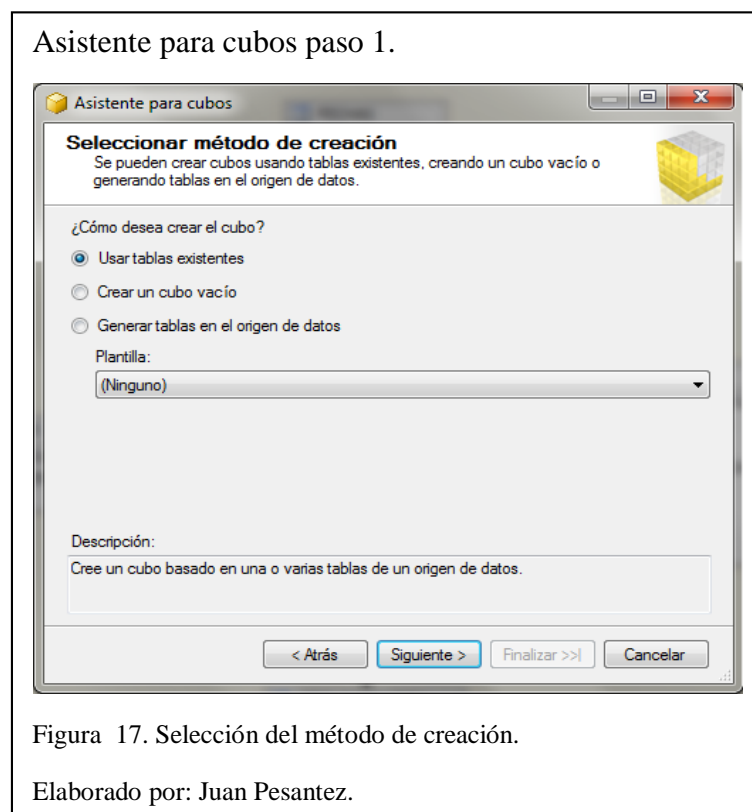
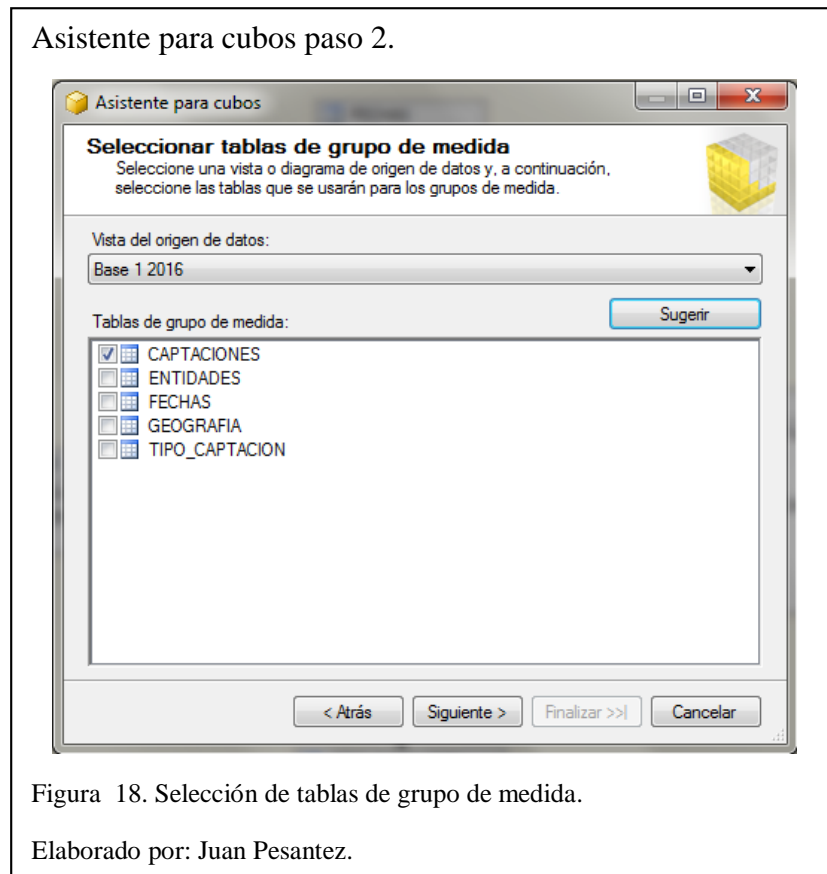


Figura 17. Selección del método de creación.

Elaborado por: Juan Pesantez.

El siguiente paso es seleccionar la tabla que es para el proyecto propuesta como una tabla de hechos de entre las tablas compuestas por la base de datos, aquellas tablas no seleccionadas, el software las analiza y las coloca como tablas de dimensiones de manera automática, como lo muestra la Figura 18.



El paso siguiente del asistente de creación de cubos, es la selección de las medidas, las que son tomadas de las tablas de hechos antes seleccionadas y que pertenecen al data mart de captaciones (ver Figura 19).

Asistente para cubos paso 3.

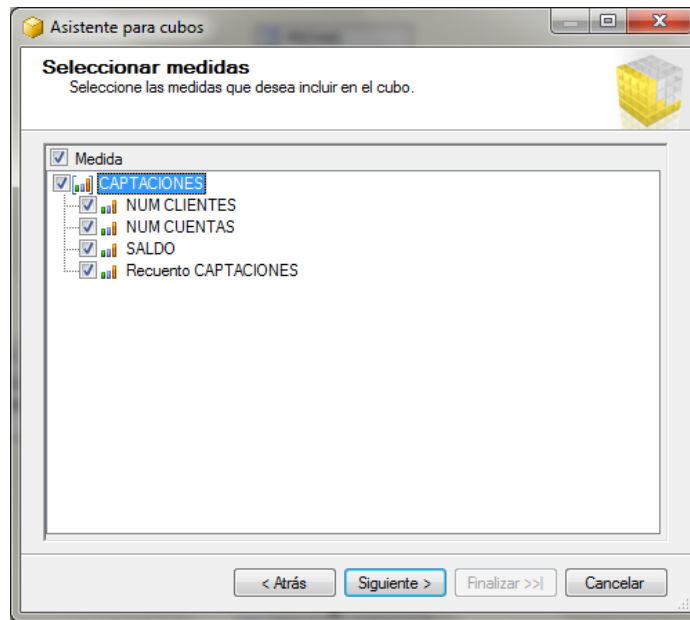


Figura 19. Selección de medidas.

Elaborado por: Juan Pesantez.

A continuación, el asistente de creación de cubos, muestra una pantalla donde se puede elegir el grupo de tablas de dimensiones para el cubo, tal y como se muestra en la Figura 20.

Asistente para cubos paso 4.

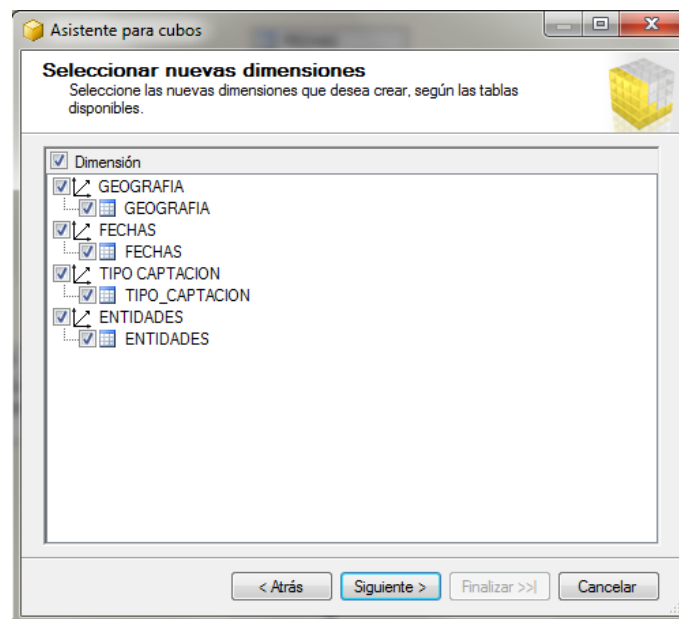
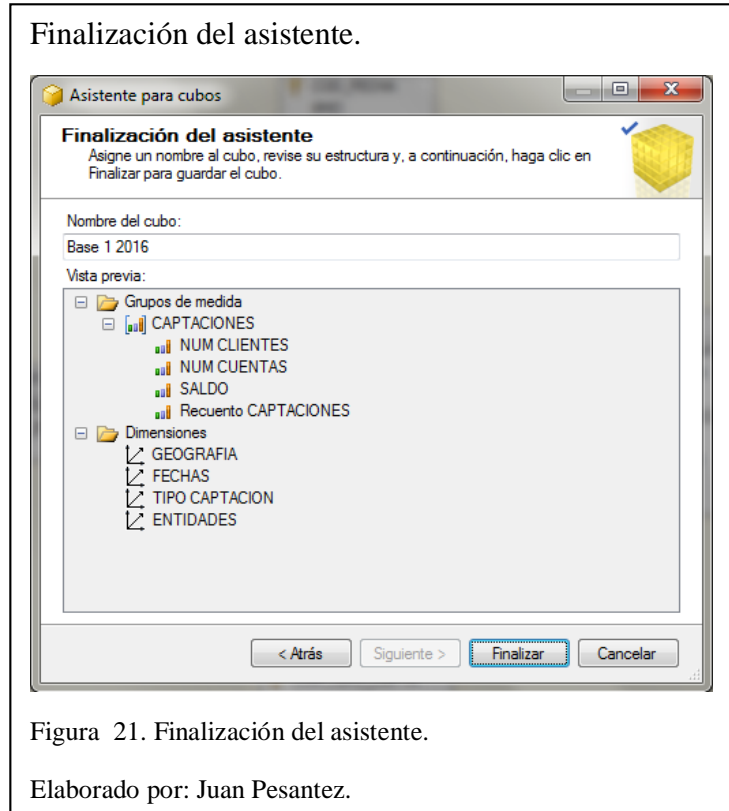


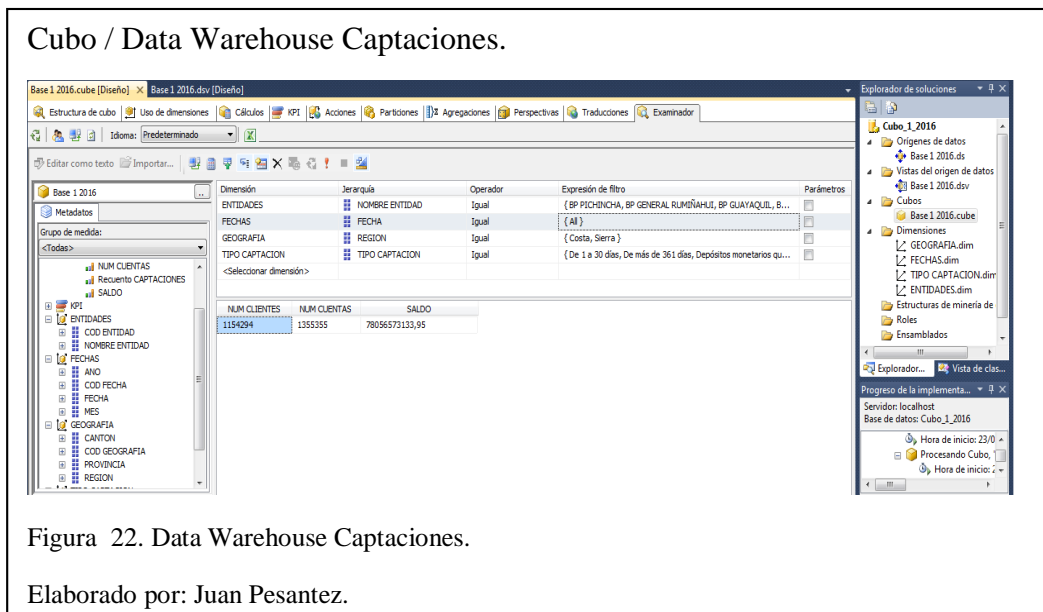
Figura 20. Selección de nuevas dimensiones.

Elaborado por: Juan Pesantez.

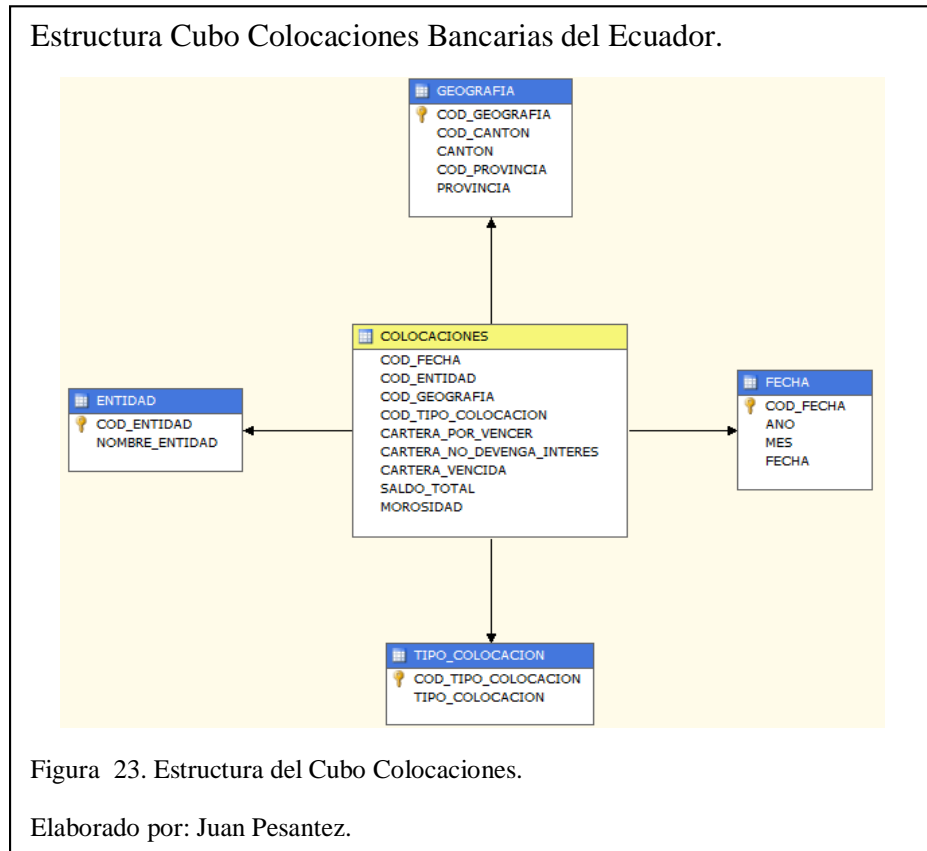
Una vez realizado todos los pasos anteriores el asistente de creación de cubos, muestra una pantalla como resumen de la tabla de hechos y las tablas de dimensiones que serán utilizadas para la creación del cubo (Ver Figura 21).



Con la ejecución de este último paso, se tiene creado el cubo dimensional o data warehouse con las tablas de dimensiones y tablas de hechos correspondientes a captaciones de los bancos privados del Ecuador desde el año 2011 hasta el año 2015, como lo muestra la Figura 22.



Para la construcción del cubo concerniente a colocaciones de los bancos privados del Ecuador, se realiza el mismo proceso mencionado previamente, obteniendo como resultado un data warehouse con tablas de dimensiones y tablas de hechos pertenecientes al data mart de colocaciones, previamente ingresados todos los datos de los distintos tipos de colocaciones existentes, para tener como resultado la Figura 23, que representa la estructura del cubo de Colocaciones.



Dentro del examinador del cubo, se puede ingresar distintos tipos de consultas para verificar la existencia, veracidad y coherencia entre la información ingresada en el motor de base de datos, así como la información que se encuentra generada por la herramienta de creación de cubos de Microsoft como lo muestra la Figura 24.

Examinador para el Cubo Colocaciones.

Dimensión	Jerarquía	Operador	Expresión de filtro	Parámetros
ENTIDAD	NOMBRE ENTIDAD	Igual	{ All }	<input type="checkbox"/>
FECHA	COD FECHA	Igual	{ All }	<input type="checkbox"/>
GEOGRAFIA	CANTON	Igual	{ All }	<input type="checkbox"/>
TIPO COLOCACION	TIPO COLOCACION	Igual	{ All }	<input type="checkbox"/>
<Seleccionar dimensión>				

CARTERA NO DEVENGA INTERES	CARTERA POR VENCER	CARTERA VENCIDA	MOROSIDAD	SALDO TOTAL
21368749261,1702	946428619409,71	9127704252,23002	76961859427,8257	848089527714,195

Figura 24. Examinador del Cubo Colocaciones.
Elaborado por: Juan Pesantez.

3.6. Minería de datos.

3.6.1. Algoritmo de clústeres de Microsoft.

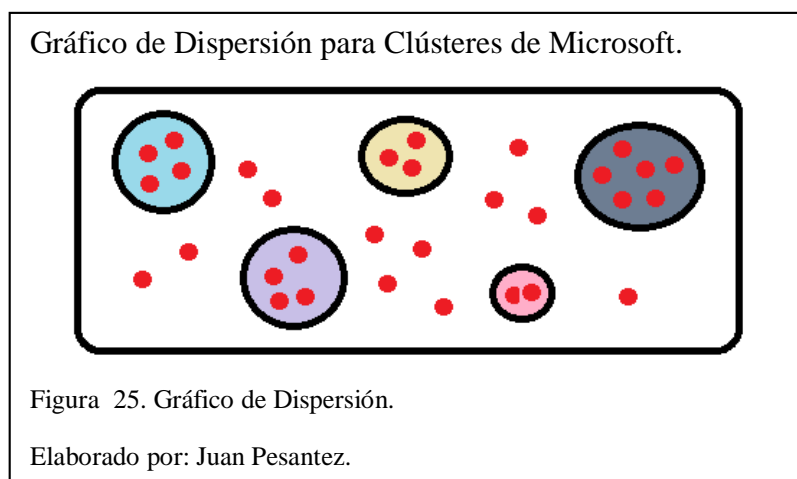
El algoritmo de clústeres de Microsoft, es un algoritmo de fraccionamiento suministrado por Analysis Services. El algoritmo utiliza métodos iterativos para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones. (Microsoft, 2016)

Los modelos de agrupación en clústeres identifican las relaciones en un conjunto de datos que no se podrían derivar lógicamente a través de la observación casual. (Microsoft, 2016)

El algoritmo de clústeres se diferencia de otros algoritmos de minería de datos, como el algoritmo de árboles de decisión de Microsoft, en que no se tiene que designar una columna de predicción para generar un modelo de agrupación en clústeres. El algoritmo de clústeres entrena el modelo de forma estricta a partir de las relaciones que existen en los datos y de los clústeres que identifica el algoritmo. (Microsoft, 2016)

3.6.1.1. Funcionamiento del algoritmo de clústeres de Microsoft

Identifica primero las relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas. Un gráfico de dispersión es una forma útil de representar visualmente el modo en que el algoritmo agrupa los datos, (ver Figura 25). El gráfico de dispersión representa todos los casos del conjunto; cada caso es un punto del gráfico. Los clústeres agrupan los puntos del gráfico e ilustran las relaciones que identifica el algoritmo. (Microsoft, 2016)



Después de definir los clústeres, el algoritmo calcula el grado de perfección con que los clústeres representan las agrupaciones de puntos y, a continuación, intenta volver a definir las agrupaciones para crear clústeres que representen mejor los datos. El algoritmo establece una iteración en este proceso hasta que ya no es posible mejorar los resultados mediante la redefinición de los clústeres. (Microsoft, 2016)

3.6.1.2. Ejecución del algoritmo clústeres de Microsoft para captaciones de crédito.

Para crear una estructura de minería de datos en SQL Server Data Tools, se utiliza el asistente de minería de datos el cual guía paso a paso para la correcta ejecución del algoritmo Clústeres de Microsoft.

La figura 26, muestra que el asistente solicita seleccionar un método para definir la estructura de minería de datos, en este caso se selecciona a partir de un cubo existente previamente construido, el mismo que facilita el proceso para la minería de datos.

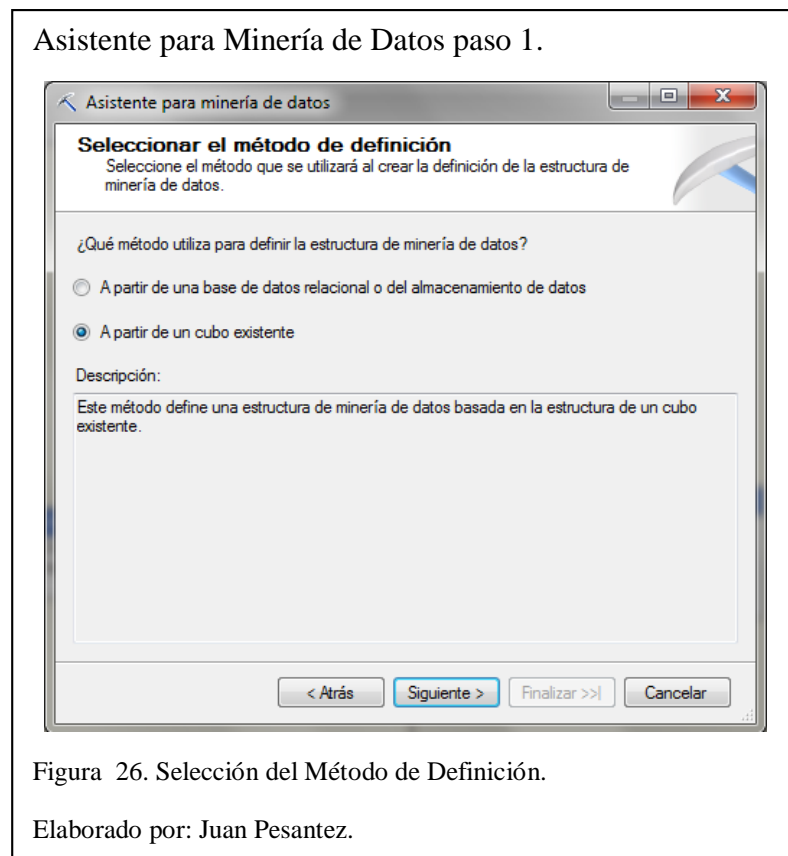


Figura 26. Selección del Método de Definición.

Elaborado por: Juan Pesantez.

El siguiente paso, es seleccionar la técnica de minería que se va a utilizar, optando entre varios algoritmos propios de la herramienta de Microsoft, para este caso práctico se selecciona el algoritmo de Clústeres de Microsoft (ver Figura 27).

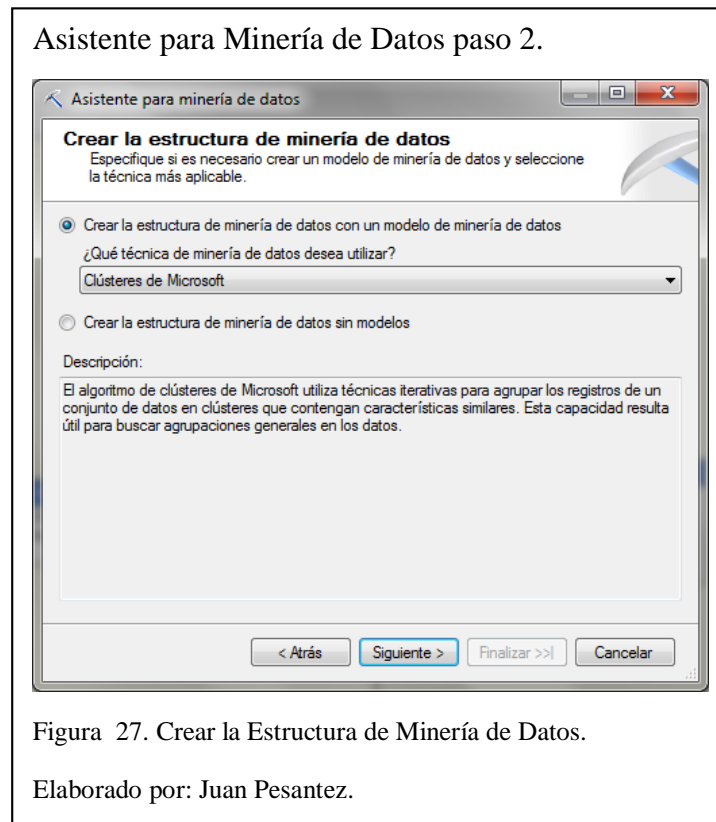


Figura 27. Crear la Estructura de Minería de Datos.

Elaborado por: Juan Pesantez.

Una vez seleccionado una de las herramientas de minería de datos, surge en el apartado inferior la descripción perteneciente a cada algoritmo, para que de esta manera se tenga una idea clara de cómo es el funcionamiento del mismo.

A continuación, se debe seleccionar la tabla de dimensión que el algoritmo toma como base para el análisis de minería de datos, la misma que se coloca como modelo de guía para la agrupación de datos propia de este algoritmo como lo muestra la Figura 28.

Asistente para Minería de Datos paso 3.

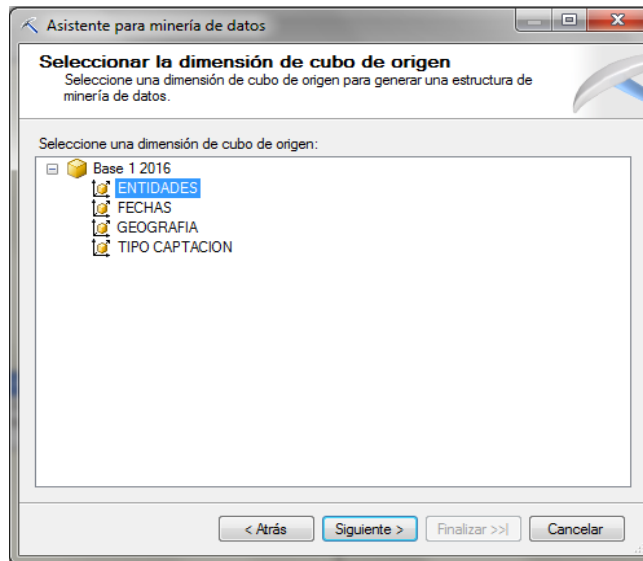


Figura 28. Selección de la Dimensión de Cubo de Origen.

Elaborado por: Juan Pesantez.

Al continuar con el asistente, este requiere seleccionar el campo de la dimensión que se ocupa para la agrupación de datos, ya que varias tablas pueden tener múltiples campos (ver Figura 29).

Data Warehouse Dimensiones.

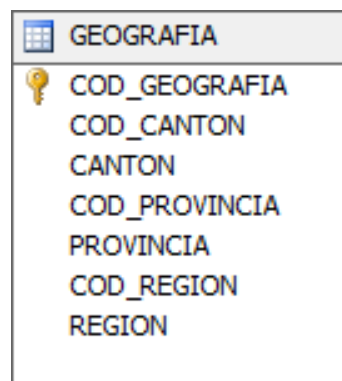
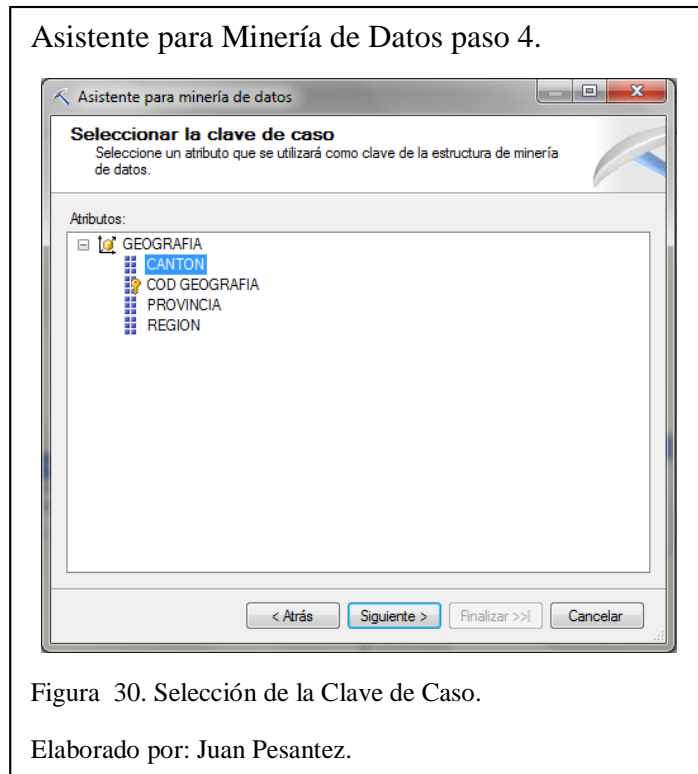


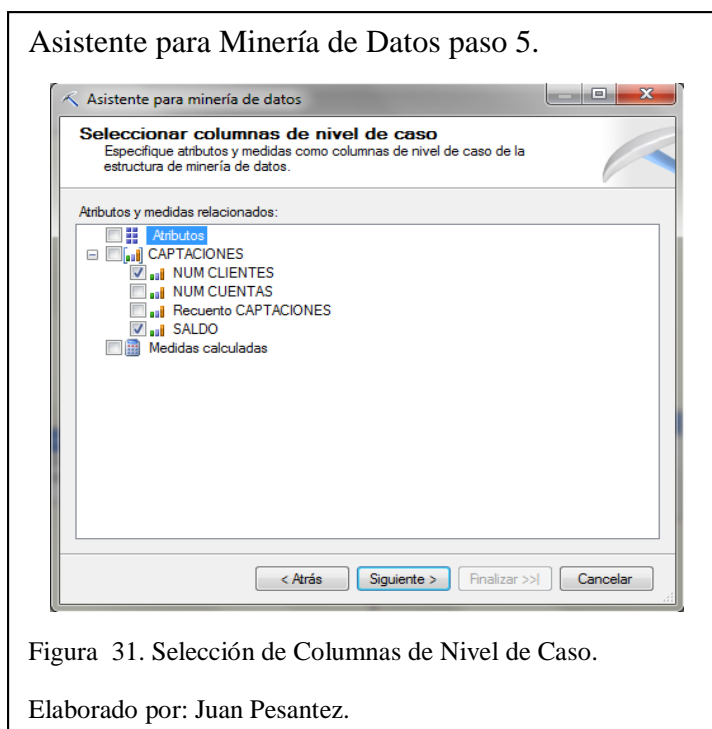
Figura 29. Dimensión geografía.

Elaborado por: Juan Pesantez

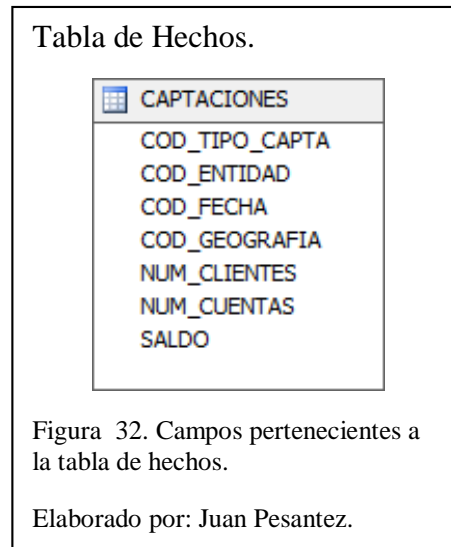
Para la ejecución del algoritmo se usa la dimensión geografía, y la agrupación por cantones tal y como muestra la Figura 30.



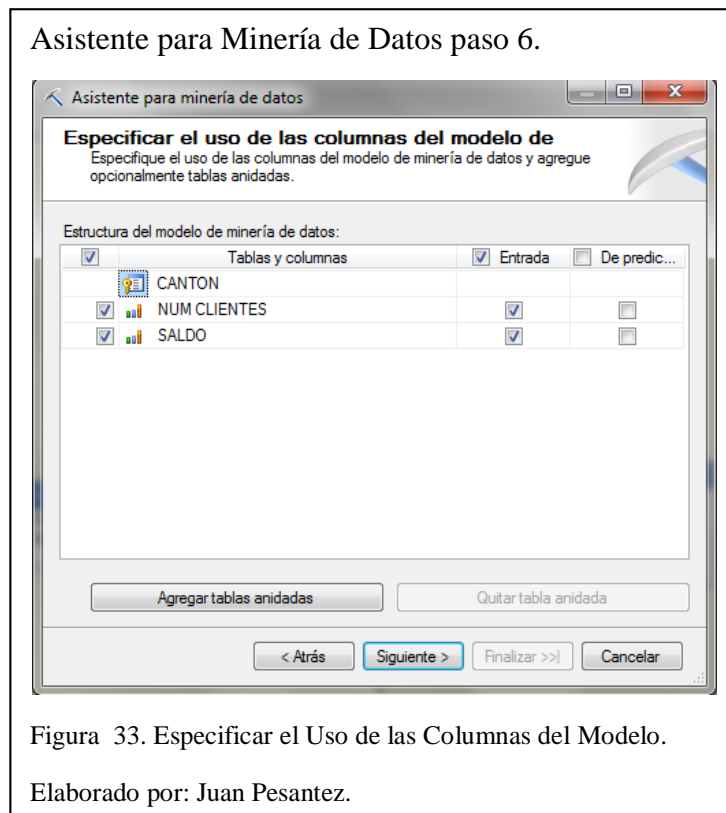
Inmediatamente, el asistente requiere seleccionar las columnas que tienen los datos de la tabla de hechos como atributos o medidas para el algoritmo, como lo muestra la Figura 31.



En la siguiente pantalla, se procede a seleccionar los campos que son usados para la predicción y una segunda agrupación de datos para el algoritmo, estos campos son pertenecientes a la tabla de hechos como lo muestra la Figura 32.



El algoritmo de Clústeres de Microsoft no obliga a seleccionar un campo específico de tal manera que, si no se selecciona un campo para predicción, el algoritmo muestra los clústeres de agrupación según los datos seleccionados y las tablas de dimensiones (ver Figura 33).



La siguiente pantalla, es para la segmentación del cubo de origen, esto se hace mediante las tablas de dimensiones y sus campos a elegir, lo cual hace que los clústeres se dividan según las expresiones seleccionadas, como la dimensión, la jerarquía, el operador y la expresión de filtro (ver Figura 34).

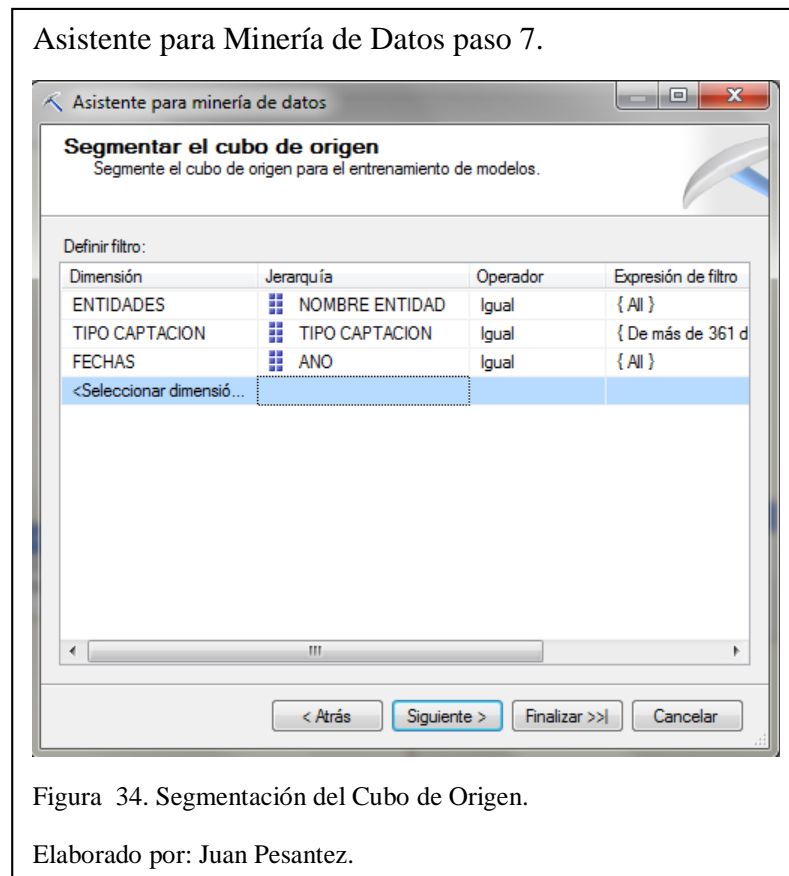
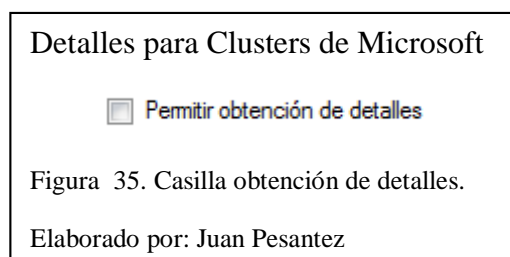


Figura 34. Segmentación del Cubo de Origen.

Elaborado por: Juan Pesantez.

Como paso final, el asistente requiere ingresar un nombre tanto a la estructura de minería de datos como al modelo de minería de datos, recordando que hay que marcar la opción de permitir la obtención de detalles (ver Figura 35).



Es importante en muchos de los algoritmos de minería de datos habilitar este campo, ya que es la única vía para que el usuario pueda visualizar los elementos que existen dentro de cada clúster como se muestra en la Figura 36.

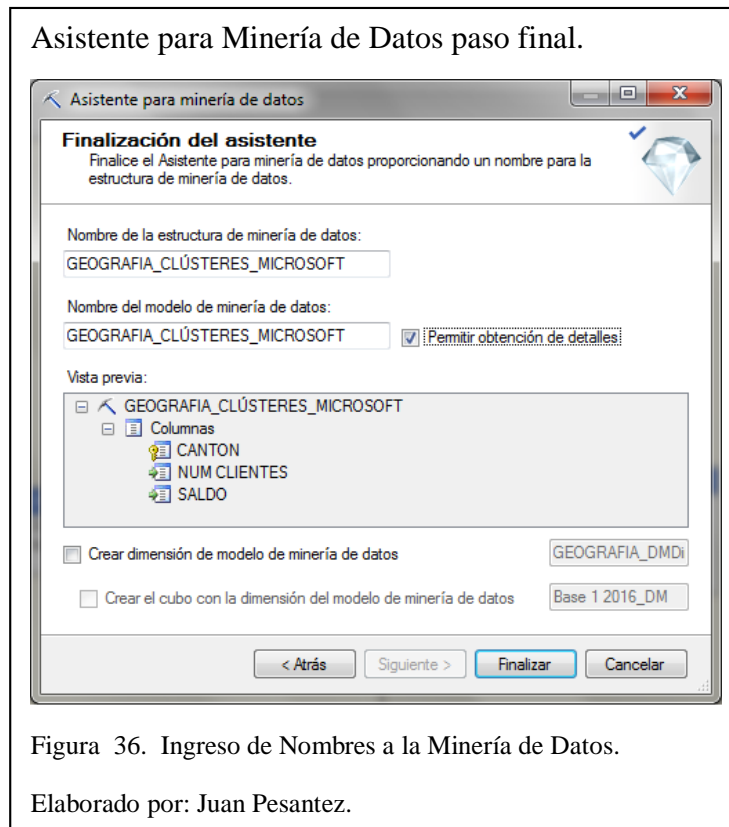


Figura 36. Ingreso de Nombres a la Minería de Datos.

Elaborado por: Juan Pesantez.

Para poder visualizar el resultado de la minería de datos o su predicción, una vez terminado el asistente se debe iniciar la depuración de los datos o igualmente presionar la tecla “F5”, de tal modo que la herramienta procese toda la información y muestre los clústeres en el visor de modelos de minería de datos, como se muestra en la Figura 37.

Resultado Final de la Minería Clústeres de Microsoft.

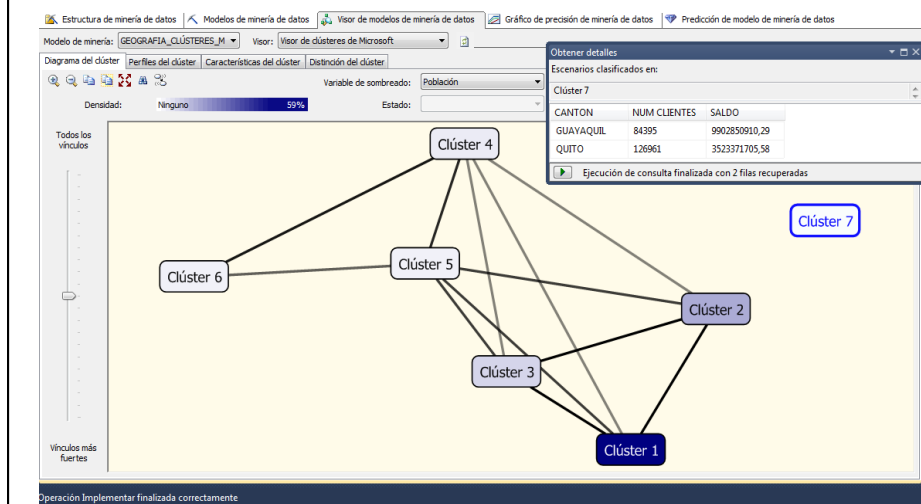


Figura 37. Visor de Modelos de Minería de Datos.

Elaborado por: Juan Pesantez.

Para interpretar este algoritmo se debe observar los vínculos que hay entre clústeres y la densidad del color de cada uno, para este caso práctico el “Clúster 1”, es aquel que más datos agrupados posee y coexistirían los cantones que tienen mayor similitud tanto para el número de clientes como para el saldo. El “Clúster 7”, que no presenta vínculo alguno es aquel que se constituye por los cantones donde su similitud es única de entre todo el universo de datos, en la Figura 38 se observan varios de los detalles incluidos en el “Clúster 1” y el “Clúster 2”, que son aquellos que tienen mayor densidad y que poseen mayor número de vínculos.

Representación de Clústeres 1 y 2.

CANTON	NUM CLIENTES	SALDO
PEDRO VICENTE MALDONADO	17	205047,83
PENIPE	0	0
PICHINCHA	24	512000
PIMAMPIRO	9572	8000529,75
PINDAL	13008	4751406,94
PIÑAS	446	24289028,13
PLAYAS	121	1237499,72
PORTOVELO	18	1129730,73
PORTOVIEJO	9406	71929518,67
PUCARA	16	238631,46
PUEBLOVIEJO	30	456211,27
PUERTO LÓPEZ	38	857387,01
PUERTO QUITO	0	0
PUJILI	1	13795,28
PUTUMAYO	0	0

CANTON	NUM CLIENTES	SALDO
EL TRIUNFO	244	1889956,33
GENERAL ANTONIO ELIZALDE	75	2381734,24
GUALAQUIZA	86	1379500
HUAQUILLAS	243	4425433,84
JIPJIAPA	85	1391967
LA CONCORDIA	123	2317448,4
LA TRONCAL	143	2185680,77
MORONA	103	2012618,28
NARANJAL	108	2076876,21
PALTAS	85	1551376,27
PASTAZA	176	3419841,69
PAUTE	64	1494769,02
PEDERNALES	59	1447070,89
PLAYAS	121	1237499,72
QUININDÉ	17	1145818,12

Figura 38. Detalles de Clústeres 1 y 2.

Elaborado por: Juan Pesantez.

El gráfico anterior, muestra que, en los clústeres seleccionados hay varios cantones que poseen a un determinado número de clientes dentro de un rango dado por el algoritmo y de igual manera con el saldo de las cuentas de los clientes, los clústeres mostrados indican agrupaciones por cantón, número de clientes y saldos, el algoritmo realiza comparaciones en donde se encuentran grandes cantidades de información y procede a agruparla, para este ejemplo el algoritmo toma como referencia la columna de saldo y procede a agrupar según rangos o especificaciones propias de la herramienta, haciendo que de esta manera se pueda visualizar de mejor manera como se encuentran relacionadas, agrupadas y seleccionadas cada una de las filas mostradas.

3.6.1.3. Ejecución del algoritmo clústeres de Microsoft para colocaciones.

Para la correcta ejecución de este algoritmo, se debe partir de un cubo multidimensional previamente creado, el cual va a ser quien se encargue de otorgar los datos para su futuro análisis. El proceso comienza con el asistente para minería de datos, donde se selecciona el algoritmo que va a ser utilizado para la predicción de los datos, para este ejemplo se selecciona “Clústeres de Microsoft”, como lo indica la Figura 39.

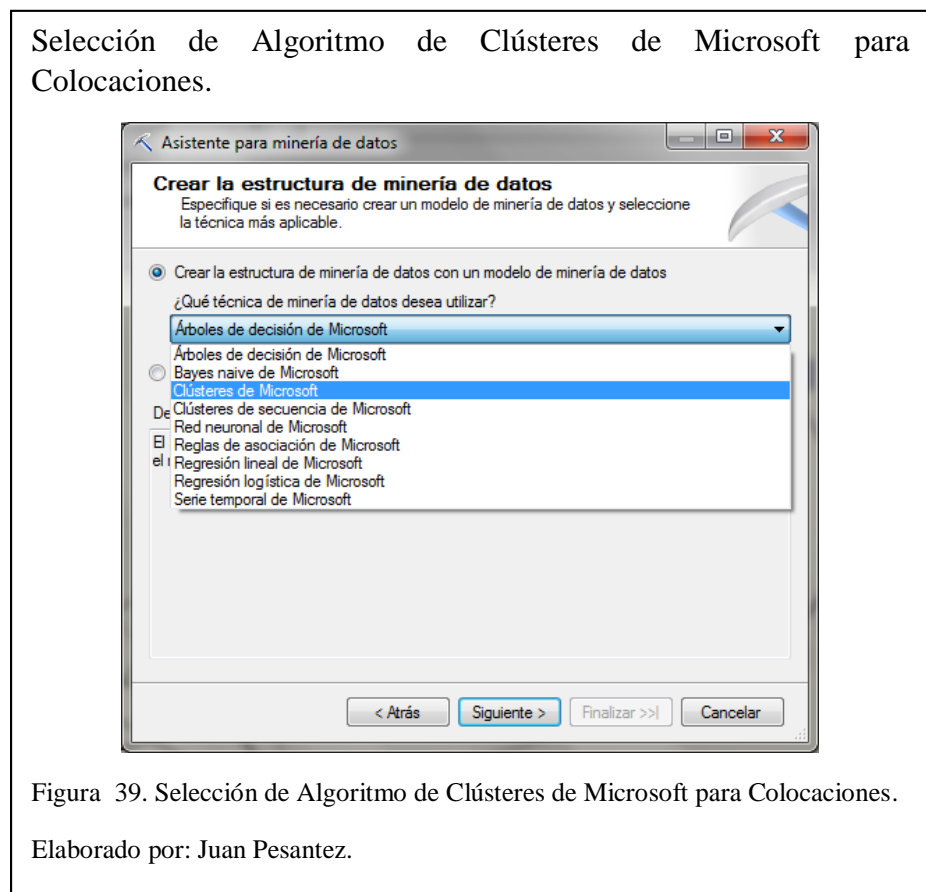
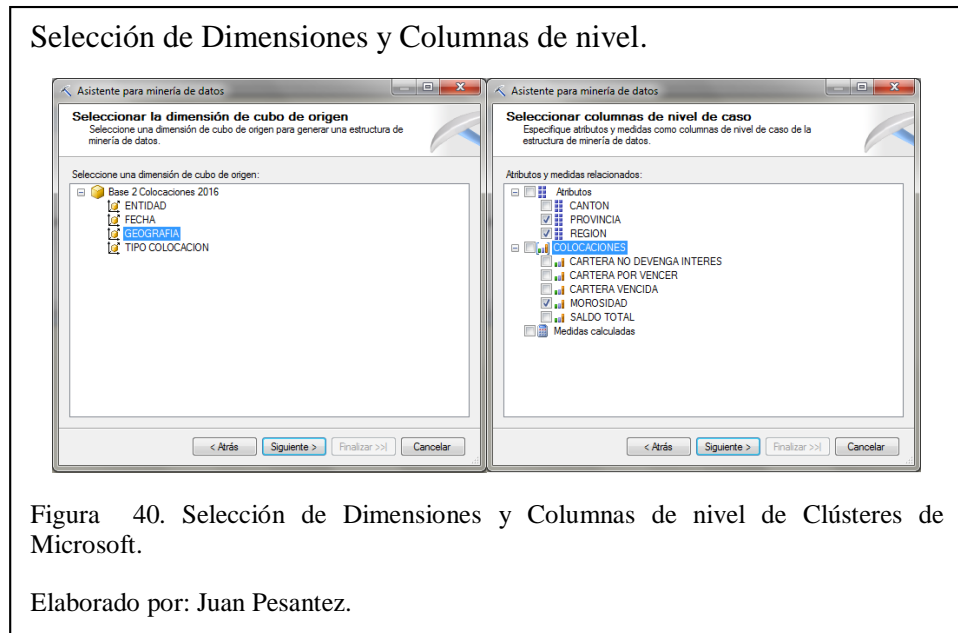


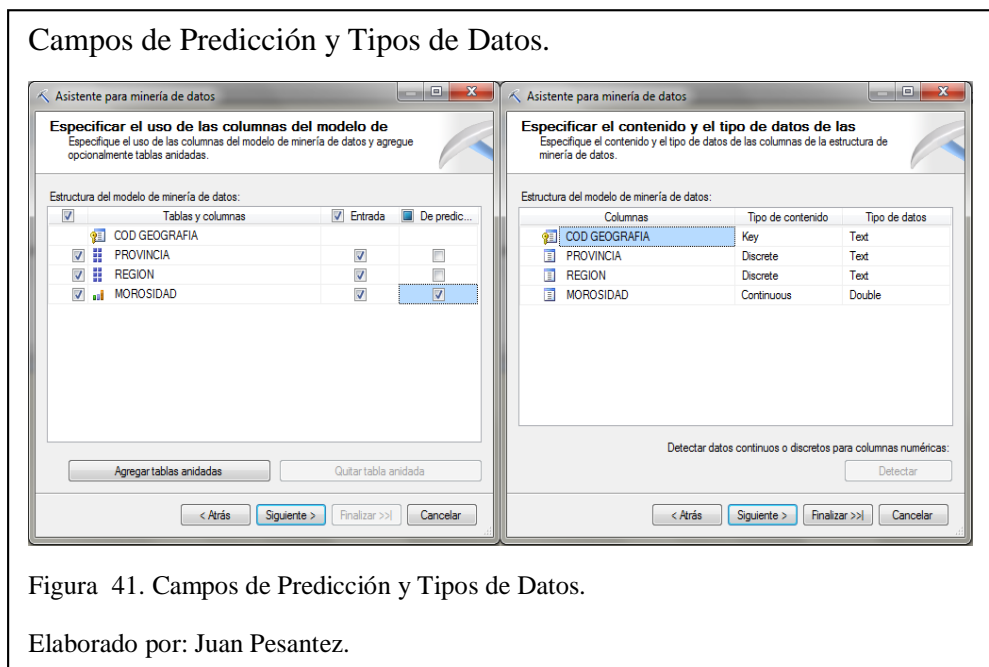
Figura 39. Selección de Algoritmo de Clústeres de Microsoft para Colocaciones.

Elaborado por: Juan Pesantez.

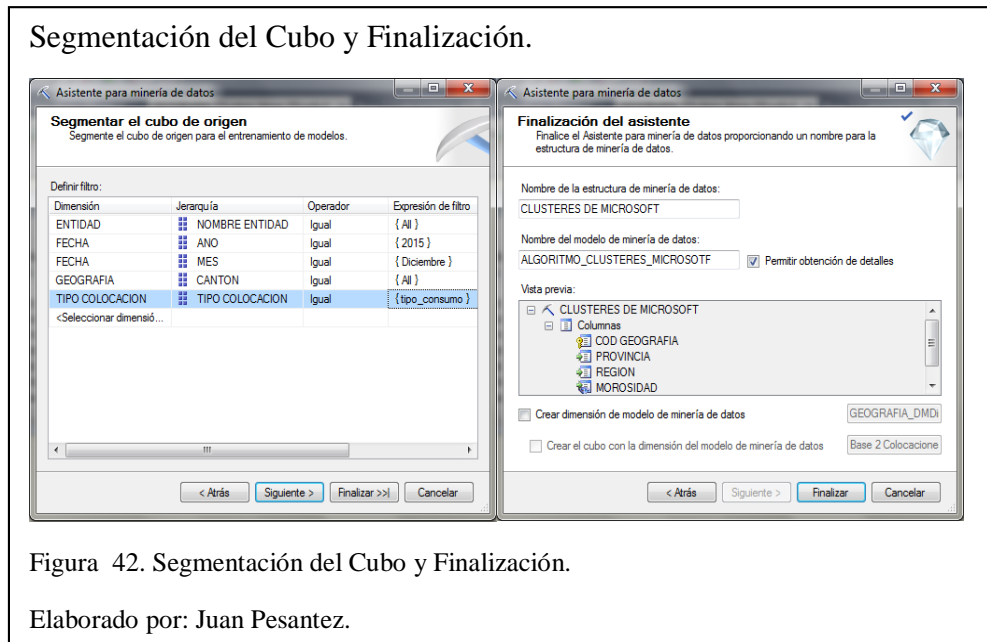
El siguiente paso es seleccionar la o las dimensiones y las columnas de nivel de caso que contienen los valores o datos a ser evaluados por el algoritmo y las columnas que son presentadas dentro de los resultados futuros, como se muestra en la Figura 40.



En seguida, surge una pantalla donde permite seleccionar uno de los campos de la tabla de hechos para ser tomado como “De predicción”, y los campos que son los datos de entrada que se muestran en los resultados, además de la opción de especificar el tipo de datos que contienen los campos para que el modelo de minería obtenga resultados con mayor precisión, como se muestra en la Figura 41.



El siguiente paso, es elegir los campos que el algoritmo toma como variables de agrupación, de esta manera seleccionando valores o términos comunes y agrupándolos dentro de clústeres según los parámetros establecidos que tiene el algoritmo de “Clústeres de Microsoft”, como paso final también el asistente de minería de datos muestra un informe donde se puede visualizar los valores preseleccionados, para la correcta ejecución del algoritmo, como se ve en la Figura 42.



Como resultado del asistente para minería de datos, se obtiene un diagrama basado en clústeres, el cual ha agrupado todos los posibles elementos comunes encontrados dentro de los parámetros ingresados previamente.

Cada uno de los clústeres que se presentan contienen varios elementos de las colocaciones de bancos privados, estos elementos dan a conocer aquellos que tienen valores altos, medios o bajos según la agrupación obtenida, dichas agrupaciones muestran un volumen o densidad en porcentajes altos según la cantidad de elementos que se agruparon dentro de cada clúster, así como la densidad de las relaciones entre cada clúster, como se puede observar en la Figura 43.

Clústeres de Microsoft para Colocaciones.

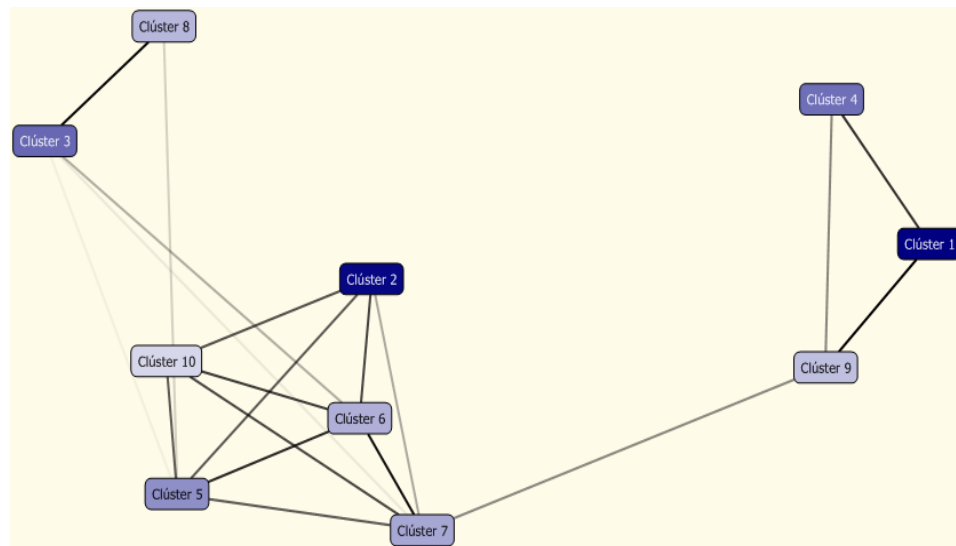


Figura 43. Clústeres de Microsoft para Colocaciones.

Elaborado por: Juan Pesantez.

Al aplicar este algoritmo en base a la dimensión geografía y sus campos provincia y región, se observa la agrupación en clústeres por morosidad referente a los créditos de consumo aplicados para el mes de diciembre y para el año 2015. Los clústeres con mayor densidad, son aquellos que se encuentran resaltados con un color más oscuro, lo que demuestra que contienen una gran cantidad de elementos cuyos términos son comunes y que poseen uno o varios vínculos fuertes entre sí, en la Figura 44, se indica los clústeres con mayor densidad y varios de los elementos que se encuentran dentro de cada uno. Cabe destacar que este mismo ejercicio se puede aplicar a cualquier tipo de combinación entre los elementos de cada una de las dimensiones, así como para otros los tipos de créditos ya sean:

- Crédito comercial.
- Microcrédito.
- Créditos para vivienda.

Clústeres con mayor densidad.

Clúster 1				Clúster 2			
COD GEOGRAF...	PROVINCIA	REGION	MOROSIDAD	COD GEOGRAF...	PROVINCIA	REGION	MOROSIDAD
020802	ESMERALDAS	Costa	0	031604	PASTAZA	Amazonica	0
020803	ESMERALDAS	Costa	0	011701	PICHINCHA	Sierra	2,372498221
020804	ESMERALDAS	Costa	0,075851654	011702	PICHINCHA	Sierra	0,193086276
020807	ESMERALDAS	Costa	0	011703	PICHINCHA	Sierra	0,223882953
020902	GUAYAS	Costa	0	011704	PICHINCHA	Sierra	0
020903	GUAYAS	Costa	0	011705	PICHINCHA	Sierra	0,339153201
020904	GUAYAS	Costa	0,122057772	011707	PICHINCHA	Sierra	0,043252836
020905	GUAYAS	Costa	0	011708	PICHINCHA	Sierra	0,056313422
020908	GUAYAS	Costa	0,135250785	011709	PICHINCHA	Sierra	0
020911	GUAYAS	Costa	0,028054192	011801	TUNGURAHUA	Sierra	2,279943161
020912	GUAYAS	Costa	0,031157273	011802	TUNGURAHUA	Sierra	0,123608206
020913	GUAYAS	Costa	0	011803	TUNGURAHUA	Sierra	0
020914	GUAYAS	Costa	0	011804	TUNGURAHUA	Sierra	0
020918	GUAYAS	Costa	0	011805	TUNGURAHUA	Sierra	0
020919	GUAYAS	Costa	0	011806	TUNGURAHUA	Sierra	0
020920	GUAYAS	Costa	0	011807	TUNGURAHUA	Sierra	0,088502013
020922	GUAYAS	Costa	0	011808	TUNGURAHUA	Sierra	0,019781315
020923	GUAYAS	Costa	0,00067061	011809	TUNGURAHUA	Sierra	0
020924	GUAYAS	Costa	0	031901	ZAMORA CHI...	Amazonica	0,043324022
020925	GUAYAS	Costa	0	031902	ZAMORA CHI...	Amazonica	0

Figura 44. Clústeres con mayor densidad.

Elaborado por: Juan Pesantez.

Estos clústeres, indican cuales son las regiones y/o provincias que tienen mayores indicadores de morosidad, agrupándolos de manera que las entidades bancarias tengan una visión de cuáles son las regiones y/o provincias indicadas para ofertar préstamos de tipo consumo y que los mismos no bajen a los niveles de morosidad.

3.6.2. Algoritmo Serie Temporal de Microsoft.

El algoritmo de serie temporal de Microsoft, suministra los algoritmos de retracción que se perfeccionan para la previsión en el tiempo de valores continuos tales como las ventas de productos. Mientras que otros algoritmos de Microsoft, como por ejemplo los árboles de decisión, requieren columnas adicionales de nueva información como entrada para predecir una tendencia, los modelos de serie temporal no las necesitan. Un modelo de serie temporal puede predecir tendencias basadas únicamente en el conjunto de datos original utilizado para crear el modelo. Es posible también agregar nuevos datos al modelo al realizar una predicción e incorporar automáticamente los nuevos datos en el análisis de tendencias. (Microsoft, 2016)

Una característica importante del algoritmo de serie temporal de Microsoft es su capacidad para llevar a cabo predicciones cruzadas. Si entrena el algoritmo con dos series independientes, pero relacionadas, puede utilizar el modelo generado para predecir el resultado de una serie basándose en el comportamiento de la otra. La predicción cruzada también es útil para crear un modelo general que se puede aplicar a múltiples series. Por ejemplo, las predicciones para una región determinada son inestables debido a que la serie no dispone de datos de buena calidad. Se puede entrenar un modelo general sobre la media de las cuatro regiones y, a continuación, aplicar el modelo a las series individuales para crear predicciones más estables para cada región. (Microsoft, 2016)

3.6.2.1. ¿Cómo funciona el algoritmo de serie temporal de Microsoft?

El algoritmo de serie temporal de Microsoft, manipula una composición de los dos algoritmos al examinar patrones y realizar predicciones. El algoritmo adiestra dos modelos autónomos sobre los mismos datos: uno de los modelos utiliza el algoritmo ARTXP, el cual se utiliza para predicciones a corto plazo y el otro modelo utiliza el algoritmo ARIMA, el mismo que se lo utiliza para predicciones a largo plazo. (Microsoft, 2016)

Posteriormente, el algoritmo combina los resultados de los dos modelos para obtener la mejor predicción sobre un número variable de intervalos de tiempo. Dado que ARTXP obtiene mejores resultados en las predicciones a corto plazo, se le da mayor importancia al principio de una serie de predicciones. Sin embargo, a medida que los intervalos de tiempo que se están prediciendo se adentran en el futuro, se va dando más importancia a ARIMA. (Microsoft, 2016)

Es posible también controlar la mezcla de algoritmos para favorecer la predicción a corto o a largo plazo en las series temporales. (Microsoft, 2016)

3.6.2.2. Ejecución del algoritmo serie temporal de Microsoft.

Para este algoritmo se utiliza el asistente para minería de datos haciendo énfasis en los distintos pasos entre cada modelo de minería de datos.

Lo primero que se debe hacer, es seleccionar el algoritmo que se va a utilizar, para este caso “Serie Temporal de Microsoft”, como lo muestra la Figura 45.

Asistente para minería de datos paso 2.

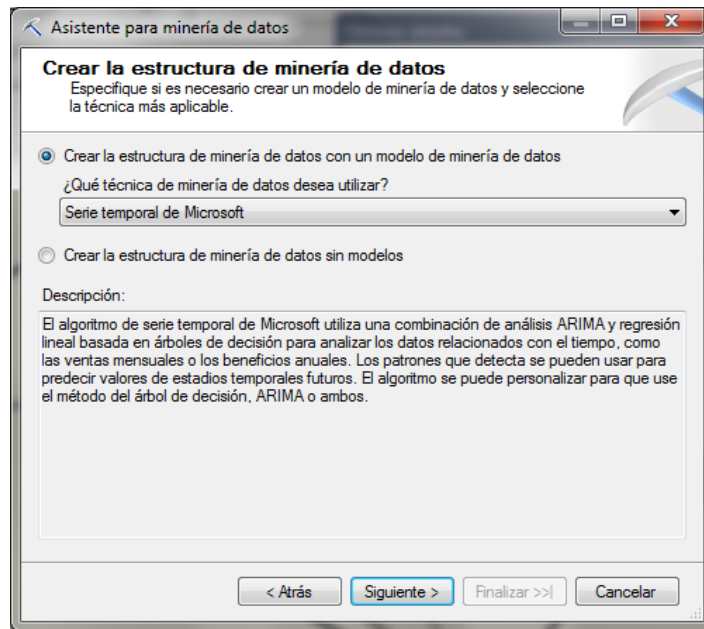


Figura 45. Crear la Estructura de Minería de Datos.

Elaborado por: Juan Pesantez.

El algoritmo de “Serie Temporal de Microsoft” solicita que, en su dimensión de cubo de origen, sea aquella tabla o dimensión que tenga como campo fecha o datetime indispensable para poder de esta manera crear la serie temporal, como lo muestra la Figura 46.

Asistente para minería de datos paso 3.

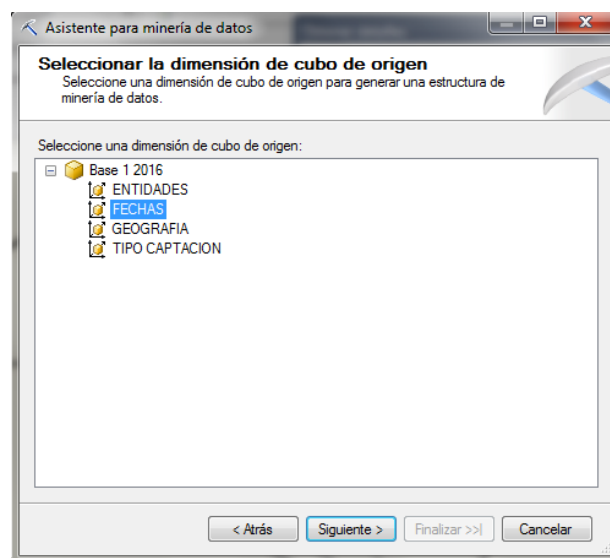
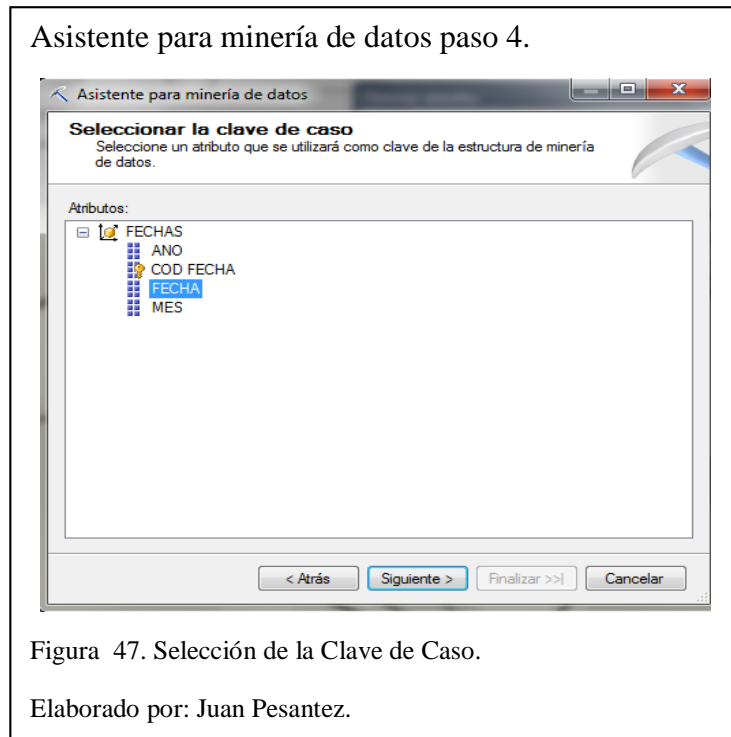


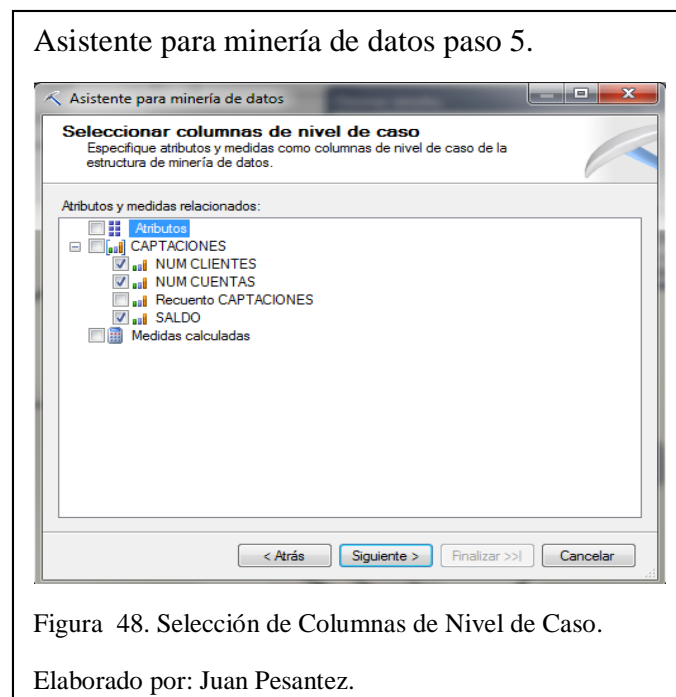
Figura 46. Selección de la Dimensión del Cubo de Origen.

Elaborado por: Juan Pesantez.

La Figura 47 muestra el contenido de la dimensión FECHAS elegida en el paso anterior y la selección de la columna que contiene datos de tipo datetime.



Para la selección de las columnas de nivel de caso, en este algoritmo de minería de datos se selecciona, tanto el número de clientes, número de cuentas y saldo, como lo muestra la Figura 48.



Posteriormente, se seleccionan los campos que son vistos por el algoritmo como las predicciones a realizar, dentro de la tabla de hechos para captaciones los campos disponibles que se tienen son: número de clientes, número de cuentas y el saldo correspondiente a cada uno como se muestra en la Figura 49.

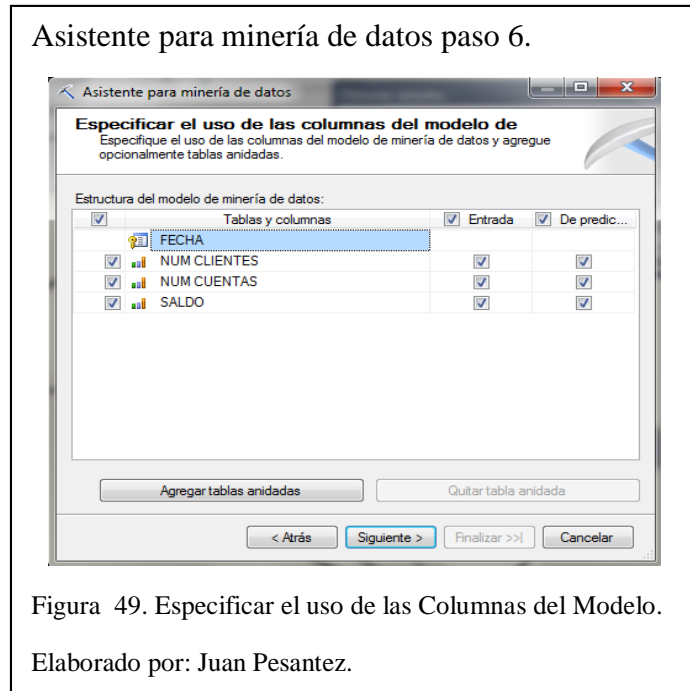


Figura 49. Especificar el uso de las Columnas del Modelo.

Elaborado por: Juan Pesantez.

Siguiendo con el asistente, se selecciona los campos de las tablas de dimensiones que se utilizan para la segmentación del algoritmo, como se observa en la Figura 50.

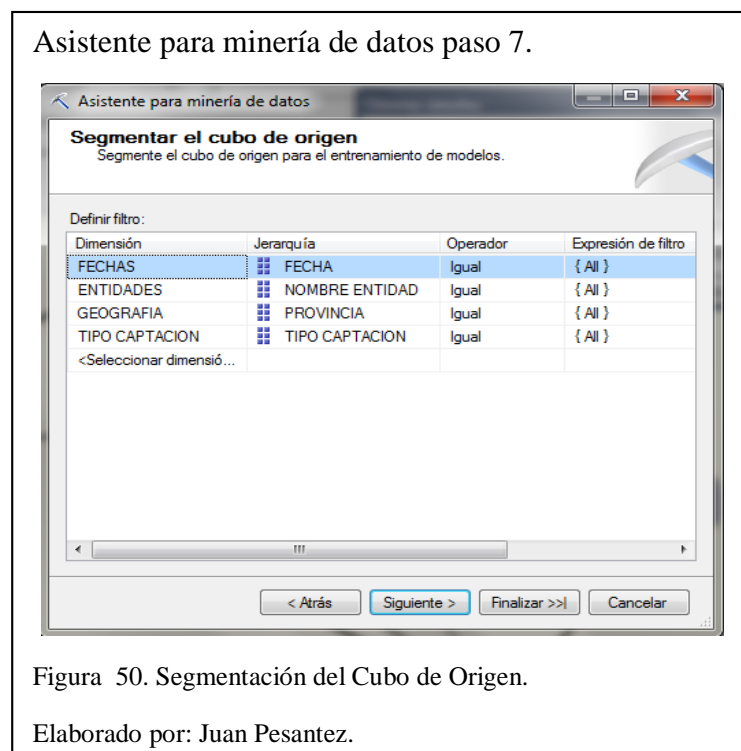
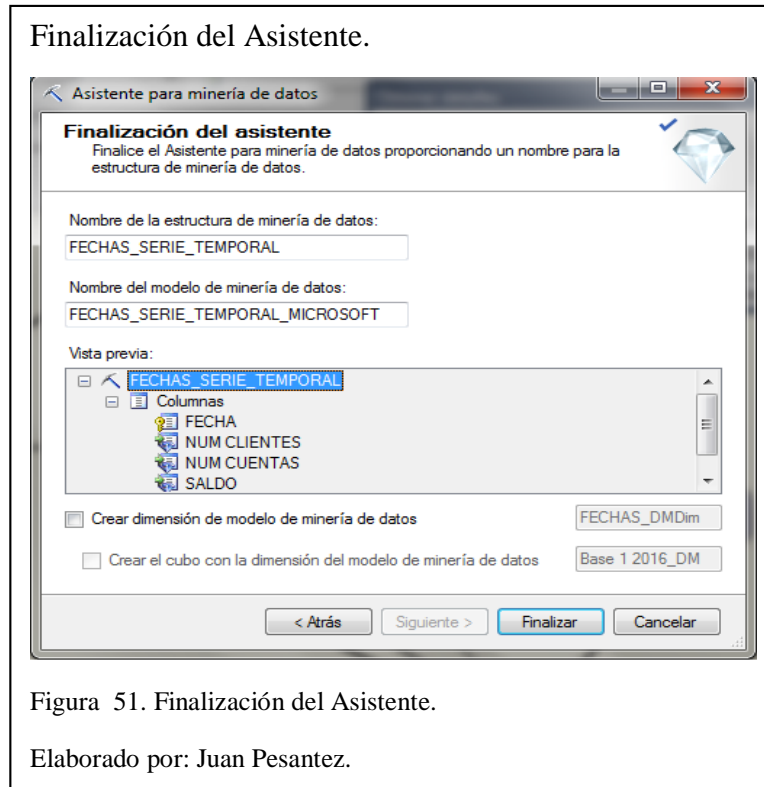


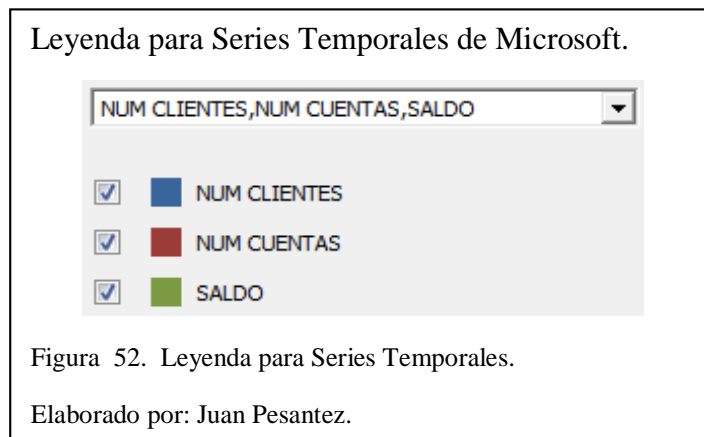
Figura 50. Segmentación del Cubo de Origen.

Elaborado por: Juan Pesantez.

Por último, se ingresa un nombre tanto a la estructura como al modelo de minería de datos y se verifica que los campos sean los indicados para terminar correctamente con el asistente de minería de datos (ver Figura 51).



Como resultado en la Figura número 53, se puede visualizar un gráfico donde se muestra los valores en el transcurso histórico sobre número de cuentas, número de clientes y el saldo correspondiente para cada una de estos datos, el visor para este modelo de minería de datos muestra para cada uno de sus gráficos una leyenda correspondiente, para estar al tanto de la pertenencia de cada línea, su color y equivalencia como lo muestra la Figura número 52.



Resultado de la minería de datos Serie Temporal de Microsoft.

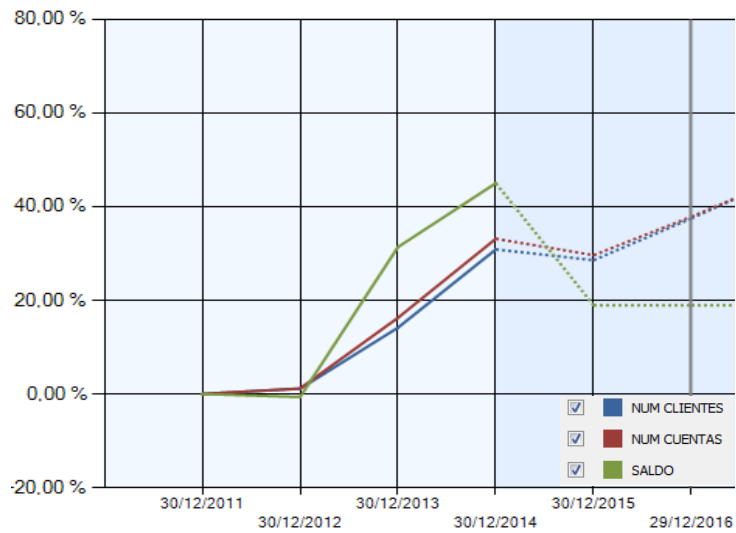


Figura 53. Visor del Modelo de Minería de Datos Serie Temporal.

Elaborado por: Juan Pesantez.

El resultado de aplicar el algoritmo de Series temporales de Microsoft, presenta un gráfico que visualiza el histórico del número de clientes, número de cuentas y saldos de captaciones bancarias del Ecuador. Con toda esta información el algoritmo realiza una predicción estimada en donde se puede ver las tendencias futuras para el año 2016 como lo muestra la Figura 54.

Predicción del Algoritmo Serie Temporal de Microsoft.

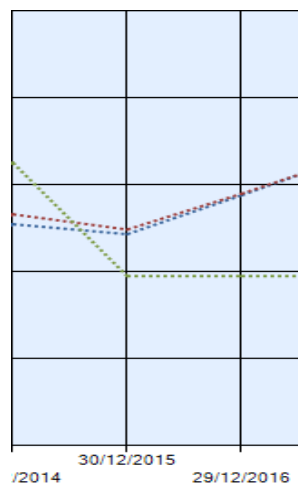


Figura 54. Predicción del Modelo de Minería de Datos.

Elaborado por: Juan Pesantez.

En el gráfico anterior, fruto de aplicar el algoritmo de “Series Temporales”, se observa que para el año 2016 habrá una caída de número de clientes, número de cuentas y saldos. Esta información es valiosa y le sirve a cualquier organización para observar el mercado y en función de eso realizar campañas de captación de clientes y depósitos colocando énfasis en provincias donde los depósitos no obtienen un nivel esperado para las entidades bancarias.

3.6.3. Algoritmo Bayes naive de Microsoft.

El algoritmo Bayes naive de Microsoft, es un algoritmo de categorización que proporciona Microsoft SQL Server Analysis Services para el esculpido predictivo. Este algoritmo calcula la probabilidad condicional entre columnas de entrada y de predicción y supone que las columnas son independientes. Esta suposición de independencia conduce al nombre Bayes naive. (Microsoft, 2016).

El algoritmo solo admite atributos discretos o atributos de datos discretos, y considera independientes todos los atributos de entrada, según el atributo de predicción. (Microsoft, 2016)

3.6.3.1. ¿Cómo funciona el algoritmo Bayes naive de Microsoft?

El algoritmo Bayes naive de Microsoft, calcula la probabilidad de cada estado de cada columna de entrada, dado cada posible estado de la columna de predicción.

El Visor Bayes naive de Microsoft muestra las columnas de entrada del conjunto de datos e indica cómo se distribuyen los estados de cada columna, dado cada estado de la columna de predicción. Se puede usar esta vista para identificar las columnas de entrada que son importantes para diferenciar los distintos estados de la columna de predicción. El algoritmo utiliza la información numérica derivada de un dato para predecir su continuidad o futura adquisición de determinado bien o producto.

3.6.3.2. Ejecución del algoritmo Bayes naive de Microsoft.

Para iniciar con este algoritmo, se selecciona un cubo o data warehouse construido previamente, perteneciente a colocaciones bancarias, una vez realizado el paso anterior es necesario escoger, cuál va a ser la estructura de

minería de datos a ser selecta, para este caso se utiliza el algoritmo de “Bayes naive de Microsoft”, como lo muestra en la Figura 55.

Creación de la Estructura de Minería de Datos “Bayes naive”.

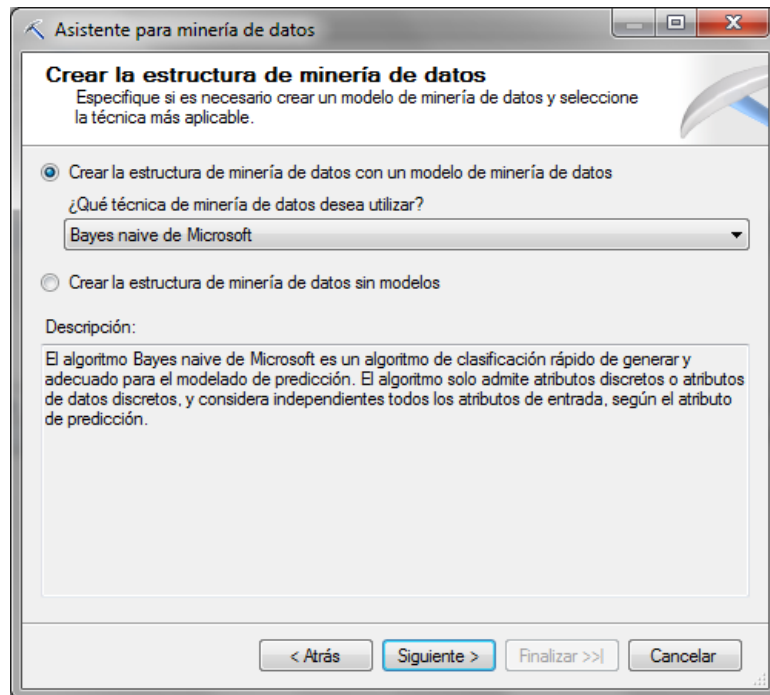


Figura 55. Creación de la Estructura de Minería de Datos "Bayes naive".

Elaborado por: Juan Pesantez.

A continuación, el asistente de creación solicita que se seleccione la dimensión base o primaria para el cubo de origen, y posteriormente solicita el ingreso de un campo que se toma como clave primaria para la segmentación de los datos del cubo de origen, como lo muestra la Figura 56.

Creación de la Dimensión y clave de Caso para Bayes naive.

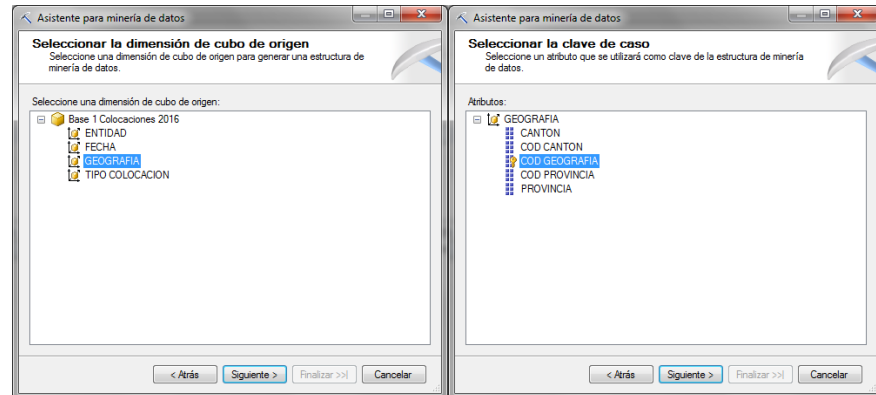


Figura 56. Dimensión y clave de Caso para "Bayes naive".

Elaborado por: Juan Pesantez.

Siguiendo con el asistente de minería de datos, solicita seleccionar los atributos y medidas de nivel de caso, estos datos son agrupados según el algoritmo utilizado para la secuencia de minería de datos, una vez seleccionadas las medidas de nivel de caso, se procede a escoger los campos que estarían selectos como futuras predicciones para este modelo, como lo muestra la Figura 57.

Columnas de nivel de Caso, Columnas de Predicción "Bayes naive".

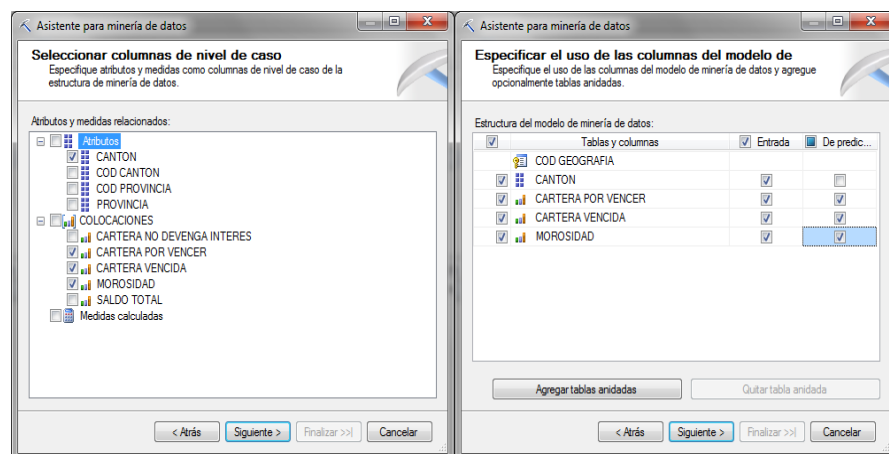
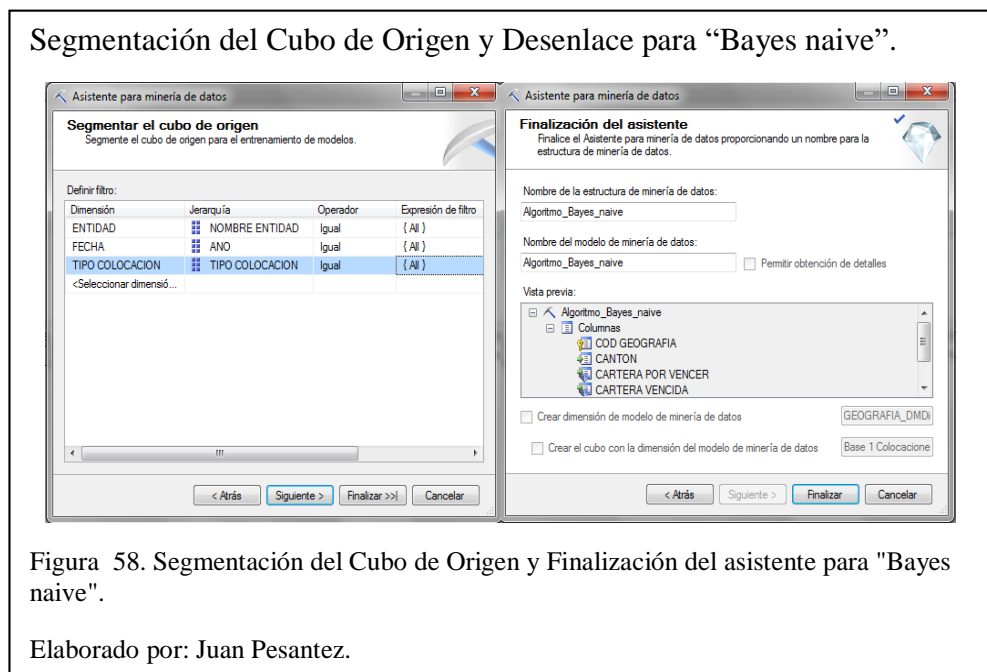


Figura 57. Columnas de Nivel de Caso y Columnas de Predicción para "Bayes naive".

Elaborado por: Juan Pesantez.

El paso siguiente, es proceder a la segmentación del cubo de origen, seleccionando uno o varios campos de las tablas de dimensiones, que son los encargados de realizar la división y agrupación de la data del cubo de colocaciones bancarias del Ecuador, con los parámetros anteriormente seleccionados, por último, el asistente muestra un resumen de las medidas ingresadas para la ejecución del algoritmo bayesiano, como lo muestra la Figura 58.



Como resultado del asistente para la minería de datos, se puede observar la relación que se tiene entre los tipos de variables o campos que se ingresaron para ser parte de la predicción, indicando todos los vínculos que se tiene dentro de las predicciones, el algoritmo muestra una leyenda para dar a entender el significado de cada color de la red de dependencia, para los atributos de predicción del algoritmo basados en la agrupación de datos y sus futuras tendencias, como se puede ver en la Figura 59.

Red de Dependencias de "Bayes naive".

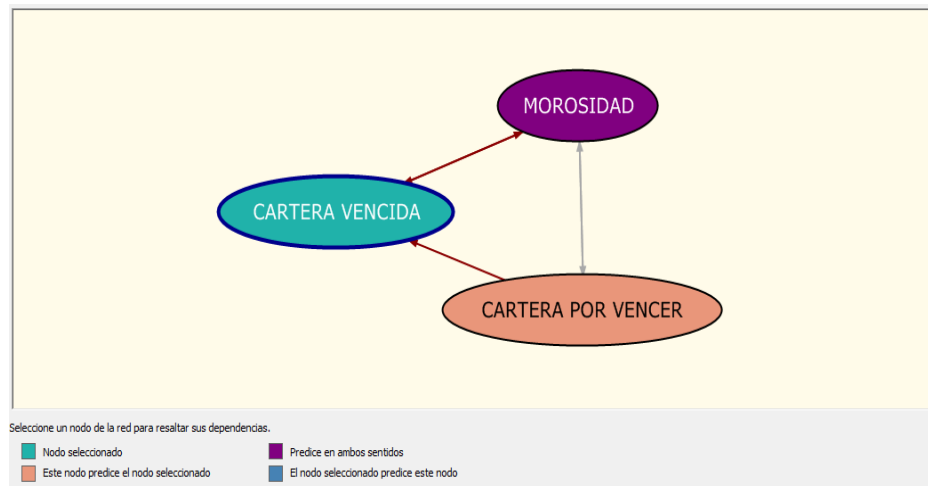


Figura 59. Red De Dependencias de "Bayes naive".

Elaborado por: Juan Pesantez.

Leyenda suministrada por el visor de Bayes naive, se observa que al seleccionar cada uno de los nodos se muestra en forma de flechas el sentido en que se predice el nodo antecesor, así como también el sentido de la predicción.

Para una mejor interpretación de este algoritmo se continua a la siguiente pestaña que indica cada uno de los perfiles de atributo pertenecientes a los nodos que predicen ya sea en uno o varios sentidos al nodo seleccionado, mediante la agrupación de valores similares, esta pestaña muestra una leyenda para cada nodo otorgando color, significado y distribución para cada agrupación de datos, como lo muestra la Figura 60, la cual pertenece a la morosidad.

Leyenda de Morosidad para "Bayes naive".

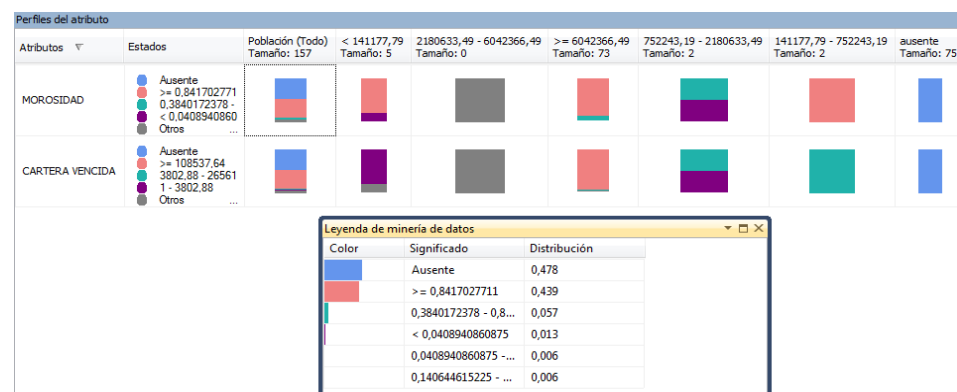
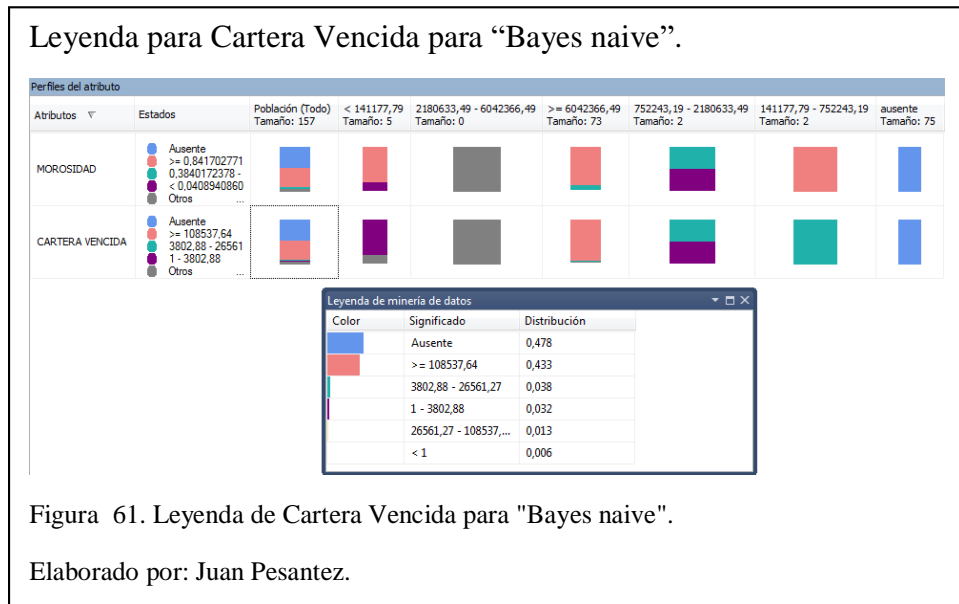


Figura 60. Leyenda de Morosidad para "Bayes naive".

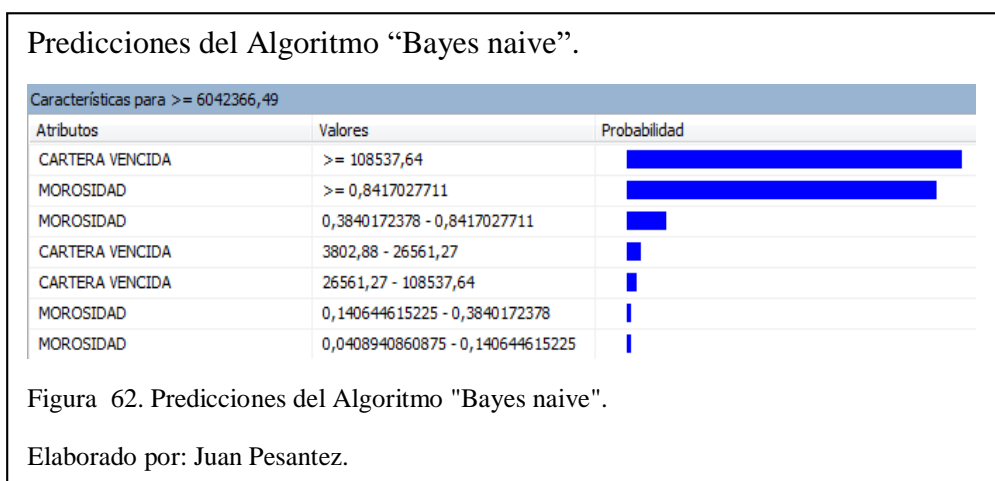
Elaborado por: Juan Pesantez.

Para la Figura 61, se observan los colores, significado y distribución perteneciente a la cartera vencida, cada significado muestra un límite o rango en valores numéricos el cual sirve como barrera antes de pasar a una nueva distribución



Una vez comprendido el nivel de pertenencia entre color y significado se puede obtener cada una de las distribuciones para los nodos y sus agrupaciones de datos, estas distribuciones son aquellos porcentajes que miden las futuras predicciones para cada nodo y sus valores.

Para cada uno de los nodos, se pueden observar características de atributo ya sea para cartera vencida o para la morosidad, el algoritmo muestra valores distintos para los mismos atributos, pero obteniendo diferentes niveles de probabilidad dados estos de forma que se prediga un nivel de éxito al estar dentro de los valores indicados, como se muestra en la Figura 62.



3.6.4. Algoritmo Reglas de asociación de Microsoft.

El algoritmo Reglas de asociación de Microsoft es una implementación sencilla del conocido algoritmo Apriori.

Tanto el algoritmo Árboles de decisión de Microsoft como el algoritmo Reglas de asociación de Microsoft se pueden usar para analizar asociaciones, pero las reglas encontradas por cada uno de ellos pueden diferir. En un modelo de árboles de decisión, las divisiones que conducen a reglas específicas están basadas en la obtención de información, mientras que, en un modelo de asociación, las reglas están basadas totalmente en la confianza. Por tanto, en un modelo de asociación, una regla segura, o una que tenga una confianza alta, puede que no tenga ningún interés, dado que no proporciona nueva información. (Microsoft, 2016)

Este algoritmo de Microsoft es un algoritmo de asociación suministrado por Analysis Services, útil para los motores de recomendación. Un motor de recomendación, recomienda productos a los clientes basándose en los elementos que ya han adquirido o en los que tienen interés. El algoritmo de asociación de Microsoft también resulta útil para el análisis de la cesta de compra. (Microsoft, 2016)

3.6.4.1. ¿Cómo funciona el algoritmo Reglas de asociación de Microsoft?

El algoritmo de asociación de Microsoft recorre un conjunto de datos para hallar elementos que aparezcan juntos en un caso. A continuación, agrupa en conjuntos de elementos todos los elementos asociados que aparecen, como mínimo, en el número de casos especificado en el parámetro MINIMUM_SUPPORT. El algoritmo generará reglas a partir de los conjuntos de elementos. Estas reglas se usan para predecir la presencia de un elemento en la base de datos, basándose en la presencia de otros elementos específicos que el algoritmo ha identificado como importantes. Por ejemplo, una regla puede ser "if Touring 1000=existing and Road bottle cage=existing, then Water bottle=existing", y puede tener una probabilidad de 0.812. En este ejemplo, el algoritmo identifica que la presencia en la cesta del neumático Touring 1000 y del soporte de la botella de agua predice que probablemente la cesta de compra incluirá también una botella de agua. (Microsoft, 2016)

3.6.4.2 Ejecución del algoritmo Reglas de asociación de Microsoft.

Para la correcta ejecución de este algoritmo, se debe iniciar el asistente para minería de datos, que guía al usuario paso por paso, describiendo cada una de las pantallas y su correcta manipulación.

La pantalla inicial que indica el asistente, es la selección del tipo de algoritmo que se desea utilizar, para este caso se opta por el algoritmo de “Series de Asociación de Microsoft”, el cual va a partir de un cubo generado previamente sobre colocaciones bancarias del Ecuador, tal como muestra la Figura 63.

Creación de Estructura de minería de datos para “Series Temporales de Microsoft”.

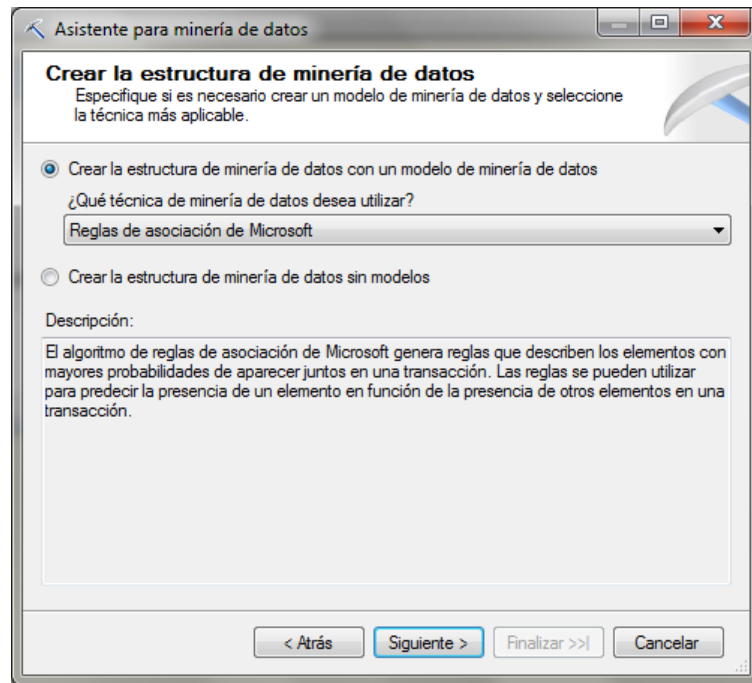


Figura 63. Creación de Estructura para "Series Temporales".

Elaborado por: Juan Pesantez.

El siguiente paso, es elegir cual va a ser la dimensión base para la segmentación de la data, esta dimensión es tomada como el punto de partida de este algoritmo de minería de datos, y siguiendo con el asistente se debe elegir un atributo que se es tomado como clave primaria para el algoritmo, como lo muestra la Figura 64.

Dimensión de origen y clave de caso para Series Temporales.

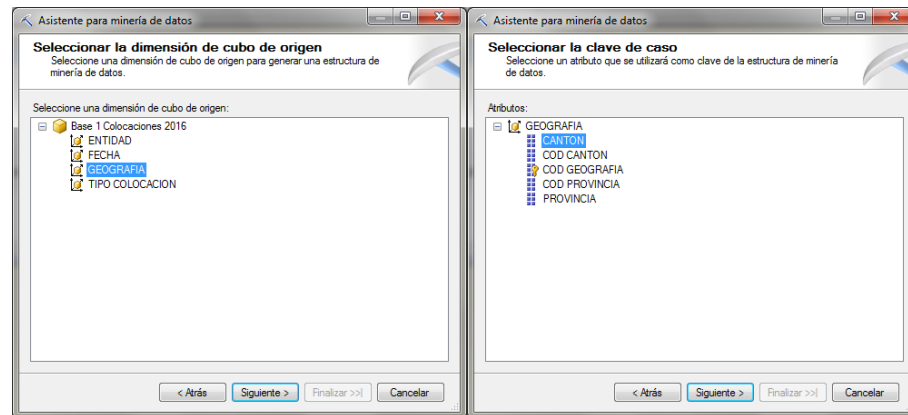


Figura 64. Dimensión de origen y clave de caso para "Series Temporales".

Elaborado por: Juan Pesantez.

Continuando con el asistente, se debe elegir los campos de la tabla de hechos que se toman como medidas de nivel de caso, así como los campos que están sugeridos para realizar la predicción del algoritmo de minería de datos, presentados en la Figura 65.

Asignación de Columnas de Nivel de Caso y Columnas de Predicción para "Series Temporales de Microsoft".

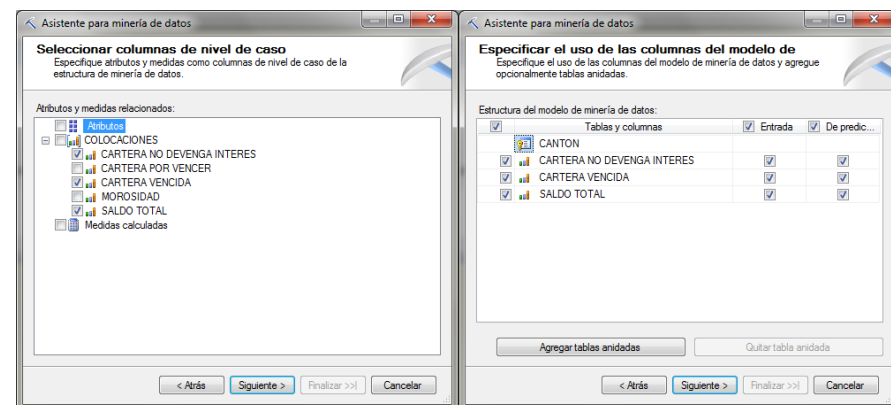


Figura 65. Columnas de Nivel de Caso y Columnas de Predicción para "Series Temporales".

Elaborado por: Juan Pesantez.

El siguiente paso, es seleccionar uno a uno los campos o columnas de las tablas de dimensiones que utiliza el algoritmo para crear las series temporales, premisas y conjuntos de datos propios de este algoritmo, como lo muestra la Figura 66.

Segmentación del Cubo de Origen para Reglas de Asociación.

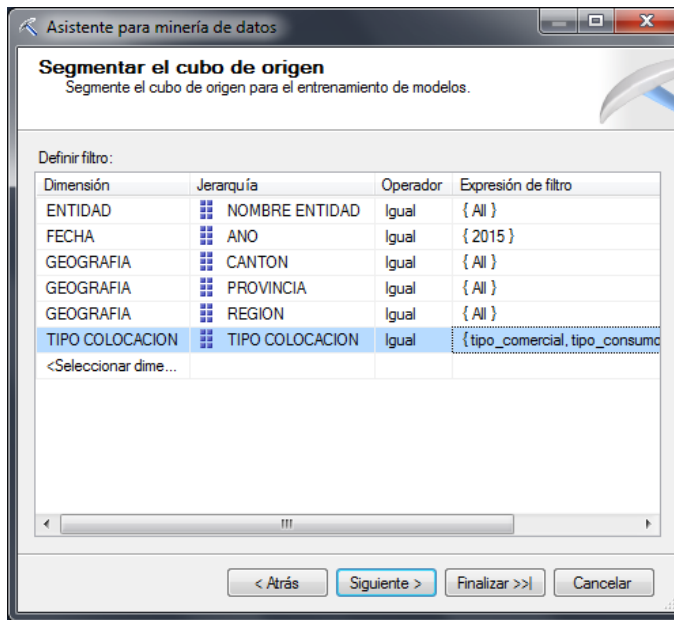


Figura 66. Segmentación del Cubo de Origen para "Reglas de Asociación".

Elaborado por: Juan Pesantez.

Por último, el asistente para minería de datos muestra un resumen de lo seleccionado previamente, haciendo que el usuario se asegure de que los parámetros ingresados sean los indicados para la correcta ejecución del algoritmo, como se muestra en la Figura 67.

Paso final y verificación de parámetros para "Reglas de Asociación".

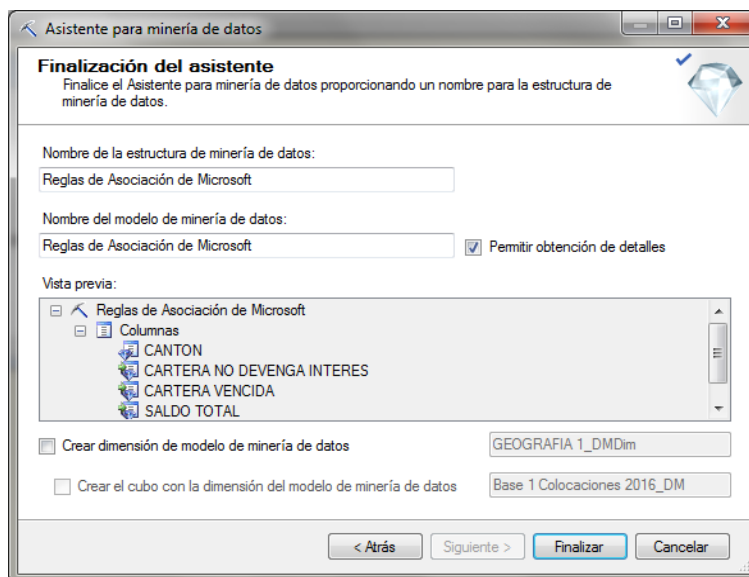


Figura 67 Finalización del Asistente para "Reglas de Asociación de Microsoft".

Elaborado por: Juan Pesantez.

Como resultado de la ejecución del asistente para minería de datos, se observa que el algoritmo crea 9 reglas o premisas basadas en la información proporcionada, cada una de estas reglas poseen una probabilidad de nivel de éxito, así como también se marca un nivel de importancia para cada regla, estas premisas contienen rangos numéricos para cada una de estas reglas, las cuales indican límites y clases para que se cumpla la premisa como se muestra en la Figura 68.

Conjunto de Reglas, Probabilidades e Importancia del Algoritmo Reglas de Asociación.

▼	Probabilidad	Importancia	Regla
	1,000	0,242	CARTERA VENCIDA = 11101914,1225251 - 25332898,4123638 -> CARTERA NO DEVENGA INTERES >= 259919,14
	1,000	0,751	CARTERA NO DEVENGA INTERES < 14182,23 -> CARTERA VENCIDA < 522780,974094746
	1,000	0,786	CARTERA NO DEVENGA INTERES = 14182,23 - 74644,89 -> CARTERA VENCIDA < 522780,974094746
	1,000	0,302	CARTERA VENCIDA >= 25332898,4123638 -> CARTERA NO DEVENGA INTERES >= 259919,14
	1,000	0,326	CARTERA VENCIDA = 4471090,10470994 - 11101914,1225251 -> CARTERA NO DEVENGA INTERES >= 259919,14
	1,000	0,706	CARTERA NO DEVENGA INTERES = 74644,89 - 259919,14 -> CARTERA VENCIDA < 522780,974094746
	1,000	0,512	CARTERA VENCIDA = 522780,974094746 - 4471090,10470994 -> CARTERA NO DEVENGA INTERES >= 259919,14
	0,625	0,145	CARTERA VENCIDA < 522780,974094746 -> CARTERA NO DEVENGA INTERES >= 259919,14
	0,521		CARTERA NO DEVENGA INTERES >= 259919,14 -> CARTERA VENCIDA = 522780,974094746 - 4471090,10470994

Figura 68. Premisas del Algoritmo "Reglas de Asociación de Microsoft".

Elaborado por: Juan Pesantez.

Se puede visualizar las premisas orientadas a base de conjuntos de elementos, observando el nivel de soporte para cada regla, que indica los niveles de aceptación entre cada una de las reglas establecidas dadas por la agrupación de elementos, como se muestra en la Figura 69.

Soporte y Tamaño de conjuntos de elementos para "Reglas de Asociación de Microsoft".

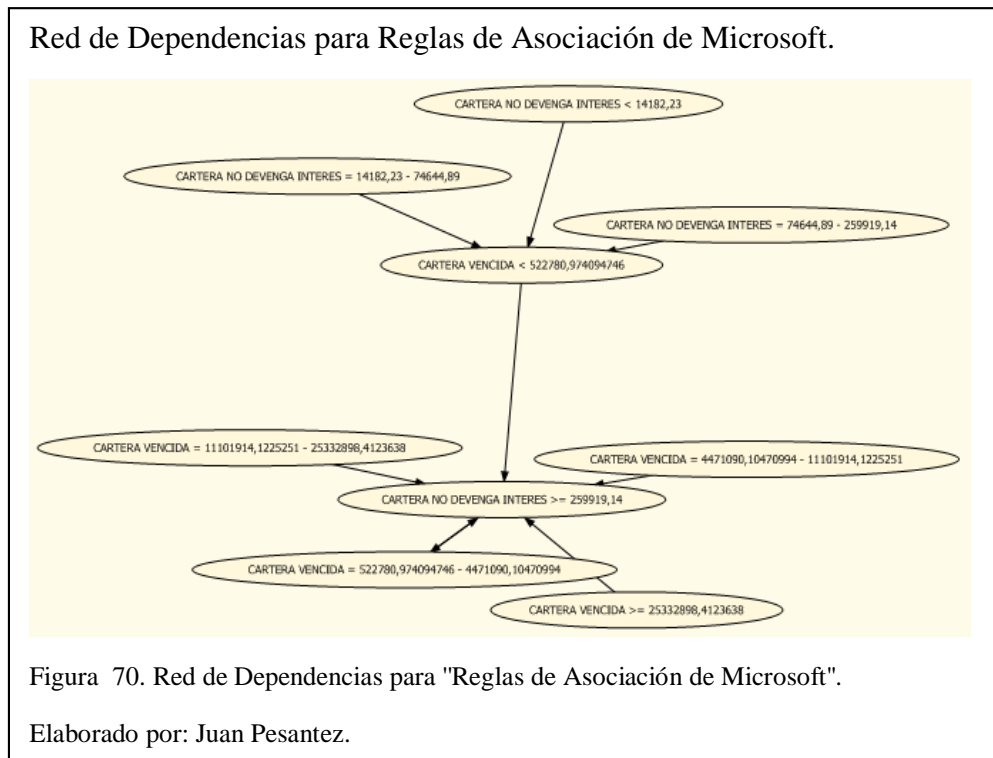
▼	Soporte	Tamaño	Conjunto de elementos
	73	1	CARTERA NO DEVENGA INTERES >= 259919,14
	38	1	CARTERA VENCIDA = 522780,974094746 - 4471090,10470994
	38	2	CARTERA VENCIDA = 522780,974094746 - 4471090,10470994, CARTERA NO DEVENGA INTERES >= 259919,14
	24	1	CARTERA VENCIDA < 522780,974094746
	15	2	CARTERA VENCIDA < 522780,974094746, CARTERA NO DEVENGA INTERES >= 259919,14
	10	1	CARTERA VENCIDA = 4471090,10470994 - 11101914,1225251
	10	2	CARTERA VENCIDA = 4471090,10470994 - 11101914,1225251, CARTERA NO DEVENGA INTERES >= 259919,14
	7	1	CARTERA VENCIDA >= 25332898,4123638
	7	2	CARTERA VENCIDA >= 25332898,4123638, CARTERA NO DEVENGA INTERES >= 259919,14
	4	1	CARTERA NO DEVENGA INTERES = 14182,23 - 74644,89
	4	2	CARTERA NO DEVENGA INTERES = 14182,23 - 74644,89, CARTERA VENCIDA < 522780,974094746
	3	1	CARTERA NO DEVENGA INTERES < 14182,23
	3	1	CARTERA VENCIDA = 11101914,1225251 - 25332898,4123638
	3	2	CARTERA VENCIDA = 11101914,1225251 - 25332898,4123638, CARTERA NO DEVENGA INTERES >= 259919,14
	3	2	CARTERA NO DEVENGA INTERES < 14182,23, CARTERA VENCIDA < 522780,974094746
	2	1	CARTERA NO DEVENGA INTERES = 74644,89 - 259919,14
	2	2	CARTERA NO DEVENGA INTERES = 74644,89 - 259919,14, CARTERA VENCIDA < 522780,974094746

Conjuntos de elementos: 17

Figura 69. Soporte y Tamaño de conjuntos de elementos para "Reglas de Asociación de Microsoft".

Elaborado por: Juan Pesantez.

El visor de Reglas de asociación puede también indicar al usuario la red de dependencia generada por los parámetros previamente seleccionados, agrupando por medio de nodos cada una de las reglas generadas y sus valores, cada nodo contiene un rango o regla que marca un índice de predicción para un nodo predecesor, como se muestra en la Figura 70.



3.7. Reportes.

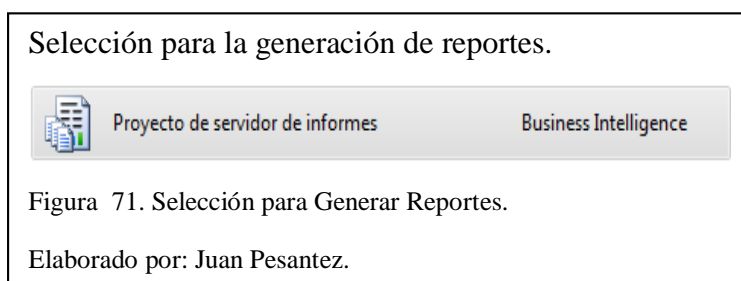
3.7.1. Reporte para captaciones de bancos privados del Ecuador.

Un reporte es un informe o una determinada noticia, este documento pretende transmitir información hacia una o varias personas hacia quienes se encuentra dirigido.

Un reporte pretende generar conocimiento de una gran cantidad de datos dispersos, este documento facilita la comprensión del conjunto de datos agrupando por medio de filas y columnas la información para la correcta representación del conocimiento.

3.7.1.1. Construcción de reportes.

Para la generación de los reportes basados en los cubos construidos posteriormente se utiliza la herramienta SQL Server Data Tools y seleccionando “Proyecto de Servidor de Informes” de Bussines Intelligence como se muestra en la Figura 71.



El siguiente paso, es seleccionar un origen de datos compartidos, que son el medio o vinculo que se tiene entre el proyecto de servidor de informes y los cubos generados en el proyecto multidimensional, al ingresar un nuevo origen de datos el asistente solicita colocar un nombre y un tipo, que para este caso es Microsoft SQL Server Analysis Services, como lo muestra la Figura 72.

Origen de Datos Compartidos para Reportes.

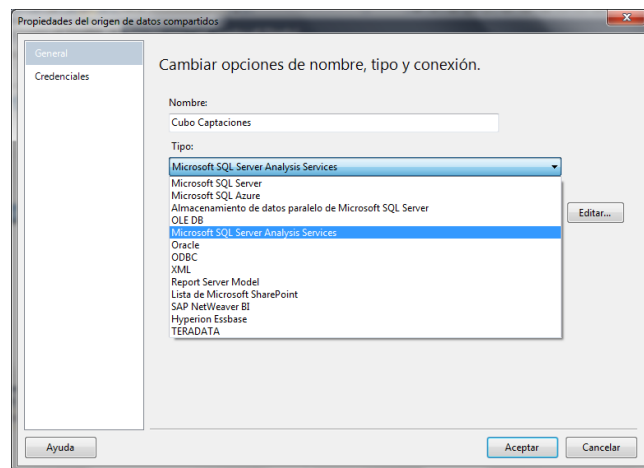


Figura 72. Origen de Datos Compartidos para Reportes.

Elaborado por: Juan Pesantez.

Posteriormente, se debe editar el origen de datos para completar las propiedades de la conexión y tener conformado un vínculo a los cubos generados como lo muestra la Figura 73.

Propiedades de la conexión del Servidor de Reportes.

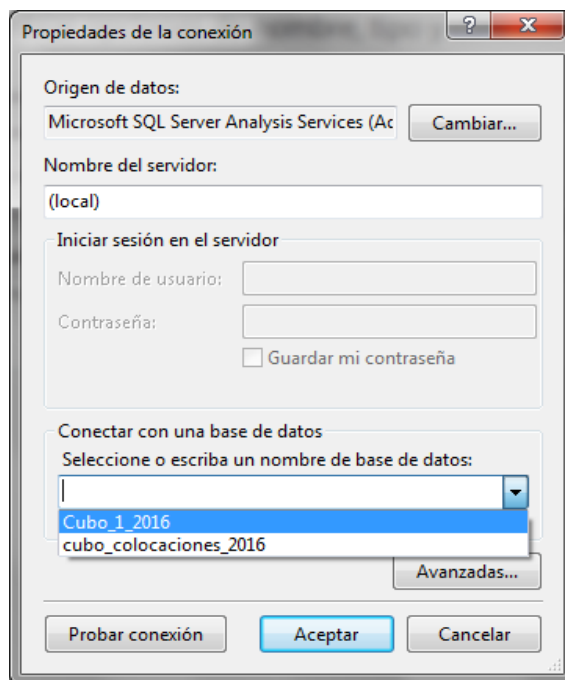


Figura 73. Propiedades de la Conexión del Servidor de Reportes.

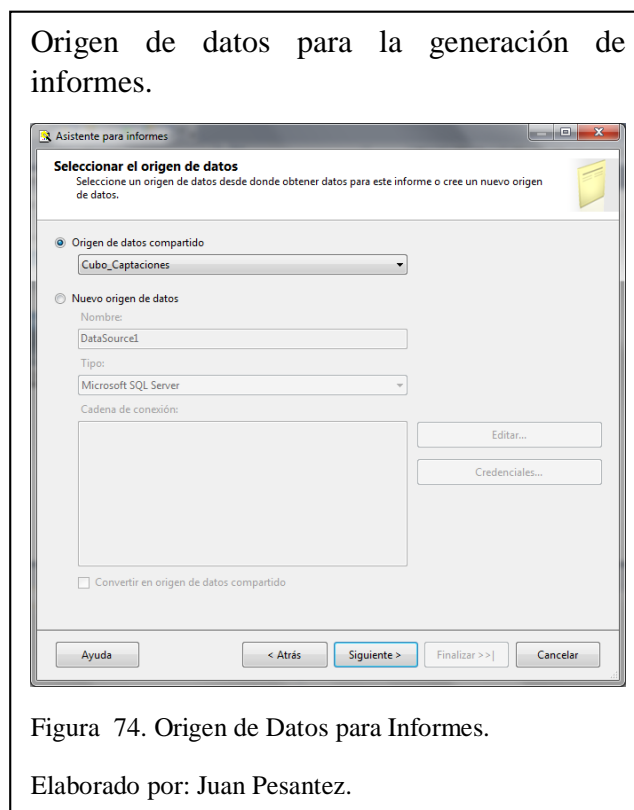
Elaborado por: Juan Pesantez.

Una vez seleccionada la fuente de datos que conecta el generador de informes con los cubos creados, se puede verificar si la conexión se encuentra debidamente realizada, presionando el botón de probar conexión y solo resta generar los informes con los parámetros que seleccione el usuario.

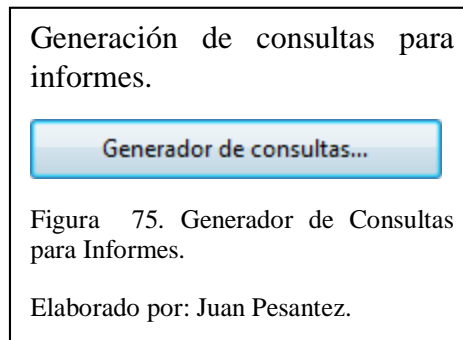
Ya creada la conexión, que es la fuente de donde se procede a extraer los datos, lo siguiente es crear un nuevo informe, ya sea este para captaciones o para colocaciones.

El procedimiento se lo realiza a través del asistente para informes, el cual guía al usuario paso a paso para la generación y presentación de la información en este documento.

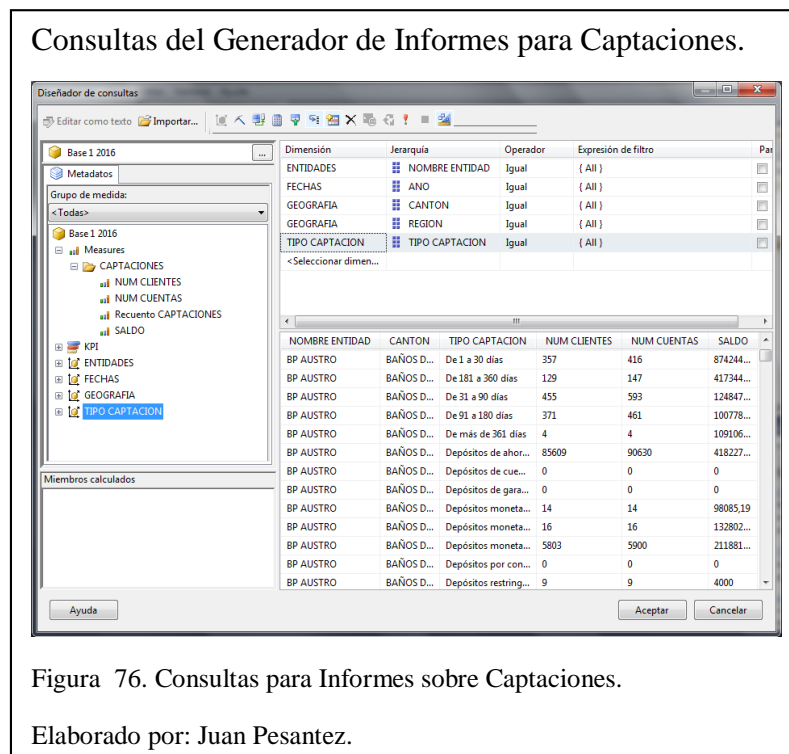
Al iniciar el asistente, muestra una pantalla en donde se debe seleccionar cual es el origen de datos compartido, que se toma como base para la captura de datos como lo muestra la Figura 74.



El siguiente paso es generar varias consultas para que de ese modo se organice la información deseada a presentar en el informe, presionando el botón del generador de consultas que se muestra en la Figura 75.



A continuación, aparece el examinador de cubos en donde se procede a colocar los campos de la tabla de hechos y de las tablas de dimensiones que se desea que sean incluidos en el informe final, estas consultas, campos y datos, difieren de los cubos que se generaron anteriormente, tanto de captaciones como de colocaciones de bancos privados del Ecuador, en la Figura 76 se muestra las consultas del cubo de captaciones.



En la figura 77 se muestran las consultas generadas para el cubo de colocaciones.

Consultas del Generador de Informes para Colocaciones.

Figura 77. Consultas para Informes sobre Colocaciones.

Elaborado por: Juan Pesantez.

Si se desea ver en lenguaje SQL como sería la consulta creada, el asistente muestra en pantalla todo el código generado a través de las tablas de hechos y dimensiones que se seleccionaron previamente en el explorador de cubos, como lo muestra la Figura 78.

Consulta Creada por el Generador de Informes.

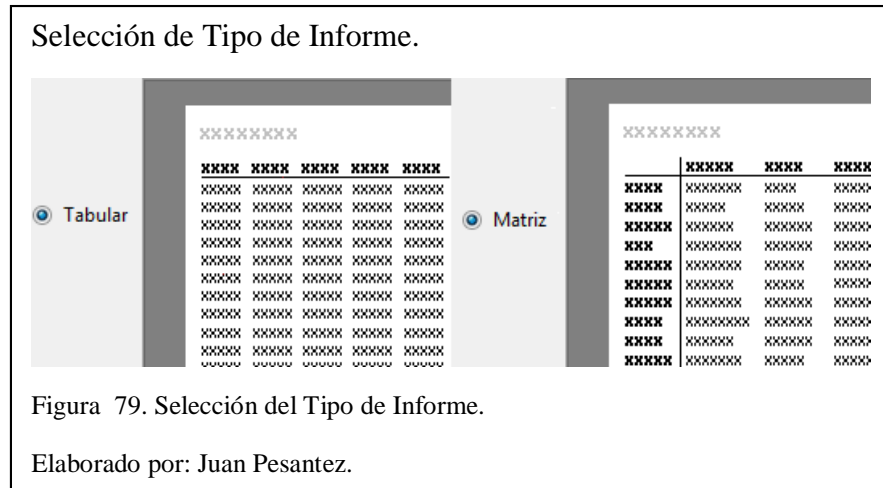
Cadena de consulta:

```
SELECT NON EMPTY { [Measures].[NUM CLIENTES], [Measures].[NUM CUENTAS], [Measures].[SALDO] } ON
COLUMNS, NON EMPTY { ([ENTIDADES].[NOMBRE ENTIDAD].[NOMBRE ENTIDAD].ALLMEMBERS *
[GEOGRAFIA].[CANTON].[CANTON].ALLMEMBERS * [TIPO CAPTACION].[TIPO CAPTACION].[TIPO
CAPTACION].ALLMEMBERS ) ; DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON
ROWS FROM ( (SELECT ({ [TIPO CAPTACION].[TIPO CAPTACION].[All] }) ON COLUMNS FROM ( (SELECT (
{ [GEOGRAFIA].[REGION].[All] }) ON COLUMNS FROM ( (SELECT ( { [GEOGRAFIA].[CANTON].[All] }) ON
COLUMNS FROM ( (SELECT ( { [FECHAS].[ANO].[All] }) ON COLUMNS FROM ( (SELECT ( { [ENTIDADES].
[NOMBRE ENTIDAD].[All] }) ON COLUMNS FROM [Base 1 2016])))) WHERE ( ([FECHAS].[ANO].[All],
[GEOGRAFIA].[REGION].[All] ) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE,
FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

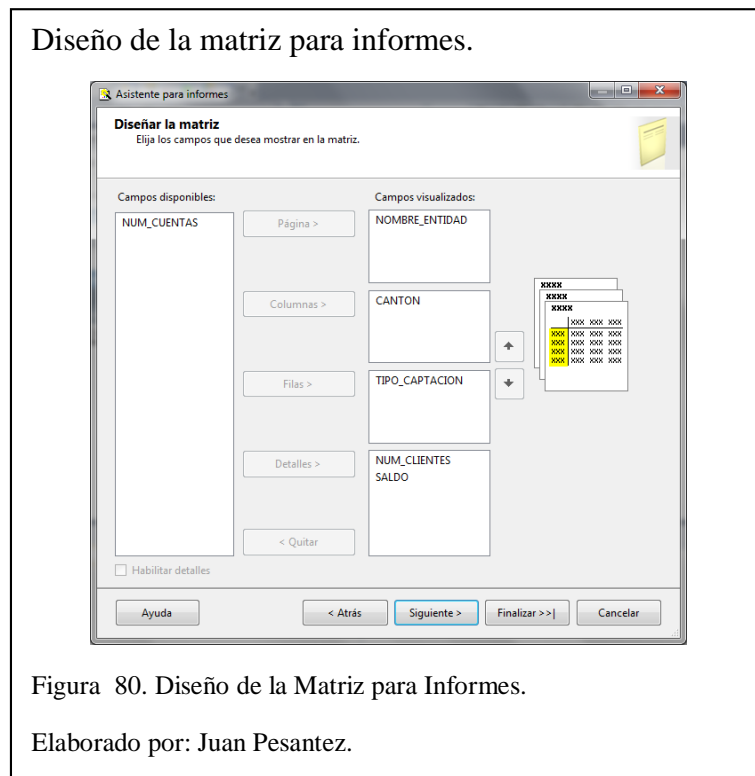
Figura 78. Consulta en Lenguaje SQL.

Elaborado por: Juan Pesantez.

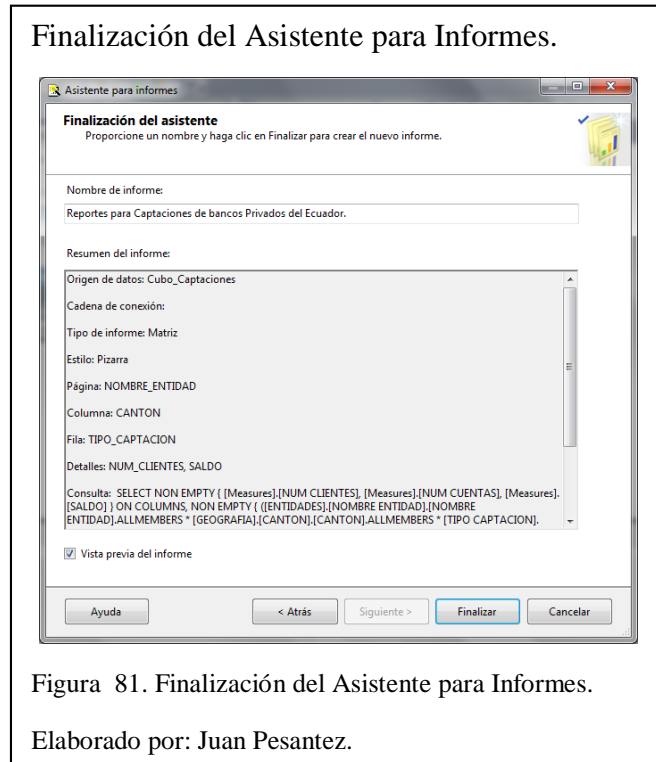
El siguiente paso es seleccionar la forma de presentación del informe, ya sean estos en tipo matriz o de tipo tabular como se ve en la Figura 79.



Una vez seleccionado el tipo de presentación, se debe colocar los campos que van a ser tomados como páginas, columnas, filas los mismos que otorgan detalles al informe, como se puede observar en la Figura 80.



Por último, el asistente muestra un detalle de todos los parámetros y variables seleccionados por el usuario y campos para ingresar un nombre al informe y la posibilidad de marcar una casilla para visualizar una perspectiva previa de la información, como se muestra en la Figura 81.



Para concluir se puede observar por medio de la vista previa como se encuentra constituido el informe y su presentación final, como se puede observar Figura 82.

Presentación de Informe Captaciones.

BP PROMERICA

	AMBATO			AZOGUES			BAÑOS DE AGUA SANTA		
	NUM CLIENTES	NUM CUENTAS	SALDO	NUM CLIENTES	NUM CUENTAS	SALDO	NUM CLIENTES	NUM CUENTAS	SALDO
De 1 a 30 días	1487	1975	130003190,85	1324	1649	52198315,64	655	749	37569637,9
De 181 a 360 días	490	542	10322009,62	2234	2780	51409520,89	783	883	17852789,7
De 31 a 90 días	1582	1774	50451089,83	1840	2410	79157416,69	789	916	45913964,2
De 91 a 180 días	1012	1100	17791158,8	2007	2549	76591093,74	633	721	27838542,6
De más de 361 días	101	102	2474804,51	112	116	1980440,08	42	42	1867368,3
Depósitos de ahorro	245906	251438	319700443,18	139817	149720	206860114,97	60483	63541	179946054,2
Depósitos de cuenta básica	268	380	34813330,5	257	343	13067483,38	167	199	13482450,7
Depósitos de garantía	0	0	0	0	0	0	0	0	0
Depósitos monetarios de instituciones financieras	2	2	20960,85	31	31	752167,31	12	12	149831,3
Depósitos monetarios	11602	11841	81412820,15	5637	5708	42139916,49	1603	1623	13526409,8

Figura 82. Informe Captaciones.
Elaborado por: Juan Pesantez.

Los pasos realizados para el informe preliminar, son similares para el informe de colocaciones, haciendo que, si el usuario sigue de la misma manera los pasos, se obtiene un informe detallado para colocaciones, como se muestra en la Figura 83.

Presentación de Informe Colocaciones.

BANCO DELBANK S.A.

GUAYAQUIL

	Abril					Agosto				
	CARTERA NO DEVENGA INTERES	CARTERA POR VENCER	CARTERA VENCIDA	MOROSIDAD	SALDO TOTAL	CARTERA NO DEVENGA INTERES	CARTERA POR VENCER	CARTERA VENCIDA	MOROSIDAD	
tipo_comercial	478126,75	17596831,98	140181,62	0,14401735	18215140,35	582248,82	19780959,7	149091,24	3241083,14441795	
tipo_consumo	680631,86	10012044,29	70776,8	0,382861967	9824938,36409089	914114,11	11494447,77	86924,33	1278171,04206706	
tipo_microcredito	163513,64	966263,12	34831,72	0,996511744	1164608,48	254672,9	1092783,47	50106,95	40393,818077996	
tipo_vivienda	415503,61	9294387,5	28119,52	0,258061269	9738010,63	387877,22	9796081,46	49560,29	1641197,42363424	

Figura 83. Presentación de Informe Colocaciones.

Elaborado por: Juan Pesantez.

CONCLUSIONES.

- Mediante el análisis, desarrollo e implementación del presente proyecto, se logra dar una perspectiva de lo rentables que son las herramientas informáticas para la inteligencia de negocios, haciendo de esta manera una correlación entre las tecnologías de la información y la banca ecuatoriana aumentando beneficios y disminuyendo costos.
- Para la aplicación de los algoritmos de minería de datos, es importante poseer de una gran plataforma de información concentrada en un gestor de base de datos, la cual mediante su calidad y extensión de información permitió una predicción acertada de la data.
- La elección de los diferentes algoritmos de minería de datos, depende de cómo se desea agrupar, clasificar u obtener los datos para su uso posterior.
- Si bien los algoritmos presentan varios resultados, estos deberían ser validados por personas expertas en el negocio, con el fin de verificar los resultados obtenidos producto de la aplicación de los algoritmos de minería de datos.

RECOMENDACIONES.

- Procurar mantener actualizada la data por parte de los archivos de la Superintendencia de Bancos del Ecuador.
- Se debe capacitar a personal específico para manejo de herramientas, análisis e interpretación de minería de datos y sus predicciones.
- Elegir una herramienta adecuada tanto para el análisis de información, así como para la aplicación de algoritmos de minería de datos, ya que los resultados de predicción pueden ser difusos entre múltiples herramientas.
- No se puede pasar por alto el uso de metodologías para la construcción de data warehouse, así como para los procesos de extracción, transformación y carga (“E.T.L.”), que manipulan los datos y la información de forma eficiente, otorgando mayor alcance e intuición para no pasar por alto ningún detalle para el proyecto.

REFERENCIAS.

- Curto, D. J. (2010). *Introducción al Business Intelligence*. España: Editorial UOC.
- Hueso Ibañez, Luis. (2000). *Gestión de bases de datos (2a. ed.)*. RA-MA Editorial.
- Jiménez Capel, María Yolanda. (2014). *Bases de datos relacionales y modelado de datos (UF1471)*. Madrid, ESPAÑA: IC Editorial.
- López Porrero, Beatriz E, Pérez Vázquez, Ramiro A. (2009). *Limpieza de datos*. Editorial Feijóo.
- Microsoft. (24 de 02 de 2016). <https://msdn.microsoft.com/es-es>. Obtenido de <https://msdn.microsoft.com/es-es>.
- Oppel, Andy. (2010). *Fundamentos de bases de datos*. McGraw-Hill Interamericana.
- Presser Carne, Cynthia. (2000). *Data mining*. El Cid Editor | apuntes.
- Ralph Kimball, Margy Ross. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*. Indiana: John Wiley & Sons, Inc.
- Reinosa, E. J., Maldonado, C. A., & Muñoz, R. (2012). *Bases de datos*. México: Alfaomega Grupo Editor.
- Rodríguez Parrilla, J. M. (2014). *Cómo hacer inteligente su negocio: business intelligence a su alcance*. México: Grupo Editorial Patria.
- Superintendencia de Bancos del Ecuador, (. (2016). *Bienvenidos a la Superintendencia de Bancos del Ecuador*. Obtenido de <http://www.sbs.gob.ec/>
- Trujillo, J. C. (2013). *Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional*. España: : ECU.
- Valderrey Sanz, Pablo. (2000). *Gestión de bases de datos*. RA-MA Editorial.

ANEXOS.

Anexo 1. Módulo Región.

```
Sub macro_region() 'inicio, nombre del módulo

Dim i As Long
Dim j As Integer '-----> creación de variables

For i = 1 To 71292 'inicio ciclo for
  For j = 2 To 224 'inicio ciclo for
    If (Sheets(1).Cells(i, "AX") = Sheets(2).Cells(j, "E")) Then
      Sheets(1).Cells(i, "AW") = Sheets(2).Cells(j, "G")
    End If 'fin If
  Next j 'aumento de la variable j segundo for
Next i 'aumento de la variable i primer for

End Sub 'fin modulo
```

Anexo 2. Módulo Provincias.

```
Sub macro_provincias() 'inicio, nombre del módulo

Dim i As Long
Dim j As Integer '-----> creación de variables

For i = 1 To 71292 'inicio ciclo for 1
  For j = 2 To 224 'inico ciclo for 2
    If (Sheets(1).Cells(i, "AY") = Sheets(2).Cells(j, "C")) Then
      Sheets(1).Cells(i, "AX") = Sheets(2).Cells(j, "E")
    End If 'fin if
  Next j 'avanza variable j para for 2
Next i 'avanza variable j para for 1

End Sub 'Fin módulo
```

Anexo 3. Módulo Cantón.

```
Sub macro_canton() 'inicio, nombre del módulo

Dim valor1 As String
Dim valor2 As String
Dim valor3 As String
Dim entidad As String
Dim inicio As Long
Dim i As Long           '-----> creación de variables
Dim j As Long
Dim k As Integer
Dim cuenta As Integer
Dim acumulador As Long
Dim aumento As Long

aumento = 1 'aignación
acumulador = 156
cuenta = 0 'inicializa la variable cuenta
inicio = 4 'donde se encuentra el primer valor
valor1 = Cells(inicio, "D").Value 'toma el valor de la celda asignada
valor2 = Cells(5, "D").Value 'toma el valor de la celda asignada

For i = 1 To 71292 'inicio ciclo for 1
    valor3 = "Total " & Right(valor1, cuenta) 'asigno valor3 con número de caracteres de la variable cuenta
    If valor1 = valor2 Or valor1 = valor3 Then 'condición para verificar si es o no una entidad en la celda
        inicio = inicio + 1 'aumenta la cuenta para la variable inicio
        valor1 = Cells(inicio, "D").Value 'toma el valor de la celda asignada
    Else
        entidad = valor1 'asigno valor de variable a una auxiliar
        For j = aumento To acumulador 'inicio ciclo for 2
            Cells(j, "AY") = entidad 'Celdas donde se colocaran los datos
        Next j 'siguiente valor para la variable j del for 2
        aumento = aumento + 156 'aumenta segun el número de tipos de deposito
        acumulador = acumulador + 156 'aumento variable acumulador
        inicio = inicio + 1 'aumenta la cuenta para la variable inicio
        cuenta = Len(valor1) 'asigno el numero de caracteres correspondientes al nombre de la entidad
        valor1 = Cells(inicio, "D").Value 'toma el valor de la celda asignada
    End If 'fin if
Next i 'siguiente valor para la variable j del for 1

End Sub 'fin módulo
```

Anexo 4. Módulo Tipo Depósito.

```
Sub macro_Tipo_Deposito() 'inicio, nombre del módulo

Dim inicio As Long
Dim i As Integer
Dim tipo As String      '-----> creación de variables
Dim acum As Long
Dim acum2 As Long

acum = 1
acum2 = 0                '-----> Asignación de variables
inicio = 4

Do While acum2 <> 2000 'ciclo do while inicio

For i = 4 To 16 'inicio ciclo for 1
    tipo = Sheets(1).Cells(i, "E").Value
    For j = 1 To 12 'inicio ciclo for 2
        Cells(acum, "AZ") = tipo 'Celdas donde se colocaran los datos
        acum = acum + 1 'aumento variable acum
    Next j 'siguiente valor a variable j for 2
Next i 'siguiente valor a variable i for 1
acum2 = acum2 + 5 'aumento variable acum2
Loop 'fin ciclo do while
End Sub 'fin módulo
```

Anexo 5. Módulo Número de Clientes.

```
Sub macro_numclientes() 'inicio, nombre del módulo

Dim acum As Long
Dim acum2 As Integer
Dim vall As String
Dim inicio As Integer
Dim desde As Integer
Dim cond As Integer      '-----> creación de variables
Dim i As Integer
Dim j As Integer
Dim k As Integer
Dim num As Integer
Dim cont As Integer

desde = 4
cond = 1
acum = 1
inicio = 4      '-----> asignación de variables
num = 1
cont = 1

Do While cond <> 0      'inicio ciclo while

  If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
    For i = desde To 12 + desde      'columnas de seleccion para los datos
      For j = 6 To 17      'filas de eleccion para los datos
        Sheets(1).Cells(acum, "BA") = Cells(i, j) 'Celdas donde se colocaran los datos
        acum = acum + 1 'aumento variable acum
      Next j 'siguiente valor a variable j
      inicio = inicio + 1 'aumento variable inicio
    Next i 'siguiente valor variable i
    acum2 = i 'aignación variable acum2 = i
  End If 'fin ciclo if

  For k = 1 To 5 'inicio ciclo for
    If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
      desde = inicio
      cond = 1      '-----> asignación de variables
      cont = 1
    Else
      cond = 0
      inicio = inicio + 1      '-----> asignación de variables
    End If 'fin if
  Next k 'siguiente valor a variable k
  num = num + 1 'aumento variable num

Loop 'fin ciclo while

End Sub 'fin módulo
```

Anexo 6. Módulo Número de Cuentas.

```
Sub macro_num_cuentas()

Dim acum As Long
Dim acum2 As Integer
Dim vall As String
Dim inicio As Integer
Dim desde As Integer
Dim cond As Integer          '-----> creación de variables
Dim i As Integer
Dim j As Integer
Dim k As Integer
Dim num As Integer
Dim cont As Integer

desde = 4
cond = 1
acum = 1
inicio = 4          '-----> Asignación de variables
num = 1
cont = 1

Do While cond <> 0 'inicio ciclo do while

  If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
    For i = desde To 12 + desde 'columnas de seleccion para los datos
      For j = 18 To 29 'filas de eleccion para los datos
        Sheets(1).Cells(acum, "BB") = Cells(i, j) 'Celdas donde se colocaran los datos
        acum = acum + 1 'aumento de variable acumulador
      Next j 'siguiente valor para j
      inicio = inicio + 1 'aumento variable inicio
    Next i 'siguiente valor para i
    acum2 = i 'asignacion de variable i a variable acum2
  End If 'fin if
  For k = 1 To 5 'inicio ciclo for
    If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
      desde = inicio
      cond = 1          '----->'asignacion de variables
      cont = 1
    Else
      cond = 0
      inicio = inicio + 1          '----->'asignacion de variables
    End If 'fin if
  Next k 'siguiente valor de k
  num = num + 1 'aumento de variable num

Loop 'fin de ciclo do while

End Sub 'fin módulo
```

Anexo 7. Módulo Saldo.

```
Sub macro_saldo() 'inicio, nombre del módulo

Dim acum As Long
Dim acum2 As Integer
Dim vall As String
Dim inicio As Integer
Dim desde As Integer      '----->creación de variables
Dim cond As Integer
Dim i As Integer
Dim j As Integer
Dim k As Integer
Dim num As Integer
Dim cont As Integer

desde = 4
cond = 1
acum = 1
inicio = 4                '-----> asignación de variables
num = 1
cont = 1

Do While cond <> 0      'inicio ciclo do while

  If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
    For i = desde To 12 + desde      'columnas de seleccion para los datos
      For j = 30 To 41      'filas de eleccion para los datos
        Sheets(1).Cells(acum, "BC") = Cells(i, j) 'Celdas donde se colocaran los datos
        acum = acum + 1 'aumento variable acum
      Next j 'siguiente valor para variable j
      inicio = inicio + 1 'aumento variable inicio
    Next i 'siguiente valor para variable i
    acum2 = i 'asigno la variable i a la variable acum2
  End If ' fin if
  For k = 1 To 5 'inicio ciclo for
    If (Sheets(1).Cells(inicio, "E").Value = "DEPÓSITOS MONETARIOS QUE GENERAN INTERESES") Then
      desde = inicio
      cond = 1
      cont = 1          '-----> asignación de variables
    Else
      cond = 0
      inicio = inicio + 1
    End If 'fin if
  Next k 'siguiente valor para variable k
  num = num + 1 'aumento para variable num

Loop 'fin do while

End Sub 'fin módulo
```