

**UNIVERSIDAD RICARDO PALMA  
FACULTAD DE INGENIERÍA**

**PROGRAMA DE TITULACIÓN POR TESIS**

**ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA**



**SISTEMA DE INFORMACIÓN PARA LA TOMA DE DECISIONES,  
USANDO TÉCNICAS DE ANALISIS PREDICTIVO PARA LA  
EMPRESA IASACORP INTERNATIONAL S.A.**

**PARA OBTENER EL TÍTULO PROFESIONAL DE  
INGENIERO INFORMÁTICO**

**PRESENTADO POR:**

**BACH. ESPINOZA ESPINOZA BERTHA YRENE  
BACH. GUTIERREZ RIVERA NATALIA ELIZABETH**

**ASESOR: LIC. RENZO WARTHON VARELA**

**LIMA-PERÚ  
AÑO: 2015**

## **DEDICATORIA**

*A mis padres, por su constante motivación para dar este gran paso, a mi esposo y mi hija por su apoyo, confianza y amor.*

*Bertha Y. Espinoza Espinoza*

*A mis padres, por su constante apoyo moral y comprensión en todos estos meses, y la fuerza que me dieron para seguir adelante.*

*Natalia E. Gutiérrez Rivera*

## **AGRADECIMIENTOS**

*A todas las personas que nos apoyaron en la elaboración de nuestra tesis, en especial a Dios y a nuestro asesor Renzo Warthon V., por todas las enseñanzas brindadas en esta etapa.*

# ÍNDICE

<b>INTRODUCCIÓN</b>	<b>1</b>
<b>Antecedentes del Problema</b>	<b>2</b>
<b>Definición del problema</b>	<b>3</b>
<b>Propósito del proyecto</b>	<b>4</b>
<b>Importancia del proyecto</b>	<b>5</b>
<b>Justificación del Proyecto</b>	<b>5</b>
<b>Definición de Términos</b>	<b>6</b>
<b>Resumen</b>	<b>8</b>
<b>CAPÍTULO I: MARCO TEORICO</b>	<b>9</b>
<b>Introducción</b>	<b>9</b>
<b>Revisión de la Literatura</b>	<b>10</b>
<b>CAPITULO II: DESARROLLO DEL PROYECTO</b>	<b>41</b>
<b>1. Alcance del Proyecto</b>	<b>41</b>
Alcance del Producto	41
Criterios de Aceptación del Producto	42
Estructura de Desglose del Trabajo y Entregables	43
Exclusiones del proyecto	47
Restricciones del Proyecto	47
Supuestos del Proyecto	47
<b>2. Modelado del Negocio</b>	<b>48</b>

<b>3. Requerimientos de Software/Producto</b>	<b>54</b>
<b>4. Arquitectura de Software/Producto</b>	<b>56</b>
<b>5. Diseño de la Solución</b>	<b>64</b>
<b>CONCLUSIONES</b>	<b>141</b>
<b>RECOMENDACIONES</b>	<b>143</b>
<b>REFERENCIAS BIBLIOGRAFICAS</b>	<b>144</b>
<b>ANEXOS</b>	<b>146</b>
<b>Anexo 1: Script para probar el DTS de extracción de las tablas maestros y ventas</b>	<b>146</b>
<b>Anexo 2 : Script para comprobar la carga y eliminación de los registros en las tablas de trabajo dimensionales y de hechos.</b>	<b>148</b>
<b>Anexo 3 : Script para comprobar la carga o eliminación de los registros en las tablas dimensionales y de hechos en el datawarehouse.</b>	<b>149</b>

# ÍNDICE DE GRÁFICOS

Gráfico 1: Organigrama de la Empresa	2
Gráfico 2: Entorno/Competencias de la Empresa	3
Gráfico 3: Modelo de proceso CRISP-DM	12
Gráfico 4: Tareas de la metodología de Kimball	15
Gráfico 5: Ejemplo de modelo de datos en estrella	18
Gráfico 6: Ejemplo de Árbol de Decisión Costo - Beneficio de Proveedores	29
Gráfico 7: Histograma de Predicción	32
Gráfico 8: Correlación del Algoritmo	32
Gráfico 9: Diagrama de Regresión	33
Gráfico 10: Ecuación de Regresión	33
Gráfico 11: Diagrama de Estructura de Desglose de Trabajo	45
Gráfico 12 Cronograma Gantt del Proyecto	46
Gráfico 13: Actor del Negocio	48
Gráfico 14: Diagrama de Casos de Uso del Negocio	48
Gráfico 15: Trabajador del Negocio	49
Gráfico 16: Diagrama de procesos del negocio AS IS	51
Gráfico 17: Diagrama de proceso del negocio TO BE	53
Gráfico 18: Diagrama de Arquitectura del Sistema	56
Gráfico 19: Diagrama de despliegue	58
Gráfico 20: Diagrama de componentes	59
Gráfico 21: Diagrama de clases (Diseño Lógico)	62
Gráfico 22: Diagrama de Base de Datos (Diseño Físico)	63
Gráfico 23: Actores del Sistema	64
Gráfico 24: Diagrama de Paquetes del Sistema	65
Gráfico 25: Diagrama de Caso de Uso del Sistema: Proceso ETL	66
Gráfico 26: Diagrama de Caso de Uso del Sistema: Clasificación de Clientes	67

Gráfico 27: Diagrama de Caso de Uso del Sistema: Analizar Información	67
Gráfico 28: Diagrama de Clases: Extraer Información	75
Gráfico 29: Diagrama de Secuencia: Extraer Información	76
Gráfico 30: Creación de Nuevo Proyecto DTS	77
Gráfico 31: Administrador de conexiones	78
Gráfico 32: Conexión OLE DB	78
Gráfico 33: Generación Tarea Truncate Table Maestras	79
Gráfico 34: Flujo de trabajo para la tabla artículos1	80
Gráfico 35: Conexión desde la Base de Datos Transaccional de la tabla articulos1	80
Gráfico 36: Conexión a la BD de Trabajo, tabla Articulos1	81
Gráfico 37: Generación Tarea truncate Table Transaccionales	81
Gráfico 38: Flujo de trabajo para la tabla cboletav	82
Gráfico 39: Conexión desde la Base de Datos Transaccional, de la tabla cboletav	82
Gráfico 40: Conexión hacia la Base de Datos de trabajo, de la tabla cboletav	83
Gráfico 41: Diagrama de Clases: Transformar Información	90
Gráfico 42: Diagrama de Secuencia: Transformar Información	91
Gráfico 43: Administrador de Conexiones	92
Gráfico 44: Truncar las tablas de trabajo dimensional y de hechos	93
Gráfico 45: Actualización de datos Tabla Cliente	94
Gráfico 46: Flujo de trabajo para la tabla de trabajo dm_articulo_work	94
Gráfico 47: Conexión desde la base de datos de trabajo, de la tabla articulos1	95
Gráfico 48: Conexión hacia la BD de Trabajo en la tabla de dimensión dm_articulo_work	95
Gráfico 49 Flujo de trabajo de las tablas trabajo dimensionales y de hecho:	96
Gráfico 50: Diagrama de Clases: Cargar Información	103
Gráfico 51: Diagrama de Secuencia: Cargar Información	104
Gráfico 52: Administrador de Conexiones	105

Gráfico 53: Eliminación de registros de las tablas de trabajo dimensional y de hechos	106
Gráfico 54: Flujo de trabajo para la tabla dimensional dm_artículo	107
Gráfico 55: Conexión desde la base de datos de trabajo, de la tabla dm_articulo_work	107
Gráfico 56: Conexión hacia el Datawarehouse en la tabla dimensión dm_articulo	108
Gráfico 57: Flujo de trabajo para las tablas de trabajo dimensionales y de hecho	108
Gráfico 58: Diagrama de Clases: Generar Clasificación de Clientes	110
Gráfico 59: Diagrama de Secuencia: Generar Clasificación de Clientes	111
Gráfico 60: Creación de Proyecto	112
Gráfico 61: Creación al nuevo origen de datos	112
Gráfico 62: Asistente para la creación del origen de datos	113
Gráfico 63: Selección de la base de datos origen	113
Gráfico 64: Generación de la conexión del origen de datos	114
Gráfico 65: Creación de la Vista de Origen de Datos	115
Gráfico 66: Creación de la estructura de minería de datos: Por Canal de Venta y Línea	116
Gráfico 67: Asistente de minería de datos	116
Gráfico 68: Selección Base de Datos relacional	117
Gráfico 69: Selección de la técnica de minería de datos	117
Gráfico 70: Selección la vista de origen de datos creada	118
Gráfico 71: Selección del tipo de tabla v_clifrecuente	118
Gráfico 72: Selección de variables para el análisis de predicción	119
Gráfico 73: Visualización del contenido y tipos de datos	119
Gráfico 74: Asignación del nombre de la estructura y modelo de minería de datos	120
Gráfico 75: Vista de la estructura de minería de datos creada	120
Gráfico 76: Estructura de minería de datos	121
Gráfico 77: Modelo de minería de datos	121

Gráfico 78: Visor de minería de datos-Leyenda de minería de datos	122
Gráfico 79: Diagrama de Clases: Analizar información por Canal de Venta y Línea	124
Gráfico 80: Diagrama de Secuencia: Analizar información por Canal de Venta y Línea	125
Gráfico 81: Ejecución estructura de minería de datos Clifrecuente CanaldeVenta_Linea	126
Gráfico 82: Visor de modelos de minería de datos y visualizamos el árbol de decisión	126
Gráfico 83: Vista de Escenarios del análisis predictivo	127
Gráfico 84: Leyenda de Minería de datos	128
Gráfico 85: Análisis por Canal de Venta	128
Gráfico 86: Probabilidades de venta de Cliente Frecuentes por Línea	129
Gráfico 87: Leyenda de Minería de datos	130
Gráfico 88: Diagrama de Clases Analizar información por Línea y Familia	132
Gráfico 89: Diagrama de Secuencia CUS Analizar información por línea y familia	133
Gráfico 90: Ejecución estructura de minería de datos Clifrecuente Linea_Familia	134
Gráfico 91: Visor de modelos de minería de datos y visualizamos el árbol de decisión	134
Gráfico 92: Visor de Escenarios del análisis predictivo	135
Gráfico 93: Leyenda de Minería de datos	136
Gráfico 94: Análisis por Línea Bisutería	137
Gráfico 95: Leyenda de Minería de datos	138
Gráfico 96: Probabilidades de venta de Cliente Frecuentes por Línea y Familia	138
Gráfico 97: Leyenda de Minería de datos	139

# ÍNDICE DE TABLAS

Tabla 1 Funciones de Predicción	35
Tabla 2 Descripción de Estructura de Desglose de Trabajo	44
Tabla 3 Descripción del Modelo del Proceso de Negocio AS IS	50
Tabla 4 Descripción del Modelo de Proceso de Negocio TO BE	52
Tabla 5 Diccionario de Datos	60
Tabla 6 Especificación Técnica detallada del CUS: Extraer Información	71
Tabla 7 Especificación Técnica detallada del CUS: Transformar Información	86
Tabla 8 Especificación Técnica detallada del CUS: Cargar Información	99
Tabla 9 Escenarios y probabilidades de Ventas por Línea de Cliente Frecuentes	127
Tabla 10 Escenarios y Probabilidades de Ventas de Clientes Frecuentes	129
Tabla 11 Escenarios y Probabilidades de ventas por Familia de Clientes Frecuentes	135
Tabla 12 Escenarios y Probabilidades de venta por Lines Bisutería y Familia	137
Tabla 13 Escenarios y Probabilidades de Venta de Clientes Frecuentes	139

## **Resumen:**

En la actualidad, las empresas manejan una gran cantidad de información, el cual era inimaginable años atrás, la capacidad de recolectarla es muy impresionante. En consecuencia, para varias empresas esta información se ha convertido en un tema difícil de manejar. Diariamente, las empresas sea del sector, tipo o tamaño que sea, toman decisiones, las cuales la mayoría son decisiones estratégicas que pueden afectar el correcto funcionamiento de la empresa.

Es aquí, donde ingresa una de las herramientas más mencionadas en el área de TI: Business Intelligence, este término se refiere al uso de datos en una empresa para facilitar la toma de decisiones, explotar su información, y mejor aún, plantear o predecir escenarios a futuro.

El presente trabajo permitirá al área de Marketing de la empresa lasacorp International, obtener información sobre el comportamiento y hábitos de compra de los clientes, mediante técnicas de minería de datos como Árbol de Decisión y técnicas de análisis predictivo, la cual ayudará a la toma de decisiones para establecer estrategias de venta de las líneas (bisutería, complementos de vestir, accesorios de cabello, etc.) que maneja la empresa y de las próximas compras.

De acuerdo a lo planteado anterior mente, la implementación de este tipo de sistemas de información ofrece a la empresa ventajas competitivas, permite a la gerencia analizar y entender mejor la información y por consecuencia tomar mejores decisiones de negocio.

**Palabras claves:** Business Intelligence, Análisis Predictivo, Árbol de Decisión

**Abstract:**

At present, companies handle a lot of information, which was unimaginable years ago, the ability to collect it is very impressive. Consequently, for many companies this information has become a difficult issue to handle. Due to the large volume of information we have, instead of being useful you can fall in a failed attempt to give proper use.

Every day, companies in any sector, type or size, make decisions, most of which are strategic decisions that may affect the proper functioning of the company.

It's here, where we talk about the most mentioned tools in the area of IT: Business Intelligence, this term refers to the use of data in an enterprise to facilitate decision-making, exploit their information, and better yet, raise or predict scenarios future.

This work will allow the area lasacorp Marketing Company International, information on the behavior and buying habits of customers, through predictive analysis techniques, which will help the decision to establish sales strategies lines (jewelry, clothing, hair accessories, etc.) that manages the company and nearby shopping.

According to the points made above, the implementation of such information systems offers companies competitive advantages, allows management to better analyze and understand information and consequently make better business decisions.

**Key words:** Business Intelligence, Predictive Analytics, Decision Tree

# INTRODUCCIÓN

El propósito de este proyecto es obtener un sistema que desempeñe el papel de soporte para la toma de decisiones, de respuesta rápida, con información precisa para poder aprovechar las oportunidades: “estar en el lugar indicado, en el momento oportuno, con la información correcta”.

Los sistemas orientados para la toma de decisiones son los englobados por el término Business Intelligence.

Business Intelligence es uno de los puntales de la actual revolución tecnológica que estamos experimentando. La cantidad de datos generados por la sociedad de la información crece día a día, y seguirá creciendo gracias a la explosión de las redes sociales, el big data, los dispositivos móviles, etc. Este incremento exponencial del volumen de datos que se generan hace imprescindible el uso de sistemas que sean capaces de analizarlos y convertirlos en información útil.

Pero muy pocas empresas aprovechan esta información para pronosticar o bien para encontrar las agrupaciones o relaciones entre datos, que a simple vista no se logran descubrir.

Es por eso que mostramos en este proyecto que implementar soluciones de Análisis predictivo en las estrategias de Business Intelligence que pueden tener una empresa, lograr hacer realidad beneficios importantes, se estima que los análisis predictivos cobrarán mayor relevancia en el futuro inmediato en buena parte porque cada vez se tiene más información del cliente y además la nube permite que los datos se puedan manejar de una forma más escalable y flexible.

## Antecedentes del Problema

**IASACORP INTERNATIONAL S.A.** es una empresa importadora y comercializadora de productos de moda para mujer, está orientada al rubro retail, dueña de las marcas Do it!, Do it! Kids, Glitter y Franquicias como Funky Fish, estas marcas están presentes en diferentes canales de venta como centros comerciales, tiendas por departamentos y supermercados.

En el área de Marketing, actualmente las estimaciones de ventas y de compras por cada línea (bisutería, accesorio de cabello, complemento de vestir, etc.) y el análisis de ventas por línea y producto, son trabajadas en base a reportes históricos. Esta información es extraída del Sistema Transaccional de Ventas.

En base a esta información, a *supuestos y experiencia en el mercado* por parte de las integrantes del área de Marketing, se toman decisiones para las estimaciones de venta por línea y familia, estimaciones de compras, lanzamiento de ofertas, descuentos o nuevas líneas de negocio.

## Organigrama de la Empresa

A continuación se muestra el Organigrama de la empresa IASACORP INTERNATIONAL S.A, las áreas están distribuidas por Gerencias. Ver Gráfico 1.



Gráfico 1: Organigrama de la Empresa

## Entorno/Competencias de la Empresa

La competencia está integrada por las empresas que actúan en el mismo mercado y realizan la misma función dentro de un mismo grupo de clientes con independencia de la tecnología empleada para ello.

El entorno y las competencias de la empresa IASACORP INTERNATIONAL se muestra a continuación en el Gráfico 2:



Gráfico 2: Entorno/Competencias de la Empresa

## Definición del problema

### Problema Central

La empresa no puede realizar pronósticos de Compra y Venta correctos pues estos se basan en supuestos y estimaciones del comportamiento histórico del mercado.

### Problemas Específicos

El área de Marketing presenta las siguientes dificultades:

1. Las ventas reales de la compañía pueden verse afectadas al utilizar un plan de ventas que es basado en supuestos, conocimientos del mercado e información histórica.

2. El proceso de elaboración del plan de compras puede no ser tan exacto al utilizar un plan de ventas que fue elaborado como se indica en el punto 1.
3. La información obtenida del sistema transaccional de ventas no es de gran ayuda para analizar los hábitos de consumo de los clientes preferenciales, por lo tanto las estrategias de promociones, ofertas y descuentos para los clientes no es la mejor.

## **Propósito del proyecto**

### **Objetivo General**

Desarrollar un sistema de información para el apoyo a la toma de decisiones del área de Marketing, que mediante técnicas de Análisis Predictivo permita realizar las proyecciones de venta y compra de las líneas y familias de productos de manera exacta, así mismo conocer el comportamiento de consumo de los clientes preferenciales para establecer estrategias de marketing más efectivas.

### **Objetivos Específicos**

1. Identificar dentro del Sistema transaccional de ventas, la información comercial del negocio, la cual incluye: datos de clientes preferenciales, sus transacciones de venta, el detalle de los productos, etc. la cual nos servirá como fuente de origen de datos.
2. Construir un modelo de datos multidimensional que mediante un proceso de ETL (Extracción, Transformación y Carga) se registre la información comercial identificada en el objetivo específico 1 y que permita ser explotado por la herramienta de análisis predictivo.

3. Implementar el algoritmo de árbol de decisión dentro de la herramienta de análisis predictivo, la cual generará la información que servirá de apoyo para la proyección de ventas, estimaciones de compras y lanzamiento de estrategias comerciales.

### **Importancia del proyecto**

Implementar una herramienta de Análisis Predictivo ayudará a conocer de antemano qué va a suceder y establecer qué tenemos que hacer ante esa situación determinada, se trata de minimizar los riesgos y optimizar estrategias.

Esta herramienta ayudará al área de Marketing a estimar con mayor exactitud los planes de venta y compra trayendo como consecuencia la mejora del presupuesto de compra, minimizar los saldos de inventario no vendido, etc. Asimismo, identificar las necesidades de los clientes en base a sus hábitos de consumo, ofrecer nuevos productos, etc.

### **Justificación del Proyecto**

Debido a las dificultades identificadas en la problemática del proyecto, la implementación de una herramienta de análisis de información permitirá a la Gerencia de Marketing tomar decisiones, que ayudarán a mejorar:

1. Las estimaciones de ventas por línea y familia de productos con mayor grado de precisión.
2. El proceso de elaboración del plan de compras de manera exacta.
3. Establecer estrategias de marketing (ofertas, descuentos, promociones) más efectivas en base al hábito de consumo de los clientes preferenciales.

## Definición de Términos

**Retail (el detal o venta al detalle)** es un sector económico que engloba a las empresas especializadas en la comercialización masiva de productos o servicios uniformes a grandes cantidades de clientes. Es el sector industrial que entrega productos al consumidor final.

**BI (Business Intelligence)**, conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

**DW (Datawarehouse)**, es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

**DM (Data Mart)**, son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

**ETL (Extract Transformation Load)**, proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, datamart, o datawarehouse para analizar.

**OLAP (On-Line Analytical Processing)**, Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

**Data Mining (minería de datos)**, es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos.

**Análisis Predictivo**, agrupa una variedad de técnicas estadísticas de modelización, aprendizaje automático y minería de datos que analiza los datos actuales e históricos reales para hacer predicciones acerca del futuro o acontecimientos no conocidos.

**Algoritmo**, es un conjunto prescrito de instrucciones o reglas bien definidas, ordenadas y finitas que permite realizar una actividad mediante pasos sucesivos que no generen dudas a quien deba realizar dicha actividad.

**Analysis Services**, es un motor de datos analíticos en línea que se usa en soluciones de ayuda a la toma de decisiones y Business Intelligence (BI), y proporciona los datos analíticos para informes empresariales.

**Árbol de Decisión**, un árbol de decisión indica las acciones a realizar en función del valor de una o varias variables. Es una representación en forma de árbol cuyas ramas se bifurcan en función de los valores tomados por las variables y que terminan en una acción concreta.

De forma más concreta, refiriéndonos al ámbito empresarial, podemos decir que los árboles de decisión son diagramas de decisiones secuenciales nos muestran sus posibles resultados. Éstos ayudan a las empresas a determinar cuáles son sus opciones al mostrarles las distintas decisiones y sus resultados.

## **Resumen**

La tecnología de Business Intelligence resulta sumamente importante para las empresas, estas herramientas aseguran que las decisiones tomadas sean siempre las mejores.

Esto gracias a la capacidad de explotar su información, con la intención de poder manipularlos de una manera más sencilla y entender el porqué de nuestro desempeño o, mejor aún, plantear escenarios a futuro, lo cual ayudará a tomar mejores decisiones.

Si se tiene suficiente cantidad de datos, se pueden empezar a identificar patrones y a construir modelos. Y una vez que se ha construido un modelo, ya se puede predecir. Existen una gran variedad de métodos para hacer predicciones sobre el futuro. De todas formas, es importante entender que el objetivo del análisis predictivo es hacer predicciones de lo que puede pasar con una cierta probabilidad.

Es por eso que el objetivo de este proyecto se basa en implementar un sistema de información aplicando Análisis Predictivo, con el fin de apoyar al área de Marketing en analizar la información de una manera más rápida y efectiva para la toma de decisiones.

# CAPÍTULO I: MARCO TEORICO

## Introducción

Este marco teórico tuvo como propósito investigar los temas asociados a este proyecto que permitió su desarrollo. Para mantenerse competitiva una empresa, las personas que toman decisiones requieren de un acceso rápido y fácil a información útil y valiosa de la empresa. Una forma de solucionar este problema es por medio del uso de Business Intelligence o Inteligencia de Negocios.

Algo peor que no tener información disponible es tener mucha información y no saber qué hacer con ella. La Inteligencia de Negocios (BI) es la solución a ese problema, pues por medio de dicha información puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva. La clave para BI es la información y uno de sus mayores beneficios es la posibilidad de utilizarla en la toma de decisiones.

En la actualidad hay una gran variedad de software de BI con aplicaciones similares que pueden ser utilizados en las diferentes áreas de la empresa, tales como, ventas, marketing, finanzas, etc. Son muchas las empresas que se han beneficiado por la implementación de una sistema de BI, además se pronostica que con el tiempo se convertirá en una necesidad de toda empresa.

Muchas empresas logran convertir su información en conocimiento al hacer un análisis estadístico de lo sucedido, pero muy pocas empresas aprovechan esta información para pronosticar o bien para encontrar las agrupaciones o relaciones entre datos, que a simple vista no se logran descubrir. El Análisis Predictivo es el proceso de seleccionar, explorar, y modelar grandes volúmenes de datos que develen información previamente desconocida para beneficio del negocio. Dicho en otras palabras, es descubrir por medio de un proceso cíclico nuevas correlaciones, patrones y tendencias significativas, filtrando y depurando grandes volúmenes de datos, utilizando técnicas de reconocimiento de comportamiento, así como modelos matemáticos y estadísticos.

## Revisión de la Literatura

### 1. BUSINESS INTELLIGENCE

#### Definiciones

Las aplicaciones de Business Intelligence (BI) son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa. Estas aplicaciones proporcionan a los usuarios un mayor entendimiento que les permite identificar las oportunidades y los problemas de los negocios. (CherryTree & Co., 2000)

Una interesante definición para inteligencia de negocios o BI, por sus siglas en inglés, según el Datawarehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que nos permiten transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados.

#### **¿Por qué Inteligencia de Negocios? ¿Cuáles son algunos de los padecimientos que enfrentan las empresas hoy día?**

•**Tenemos datos pero carecemos de información**, es importante almacenar los datos de clientes, empleados, departamentos, compras, ventas, entre otros en aplicaciones, sistemas financieros o fuentes de datos. Si queremos que nuestra empresa tenga mayor ventaja sobre la competencia esta gestión no es suficiente.

•**Fragmentación**, poseen aplicaciones independientes a través de todos los departamentos pero se carece de una visión global de la empresa. Esto limita a la empresa a tomar decisiones importantes sin tener todos los elementos imprescindibles a la mano. Esta fragmentación conduce a lo que se llama diferentes versiones de la verdad. Los gerenciales solicitan informes a los distintos departamentos obteniendo diferentes resultados del mismo informe.

•**Manipulación de datos**, la necesidad de generar análisis de negocios e informes nos ha llevado a utilizar herramientas de BI y/o de reportes que no son las más confiables. Esta práctica conlleva la exportación de datos a distintas herramientas que resultan en un proceso lento, costoso, duplicación de trabajo, poca confiabilidad en los informes, propenso a errores y sujetos a la interpretación individual.

•**Poca agilidad**, debido a la carencia de información, la fragmentación y la manipulación de datos se mantiene en un nivel de rendimiento bajo. Como dice el dicho: “Justo cuando me aprendí las respuestas me cambiaron las preguntas...”. Necesitamos de una herramienta lo suficientemente ágil que se ajuste a las necesidades del negocio.

### **Beneficios de Inteligencia de Negocios**

Dentro del marco de beneficios que representa una solución de inteligencia de negocios podemos mencionar que esta nos permite:

•**Manejar el crecimiento**, el reto para las empresas es evolucionar, es crecer y esto significa “cambio”. Que tan ágiles son mis procesos para enfrentar los cambios y las necesidades puntuales de la empresa.

•**Control de costos**, el manejo de costos es el detonador que fuerza a muchas empresas considerar una solución de inteligencia de negocios, para tener la capacidad de medir gastos y ver esto a un nivel de detalle que identifique la línea de negocio, producto, centro de costo, entre otras.

•**Entender mejor los clientes**, las empresas almacenan grandes cantidades de información valiosa relacionada a sus clientes. El reto es transformar esta información en conocimiento y este conocimiento dirigido a una gestión comercial que represente algún tipo de ganancia para la empresa.

•**Indicadores de gestión**, los indicadores de desempeño permiten representar medidas enfocadas al desempeño organizacional con la capacidad de representar

la estrategia organizacional en objetivos, métricas, iniciativas y tareas dirigidas a un grupo y/o individuos en la organización.

### **Metodología para proyectos de Inteligencia de Negocio (BI)**

Para implementar una tecnología en un negocio, se requiere de una metodología. La mayoría de las consultoras especializadas en alguna tecnología cuentan, con por lo menos, una metodología, según los tipos de proyectos que aborden. Estos métodos son definidos a partir de sus experiencias y tomando lo mejor de los procedimientos más exitosos o populares. Contar con una metodología, se ha convertido tan importante y necesario como la carta de presentación de las empresas.

**Metodología CRISP-DM**, en sus primeros años de divulgación tenía apoyo de empresas privadas y organismos públicos, pero poco a poco ha ido perdiendo uno que otro “*Project Partner*”. Se desconoce el motivo de esta aparente pérdida de apoyo, pero se está seguro que no corresponde a la falta de calidad o efectividad del método, porque ha sido adoptado por otros organismos y empresas. A continuación se muestra el modelo de proceso CRISP-DM. Ver Gráfico 3.

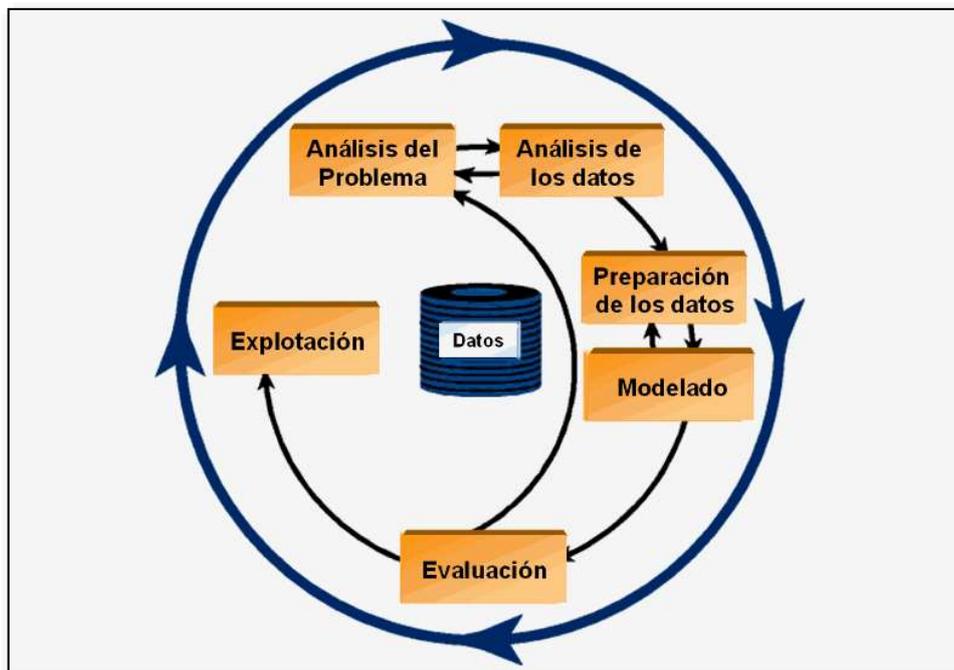


Gráfico 3: Modelo de proceso CRISP-DM

El estándar incluye un modelo y una guía, estructurados en seis fases, algunas de estas fases son bidireccionales, lo que significa que algunas fases permitirán revisar parcial o totalmente las fases anteriores. (Aníbal Goicochea, 2009)

**a. Comprensión del negocio** (Objetivos y requerimientos desde una perspectiva no técnica)

Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito)

Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio,...)

**b. Comprensión de los datos** (Familiarizarse con los datos teniendo presente los objetivos del negocio)

Recopilación inicial de datos

Descripción de los datos

Exploración de los datos

Verificación de calidad de datos

**c. Preparación de los datos** (Obtener la vista minable o dataset)

Selección de los datos

Limpieza de datos

Construcción de datos

Integración de datos

Formateo de datos

**d. Modelado** (Aplicar las técnicas de minería de datos a los dataset)

Selección de la técnica de modelado

Diseño de la evaluación

Construcción del modelo

Evaluación del modelo

**e. Evaluación de Resultados** (De los modelos de las fases anteriores para determinar si son útiles a las necesidades del negocio)

Evaluación de resultados

Revisar el proceso

Establecimiento de los siguientes pasos o acciones

**f. Despliegue** (Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización)

Planificación de despliegue

Planificación de la monitorización y del mantenimiento

Generación de informe final

Revisión del proyecto

### **Metodología de Kimball**

Ralf Kimball (1944) es considerado el inventor del Modelo Dimensional y pionero en Datawarehouse e Inteligencia de Negocios. Define un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un datawarehouse no era más que: "la unión de todos los Data Marts de una entidad". Defiende por tanto una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos.

La Metodología Kimball, es una metodología empleada para la construcción de un almacén de datos (datawarehouse, DW) no es más que, una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad.

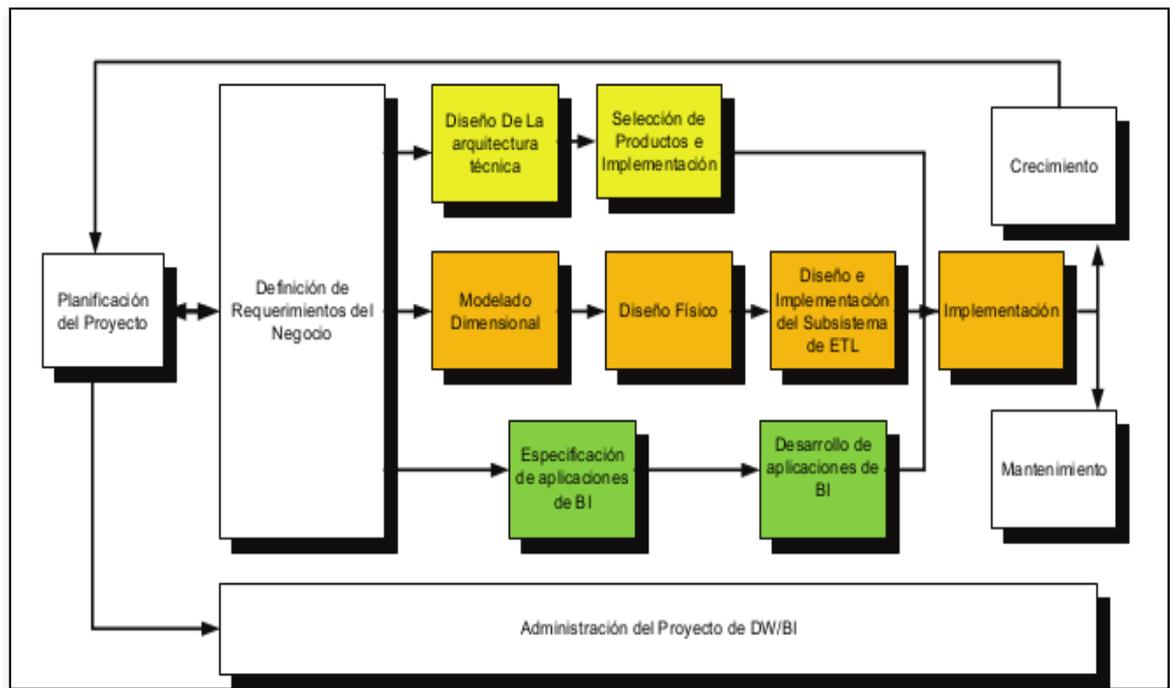


Gráfico 4: Tareas de la metodología de Kimball

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) (Kimball et al 98, 08, Mundy & Thornthwaite 06). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos, tal como se detalla en el Gráfico 4.

- **Centrarse en el negocio:** Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- **Construir una infraestructura de información adecuada:** Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de

aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.

- **Ofrecer la solución completa:** proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

### **Importancia De Inteligencia De Negocios En Las Organizaciones**

El exceso de información no es poder, pero el conocimiento si lo es. Con demasiada frecuencia, la transformación y el análisis de toda la información y los datos que las propias compañías generan se convierte en un verdadero problema y, por lo tanto, la toma de decisiones se vuelve desesperadamente lenta. Las tecnologías de BI intentan ayudar a las personas a entender los datos más rápidamente a fin de que puedan tomar mejores y más rápidas decisiones y, finalmente, mejorar sus movimientos hacia la consecución de objetivos de negocios.

## 2. DATAWAREHOUSE

### Definiciones

“Un datawarehouse es una colección de datos orientada a un dominio, integrado, no volátil y variable en el tiempo que ayuda en la toma de decisiones en una organización” (*Bill Inmon, 1992*)

Para este autor el datawarehouse es una parte del sistema de la inteligencia de negocios. Los datamarts se crean después de diseñar el DW y obtienen la información de éste, no siendo posible la consulta directa de información (se dice que la información no está almacenada de forma dimensional). Esta aproximación es conocida como “Top-down”.

*“Un datawarehouse es una copia de los datos de transaccionales especialmente estructurada para la consulta y el análisis” (Ralph Kimball, 1992)*

Para Kimball el datawarehouse es el conjunto de todos los data marts existentes en la empresa, aunque normalmente se almacenen de forma separada. Bajo esta concepción la información está almacenada en un modelo dimensional y por tanto está lista para ser consultada. Esta aproximación se conoce como “Bottom-up”, una metodología ascendente a la hora de diseñar un almacén de datos.

A lo largo de los años se han ido obteniendo definiciones más completas y detalladas:

“Un datawarehouse es un conjunto integrado de bases de datos que se diseña y utiliza para apoyar en la toma de decisiones y en él cada unidad de datos es relevante en algún instante de tiempo, además, contiene información no sólo de bases de datos relacionales, sino de otras fuentes relacionadas con la actividad de la organización y cuya finalidad no sólo se centra en el almacenamiento de esos datos, sino en su análisis y procesamiento mediante los procesos encargados de su gestión para la obtención de información estructurada y en definitiva útil para la toma de decisiones” (*Delgado, 1999*)

## Modelos de tablas

A la hora de diseñar un datawarehouse, uno de los elementos esenciales para su diseño es el modelo de tablas a escoger. Actualmente el modelo de tablas que normalmente se sigue en el diseño de un datawarehouse suele ser el modelo multidimensional propuesto por Kimball, ya que se ha impuesto con el tiempo al modelo relacional de Inmon, salvo algunas excepciones. El modelo a seguir se denomina esquema en estrella, el cual se muestra en el Gráfico 5.

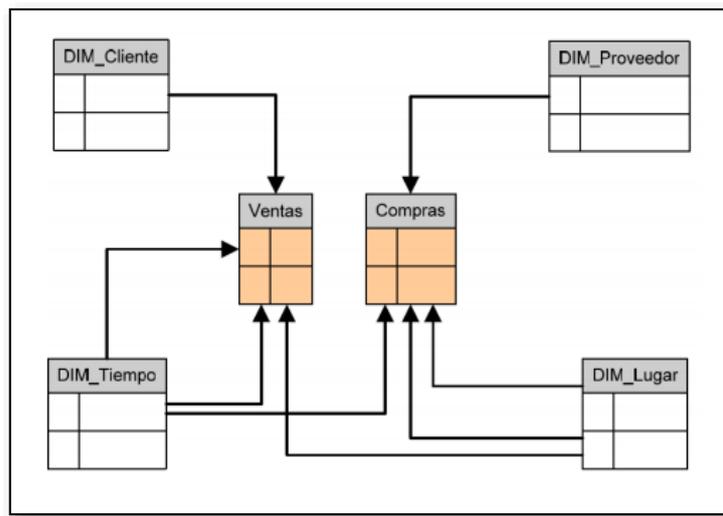


Gráfico 5: Ejemplo de modelo de datos en estrella

Este modelo permite el análisis de información a partir de una base de datos relacional, estructurando los datos y organizándolos de forma multidimensional.

Este modelo está compuesto de una serie de tablas que se clasifican en:

**Tablas de hechos:** Es la tabla central o nuclear (tablas en color naranja tal como se muestra en el Gráfico 5) de un esquema multidimensional. Esta tabla contiene los datos a medir, aquello que es cuantificable, datos que suelen ser las mediciones numéricas del negocio.

**Tablas de dimensiones:** Estas tablas (tablas en color blanco tal como se muestra en el Gráfico 5) se sitúan alrededor de la tabla de hechos, relacionándose con ella. Contienen información dimensional que permite filtrar, organizar y alimentar la información almacenada en la tabla de hechos.

### 3. PROCESO ETL (Extract, Transform and Load)

*Extract, Transform and Load* (Extraer, Transformar y Cargar). Los procesos ETL son los encargados de extraer datos de múltiples fuentes, darles formato y presentación, convertirlos en información útil y organizada, y cargarlos y almacenarlos en un almacén de datos o datamart para su posterior análisis a través de las herramientas expuestas anteriormente.

**a. Extraer:** Este paso se basa en extraer e integrar la información de diferentes fuentes (ERP, CRM, Excel) en el datawarehouse.

**b. Transformar:** En esta fase se ponen en práctica una serie de reglas de negocio para seleccionar únicamente la información necesaria para el datawarehouse. Utilizando técnicas de filtrado, manipulación de datos y cálculos evitaremos almacenar información no necesaria, redundante o errónea.

**c. Cargar:** En esta última fase los datos ya formateados, integrados y seleccionados se almacenan en el datawarehouse. Cabe destacar que esta carga no siempre se realiza de la misma manera, ya que hay organizaciones que optan por borrar todo el contenido del Datawarehouse y cargarlo de nuevo, y otras que optan por actualizar el datawarehouse únicamente con la información que ha llegado nueva.

La idea es que una aplicación ETL lea los datos primarios de unas bases de datos de sistemas principales, realice transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos son disponibles para analizar por los usuarios.

#### **Las más populares herramientas y aplicaciones ETL del mercado:**

- IBM Websphere DataStage
- Pentaho Data Integration (Kettle ETL) - Herramienta Open Source
- Oracle Warehouse Builder
- Cognos Decisionstream
- BusinessObjects Data Integrator (BODI)
- Microsoft SQL Server Integration Services (SSIS)

## 4. DATA MINING

### Conceptos

La minería de datos o exploración de datos es un campo de las ciencias de la computación, referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

La minería de datos es el proceso de detectar la información de grandes conjuntos de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

### Técnicas de Minería de Datos

Como ya se ha comentado, las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados. Las técnicas más representativas son:

- ✓ **Redes neuronales.**- Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales.
  
- ✓ **Regresión lineal.**- Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

- ✓ **Árboles de decisión.**- Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial y el análisis predictivo, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.
  
- ✓ **Modelos estadísticos.**- Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.
  
- ✓ **Agrupamiento o Clustering.**- Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.
  
- ✓ **Reglas de asociación.**- Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados (Weiss y Indurkha, 1998):

- ✓ **Algoritmos supervisados (o predictivos):** predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.
  
- ✓ **Algoritmos no supervisados (o del descubrimiento del conocimiento):** se descubren patrones y tendencias en los datos.

## **Tendencias**

La Minería de Datos ha sufrido transformaciones en los últimos años de acuerdo con cambios tecnológicos, de estrategias de marketing, la extensión de los modelos de compra en línea, etc. Los más importantes de ellos son:

- La importancia que han cobrado los datos no estructurados (texto, páginas de Internet, etc.).
- La necesidad de integrar los algoritmos y resultados obtenidos en sistemas operacionales, portales de Internet, etc.
- La exigencia de que los procesos funcionen prácticamente en línea (por ejemplo, en casos de fraude con una tarjeta de crédito).
- Los tiempos de respuesta. El gran volumen de datos que hay que procesar en muchos casos para obtener un modelo válido es un inconveniente; esto implica grandes cantidades de tiempo de proceso y hay problemas que requieren una respuesta en tiempo real.

## 5. ANALISIS PREDICTIVO

El Business Intelligence tradicional ha permitido a las organizaciones contar con potentes herramientas de reporting y query. El modo de tomar decisiones se transformó gracias a la estructuración de la información y a la perspectiva que la inteligencia de negocio proporcionaba. Sin embargo, generalmente, se trataba de soluciones que basaban su funcionamiento en la consulta de datos históricos.

Las soluciones de análisis predictivo van un paso más allá, complementando esta visión, ya que amplían las posibilidades de la inteligencia de negocio más extendida, otorgándole la capacidad de establecer posibles escenarios futuros en un análisis de "qué pasaría si...".

La capacidad de anticiparse al mercado es la característica fundamental de la ventaja competitiva y, para lograr este poder de previsión, se apoya principalmente en tres factores:

- Contar con la información necesaria.
- Disponer de la herramienta de análisis adecuada, respaldada por tecnología ad hoc.
- Actuar con el mayor margen de tiempo posible.

El Análisis predictivo utiliza estadística junto con algoritmos de minería de datos. Se basan en el análisis de los datos actuales e históricos para hacer predicciones sobre futuros eventos. Dichas predicciones raramente suelen ser afirmaciones absolutas, pareciéndose más a eventos y su probabilidad de que sucedan en el futuro.

En el mundo de los negocios los modelos predictivos explotan los patrones de comportamiento encontrados en el pasado para poder identificar riesgos y oportunidades. Los modelos capturan las relaciones entre muchos factores permitiendo capturar riesgos potenciales asociados a un conjunto de condiciones, guiando así a la toma de decisiones.

En la banca, típicamente, antes de conceder un crédito, préstamo o hipoteca, evalúan el perfil de riesgo de la persona usando un modelo de puntuación. Los modelos de puntuación tienen en cuenta el comportamiento histórico del cliente,

como puede ser el saldo de su cuenta bancaria a lo largo del tiempo, descubiertos, impagados, así como datos “estáticos” del cliente.

El análisis predictivo se utiliza en multitud de campos, aseguradoras, telecomunicaciones, agencias de viaje, farmacéuticas, médicas, etc.

## **Tipos De Análisis Predictivos**

Para llevar a cabo el análisis predictivo existen 3 modelos que pasaremos a describir:

### ✓ **Modelos predictivos:**

Los modelos predictivos analizan los resultados anteriores para evaluar qué probabilidad tiene un cliente para mostrar un comportamiento específico en el futuro con el fin de mejorar la eficacia de marketing. Esta categoría también incluye modelos que buscan patrones discriminatorios de datos para responder a las preguntas sobre el comportamiento del cliente, tales como la detección de tipos de fraudes.

### ✓ **Modelos descriptivos:**

Los modelos descriptivos describen las relaciones en los datos para poder clasificar a los clientes en grupos. A diferencia de modelos de predicción que se centran en predecir el comportamiento de un único cliente (como el riesgo de crédito), los Modelos descriptivos identifican diferentes relaciones entre clientes o productos. Pero los modelos descriptivos no clasifican a los clientes según su probabilidad de tomar una acción en particular.

### ✓ **Modelos de decisión:**

Los modelos de decisión describen la relación entre todos los elementos de una decisión y los datos conocidos (incluidos los resultados de los modelos de predicción), la decisión y el plan de variables y valores que determinan la decisión con el fin de predecir los resultados de las decisiones de muchas variables. Estos modelos pueden ser utilizados en optimización.

## **Minimización del riesgo y análisis predictivo**

Los métodos actuariales que permiten que una compañía de seguros lleve a cabo su negocio principal realizan exactamente la misma función que los modelos predictivos: clasificar a los clientes en función de la probabilidad de resultado positivo o negativo. Los modelos predictivos mejoran los métodos actuariales estándar, puesto que:

- Incorporan la automatización de análisis adicionales.
- Abarcan un conjunto más amplio de variables de cliente.

Para afianzar estas ventajas, muchas compañías de seguros están ampliando sus prácticas al integrar los análisis predictivos con el fin de mejorar sus decisiones en cuanto a selección y tarificación. Al igual que en el sector de las aseguradoras, todo negocio consiste, básicamente, en un ejercicio de gestión de riesgos.

Cualquier decisión que tome una empresa, cualquier paso que dé, repercute en los riesgos que ésta debe afrontar, tales como el abandono de un cliente, la no respuesta a una cara y atractiva publicidad, el gasto que supone un descuento de permanencia aun cuando no logre su fin, el de no ser seleccionado para una solicitud telefónica que hubiera comportado una venta, la comisión de fraude, convertirse en un "cliente perdido" al modo de un mal deudor o un titular de póliza de seguros con un elevado número de quejas.

## **Algoritmos De Análisis Predictivo**

El Análisis Predictivo (predictive analytics) requiere de un modelo predictivo pero enfrentarse a él es misión imposible sin el establecimiento de algoritmos. Su determinación no puede llevarse a cabo en cualquier momento, es preciso seguir un orden que puede estructurarse en torno a tres acciones:

1. Definir los objetivos del modelo predictivo.
2. Seleccionar los datos que se van a emplear.
3. Determinar los algoritmos a aplicar para el predictive analytics.

Llegados a este punto, es importante tener en cuenta que no todos los algoritmos usados en un modelo predictivo tienen la misma función. Dependiendo de ella se pueden clasificar en tres grandes grupos:

- Algoritmos desarrollados para resolver problemas de negocio específicos.
- Algoritmos cuya meta es mejorar los algoritmos existentes.
- Algoritmos que proveen de nuevas capacidades.

Conocer las diferencias entre unos y otros facilita la tarea de escoger el que mejor se adapta a los propósitos de la organización. Aunque, siempre, lo más importante es tener claro los objetivos de negocio a alcanzar, entre los que podrían enumerarse los siguientes, por ser los más representativos:

- **Segmentación de clientes:** en estos casos es necesario utilizar algoritmos de agrupamiento, que también permiten llevar a cabo la detección de grupos o categorías en el ámbito social.
- **Retención de clientes:** lo idónea para conseguir hacer realidad este propósito es aplicar algoritmos de clasificación. Estos algoritmos pueden destinarse a un doble uso, como es el desarrollo de un sistema de recomendación, muy extendido en el sector del comercio minorista.
- **Predicción de eventos:** los algoritmos de regresión cumplen esta misión, que en todo caso va asociada al factor tiempo, que permite adquirir un mayor control.

Los algoritmos que intervienen en los procesos de análisis predictivo son los responsables de que las empresas puedan disfrutar del grado de conocimiento necesario para tomar acciones precisas y minimizando el riesgo. Algunos ejemplos son:

**a. Algoritmos de agrupamiento:** aplicar la función de aglutinar a los diferentes clientes determina:

- El modo de orientar una campaña de marketing.

- El tiempo de mantener una oferta o una promoción específica.
- La forma de evitar el fraude.

Se identifica el Algoritmo de Clustering.

**b. Algoritmos de clasificación:** gracias a ellos se puede conocer mejor los hábitos y comportamientos de un determinado cliente tipo. De esta forma, es posible:

- Mejorar el servicio ofrecido a un tipo de cliente, por ejemplo a los más fieles o a los más asiduos.
- Diseñar ofertas personalizadas que satisfagan las necesidades de los clientes, fortaleciendo el vínculo existente entre ellos y la marca o el producto.

Entre los algoritmos usados dentro de este grupo están: Algoritmo Neural Network Algoritmo Naive Bayes y Algoritmo de Arboles de Decisión

**c. Algoritmos de regresión:** la predicción en estado puro se materializa en pronósticos fiables basados en datos objetivos que hacen posible:

- Optimizar la gestión de stocks.
- Prever la demanda para una campaña concreta.
- Determinar la inversión más rentable.

Identificar las variables, contar con una visión en detalle, construir modelos apropiados para cubrir las necesidades del negocio (ya sean de consulta, de escenarios simulados, árboles de decisión o un análisis matemáticos avanzados) son cuestiones vitales que dependen de algoritmos. Son fruto de la combinación entre arte y ciencia, data science en estado puro que aumenta sus posibilidades de éxito al conjugar experiencia, habilidad y conocimiento del negocio.

Se identifica el Algoritmo de Regresión Lineal y Algoritmo de Regresión Logística.

## **6. ARBOLES DE DECISION**

Los árboles de decisión son un modelo predictivo que construye diagramas de construcciones lógicas a partir de una base de datos. Esta es una técnica utilizada por la minería de datos para el análisis de datos con el fin de tomar decisiones convenientes.

Los árboles de decisiones son representaciones gráficas y analíticas de datos introducidas por medio de una base de datos ya establecidas. Estos ayuda en la toma de decisiones, desde un punto de vista probabilístico con el fin de elegir la opción más conveniente, los arboles de decisiones se caracterizan por ser una técnica utilizada por la minería de datos, donde se busca solucionar problemas de predicción clasificación y segmentación, una vez analizados todos los datos.

### **Características de un árbol de decisión**

- Plantea el problema desde distintas perspectivas de acción.
- Permite analizar de manera completa todas las posibles soluciones. Provee de un esquema para cuantificar el costo del resultado y su probabilidad de uso.
- Ayuda a realizar las mejores decisiones con base a la información existente y a las mejores suposiciones.
- Su estructura permite analizar las alternativas, los eventos, las probabilidades y los resultados.

### **Cómo hacer un árbol de decisión**

#### **Paso 1. Identifica todas las variables del problema central**

Una vez que sepas cuál es tu problemática, comienza a enumerar todos los factores que la componen. Por ejemplo: La contratación de un nuevo proveedor. Enlista: costos, trámites, tiempo de llegada del producto, daños mercancía, fechas de pago, etc.

## Paso 2. Priorizar

Es importante comenzar a limitar tus criterios de decisión, es por ello que de la lista anterior separes cuáles son aquellos factores que son primordiales para tu empresa. Si de la lista de 15, puedes limitarlo a 7, quiere decir que estos criterios con mayor relevancia.

## Paso 3. Identifica la importancia de los criterios

Se trata de buscar y enumerar qué de estos siete criterios son más importantes, enumera. Por ejemplo: en el cambio de proveedor. El costo estaría como número uno, en segundo lugar el tiempo de llegada del producto, en tercero los trámites, y así sucesivamente.

## Paso 4. Comienza a ver variables

Cada uno de estos factores debe contar con fortalezas o debilidades, asegúrate de comenzar la ramificación apartando tres o cuatro alternativas para cada criterio. Lo que buscamos es generar supuestos de manera objetiva, si de los criterios salieron cuatro ramificaciones y de esas cuatro salieron 12 más, esto te ayudará a profundizar en lo más conveniente. Tal como se detalla en el Gráfico 6, en este ejemplo donde se evalúa el costo beneficio de proveedores.

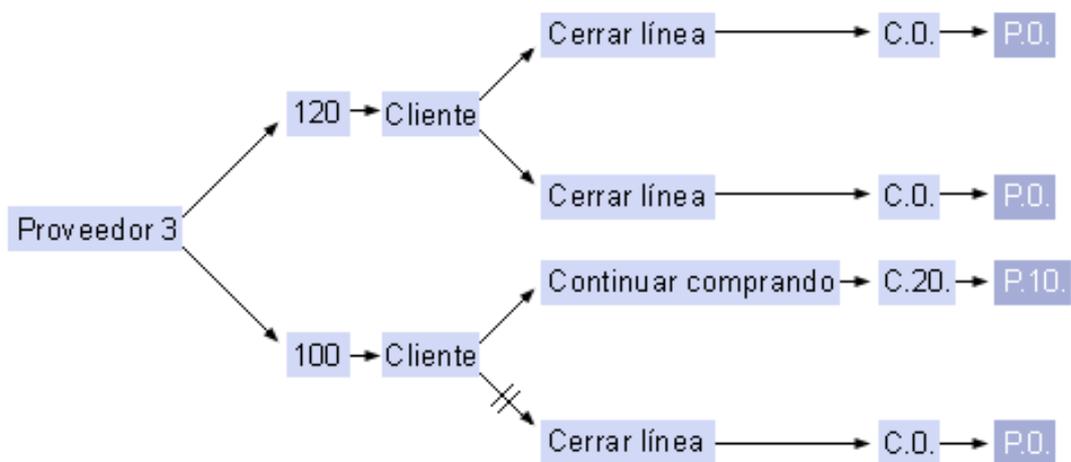


Gráfico 6: Ejemplo de Árbol de Decisión Costo - Beneficio de Proveedores

### **Paso 5. Selecciona una alternativa**

Analiza cuál es la opción más conveniente de acuerdo al árbol de decisiones, siempre toma en cuenta la importancia de los criterios y cada una de sus alternativas.

### **Paso 6. Implementa la alternativa**

Recuerda, realizar un análisis es muy distinto a implementar soluciones, cuida que se siga paso a paso cada detalle, una pequeña falla podría dar resultados no deseados y esto perjudicaría todo el proceso de selección.

### **Paso 7. Evalúa la efectividad de la decisión**

Siempre es bueno saber qué se hizo bien y qué se hizo mal, analiza con tu equipo cada paso en la evaluación de alternativas, esto ayudará a mejorar tomas de decisiones a futuro.

En el presente trabajo, se utilizó el Algoritmo de Árboles de Decisión de Microsoft, que viene como parte de las técnicas de minerías de datos dentro de la herramienta de SQL Analysis Services.

### **Algoritmo de Árboles de Decisión de Microsoft**

Es un algoritmo de clasificación y regresión para el modelado de predicción de atributos discretos y continuos. Para los atributos discretos, el algoritmo hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores, conocidos como estados, de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Específicamente, el algoritmo identifica las columnas de entrada que se correlacionan con la columna de predicción.

Para los atributos continuos, el algoritmo usa la regresión lineal para determinar dónde se divide un árbol de decisión.

Si se define más de una columna como elemento de predicción, o si los datos de entrada contienen una tabla anidada que se haya establecido como elemento de predicción, el algoritmo genera un árbol de decisión independiente para cada columna de predicción.

## ¿Cómo funciona el algoritmo de árbol de decisión?

El algoritmo de árboles de decisión genera un modelo de minería de datos mediante la creación de una serie de divisiones en el árbol. Estas divisiones se representan como *nodos*. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción. La forma en que el algoritmo determina una división varía en función de si predice una columna continua o una columna discreta.

## Datos requeridos para los modelos de árboles de decisión

Los requisitos para un modelo de árboles de decisión son los siguientes:

- **Una columna key:** cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.
- **Una columna de predicción.** Se requiere al menos una columna de predicción. Puede incluir varios atributos de predicción en un modelo y pueden ser de tipos diferentes, numérico o discreto. Sin embargo, el incremento del número de atributos de predicción puede aumentar el tiempo de procesamiento.
- **Columnas de entrada.** Se requieren columnas de entrada, que pueden ser discretas o continuas. Aumentar el número de atributos de entrada afecta al tiempo de procesamiento.

El algoritmo de árboles de decisión de Microsoft utiliza la selección de características para guiar la selección de los atributos más útiles. Todos los algoritmos de minería de datos de Analysis Services utilizan la selección de características para mejorar el rendimiento y la calidad del análisis.

## Predecir columnas discretas

La forma en que el algoritmo de árboles de decisión de Microsoft genera un árbol para una columna de predicción discreta puede mostrarse mediante un histograma. El Gráfico 7, se muestra un histograma que traza una columna de predicción, Bike Buyers, según una columna de entrada, Age. El histograma muestra que la edad de una persona ayuda a distinguir si esa persona comprará una bicicleta.

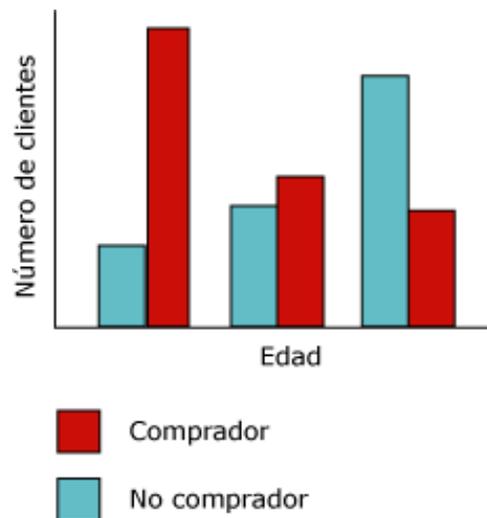


Gráfico 7: Histograma de Predicción

La correlación que aparece en el diagrama del Gráfico 8, hará que el algoritmo de árboles de decisión de Microsoft cree un nuevo nodo en el modelo.

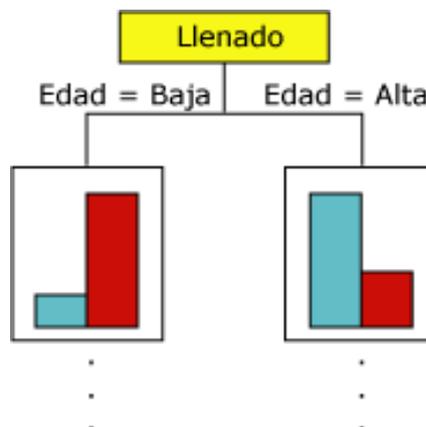


Gráfico 8: Correlación del Algoritmo

A medida que el algoritmo agrega nuevos nodos a un modelo, se forma una estructura en árbol. El nodo superior del árbol describe el desglose de la columna de predicción para la población global de clientes. A medida que el modelo crece, el algoritmo considera todas las columnas.

### Predecir columnas continuas

Cuando el algoritmo de árboles de decisión de Microsoft genera un árbol basándose en una columna de predicción continua, cada nodo contiene una fórmula de regresión. Se produce una división en un punto de no linealidad de la fórmula de regresión. Por ejemplo, considere el siguiente diagrama del Gráfico 9.

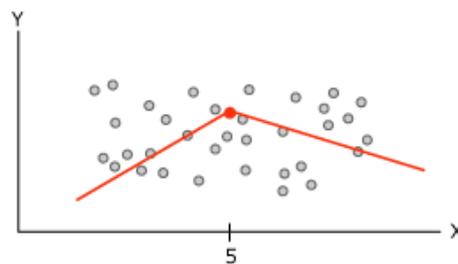


Gráfico 9: Diagrama de Regresión

El diagrama contiene los datos que pueden modelarse utilizando una sola línea o dos líneas conectadas. Sin embargo, una sola línea realizará un pobre trabajo en la representación de los datos. En su lugar, si se usan dos líneas, el modelo hará un mejor trabajo en la aproximación a los datos. El punto donde las dos líneas se unen es el punto de no linealidad y donde se dividiría un nodo de un modelo de árbol de decisión. Por ejemplo, el nodo que corresponde al punto de no linealidad del gráfico anterior podría representarse mediante el siguiente diagrama. Las dos ecuaciones representan las ecuaciones de regresión de las dos líneas, tal como se detalla en el Gráfico 10.

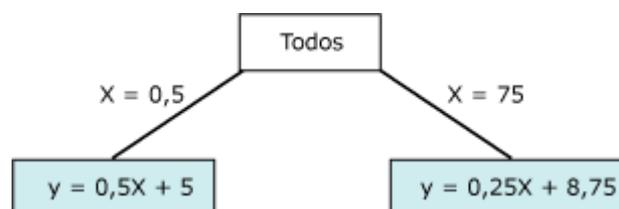


Gráfico 10: Ecuación de Regresión

## **Realizar predicciones mediante un modelo de árbol de decisión**

Dado que los árboles de decisión se pueden utilizar para diversas tareas, como son la clasificación, la regresión e incluso la asociación, al crear una consulta de predicción en un modelo de árbol de decisión se tienen muchas opciones disponibles. Se debe conocer el propósito para el que se creó el modelo si desea entender los resultados de la predicción. Los ejemplos de consulta siguientes ilustran tres escenarios diferentes:

Devolver una predicción para un modelo de clasificación, junto con la probabilidad de que la predicción sea correcta, y después filtrar los resultados por probabilidad;  
Crear una consulta singleton para predecir asociaciones.

Recuperar la fórmula de regresión para una parte de un árbol de decisión en el que la relación entre la entrada y la salida es lineal.

## Lista de funciones de predicción

Todos los algoritmos de Microsoft son compatibles con un conjunto común de funciones. Sin embargo, el algoritmo de árboles de decisión de Microsoft admite las funciones adicionales que se enumeran en la Tabla 1.

Función de predicción	Uso
<a href="#">IsDescendant (DMX)</a>	Determina si un nodo es un elemento secundario de otro nodo del modelo.
<a href="#">IsInNode (DMX)</a>	Indica si el nodo especificado contiene el caso actual.
<a href="#">PredictAdjustedProbability (DMX)</a>	Devuelve la probabilidad ponderada.
<a href="#">PredictAssociation (DMX)</a>	Predice los miembros de un conjunto de datos asociativo.
<a href="#">PredictHistogram (DMX)</a>	Devuelve una tabla de valores relacionados con el valor de predicción actual.
<a href="#">PredictNodeID (DMX)</a>	Devuelve el Node_ID de cada caso.
<a href="#">PredictProbability (DMX)</a>	Devuelve la probabilidad del valor de predicción.
<a href="#">PredictStdev (DMX)</a>	Devuelve la desviación estándar predicha para la columna especificada.
<a href="#">PredictSupport (DMX)</a>	Devuelve el valor de soporte de un estado especificado.
<a href="#">PredictVariance (DMX)</a>	Devuelve la varianza de una columna especificada.

Tabla 1 Funciones de Predicción

## Herramientas de Business Intelligence en el mercado

En el ambiente de los negocios se ha empezado a considerar las aplicaciones de Inteligencia de negocios como una herramienta fundamental para la toma de decisiones. Para llevar esto a cabo es necesario el uso de una plataforma que soporte la extracción y explotación de datos. Dentro del ámbito peruano se encuentran disponibles herramientas de extracción de datos; entre los principales están:

**Business Objects.**- Cuenta con la herramienta Business Objects Data Integrator, la cual es una plataforma de integración de datos que puede abarcar a toda la organización. Permite explorar, extraer, transformar y distribuir datos para que los Usuarios puedan disponer de información precisa, puntual y fiable. [Bus 2009]

**Microsoft SQL Server.**- Cuenta con una plataforma SSIS (SQL Server Integration Services, Servicios de integración de servidores); esta plataforma permite generar Soluciones de integración de datos de alto rendimiento, entre las que se incluyen Paquetes de extracción, transformación y carga para el almacenamiento de datos. [Mic 2009]

**IBM DataStage.**- Es una herramienta de IBM que permite crear y mantener fácil y rápidamente data marts y Datawarehouses.

**Pentaho Kettle** – programa ETL que incluye un conjunto de herramientas para realizar la extracción y transformación de datos. Uno de sus objetivos es que el proyecto ETL sea fácil de generar, mantener y desplegar. [Pen 2010]

Así también se tiene diversas herramientas de explotación tales como:

**Oracle Discoverer.**- Es la herramienta de análisis y consulta de informes que proporciona un acceso inmediato a la información del Datawarehouse o data mart. La información en Oracle Discoverer se organiza de un modo similar que en Microsoft Excel. [Ora 2009]

**SAS-** Ofrece la herramienta de explotación, predice y optimiza su cadena entera del servicio para transformar operaciones after-sales de la ayuda de centros de costo desconectados en centros de beneficio integrados. El sistema de la solución permite a organizaciones ganar una ventaja competitiva conservando a clientes satisfechos y mejorando imagen de marca de fábrica. [Sas 2009]

**Cognos PowerPlay** – Cognos en el mercado BI presenta una amplia gama de funcionalidades orientadas al usuario, una mayor flexibilidad y mejoras técnicas. Esta nueva versión optimiza las capacidades de Cognos Enterprise Business Intelligence Series que integra a PowerPlay, Cognos ReporNet y Metrics Manager acercando los datos críticos de negocio a una extensa variedad de usuarios. Posee un rendimiento de misión crítica que incluye análisis drag-and-drop (arrastrar y soltar), además de estar integrado estrechamente con las últimas funcionalidades para cuadros de mando y reporting. [Cog 2009]

**IBM–SurfAid** – Se emplea para atender a las demandas de gran complejidad y volumen para el rápido crecimiento del web site. Algunas herramientas y métricas adicionales incluyen acceso al banco de datos relacional, permitiendo consultas ad hoc e informes dinámicos. El SurfAid para análisis incluye acceso a los consultores para asistencia en la búsqueda de métodos de extracción de valor avanzado en base a los datos del cliente. [IBM 2009]

## 7. METODOLOGIA RUP

Es una metodología cuyo fin es entregar un producto de software. Se estructura todos los procesos y se mide la eficiencia de la organización. Es un proceso de desarrollo de software el cual utiliza el lenguaje unificado de modelado UML, constituye la metodología estándar más utilizada para el análisis, implementación y documentación de sistemas orientados a objetos. El RUP es un conjunto de metodologías adaptables al contexto y necesidades de cada organización. Describe cómo aplicar enfoques para el desarrollo del software, llevando a cabo unos pasos para su realización. Se centra en la producción y mantenimiento de modelos del sistema.

El ciclo de vida RUP es una implementación del Desarrollo en espiral. Fue creado ensamblando los elementos en secuencias semi-ordenadas. El ciclo de vida organiza las tareas en fases e iteraciones.

RUP divide el proceso en cuatro fases, dentro de las cuales se realizan varias iteraciones en número variable según el proyecto y en las que se hace un mayor o menor hincapié en las distintas actividades.

### **Fases del ciclo de vida del RUP:**

1. **Fase de Inicio:** Esta fase tiene como propósito definir y acordar el alcance del proyecto con los patrocinadores, identificar los riesgos asociados al proyecto, proponer una visión muy general de la arquitectura de software y producir el plan de las fases y el de iteraciones posteriores.

2. **Fase de elaboración:** En la fase de elaboración se seleccionan los casos de uso que permiten definir la arquitectura base del sistema y se desarrollaran en esta fase, se realiza la especificación de los casos de uso seleccionados y el primer análisis del dominio del problema, se diseña la solución preliminar.

3. **Fase de Desarrollo:** El propósito de esta fase es completar la funcionalidad del sistema, para ello se deben clarificar los requerimientos pendientes,

administrar los cambios de acuerdo a las evaluaciones realizados por los usuarios y se realizan las mejoras para el proyecto.

**4. Fase de Cierre:** El propósito de esta fase es asegurar que el software esté disponible para los usuarios finales, ajustar los errores y defectos encontrados en las pruebas de aceptación, capacitar a los usuarios y proveer el soporte técnico necesario. Se debe verificar que el producto cumpla con las especificaciones entregadas por las personas involucradas en el proyecto.

### **Disciplina de desarrollo de RUP**

Determina las etapas a realizar durante el proyecto de creación del software.

**Ingeniería o modelado del negocio:** Analizar y entender las necesidades del negocio para el cual se está desarrollando el software.

**Requisitos:** Proveer una base para estimar los costos y tiempo de desarrollo del sistema.

**Análisis y diseño:** Trasladar los requisitos analizados anteriormente a un sistema automatizado y desarrollar una arquitectura para el sistema.

**Implementación:** Crear software que se ajuste a la arquitectura diseñada y que tenga el comportamiento deseado.

**Pruebas:** Asegurarse de que el comportamiento requerido es correcto y que todo lo solicitado está presente.

**Despliegue:** Producir distribuciones del producto y distribuirlo a los usuarios.

## Conclusiones

En la bibliografía consultada, podemos concluir que el ambiente del mundo de los negocios de hoy exige una aplicación cada vez más eficiente de la información disponible. Business Intelligence, genera un conocimiento al negocio, que se deriva de la correcta utilización de la información generada dentro y fuera de la empresa.

Son múltiples los beneficios que ofrece a las empresas, entre ellos se encuentra la generación de una ventaja competitiva. Hay una gran variedad de soluciones de BI que en suma, son muy similares, pero para que se considere completa debe reunir cuatro componentes: multidimensionalidad, data mining, agentes y Datawarehouse.

Así mismo, el análisis predictivo proporciona abundantes oportunidades para la evolución de la empresa. Se trata de minimizar los riesgos y optimizar estrategias, si bien son pocas las organizaciones que recurren a ella en sus estrategias de marketing, es una tendencia que va ganando adeptos poco a poco. Las empresas deben ser conscientes del gran valor que tienen en su información y mediante estas técnicas pueden evitar que esos errores en la toma de decisiones no se vuelvan a repetir.

Para complementar, el árbol de decisión es una de las técnicas de Data Mining más utilizada en todo el mundo. Es sumamente útil en las áreas de negocios de las principales compañías. Otra ventaja de los árboles de decisión es que las opciones posibles de una determinada condición son excluyentes. Esto permite que luego de analizar la situación se pueda tomar una acción o decisión específica.

Finalmente todas estas herramientas sirven de apoyo para generar valor y rentabilidad a la empresa generando conocimiento en base a su propia información de una manera apropiada y estructurada.

## **CAPITULO II: DESARROLLO DEL PROYECTO**

En esta etapa se definen las actividades, alcance incluyendo recursos y tiempo. Desde la etapa de planificación del proyecto se establece una relación con los miembros del área de marketing, quienes interactúan en conjunto con el presente equipo de tesis permitiendo el desarrollo del plan de proyecto y su seguimiento. Es la integración de una serie de procedimientos y actividades haciendo uso de una metodología definida que permita lograr los objetivos y metas de la manera más eficiente y efectiva.

### **1. Alcance del Proyecto**

Este proyecto abarcará los siguientes aspectos:

- ✓ Se implementó un sistema de información que permitirá obtener reportes por Línea, Familia y Canal de Venta.
- ✓ La herramienta analizó la información sólo de la Base de Datos de Cliente Preferencial.
- ✓ Como herramienta de implementación para el modelo predictivo se usó SQL Server Analysis Services 2008.

### **Alcance del Producto**

- ✓ Se tomó como fuente de datos, la base de datos transaccional de ventas (Sistema Facturador) en la cual todas las tiendas registran las ventas, esta base de datos contiene información tal como : Clientes, productos, tiendas, ubicación de la tienda, etc.
- ✓ Se construyó procedimientos almacenados (stored procedures) que se encargarán del proceso ETL (extraer la información, esta será formateada o transformada y se cargará en el datawarehouse)
- ✓ Se construyó el modelo de datos en base al mapeo de datos de la BD Transaccional (fuente u origen), este modelo de datos será el Datawarehouse (destino).

- ✓ Se implementó la herramienta SQL Server Analysis Services 2008 la cual se modeló de acuerdo a la fuente del datawarehouse, **sobre esta herramienta se aplicará el algoritmo de árboles de decisión.**
- ✓ Como output, de acuerdo al requerimiento del usuario, se obtiene información para la toma de decisiones, reportes de predicción que apoyen en las estimaciones de venta y compra, y a su vez estimen las probabilidades de venta al cliente preferencial por línea, familia y por canal de venta mediante el análisis de su patrón o hábitos de consumo.

### **Criterios de Aceptación del Producto**

Se acordó con el Gerente de Marketing, los siguientes criterios bajo los cuales el sistema cumpla con las especificaciones exigidas:

- El Sistema debe contar con una arquitectura compuesta: servidor de Base de Datos, servidor de aplicaciones y clientes finales.
- El sistema deberá funcionar sobre el motor de base de datos SQL Server.
- El Sistema debe usar como herramienta de diseño a SQL Server 2008 Analysis Services
- Garantizar la confiabilidad, seguridad y desempeño del sistema a los diferentes usuarios. La información almacenada podrá ser consultada y actualizada permanentemente, sin que se afecte el tiempo de respuesta.
- El sistema realizará los procesos de extracción (base de datos transaccional), transformación (en las tablas trabajo dimensionales) y carga (tablas de hechos en el datawarehouse).
- El Sistema ayudará al gerente de marketing y jefes de línea a la toma de decisiones de compra de mercadería, al ejecutar la estructura de minería de datos del cliente frecuente, línea y familia de producto.

- El sistema permitirá conectarse con Microsoft Excel mediante add-in o complementos de minería de datos que puedan facilitar la interfaz gráfica, una mejor usabilidad y preparación de información.
- El sistema podrá emitir dos reportes: Reporte por Canal de Ventas y Reporte por Línea y familia con sus respectivas leyendas de minería de datos donde describe las probabilidades existentes.
- El sistema permitirá generar nuevas vistas y estructuras de minería de datos con nuevas variables de entradas y predicción, por ejemplo: por periodos de tiempo, por ubicación, sexo, etc.

### **Estructura de Desglose del Trabajo y Entregables**

A continuación en la Tabla 2, se detalla el EDT o Estructura de Desglose de Trabajo, conocido por sus siglas en inglés como WBS (Work Breakdown Structure), mostramos el listado de todas las actividades que comprenden el alcance del proyecto, necesarias para realizar la planificación y efectuar la ejecución y control.

Asimismo, se muestra en el Gráfico 11 el Diagrama de Estructura de Desglose de Trabajo.

Etapa	Actividad	Descripción	Referencia – Ubicación
<b>Inicio del Proyecto</b>	Definición del problema y objetivo	En esta actividad se describe los antecedentes del problema, la definición y objetivos generales y específicos.	Definición del Problema y Propósito del Proyecto. Página 4 y 5.
<b>Inicio del Proyecto</b>	Elaboración de la estructura detallada	Se detalla y grafica cada actividad del proyecto y entregables.	Página 42 y 43
<b>Inicio del Proyecto</b>	Elaboración del cronograma Gantt	Detalla las actividades del proyecto, horas asignadas y recursos.	Página 45
<b>Análisis y Diseño del Proyecto</b>	Análisis de Requerimiento Funcionales y No Funcionales	En esta fase se definen las características y funciones que tendrá la aplicación, para determinar las necesidades del usuario.	Página 50
<b>Análisis y Diseño del Proyecto</b>	Diseño del Modelo de datos ( DWH )	Diagramas de las relaciones entre las entidades.	Página 57 y 58
<b>Análisis y Diseño del Proyecto</b>	Diseño del proceso de ETL	Proceso de movimientos y transformación de datos para la integración de datos.	Diagrama de Paquete, Paquete Proceso ETL, Casos de Uso del Sistema. Página 60
<b>Construcción del Proyecto</b>	Documento Arquitectura del Sistema	Detalla la estructura del sistema que incluyen elementos del software, las propiedades visibles de estos elementos y la relación entre ellos, así como la interacción del cliente con esta.	Diagrama de Arquitectura del Software. Página 53
<b>Construcción del Proyecto</b>	Configuración Origen y Vista de Datos en Herramienta Analysis Services.	Detalla las configuraciones de la conexión con el datawarehouse y la creación de la vista con los atributos para la explotación de información.	Especificación de los casos de uso. Página 102
<b>Construcción del Proyecto</b>	Configuración del Algoritmo de Árbol de Decisión en Herramienta Analysis Services.	Detalla la configuración de la técnica de minería de datos Árbol de Decisión de Microsoft y los campos de entrada y predicción	Especificación de los casos de uso. Página 102
<b>Construcción del Proyecto</b>	Configuración de Reportes	Configuración de la estructura de minería de datos.	Página 118
<b>Construcción del Proyecto</b>	Req. Aceptación del producto	Se detalla los requerimientos de aceptación por parte del usuario.	Página 41
<b>Implementación del Proyecto</b>	Presentación y Análisis del Producto	Detalla los análisis de los reportes de Canal de Venta y Línea y Línea y Familia.	Especificación de los casos de uso. Página 116 y 123
<b>Implementación del Proyecto</b>	Recomendaciones del Proyecto	Se detalla las recomendaciones finales del proyecto.	Página 133
<b>Implementación del Proyecto</b>	Conclusiones del Proyecto	Se detalla las conclusiones del proyecto.	Página 132

Tabla 2 Descripción de Estructura de Desglose de Trabajo

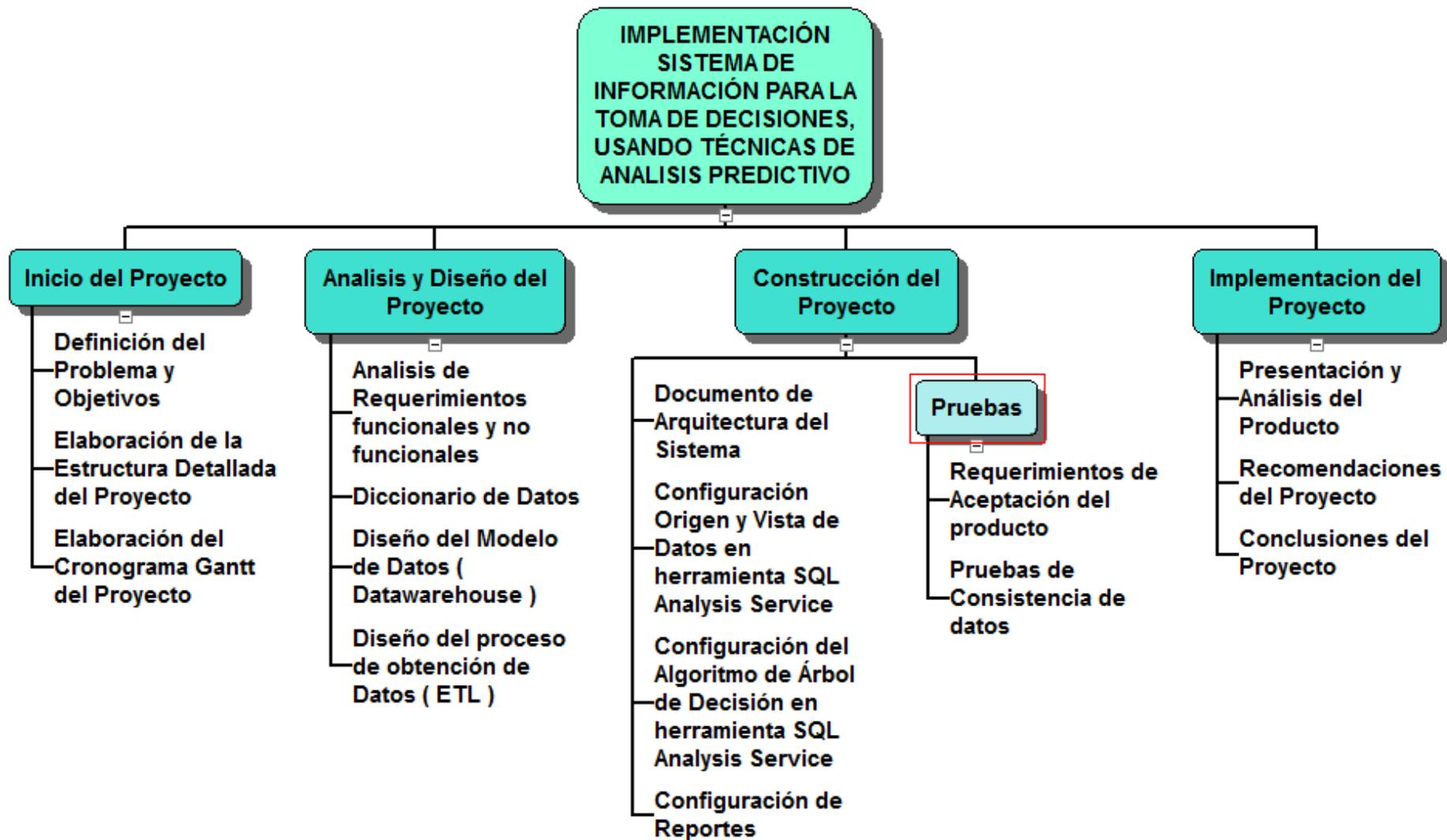


Gráfico 11: Diagrama de Estructura de Desglose de Trabajo

El cronograma Gantt del Proyecto, nos ha permitido organizar el tiempo, asignación de recursos y actividades para el desarrollo del proyecto. A continuación se muestra el cronograma en el Gráfico 12.

	Nombre de tarea	Duración	Trabajo	Comienzo	Fin	% completado	Nombres de los recursos
1	▶ <b>IMPLEMENTACIÓN SISTEMA DE INFORMACIÓN PARA LA TOMA DE DECISIONES, USANDO TÉCNICAS DE ANALISIS PREDICTIVO</b>	98 días	800.5 horas	lun 08/06/15	vie 23/10/15	100%	
2	▶ <b>INICIO DEL PROYECTO</b>	23 días	11 horas	lun 08/06/15	mié 08/07/15	100%	
3	Definición del Problema y Objetivos	0.75 días	3 horas	lun 08/06/15	lun 08/06/15	100%	Bertha Espinoza[0%],Natalia Gutierrez[0%]
4	Estructura detallada del Proyecto ( EDT )	1 día	2 horas	lun 06/07/15	lun 06/07/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
5	Cronograma Gantt del Proyecto	3 días	6 horas	lun 06/07/15	mié 08/07/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
6	▶ <b>ANALISIS y DISEÑO DEL PROYECTO</b>	28 días	352 horas	mar 21/07/15	lun 31/08/15	100%	
7	Análisis de Requerimientos funcionales y no funcionales	5 días	48 horas	lun 03/08/15	vie 07/08/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
8	Diccionario de Datos	5 días	40 horas	lun 03/08/15	vie 07/08/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
9	Diseño del Modelo de Datos ( Datawarehouse )	11 días	44 horas	lun 03/08/15	lun 17/08/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
10	Diseño del proceso de obtención de Datos ( ETL )	11 días	44 horas	lun 17/08/15	lun 31/08/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
11	▶ <b>Elaboración de Casos de Uso del Sistema</b>	20 días	176 horas	lun 24/08/15	vie 18/09/15	100%	
12	▷ Extraer Información	9 días	56 horas	lun 24/08/15	jue 03/09/15	100%	
16	▷ Transformar Información	5 días	60 horas	lun 07/09/15	vie 11/09/15	100%	
20	▷ Cargar Información	20 días	60 horas	lun 24/08/15	vie 18/09/15	100%	
24	▶ <b>CONSTRUCCION DEL PROYECTO</b>	32.88 días	357.5 horas	lun 17/08/15	mié 30/09/15	100%	
25	Documento de Arquitectura del Sistema	4 días	16 horas	mar 01/09/15	vie 04/09/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
26	Configuración Origen y Vista de datos en la herramienta SQL Analysis Service	4 días	16 horas	mié 16/09/15	lun 21/09/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
27	Configuración del Algoritmo de Árbol de Decisión	7 días	56 horas	mié 16/09/15	jue 24/09/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
28	▶ <b>Elaboración de Casos de Uso del Sistema</b>	13 días	260 horas	lun 14/09/15	mié 30/09/15	100%	
29	▷ Generar Clasificación de Clientes	7 días	188 horas	lun 14/09/15	mar 22/09/15	100%	
33	▷ Consultar Reporte	3 días	72 horas	lun 28/09/15	mié 30/09/15	100%	
37	▶ <b>PRUEBAS</b>	2 días	9.5 horas	lun 17/08/15	mar 18/08/15	100%	
38	Criterios de Aceptación del producto	2 días	9.5 horas	lun 17/08/15	mar 18/08/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
39	▶ <b>IMPLEMENTACION DEL PROYECTO</b>	10 días	80 horas	lun 12/10/15	vie 23/10/15	100%	
40	Presentación del proyecto	6 días	48 horas	lun 12/10/15	dom 18/10/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
41	Recomendaciones del Proyecto	2 días	16 horas	lun 19/10/15	mar 20/10/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]
42	Conclusiones del Proyecto	2 días	16 horas	lun 19/10/15	mar 20/10/15	100%	Bertha Espinoza[50%],Natalia Gutierrez[50%]

DIAGRAMA DE GANTT

Gráfico 12 Cronograma Gantt del Proyecto

## **Exclusiones del proyecto**

- No incluyó dentro del alcance del proyecto toda la base de datos de Clientes, sólo se tomó la base de datos de Clientes Preferenciales.
- No incluyó dentro del alcance del proyecto, la compra del software SQL Server 2008 donde se implementó el modelo de predicción.

## **Restricciones del Proyecto**

- La investigación estuvo financiada íntegramente por las investigadoras, ya que no cuentan con financiamiento externo alguno a pesar de la envergadura del trabajo.
- Elaborado dentro del cronograma establecido de 5 meses.

## **Supuestos del Proyecto**

- En el desarrollo del proyecto se contó con un asesor de tesis, para las consultas que se tenga de dicho proyecto.
- Apoyo de los stakeholder y Sponsor.
- Comunicación constante vía telefónica y correo con los participantes del proyecto.
- Se contó con el apoyo de la empresa.
- La empresa si tiene la capacidad financiera para implementar herramientas de BI.
- El proyecto es viable financieramente.
- La empresa cuenta dentro de su área de Sistemas con profesionales idóneos para la implantación del proyecto.
- Satisfacer las necesidades del cliente.
- La implementación generará nuevos proyectos BI.

## 2. Modelado del Negocio

El modelo de negocio de una empresa es una representación simplificada de la lógica de negocio. Describe lo que un negocio ofrece a sus clientes, cómo llega a ellos, y cómo se relaciona con ellos, y en fin como la empresa gana dinero. Para facilitar esta tarea.

### Actores del Negocio

A continuación se muestra el Actor del Negocio. Ver gráfico 13.



Gráfico 13: Actor del Negocio

### Descripción de los Actores del Negocio

**Gerente de Marketing**, actor principal del proceso del negocio, encargado de establecer estrategias y toma de decisiones en el proceso de estimación de ventas (línea, familia y producto) y compras de mercadería.

### Diagrama de Casos de Uso del Negocio

A continuación se muestra el caso de Uso del Negocio. Ver Gráfico 14.

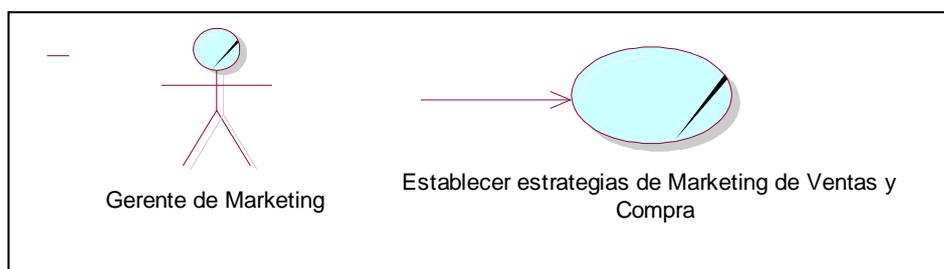
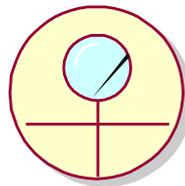


Gráfico 14: Diagrama de Casos de Uso del Negocio

## Trabajadores del Negocio

A continuación se muestra en el Gráfico 15, el trabajador del Negocio.



Jefe de Línea

*Gráfico 15: Trabajador del Negocio*

### Descripción de los Trabajadores del Negocio

**Jefe de Línea**, encargada de analizar información y generar reportes y cuadros estadísticos relacionado a la línea que tiene asignada, (Línea de Bisutería, Línea de Complemento de Vestir, Línea de Accesorios de Cabello, Línea de Cosmética, etc.), así mismo se encarga de realizar las compras de mercadería en China y/u otros países.

### Diagrama de Proceso del Negocio AS IS

A continuación, se detalla la descripción del Proceso del Negocio AS IS (“tal como está”) en la Tabla 3 y el diagrama de procesos del Negocio. Ver Gráfico 16.

Actividad	Descripción	Input	Output
<b>1</b>	En el área de Marketing, las estimaciones de compras por cada línea (bisutería, accesorio de cabello, complemento de vestir, etc.) y análisis de venta por línea y producto, son revisadas por la Gerente de Marketing y Jefes de Línea, en base a reportes que son extraídas del Sistema Transaccional de Ventas.	Sistema Transaccional de Ventas Información Histórica	Reportes Excel de Ventas por Línea, Familia, producto y unidades
<b>2</b>	Con esta información los Jefes de Línea inician el proceso de analizar la información y generar cuadros estadísticos de acuerdo a la necesidad del área para ser enviados a la Gerente de Marketing.		
<b>3</b>	La Gerente de Marketing realiza el análisis de esta información para las estimaciones de venta y compra y planeamiento de ofertas/desc.	Reportes Excel de Ventas por Línea, Familia, producto y unidades	
<b>3b</b>	En base a esta información la Gerente de Marketing obtiene: El plan de estimado de Ventas por Línea, Familia, Canal de Venta.		Reportes Excel Estimación de Ventas por Línea y Familia, Canal de Venta
<b>3b1</b>	Una vez realizado el plan de ventas, la Gerente de Marketing realiza las estimaciones de compra y se arma una matriz.	Plan Estimado de Ventas	Reportes Excel Matriz de Compra
<b>3b2</b>	Aprobada esta matriz, junto con el presupuesto de compra, es distribuida a cada Jefe de Línea para que sirva de apoyo y seguimiento en la compras de mercadería que deben realizar en los viajes hacia China o compras locales.		Reportes Excel Matriz de Compra
<b>3a</b>	En base a los cuadros estadísticos, la Gerente de Marketing realiza “supuestos” sobre el hábito de consumo de los clientes y sus preferencias, de esta manera planificar ofertas/descuentos o promociones.	Sistema Transaccional de Ventas Información Histórica Reporte de Ventas por cliente	Programa de descuentos/ofertas y promociones.

*Tabla 3 Descripción del Modelo del Proceso de Negocio AS IS*

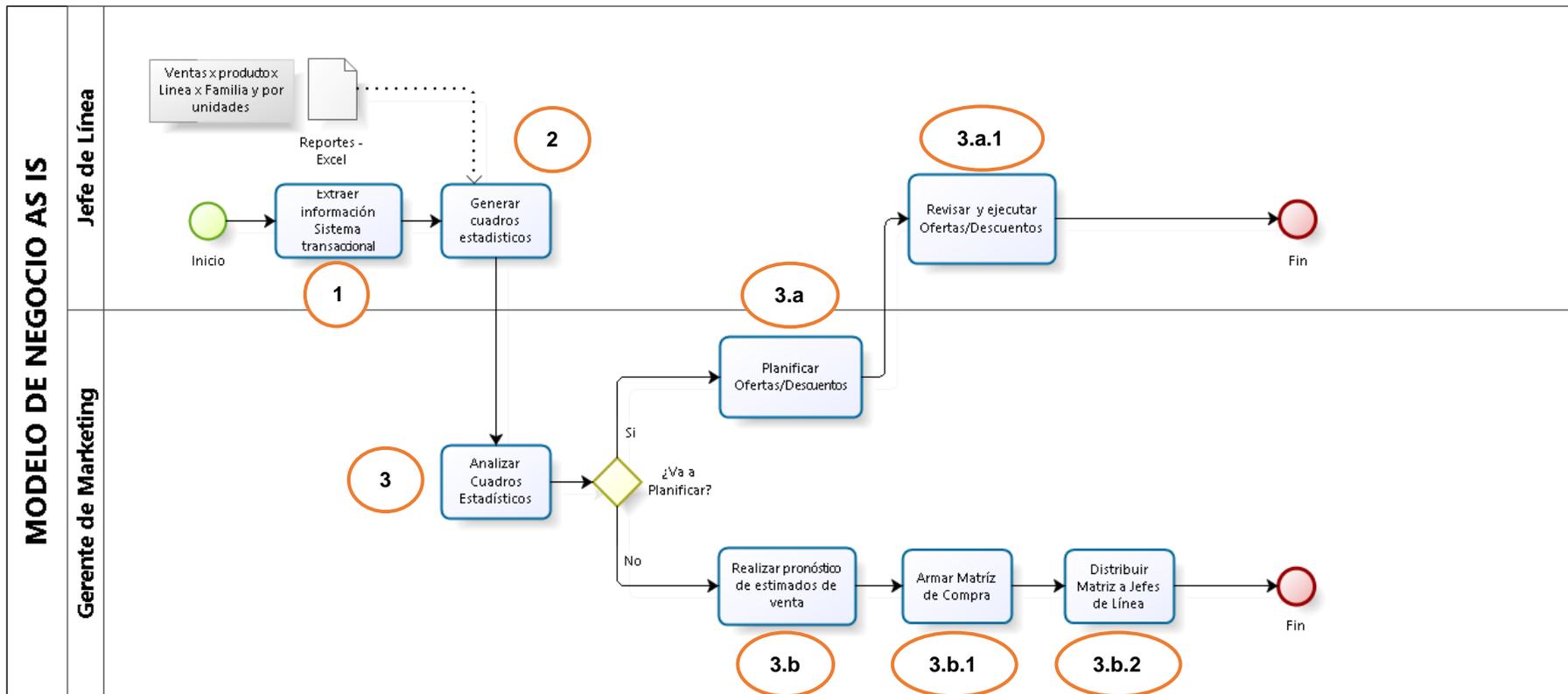


Gráfico 16: Diagrama de procesos del negocio AS IS

A continuación se detalla la descripción del Proceso del Negocio TO BE en la Tabla 4 y el diagrama de procesos .Ver Gráfico 17.

Actividad	Descripción	Input	Output
<b>1</b>	En el área de Marketing, las estimaciones de compras por cada línea (bisutería, accesorio de cabello, complemento de vestir, etc.) y análisis de venta por línea y producto son revisadas por la Gerente de Marketing y Jefes de Línea en base a reportes que son extraídas del Sistema Transaccional de Ventas.	Sistema Transaccional de Ventas Información Histórica	Reportes Excel de Ventas por Línea, Familia, producto y unidades
<b>2</b>	Con esta información las Jefes de Línea inician el proceso de analizar la información y generar cuadros estadísticos de acuerdo a la necesidad del área.		
<b>3</b>	La Gerente de Marketing realiza el análisis de esta información para las estimaciones de venta y compra y planeamiento de ofertas/desc.	Reportes Excel de Ventas por Línea, Familia, producto y unidades	
<b>3b</b>	Con la herramienta de Análisis Predictivo la Gerente de Marketing y Jefes de Línea, generan los modelos o reportes de predicción que servirán de apoyo para realizar: El plan de estimados de Ventas por Línea, Familia, Canal de Venta.	SQL Analysis Services	Reportes Excel Estimación de Ventas por Línea y Familia, Canal de Venta
<b>3b1</b>	Una vez realizado el plan de ventas, la Gerente de Marketing realiza las estimaciones de compra y se arma una matriz.	Plan Estimado de Ventas	Reportes Excel Matriz de Compra
<b>3b2</b>	La Gerente de Marketing aprueba esta matriz, junto con el presupuesto de compra, es distribuida a cada Jefe de Línea para que sirva de apoyo y seguimiento en la compras de mercadería que deben realizar en los viajes hacia China o compras locales.		Reportes Excel Matriz de Compra
<b>3a</b>	Con la herramienta de Análisis Predictivo, la Gerente de Marketing y Jefes de Línea generan los modelos o reportes de predicción que servirá de apoyo para realizar: La planificación de ofertas/descuentos y poder analizar el hábito de compras de los clientes.	SQL Analysis Services	Programa de descuentos/ofertas y promociones.
<b>3a1</b>	Las Jefes de Línea revisan la planificación de los descuentos y/u ofertas y proceden a ejecutarla informando al área de Sistemas para su configuración y a Ventas.		

Tabla 4 Descripción del Modelo de Proceso de Negocio TO BE

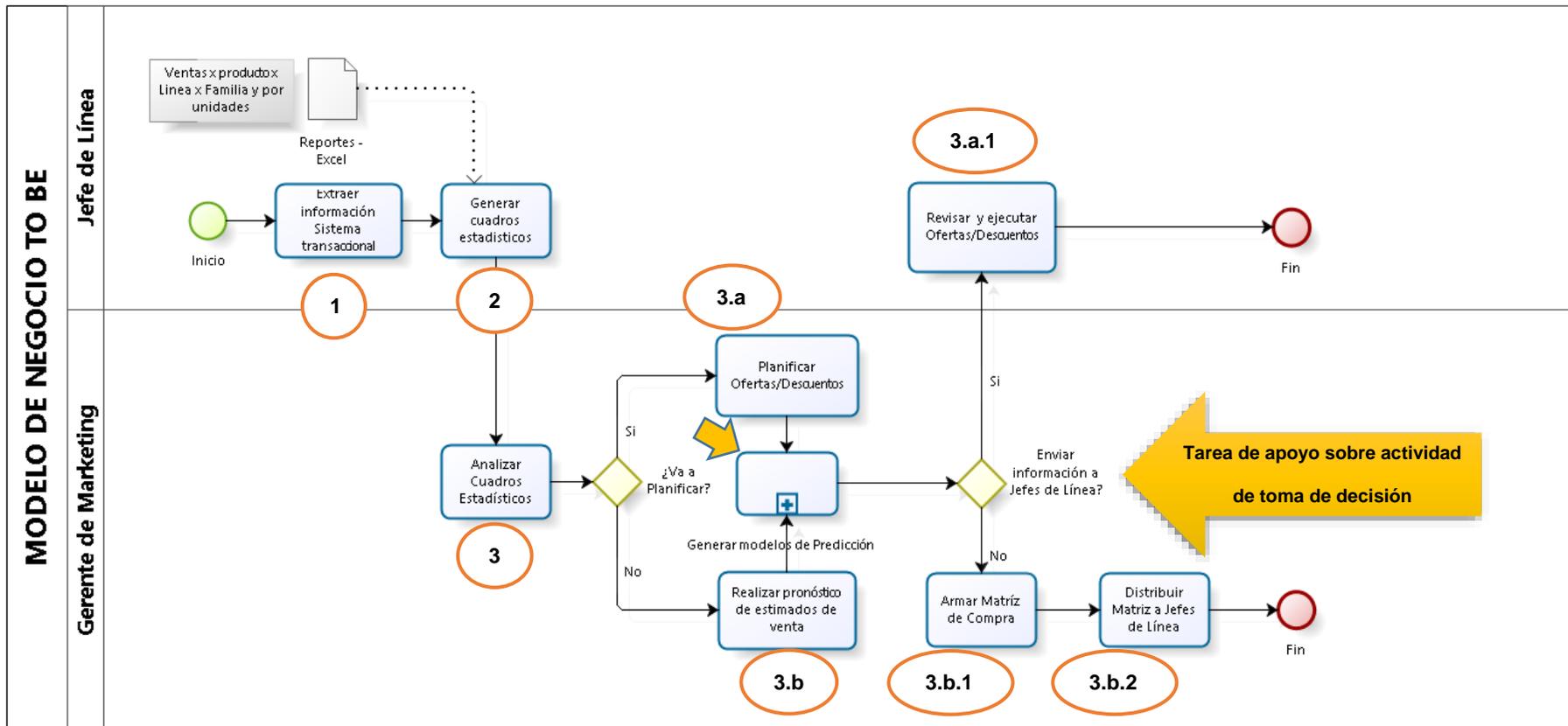


Gráfico 17: Diagrama de proceso del negocio TO BE

### **3. Requerimientos de Software/Producto**

En esta fase se definen las características y funciones que tendrá la aplicación, para determinar las necesidades del usuario.

#### **Requerimientos Funcionales**

Los requerimientos funcionales considerados para este proyecto son:

- La herramienta permitirá generar nuevas vistas y estructuras de minería de datos con nuevas variables de entrada y predicción, por ejemplo:  
Por períodos de tiempo, por ubicación, sexo, etc.
- El sistema debe ser intuitivo y fácil de usar, que capte el interés de cualquier usuario y que permita a los usuarios finales generar sus propios informes y análisis.

#### **Requerimientos No Funcionales**

##### **Desempeño:**

- Garantizar la confiabilidad, seguridad y desempeño del sistema a los diferentes usuarios. La información almacenada podrá ser consultada y actualizada permanente, sin que se afecte el tiempo de respuesta.
- El sistema debe estar en capacidad de dar respuesta al acceso de todos los usuarios con tiempo de respuesta aceptable y uniforme, en períodos de alta, media y baja demanda de uso del sistema.
- La herramienta permite conectarse con Microsoft Excel mediante add-in o complementos de minería de datos que pueden facilitar la interfaz gráfica, una mejor usabilidad y preparación de información.

##### **Disponibilidad:**

- Estar disponible en un 99.99%, las 24 horas del día.

**Escabilidad**

- El sistema debe permitir implementar nuevas funcionalidades y/o requerimientos relacionados con nuevas consultas, vistas o reportes, sin que este se vea afectado de la menor manera posible.

**Software**

- La aplicación deberá funcionar sobre el motor de base de datos SQL Server.
- Si existe la posibilidad de adquirir licencias de software.

#### 4. Arquitectura de Software/Productos

La arquitectura de software es la estructura o estructuras del sistema que incluyen elementos del software, las propiedades externamente visibles de estos elementos y la relación entre ellos, así como la interacción del cliente con esta. A continuación se detalla los elementos del diagrama de arquitectura del sistema, visualizado en el Gráfico 18:



Gráfico 18: Diagrama de Arquitectura del Sistema

**Fuentes de información**, son todos aquellos datos que reposan en diferentes fuentes o bases de datos, estos datos son almacenados por diferentes usuarios que manejan distintos sistemas y pueden encontrarse en diferentes ubicaciones, para este proyecto la fuente de información es la Base de Datos transaccional del Sistema de Ventas de la empresa, el cual se encuentra en SQL Server 2008.

**Extracción de la información (proceso ETL)**, esta etapa se realiza a partir de la extracción y transformación de la información de diferentes fuentes de datos (principalmente data histórica presente en los sistemas transaccionales de la organización), limpiarlos y cargarlos en otra base de datos o datawarehouse para analizarlo en otro sistema operacional. La herramienta utilizada para realizar este

proceso SQL Server Business Intelligence Development Studio, mediante el componente SQL Server Integration Services.

**Datawarehouse o repositorio de datos**, es la base de datos que contiene las estructuras del modelo dimensional diseñado, esta base de datos se alimenta de la carga de información ejecutada de los procesos ETL, se maneja en SQL Server 2008.

Mediante el SQL Server Analysis Services nos permitió crear, diseñar y administrar las estructuras de minerías de datos que se construyeron a partir de los datos agregados del origen de datos ( datawarehouse) mediante el uso de algoritmos de minería de datos, para este caso : Árbol de Decisión.

**Inteligencia de Negocios**, mediante los reportes o consultas propias de la herramienta del SQL Server Business Intelligence, se ilustra toda la información de una forma vistosa a los usuarios, esta es la interfaz gráfica que el usuario tiene a través de distintos dispositivos y de esta manera apoyar en la toma de decisiones en base al análisis de información.

### **Diagrama de Despliegue**

Un diagrama de despliegue modela la arquitectura en tiempo de ejecución de un sistema, se utiliza para modelar el hardware utilizando en la implementación del sistema y las relaciones entre sus componentes.

A continuación presentamos el diagrama de despliegue sobre el cual se ejecutará nuestra aplicación. Ver Gráfico 19.

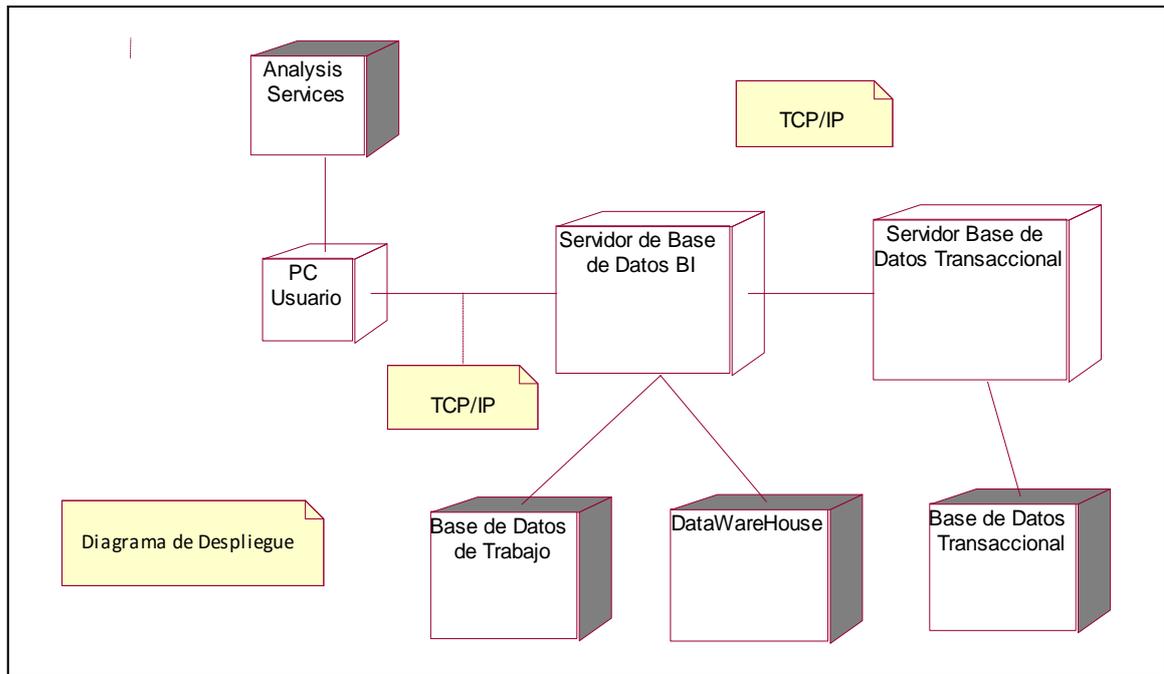


Gráfico 19: Diagrama de despliegue

Para completar el despliegue del sistema físicamente será necesario la participación de los siguientes elementos en la red:

**TCP/IP:** La arquitectura es cliente-servidor por lo que la conexión a la red será mediante este protocolo.

**Pc del cliente:** Este elemento servirá para hacer uso de la aplicación.

**Servidor de Base de Datos BI:** Servidor donde se ubicará la base de datos de la aplicación del SQL Analysis Services.

**Servidor de Base de Datos Transaccional:** Servidor donde se ubica la base de datos de Ventas y maestros, del cual se alimentará el datawarehouse.

### Diagrama de Componentes

Un diagrama de componentes muestra los elementos de un diseño de un sistema de software, permite visualizar la estructura de alto nivel del sistema y el comportamiento del servicio que estos componentes proporcionan y usan a través de interfaces. A continuación se presenta el diagrama de componentes de la aplicación. Ver Gráfico 20.

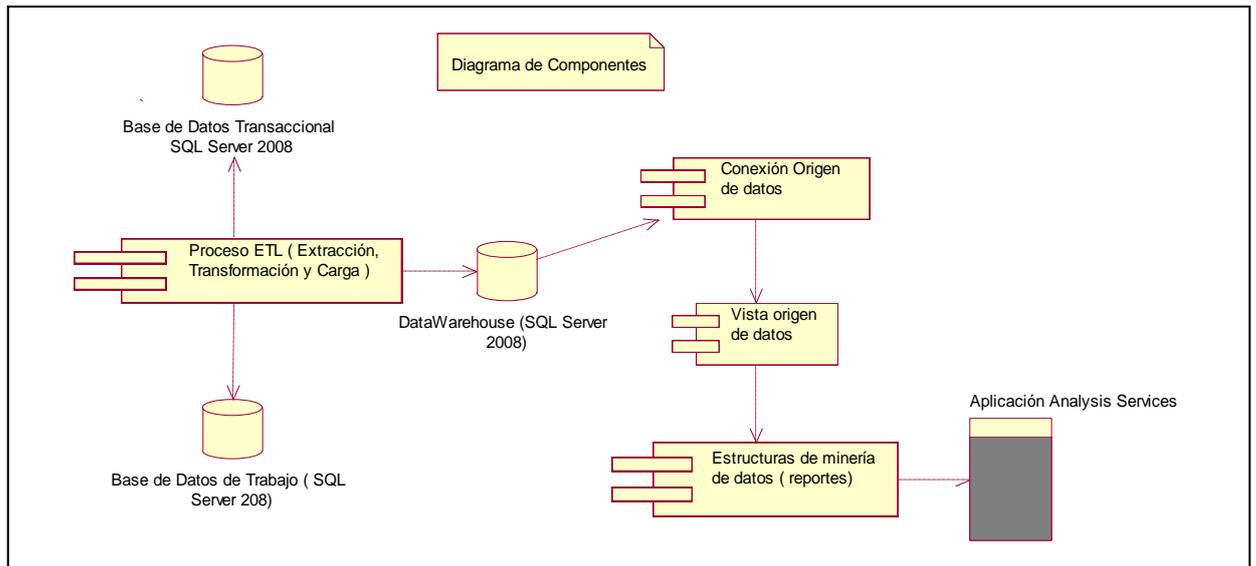


Gráfico 20: Diagrama de componentes

## Descripción del Diagrama de Componentes

- **Proceso ETL:** Representa el proceso de movimientos y transformación de datos para la integración de datos.
- **Conexión Origen de datos:** Representa la conexión al datawarehouse desde la aplicación del Analysis Services.
- **Vista origen de datos:** Representa el conjunto de componentes que contiene el modelo lógico para la estructura de minería de datos.
- **Estructura de minería de datos:** Representa el conjunto de consultas o reportes que permiten el acceso a los datos.
- **Aplicación Analysis Services:** Es el programa residente del usuario mediante el cual se puede acceder a la información.

## Diccionario de datos

El diccionario de datos es un listado organizado de todos los datos pertinentes al sistema, contiene las características lógicas, incluyendo el nombre, descripción, campos, tipo, etc. A continuación se muestra en la Tabla 5.

Tabla 5 Diccionario de Datos

Tabla	Campo	Tipo	Long	Descripción
DM_TIENDA	ptovta	int		Código punto de venta
DM_TIENDA	nompvt	varchar	10	Nombre del punto de venta
DM_TIENDA	codest	int		Código Canal de Venta
DM_TIENDA	nombre	varchar	100	Descripción Canal de Venta
DM_TIEMPO	time_id	int		Código de fecha
DM_TIEMPO	the_date	datetime		fecha
DM_TIEMPO	the_day	varchar	15	Día
DM_TIEMPO	the_month	varchar	15	Mes
DM_TIEMPO	the_year	int		Año
DM_TIEMPO	day_of_month	int		Día del Mes
DM_TIEMPO	week_of_year	float		Semana del año
DM_TIEMPO	month_of_year	int		Mes del año
DM_TIEMPO	quarter	varchar	2	Trimestre
DM_TIEMPO	Fiscal_period	varchar	255	Periodo contable
DM_CLIENTE_FRECUENTE	clifre	char	20	DNI Cliente
DM_CLIENTE_FRECUENTE	rzsfre	varchar	80	Nombres Cliente
DM_CLIENTE_FRECUENTE	edad	int		Edad
DM_CLIENTE_FRECUENTE	ubigeo	int		Código de Ubicación
DM_CLIENTE_FRECUENTE	sexo	varchar	50	Sexo
DM_CLIENTE_FRECUENTE	ocupacion	varchar	80	Ocupación
DM_UBIGEO	ubigeo	int		Código Ubicación Geográfica
DM_UBIGEO	codpais	char	2	Código país
DM_UBIGEO	despais	varchar	15	Descripción país
DM_UBIGEO	coddpto	char	2	Código Departamento
DM_UBIGEO	desdpto	varchar	25	Descripción Departamento
DM_UBIGEO	codprov	char	2	Código provincia
DM_UBIGEO	desprov	varchar	25	Descripción provincia
DM_UBIGEO	coddist	char	2	Código distrito
DM_UBIGEO	desdist	varchar	30	Descripción distrito
DM_ARTICULO	codalt	char	7	Código del producto
DM_ARTICULO	delart	varchar	60	Descripción del producto
DM_ARTICULO	codlin	char	2	Código de línea
DM_ARTICULO	delin	varchar	30	Descripción de línea
DM_ARTICULO	codfam	char	2	Código de familia

DM_ARTICULO	delfam	varchar	30	Descripción de familia
DM_ARTICULO	codsuf	char	3	Código de subfamilia
DM_ARTICULO	delsuf	varchar	30	Descripción de subfamilia
DM_ARTICULO	codbol	char	2	Código de color
DM_ARTICULO	delcol	varchar	30	Descripción de color
DM_ARTICULO	pais	char	20	Pais
DM_ARTICULO	codtip	char	2	Tipo de articulo
DM_ARTICULO	destip	char	40	Descripción tipo de articulo
DM_ARTICULO	coduneg	char	2	Código unidad de negocio
DM_ARTICULO	desuneg	char	40	Descripción unidad de nego
FC_MARKETING	ptovta	int		Código punto de venta
FC_MARKETING	clifre	char	20	DNI Cliente
FC_MARKETING	codalt	char	7	Código del producto
FC_MARKETING	time_id	int		Código de fecha
FC_MARKETING	unidades_venta	money		Unidades de venta
FC_MARKETING	valor_venta	money		Valor Venta
FC_MARKETING	importe_iva	money		Importe del IGV
FC_MARKETING	Importe_total	money		Importe Total + IGV
FC_MARKETING	costo_venta	money		Costo de Venta
FC_MARKETING	Importe_descuento	money		Importe con descuento
FC_MARKETING	Valor_venta_me	money		Valor Venta Moneda Extranjera
FC_MARKETING	importe_iva_me	money		Importe IGV moneda extranjera
FC_MARKETING	importe_total_me	money		Importe total moneda extranjera
FC_MARKETING	costo_venta_me	money		Costo venta moneda extranjera
FC_MARKETING	Importe_descuento	money		Importe descuento moneda extr.

Tabla 5 Diccionario de Datos (continuación)

### Modelo Conceptual (Diseño Lógico)

Este nivel representa el esquema lógico de los datos, reflejando su estructura y relaciones, sin entrar en detalles físicos. A continuación se muestra el modelo Conceptual. Ver Gráfico 21.

### Modelo de Datos (Diseño Físico)

Este nivel debe ser transparente para el usuario. En este nivel especifica la estructura de los datos así como el modo de almacenamiento empleado. A continuación se muestra el modelo Conceptual. Ver Gráfico 22.

## Modelo Conceptual (Diseño Lógico)

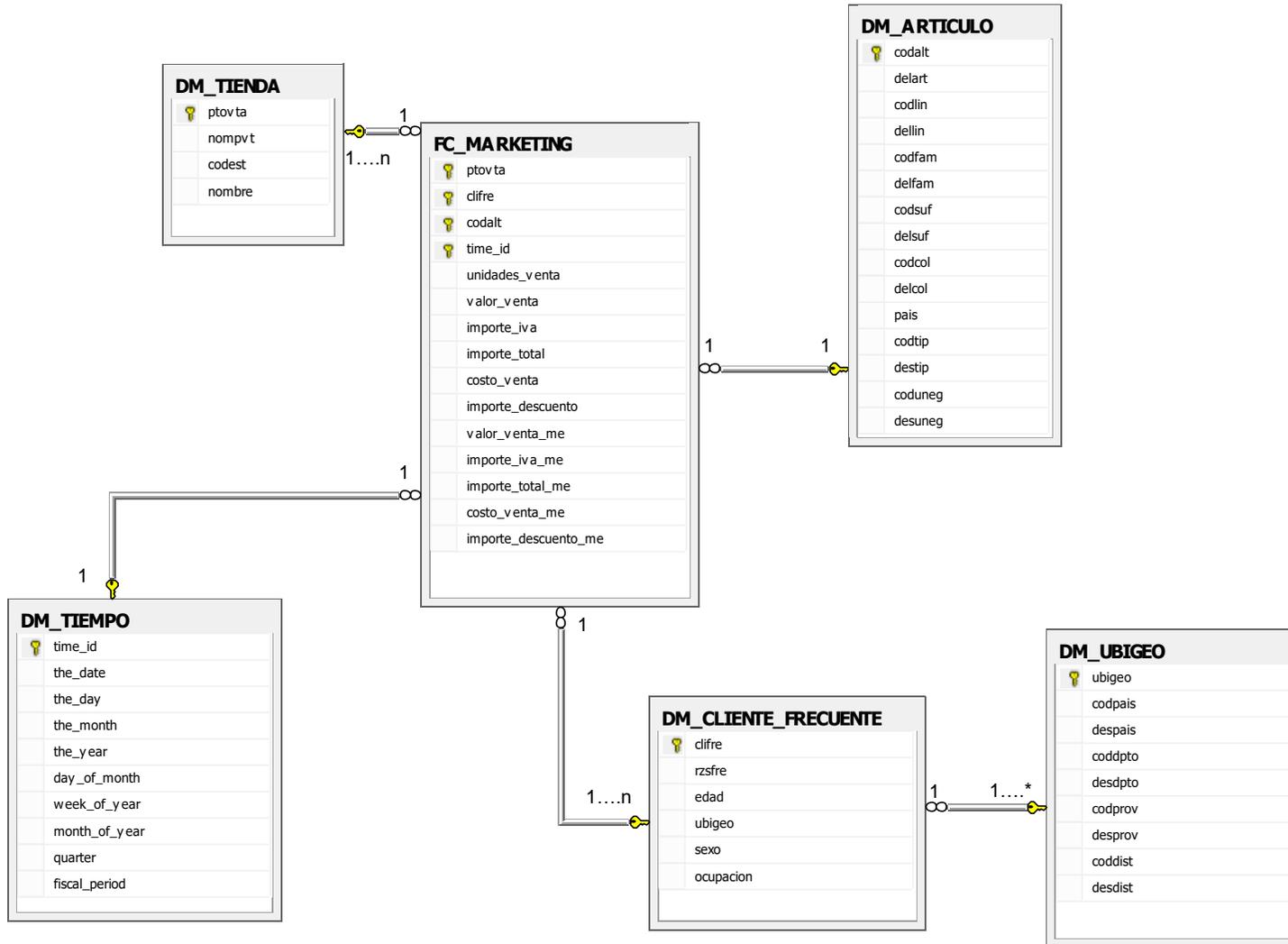


Gráfico 21: Diagrama de clases (Diseño Lógico)

## Modelo de Datos (Diseño Físico)

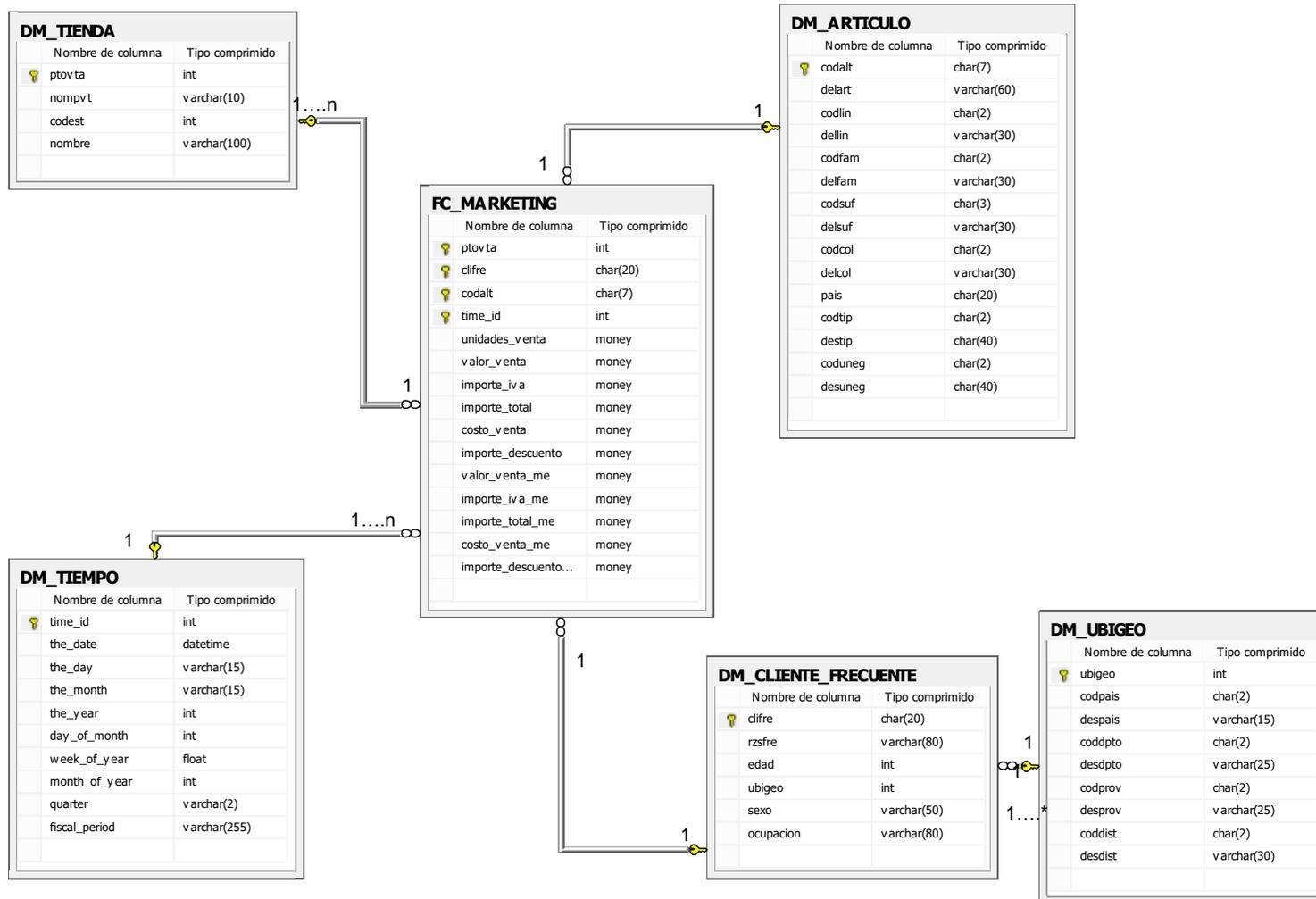


Gráfico 22: Diagrama de Base de Datos (Diseño Físico)

## 5. Diseño de la Solución

Después de determinar los procesos del negocio de la organización bajo estudio, y de describir sus flujos de trabajo mediante diagramas de actividad, se procede a realizar el modelado del sistema iniciándose con los casos de uso los cuales son identificados y estructurados a partir de las actividades de cada proceso.

### Diagrama de Actores

A continuación se muestra el diagrama de Actores del Sistema. Ver Gráfico 23.

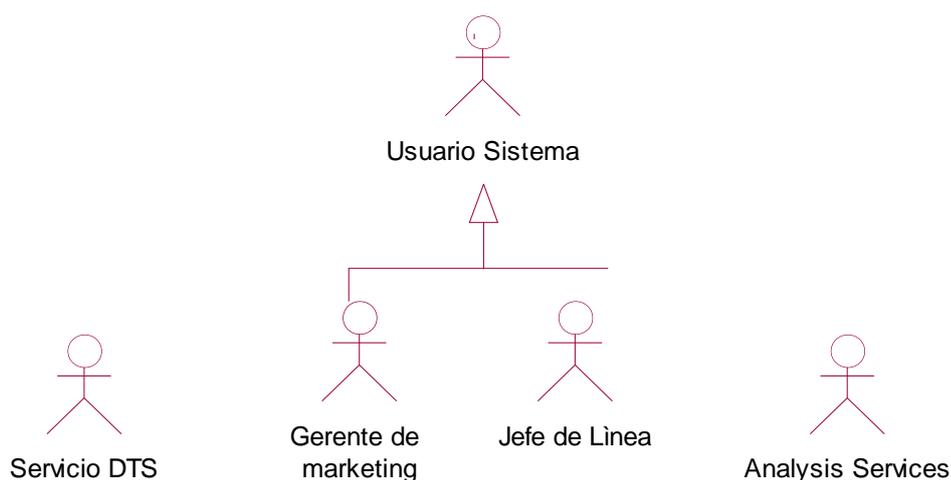


Gráfico 23: Actores del Sistema

### Descripción de Actores del Sistema

**Servicio DTS**, Servicio de Transformación de Datos, crea paquetes que procesen cubos y modelos de minería de datos, así como ejecutar consultas de pronóstico basadas en modelos de minería de datos. Estos paquetes se pueden programar para ejecutarlos automáticamente.

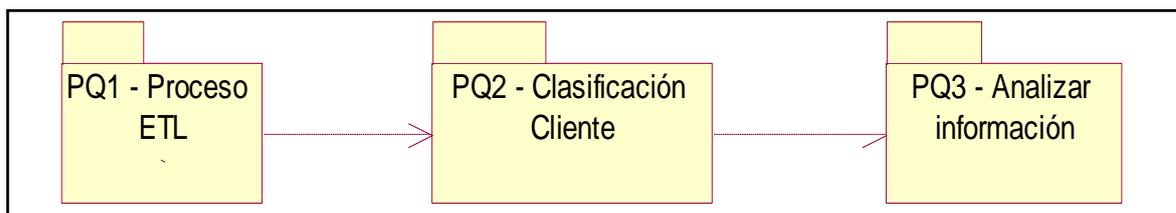
**Analysis Services**, Analysis Services es un motor de datos analíticos en línea que se usa en soluciones de ayuda a la toma de decisiones.

**Gerencia Marketing**, actor encargado de establecer estrategias y toma de decisiones en el proceso de estimación de ventas (línea, familia y producto) y compras de mercadería.

**Jefe de Línea**, actor encargado de analizar la información de su línea asignada (bisutería, complemento de vestir, etc.) y de realizar las compras de mercadería.

### Diagrama de Paquetes

A continuación, se muestra el diagrama de paquetes del sistema .Ver Gráfico 24.



*Gráfico 24: Diagrama de Paquetes del Sistema*

**Paquete Proceso ETL**, paquete que contiene los casos de uso del proceso ETL, Extraer Información, Transformar Información y Cargar Información, nos permitió mover datos desde la fuente de origen (BD transaccional), reformatearlos y limpiarlos, y cargarlos en el datawarehouse.

**Paquete Clasificación Cliente**, paquete que contiene el caso de uso Generar Clasificación de Clientes, proceso en el cual se crea la vista de origen donde se filtra sólo los clientes preferenciales, se genera la estructura de minería de datos, utilizando la técnica de minería de datos : Algoritmo de Árbol de Decisión de Microsoft.

**Paquete Analizar Información**, paquete que contiene los casos de uso Analizar información por Canal de Venta y Línea, Línea y Familia los cuales nos ayudaran a analizar la información para la toma de decisiones de estimar proyecciones de venta y compra y analizar el comportamiento o hábitos del cliente.

## Diagramas de Casos de Uso del Sistema

### Paquete: Proceso ETL

A continuación se detalla el Diagrama de Casos de Uso del Sistema.

Ver Gráfico 25.

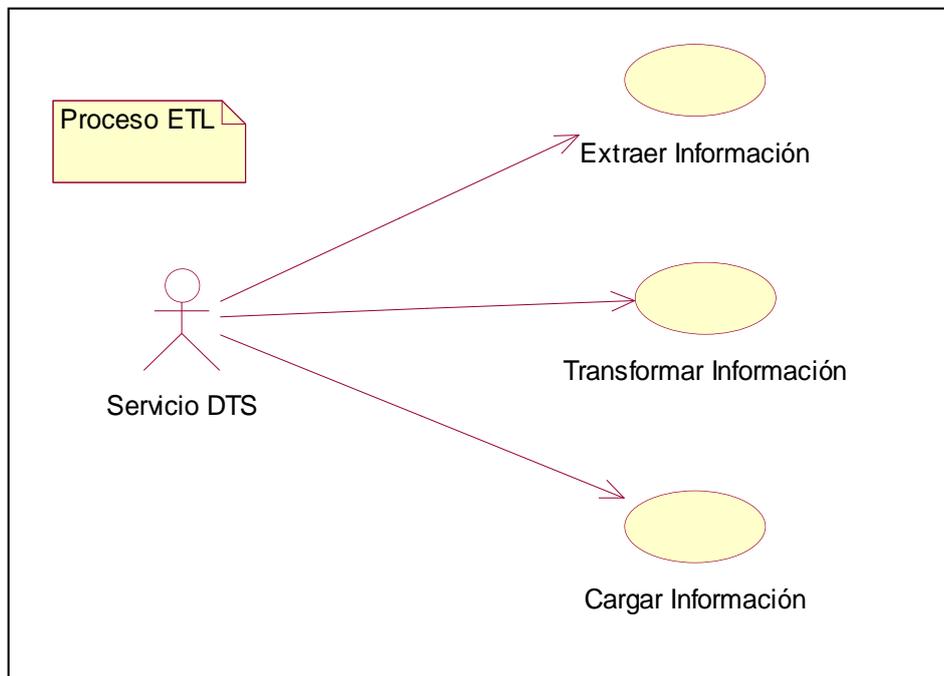


Gráfico 25: Diagrama de Caso de Uso del Sistema: Proceso ETL

**CUS\_Extraer Información:** Permite obtener la información de los datos maestros y la información de ventas que se encuentran registrados en la base de datos transaccional del Sistema Comercial de ventas.

**CUS\_Transformar información:** Permite transformar la información de los datos maestros y la información de ventas que se extrajeron de la base de datos transaccional del sistema comercial de ventas, aplicando una serie de reglas de negocio o funciones.

**CUS\_Cargar información:** La fase de carga es el momento en el cual los datos de la fase de transformación son cargados en el Datawarehouse.

## Paquete: Clasificación de Clientes (Algoritmo)

A continuación se detalla el Diagrama de Casos de Uso del Sistema.

Ver Gráfico 26.

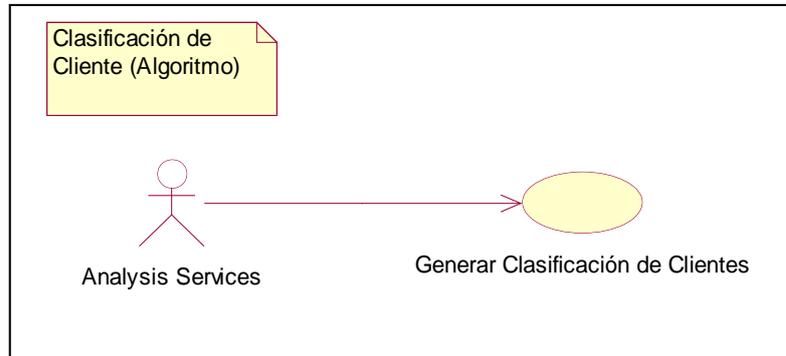


Gráfico 26: Diagrama de Caso de Uso del Sistema: Clasificación de Clientes

**CUS\_Generar clasificación de clientes:** Este proceso prepara, sondea y explora los datos para obtener la información de clasificación de clientes a través del algoritmo de Árbol de Decisión.

## Paquete: Analizar Información

A continuación se detalla el Diagrama de Casos de Uso del Sistema.

Ver Gráfico 27.

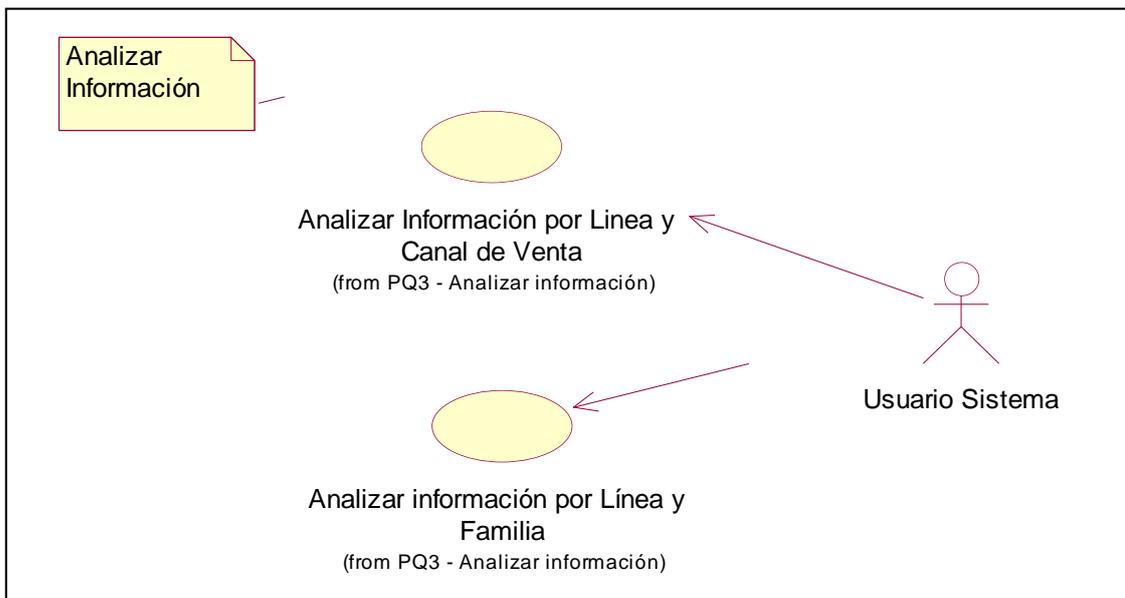


Gráfico 27: Diagrama de Caso de Uso del Sistema: Analizar Información

**CUS\_Analizar Información por Línea y Canal de Venta:** Este proceso muestra el reporte de análisis predictivo por Línea y Canal de Venta en base a la información solicitada.

**CUS\_Analizar Información por Línea y Familia:** Este proceso muestra el reporte de análisis predictivo por Línea y Familia en base a la información solicitada.

### **Especificaciones de Caso de Uso del Negocio**

El propósito de estas plantillas es presentar una guía para la construcción de Casos de Uso del negocio, ello implica los pasos para la construcción del diagrama, la especificación de los CU y la forma en cómo deben estructurarse.

A continuación se detalla la descripción del caso de Uso del Sistema: **Extraer Información.**

## DESCRIPCION DEL CASO DE USO DEL SISTEMA EXTRAER INFORMACION

### 1. DESCRIPCION CORTA

Permite obtener la información de los datos maestros y la información de ventas que se encuentran registrados en la base de datos transaccional del sistema comercial de ventas.

### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Servicio DTS

### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Ejecutar tarea de conexión al origen de datos transaccional.
2	S	Establece la conexión solicitada
3	A	Ejecutar tarea de conexión a la base de datos de trabajo
4	S	Establece la conexión solicitada
5	A	Eliminar los registros de las tablas maestras de la base de datos de trabajo
6	S	Elimina los registros de las tablas maestras
7	A	Ejecutar tarea de extracción de datos maestros
8	S	Realiza selección de las tablas maestras
9	A	Ejecutar tarea de insertar registros de datos maestros en tablas de trabajo
10	S	Inserta registros de datos maestros en la base de datos de trabajo
11	A	Eliminar los registros de las tablas de ventas de la base de datos de trabajo
12	S	Elimina los registros de las tablas de ventas
13	A	Ejecutar tarea de extracción de datos transaccionales
14	S	Realiza selección de las tablas de ventas.
15	A	Ejecutar tarea de insertar registros de datos transaccionales en tablas de trabajo
16	S	Inserta registro de datos transaccionales en la base de datos de trabajo

### 4. FLUJO ALTERNATIVO DE EVENTOS

#### 4.1. Paso 1: Si el origen de datos no está disponible

Paso	Tipo	Descripción del paso
1	S	Registrar Log de Errores de conexión

#### 4.2. Paso 3: Si el origen de datos no está disponible

Paso	Tipo	Descripción del paso
1	S	Registrar Log de Errores de conexión

#### 5. PRECONDICIONES

Ítem	Precondiciones
1	<ul style="list-style-type: none"><li>- Disponer del usuario y contraseña de la Base de Datos transaccional.</li><li>- Disponer del usuario y contraseña de la Base de Datos de Trabajo.</li></ul>

#### 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para la transformación de datos

#### 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Extraer Información tendrá una frecuencia diaria.

A continuación se detalla la Especificación Técnica detallada del Caso de Uso: **Extraer información**, tal como se visualiza en la Tabla 6:

## Especificación Técnica detallada del CUS: Extraer Información

1. El proceso de extracción inicia con establecer la conexión a la base de datos transaccional.
2. A continuación se establece la conexión a la base de datos de trabajo que permitirá almacenar temporalmente la información seleccionada de la base de datos transaccional.
3. Una vez establecida la conexión se ejecuta la tarea de extracción de los datos maestros, esta tarea realiza las siguientes acciones:
  - a. Truncar las tablas maestras
    - Truncar tabla línea
    - Truncar tabla familia
    - Truncar tabla subfamilia
    - Truncar tabla tiparticulo
    - Truncar tabla uninegocio
    - Truncar tabla color
    - Truncar tabla puntovta
    - Truncar tabla estructura
    - Truncar tabla país
    - Truncar tabla departamento
    - Truncar tabla provincia
    - Truncar tabla distrito
    - Truncar tabla articulos1
    - Truncar tabla clifrecuente
    - Truncar tabla undventa
    - Truncar tabla centros\_retail
  - b. Seleccionar información de tabla articulos1
    - Seleccionar todos los registros y todos los campos de la tabla artículos1
  - c. Insertar información de la tabla articulos1
    - Inserta todos los registros seleccionados en la tabla articulos1
  - d. Seleccionar información de tabla línea
    - Seleccionar todos los registros y todos los campos de la tabla línea
  - e. Insertar información de la tabla línea
    - Inserta todos los registros seleccionados en la tabla línea
  - f. Seleccionar información de tabla familia
    - Seleccionar todos los registros y todos los campos de la tabla familia
  - g. Insertar información de la tabla familia
    - Inserta todo los registros seleccionados en la tabla familia
  - h. Seleccionar información de tabla subfamilia
    - Seleccionar todos los registros y todos los campos de la tabla subfamilia
  - i. Insertar información de la tabla subfamilia
    - Inserta todo los registros seleccionados en la tabla subfamilia
  - j. Seleccionar información de tabla tiparticulo
    - Seleccionar todos los registros y todos los campos de la tabla tiparticulo

**Especificación Técnica detallada del CUS: Extraer Información**

- k. Insertar información de la tabla tiparticulo  
Inserta todo los registros seleccionados en la tabla tiparticulo
- l. Seleccionar información de tabla uninegocio  
Seleccionar todos los registros y todos los campos de la tabla uninegocio
- m. Inserta información de la tabla uninegocio  
Inserta todo los registros seleccionados en la tabla uninegocio
- n. Seleccionar información de tabla color  
Seleccionar todos los registros y todos los campos de la tabla color
- o. Inserta información de la tabla color  
Inserta todo los registros seleccionados en la tabla color
- p. Seleccionar información de la tabla puntovta  
Seleccionar todos los campos de la tabla puntovta
- q. Inserta información de la tabla puntovta  
Inserta todo los registros seleccionados en la tabla puntovta
- r. Seleccionar información de la tabla clifrecuente  
Seleccionar todos los registros y todos los campos de la tabla clifrecuente
- s. Inserta información de la tabla clifrecuente  
Inserta todo los registros seleccionados en la tabla clifrecuente
- t. Seleccionar información de la tabla país  
Seleccionar todos los registros y todos los campos de la tabla país
- u. Inserta información de la tabla país  
Inserta todo los registros seleccionados en la tabla país
- v. Seleccionar información de la tabla departamento  
Seleccionar todos los registros y todos los campos de la tabla departamento
- w. Inserta información de la tabla departamento  
Inserta todo los registros seleccionados en la tabla departamento
- x. Seleccionar información de la tabla provincia  
Seleccionar todos los registros y todos los campos de la tabla provincia
- y. Inserta información de la tabla provincia  
Inserta todo los registros seleccionados en la tabla provincia
- z. Seleccionar información de la tabla distrito  
Seleccionar todos los registros y todos los campos de la tabla distrito
- aa. Inserta información de la tabla distrito  
Inserta todo los registros seleccionados en la tabla distrito
- bb. Seleccionar información de la tabla estructura  
Seleccionar todos los registros y todos los campos de la tabla estructura
- cc. Inserta información de la tabla estructura  
Inserta todo los registros seleccionados en la tabla estructura
- dd. Seleccionar información de la tabla estructura undventa.

Tabla 6 Especificación Técnica detallada del CUS: Extraer Información (continuación)

## Especificación Técnica detallada del CUS: Extraer Información

- ee. Inserta información de la tabla undventa
    - Seleccionar todos los registros y todos los campos de la tabla undventa
  - ff. Seleccionar información de la tabla estructura centros\_retail
    - Inserta todo los registros seleccionados en la tabla undventa
    - Seleccionar todos los registros y todos los campos de la tabla centros\_retail
  - gg. Inserta información de la tabla centros\_retail
    - Inserta todo los registros seleccionados en la tabla centros\_retail
4. Seguidamente se ejecuta la tarea de extracción de los datos transaccionales, esta tarea realiza las siguientes acciones:
- a. Truncar las tablas de ventas
    - Truncar tabla cboletav
    - Truncar tabla dboletav
    - Truncar tabla cfactura
    - Truncar tabla dfactura
    - Truncar tabla cnotcre
    - Truncar tabla dnotcre
    - Truncar tabla cgdespacho
    - Truncar tabla dgdespacho
    - Truncar tabla cguiadevolucion
    - Truncar tabla dguiadevolucion
  - b. Seleccionar información de tabla cgdespacho
    - Seleccionar todos los campos de la tabla cgdespacho cuyos registros estén entre el 1er día del mes anterior y la fecha actual
  - c. Inserta información de la tabla cgdespacho
    - Insertar todos los registros seleccionados en la tabla cgdespacho
  - d. Seleccionar información de tabla dgdespacho
    - Seleccionar todos los campos de la tabla dgdespacho cuyos registros estén entre el 1er día del mes anterior y la fecha actual
  - e. Inserta información de la tabla dgdespacho
    - Insertar todos los registros seleccionados en la tabla dgdespacho
  - f. Seleccionar información de tabla cguiadevolucion
    - Seleccionar todos los campos de la tabla cguiadevolucion cuyos registros estén entre el 1er día del mes anterior y la fecha actual
  - g. Inserta información de la tabla cguiadevolucion
    - Insertar todos los registros seleccionados en la tabla cguiadevolucion
  - h. Seleccionar información de tabla dguiadevolucion
    - Seleccionar todos los campos de la tabla dguiadevolucion cuyos registros estén entre el 1er día del mes anterior y la fecha actual
  - i. Inserta información de la tabla dguiadevolucion

Tabla 6 Especificación Técnica detallada del CUS: Extraer Información (continuación)

**Especificación Técnica detallada del CUS: Extraer Información**

- Insertar todos los registros seleccionados en la tabla dguiadevolucion
- j. Seleccionar información de tabla cboletav  
Seleccionar todos los campos de la tabla cboletav cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- k. Inserta información de la tabla cboletav  
Insertar todos los registros seleccionados en la tabla cboletav
- l. Seleccionar información de tabla dboletav  
Seleccionar todos los campos de la tabla dboletav cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- m. Inserta información de la tabla dboletav  
Insertar todos los registros seleccionados en la tabla dboletav
- n. Seleccionar información de tabla cfactura  
Seleccionar todos los campos de la tabla cfactura cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- o. Inserta información de la tabla cfactura  
Insertar todos los registros seleccionados en la tabla cfactura
- p. Seleccionar información de tabla dfactura  
Seleccionar todos los campos de la tabla dfactura cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- q. Inserta información de la tabla dfactura  
Insertar todos los registros seleccionados en la tabla dfactura
- r. Seleccionar información de tabla cnotcre  
Seleccionar todos los campos de la tabla cnotcre cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- s. Inserta información de la tabla cnotcre  
Insertar todos los registros seleccionados en la tabla cnotcre
- t. Seleccionar información de tabla dnotcre  
Seleccionar todos los campos de la tabla dnotcre cuyos registros estén entre el 1er día del mes anterior y la fecha actual
- u. Inserta información de la tabla dnotcre  
Insertar todos los registros seleccionados en la tabla dnotcre.

Tabla 6 Especificación Técnica detallada del CUS: Extraer Información (continuación)

## **DIAGRAMA DE CLASES: EXTRAER INFORMACION (ORIGEN)**

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de trabajo de origen de la Base de Datos Transaccional. A continuación se muestra el Diagrama de Clases: Extraer Información. Ver Gráfico 28.

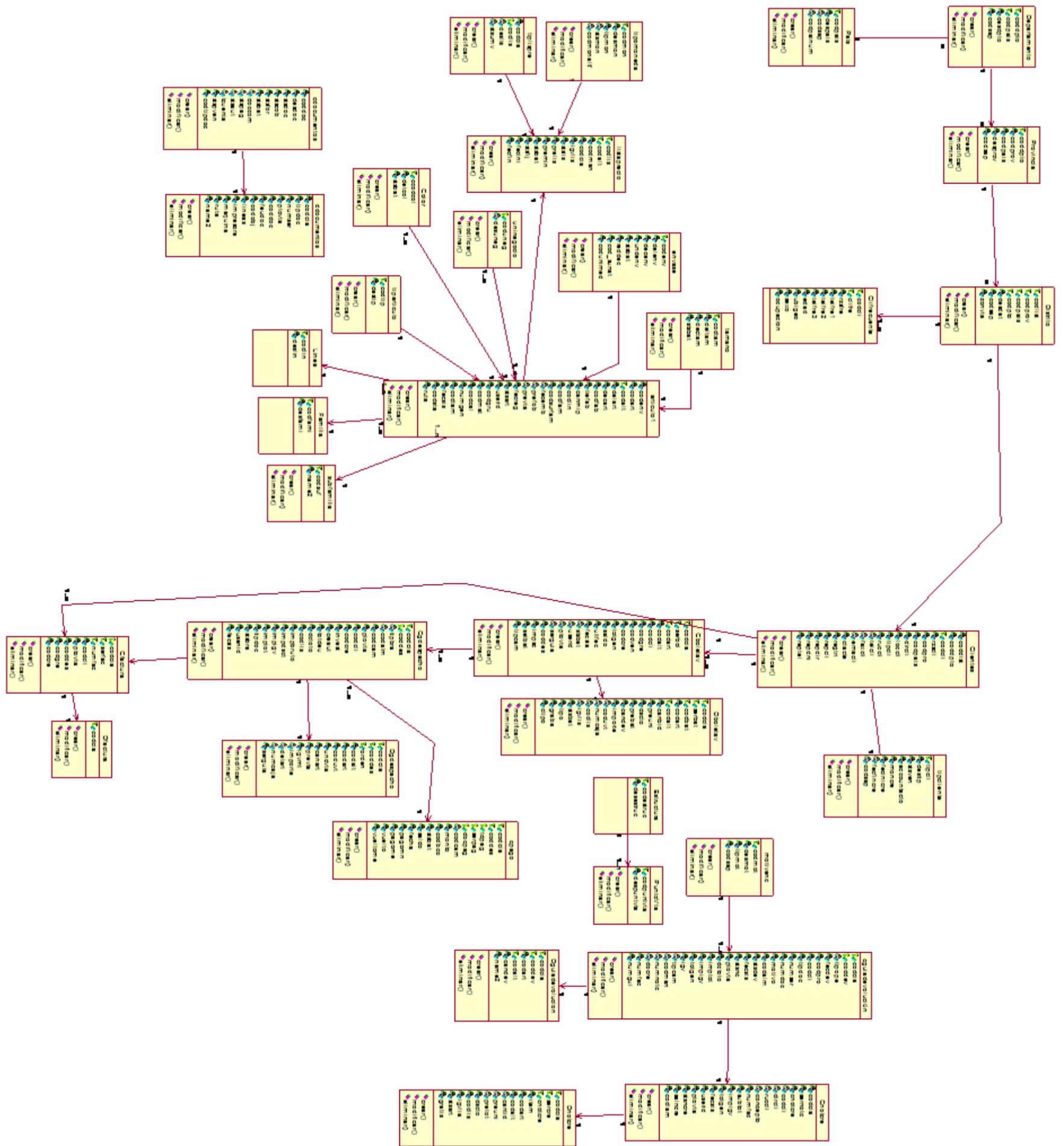


Gráfico 28: Diagrama de Clases: Extraer Información

## **DIAGRAMA DE SECUENCIA CUS: EXTRAER INFORMACION (MAESTROS Y VENTAS)**

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Extraer Información (de la tabla Maestros y la Tabla Ventas).

Ver gráfico 29.

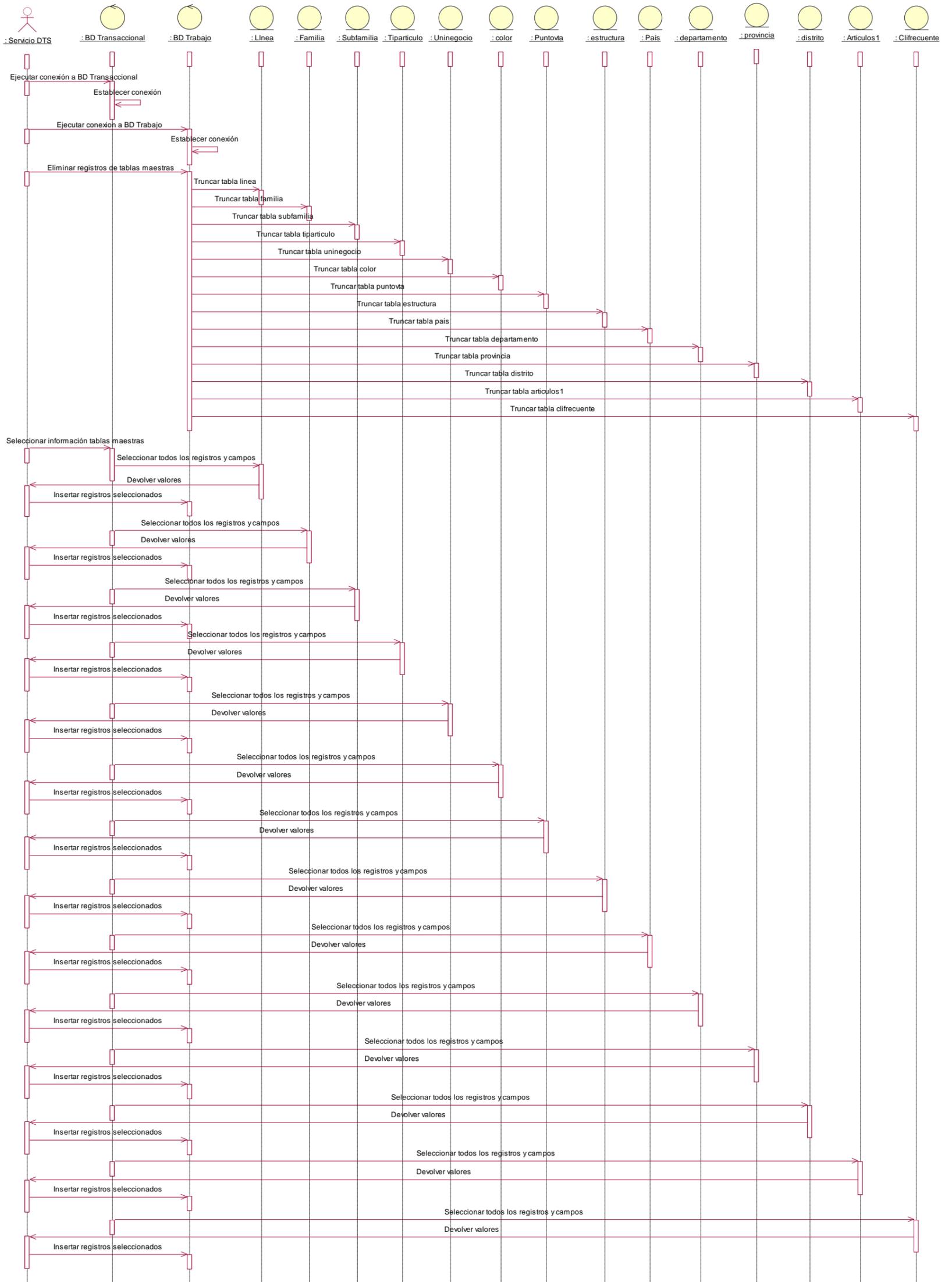


Gráfico 29: Diagrama de Secuencia: Extraer Información

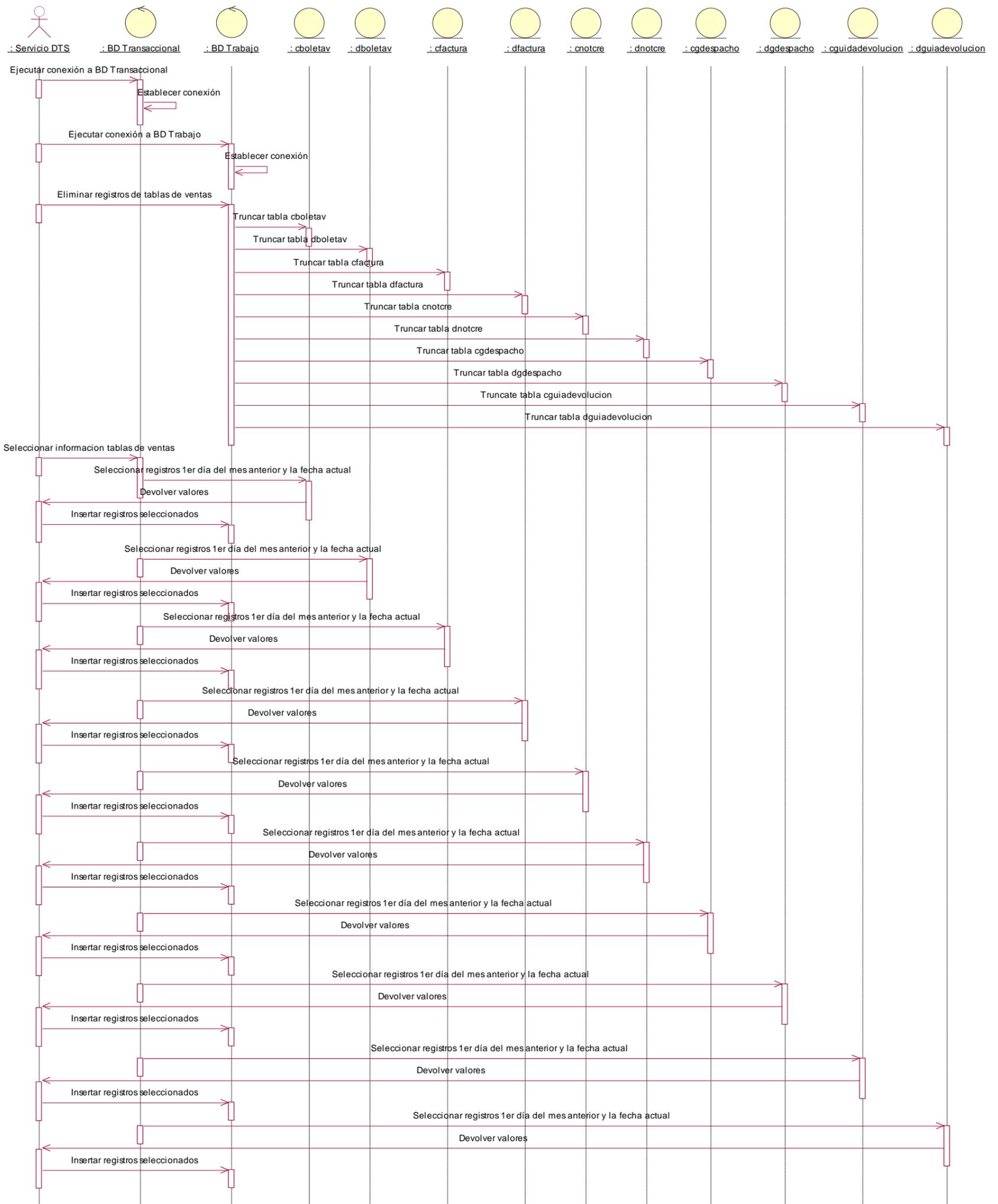


Gráfico 29: Diagrama de Secuencia: Extraer Información (continuación)

## Construcción del CUS: Extraer Información

Se usará como herramienta para ejecutar las tareas descritas en la Especificación Técnica Detallada, SQL Server Business Intelligence Development Studio, en el cual se creará un nuevo proyecto en el ambiente de desarrollo. Ver Gráfico 30:

### Integration Services (proyecto DTS)

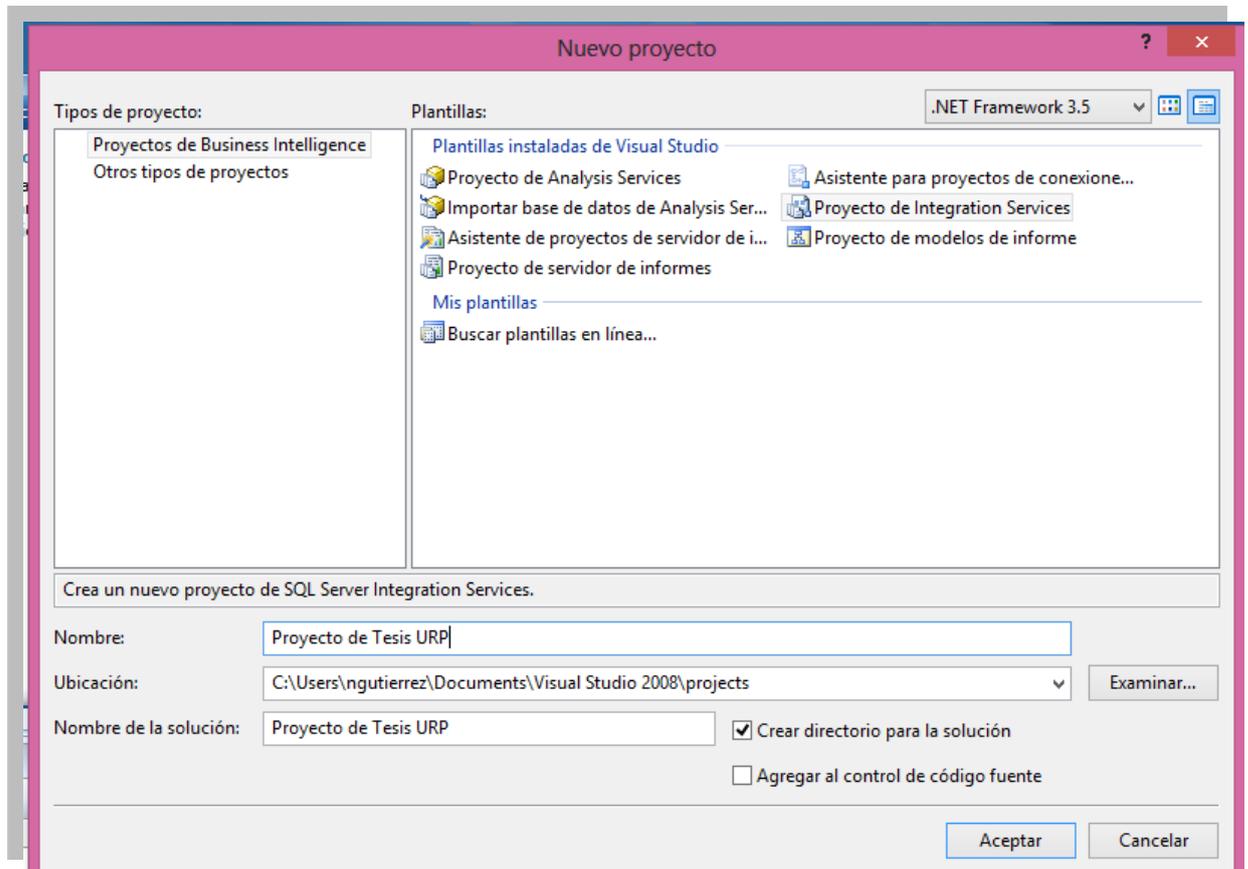


Gráfico 30: Creación de Nuevo Proyecto DTS

Se proceden a crear las conexiones de las BD en el “Administrador de conexiones”, tal como se visualiza en el Gráfico 31:

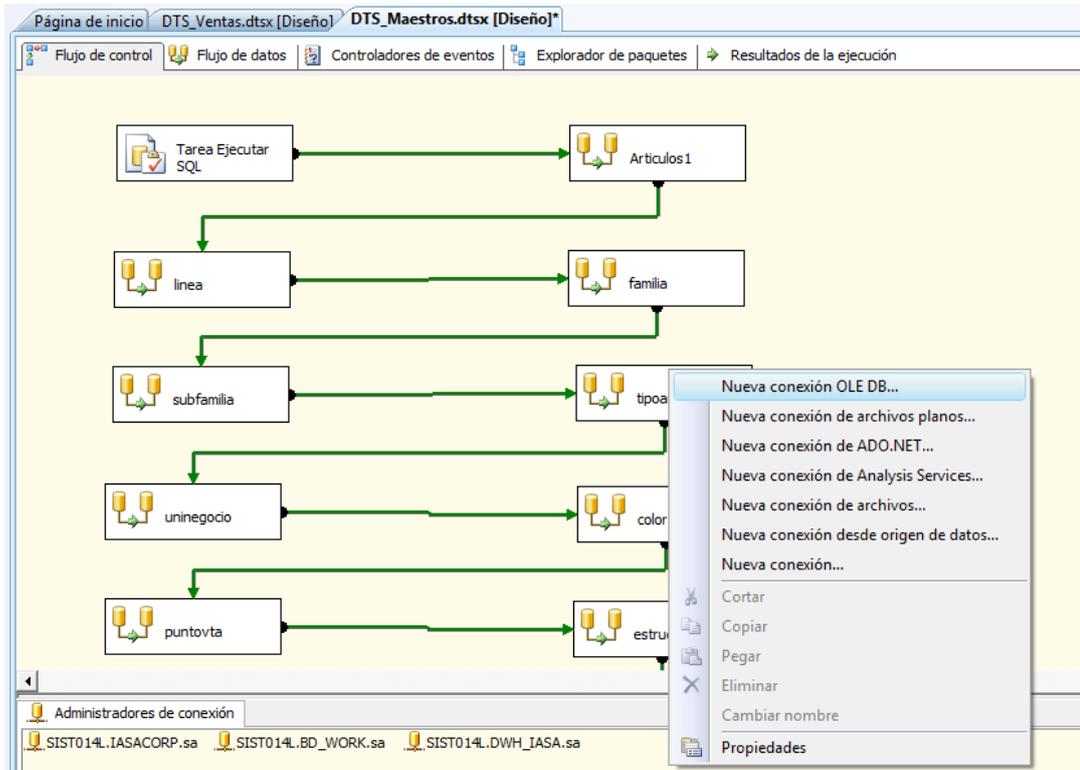


Gráfico 31: Administrador de conexiones

Se selecciona una Nueva conexión OLE DB, Ver gráfico 32.

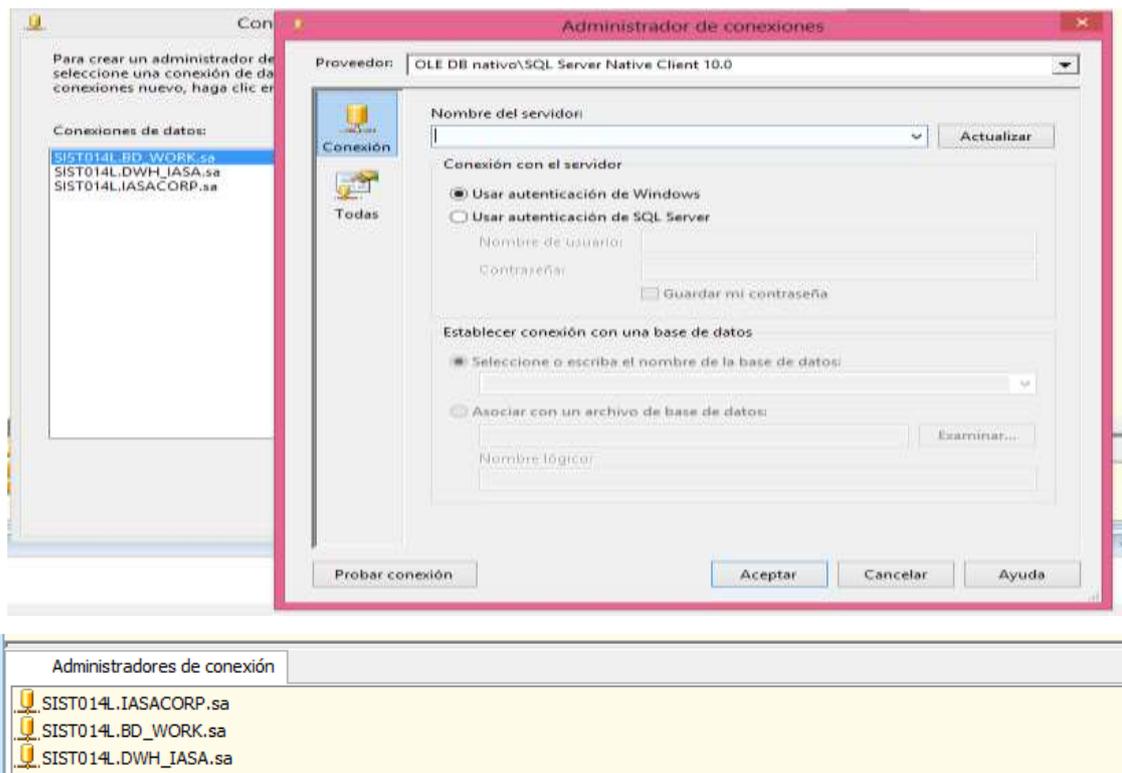


Gráfico 32: Conexión OLE DB

Luego se procede a generar el flujo de trabajo para la selección de los **datos maestros**:

1. Se genera la tarea de ejecución SQL que realiza la acción de truncar todas las tablas maestras indicadas en Especificación Técnica Detallada – numeral 4) a. Ver gráfico 33.

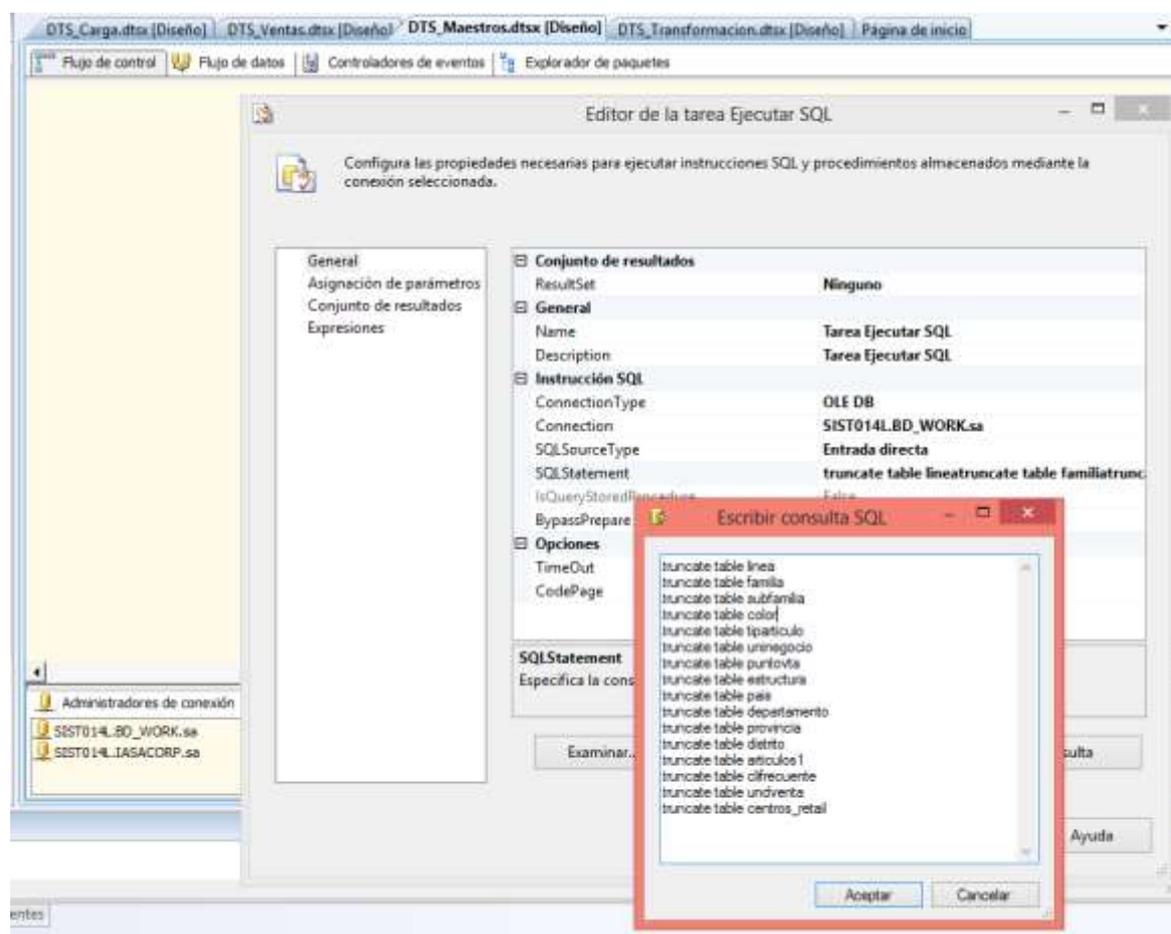


Gráfico 33: Generación Tarea Truncate Table Maestras

2. Se procede a generar el flujo de control del DTS con un nuevo flujo de trabajo para cada entidad a seleccionar según Especificación Técnica Detallada – numeral 2) de letra “b” a la letra “cc”
  - a. En el siguiente ejemplo se muestra el flujo de trabajo para la tabla artículos1. Ver gráfico 34.

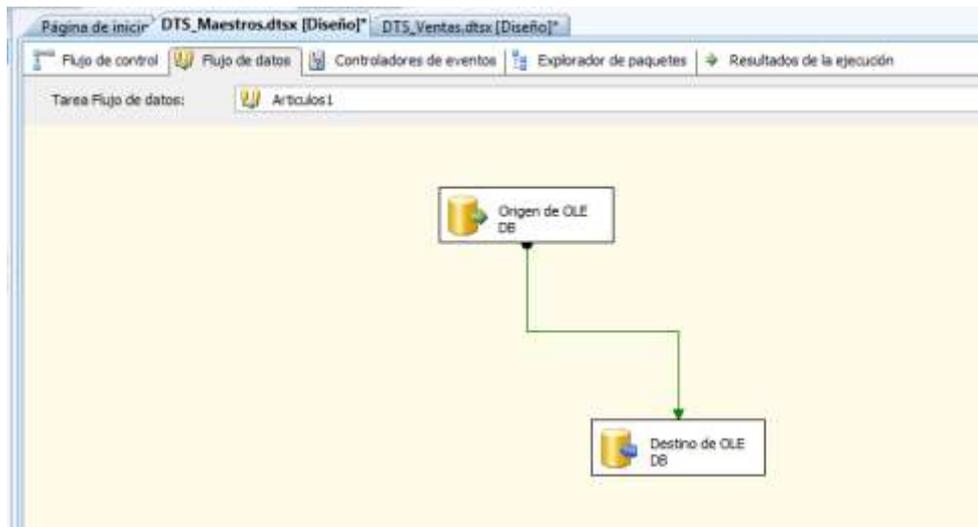


Gráfico 34: Flujo de trabajo para la tabla artículos1

- b. Se genera el origen de la conexión OLE DB, desde la Base de Datos Transaccional ( IASACORP ) de la tabla articulos1. Ver gráfico 35.

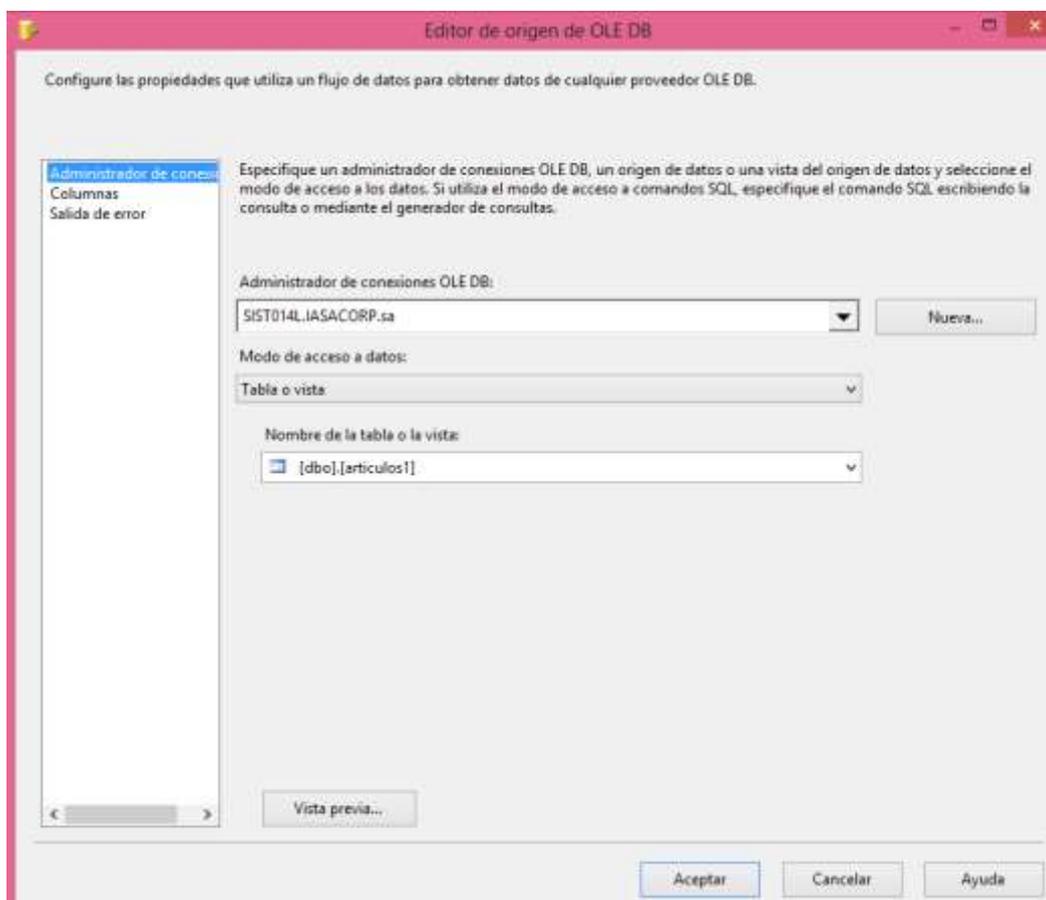


Gráfico 35: Conexión desde la Base de Datos Transaccional de la tabla articulos1

- c. Se genera el destino de la conexión OLEDB, hacia la base de datos de trabajo (BD\_WORK) de la tabla articulos1. Ver gráfico 36 :

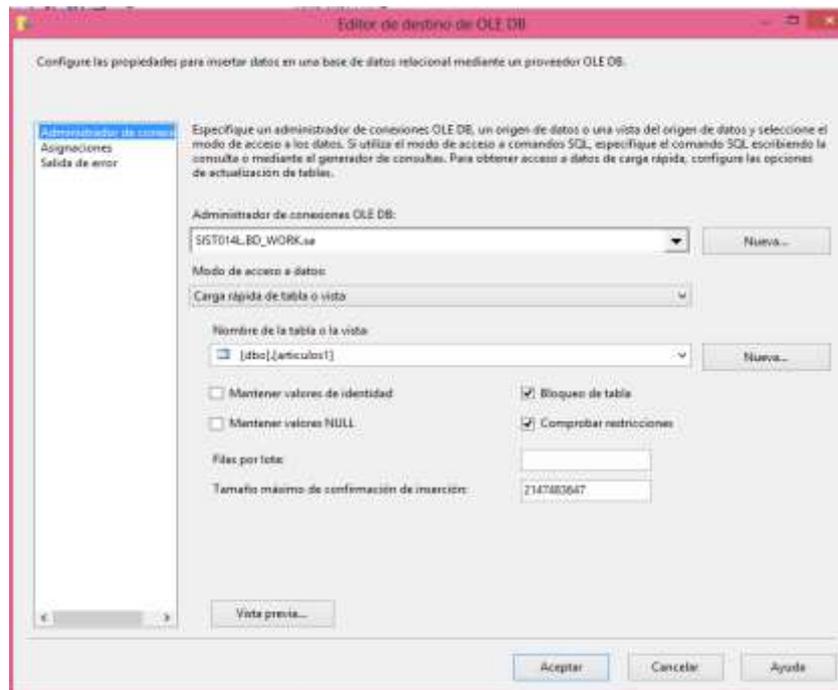


Gráfico 36: Conexión a la BD de Trabajo, tabla Articulos1

3. Luego se procede a generar el flujo de trabajo para la selección de los datos transaccionales:

- a. Se genera tarea de ejecución SQL que realiza la acción de truncar todas las tablas indicadas en Especificación Técnica Detallada – numeral 4) a. Ver gráfico 37.

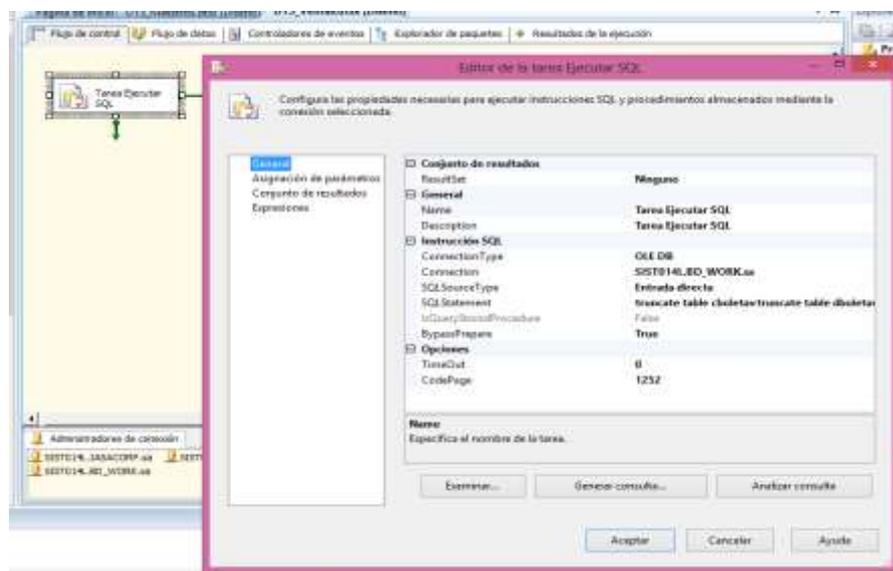


Gráfico 37: Generación Tarea truncate Table Transaccionales

4. Luego se procede a generar el flujo de control del DTS con un nuevo flujo de trabajo para cada entidad a seleccionar según Especificación Técnica Detallada – numeral 4) de letra “b” a la letra “u”
  - a. En el siguiente ejemplo se muestra el flujo de trabajo para la tabla cboletav, ver gráfico 38.

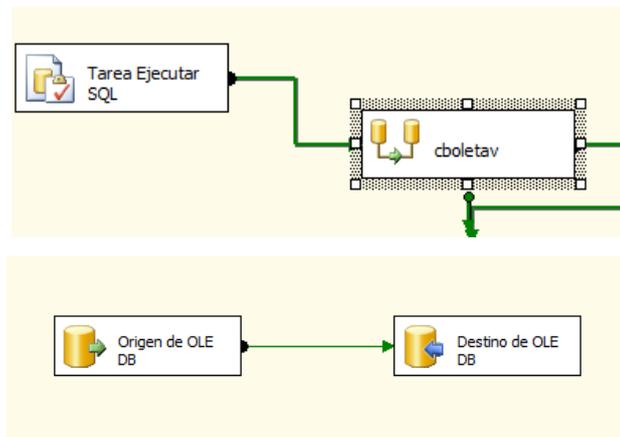


Gráfico 38: Flujo de trabajo para la tabla cboletav

5. Se genera el origen de la conexión OLE DB, desde la Base de Datos Transaccional ( IASACORP ) de la tabla cboletav, ver gráfico 39.

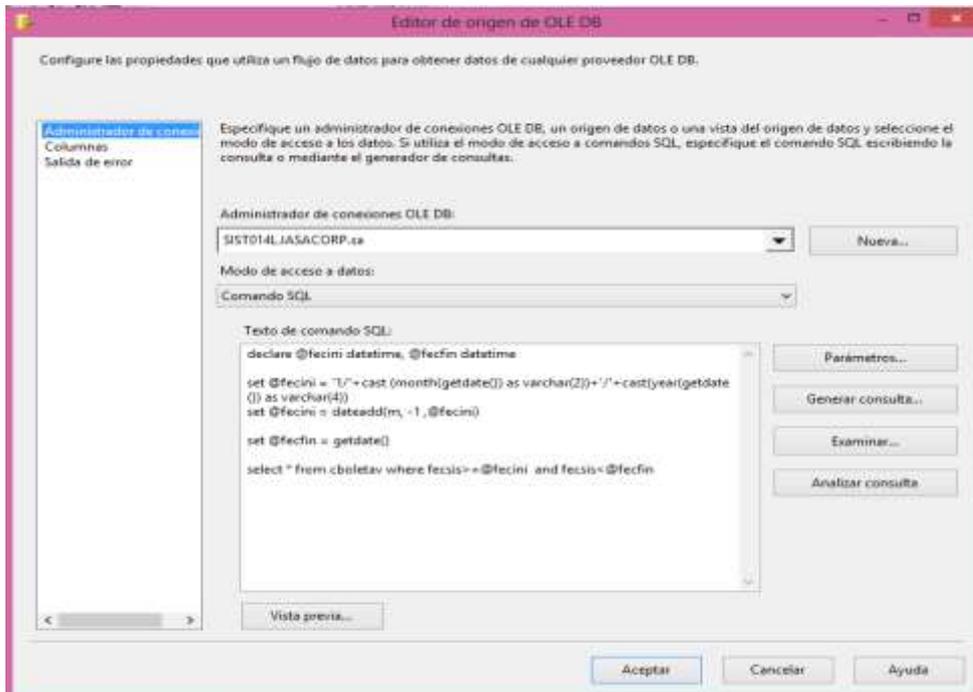


Gráfico 39: Conexión desde la Base de Datos Transaccional, de la tabla cboletav

6. Se genera el destino de la conexión OLEDB, hacia la Base de Datos de trabajo (BD\_WORK) de la tabla cboletav, ver gráfico 40.

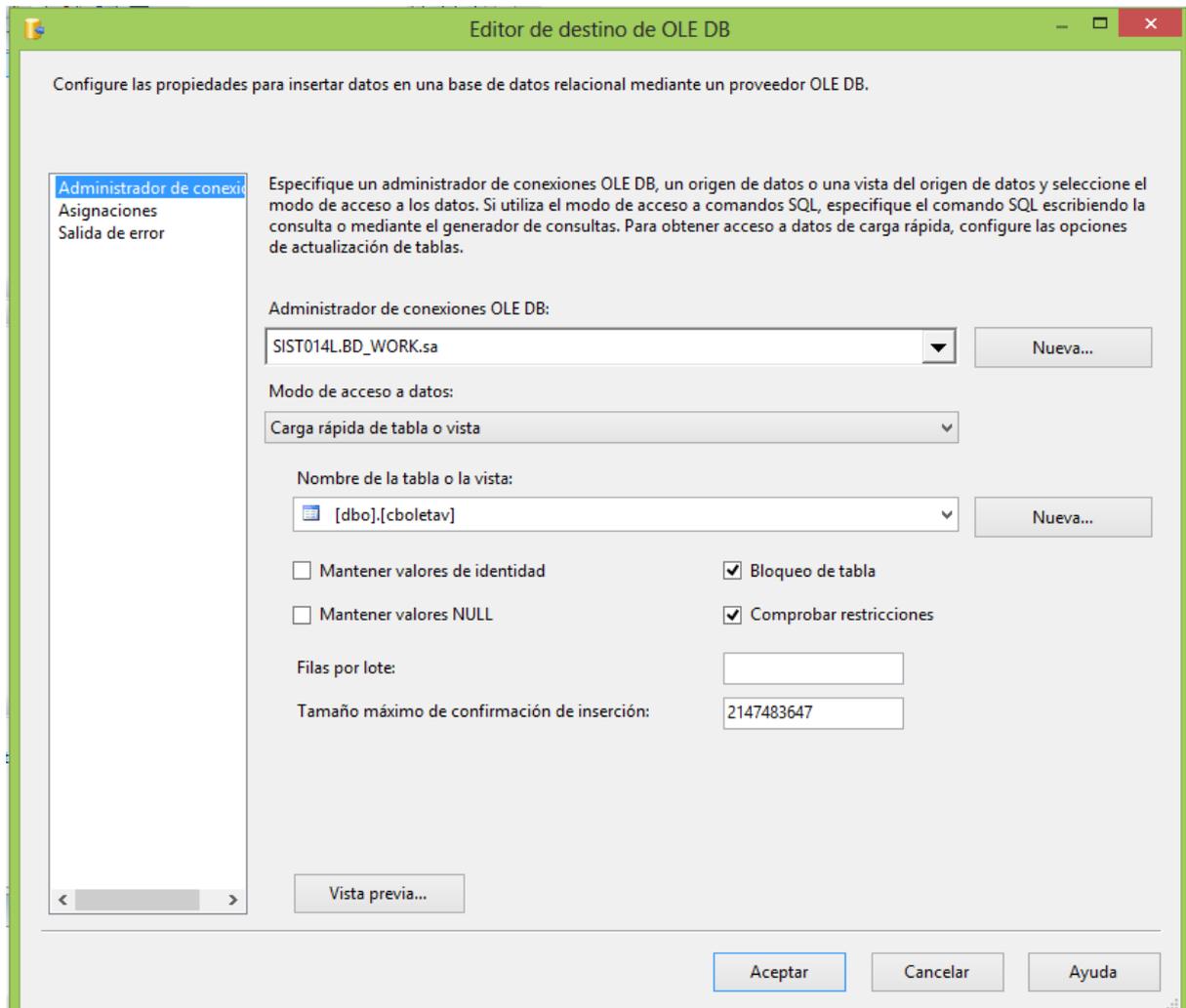


Gráfico 40: Conexión hacia la Base de Datos de trabajo, de la tabla cboletav

A continuación se detalla la descripción del caso de Uso del Sistema:  
**Transformar Información.**

## DESCRIPCION DEL CASO DE USO DEL SISTEMA TRANSFORMAR INFORMACION

### 1. DESCRIPCION CORTA

Permite transformar la información de los datos maestros y la información de ventas que se extrajeron de la base de datos transaccional del sistema comercial de ventas, aplicando una serie de reglas de negocio o funciones.

### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Servicio DTS

### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Ejecutar tarea de conexión a la base de datos de trabajo.
2	S	Establece la conexión solicitada
3	A	Eliminar los registros de las tablas dimensionales y la tabla de hechos.
4	S	Elimina los registros de las tablas dimensionales y la tabla de hechos.
5	A	Ejecutar tarea de transformación de datos maestros a la estructura de dimensiones.
6	S	Realiza transformación de las tablas maestras.
7	A	Ejecutar tarea de transformación de datos transaccionales a la estructura de tabla de hechos
8	S	Realiza extracción de las tablas de datos transaccionales.
9	A	Ejecutar tarea de almacenar dimensiones en la base de datos de trabajo
10	S	Realizar almacenamiento de dimensiones.
11	A	Ejecutar tarea de almacenar tabla de hechos en la base de datos de trabajo
12	S	Realizar almacenamiento de tabla de hechos.

### 4. FLUJO ALTERNATIVO DE EVENTOS

#### 4.1. Paso 1.1: Si el origen de datos no está disponible

Paso	Tipo	Descripción del paso
1	S	Registrar Log de Errores de conexión

## 5. PRECONDICIONES

Ítem	Precondiciones
1	Haber ejecutado el Caso de Uso: Extraer Información.

## 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para la carga de datos.

## 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Transformar Información tendrá una frecuencia diaria.

A continuación se detalla la Especificación Técnica detallada del Caso de Uso: Transformar Información, tal como se visualiza en la Tabla 7.

### **Especificación Técnica detallada del CUS: Transformar Información**

1. El proceso de transformación se inicia con establecer la conexión a la base de datos de trabajo que permitirá almacenar temporalmente la información de las dimensiones y la tabla de hechos.
2. Una vez establecida la conexión se ejecuta la tarea de transformación de los datos maestros, esta tarea realiza las siguientes acciones:

- a. Truncar las tablas de trabajo para las dimensiones y la tabla de hechos

Truncar tabla DM\_ARTICULO\_WORK

Truncar tabla DM\_CLIENTE\_FRECUENTE\_WORK

Truncar tabla DM\_TIENDA\_WORK

Truncar tabla DM\_UBIGEO\_WORK

Truncar tabla DM\_TIEMPO\_WORK

Truncar tabla FC\_MARKETING\_WORK

Actualizar datos de tabla clifrecuente para evitar caracteres especiales, por ejemplo validar que campos no contengan valores como #,\$,& y otros que puedan ocasionar errores de inserción de datos.

- b. Se inicia la selección de la información del maestro de artículos y sus atributos para generar la tabla de tipo dimensión para los artículos. La transformación se realiza con los siguientes origen de datos y destino de datos:

*Tabla 7 Especificación Técnica detallada del CUS: Transformar Información*

## Especificación Técnica detallada del CUS: Transformar Información

Tabla Origen	Campo		Tabla Destino	Campo
articulos1	codalt		DM_ARTICULO_WORK	codalt
articulos1	delart		DM_ARTICULO_WORK	delart
articulos1	país		DM_ARTICULO_WORK	país
línea	codlin		DM_ARTICULO_WORK	codlin
línea	dellin		DM_ARTICULO_WORK	dellin
familia	codfam		DM_ARTICULO_WORK	codfam
familia	delfam		DM_ARTICULO_WORK	delfam
subfamilia	codsuf		DM_ARTICULO_WORK	codsuf
subfamilia	delsuf		DM_ARTICULO_WORK	delsuf
color	codcol		DM_ARTICULO_WORK	codcol
color	delcol		DM_ARTICULO_WORK	delcol
tiparticulo	codtip		DM_ARTICULO_WORK	codtip
tiparticulo	destip		DM_ARTICULO_WORK	destip
uninegocio	coduneg		DM_ARTICULO_WORK	coduneg
uninegocio	desuneg		DM_ARTICULO_WORK	desuneg

- a. Se inicia la selección de la información del maestro de Tiendas y sus atributos para generar la tabla de tipo dimensión DM\_TIENDA. La transformación se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo		Tabla Destino	Campo
puntovta	ptovta		DM_TIENDA_WORK	ptovta
puntovta	nompvt		DM_TIENDA_WORK	nompvt
puntovta	codest		DM_TIENDA_WORK	codest
estructura	nombre		DM_TIENDA_WORK	nombre

- b. Se inicia la generación de los datos de tiempo para llenar la tabla de tipo dimensión DM\_TIEMPO. La transformación se realiza con el siguiente origen de datos y destino de datos:

Tabla 7: Especificación Técnica detallada del CUS: Transformar Información (continuación)

## Especificación Técnica detallada del CUS: Transformar Información

Tabla Origen	Campo	Tabla Destino	Campo
autogenerado		DM_TIEMPO_WORK	time_id
autogenerado		DM_TIEMPO_WORK	the_date
autogenerado		DM_TIEMPO_WORK	the_day
autogenerado		DM_TIEMPO_WORK	The_month
autogenerado		DM_TIEMPO_WORK	the_year
autogenerado		DM_TIEMPO_WORK	day_of_month
autogenerado		DM_TIEMPO_WORK	week_of_year
autogenerado		DM_TIEMPO_WORK	Month_of_year
autogenerado		DM_TIEMPO_WORK	Quarter
autogenerado		DM_TIEMPO_WORK	Fiscal_period

- a. La selección de la información del maestro de clifrecuente y sus atributos para generar la tabla de tipo dimensión DM\_CLIENTE\_FRECUENTE. La transformación se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo	Tabla Destino	Campo
clifrecuente	clifre	DM_CLIENTE_FRECUENTE_WORK	clifre
clifrecuente	rzsfre	DM_CLIENTE_FRECUENTE_WORK	rzsfre
clifrecuente	Diferencia en años entre fecha nacimiento y fecha actual	DM_CLIENTE_FRECUENTE_WORK	Edad
clifrecuente	Unión de código de país, código de departamento, código de provincia y código de distrito	DM_CLIENTE_FRECUENTE_WORK	ubigeo
clifrecuente	Cuando campo sexo es igual a 'M' es 'MASCULINO', sino cuando campo sexo es igual a 'F' es 'FEMENINO' si no es 'NO ASIGNADO'	DM_CLIENTE_FRECUENTE_WORK	sexo
clifrecuente	ocupación	DM_CLIENTE_FRECUENTE_WORK	ocupación

- b. Se inicia la selección de la información de los maestros de ubigeo y sus atributos para generar la tabla de tipo dimensión DM\_UBIGEO. La transformación se realiza con el

Tabla 7: Especificación Técnica detallada del CUS: Transformar Información (continuación)

## Especificación Técnica detallada del CUS: Transformar Información

Tabla Origen	Campo	Tabla Destino	Campo
distrito	Unión de código de país, código de departamento, código de provincia y código de distrito	DM_UBIGEO_WORK	ubigeo
distrito	codpais	DM_UBIGEO_WORK	codpais
distrito	coddpto	DM_UBIGEO_WORK	coddpto
distrito	codprov	DM_UBIGEO_WORK	codprov
distrito	coddist	DM_UBIGEO_WORK	coddist
distrito	desdist	DM_UBIGEO_WORK	desdist
pais	despais	DM_UBIGEO_WORK	despais
departamento	desdpto	DM_UBIGEO_WORK	desdpto
provincia	desprov	DM_UBIGEO_WORK	desprov

- a. Se inicia la selección de la información de las tablas de ventas y sus atributos para generar la tabla de tipo dimensión FC\_MARKETING\_WORK. La transformación se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo	Tabla Destino	Campo
cboletav,cfactura,cnotcre	ptovta	FC_MARKETING_WORK	Ptovta
cboletav,cfactura,cnotcre	fecsis	FC_MARKETING_WORK	time_id
cboletav,cfactura,cnotcre	clifre	FC_MARKETING_WORK	Clifre
dboletav,dfactura,dnotcre	codalt	FC_MARKETING_WORK	Codalt
dboletav,dfactura,dnotcre	cantid	FC_MARKETING_WORK	unidades_ventas
dboletav,dfactura,dnotcre	pretot	FC_MARKETING_WORK	valor_venta
dboletav,dfactura,dnotcre	Importe-pretot	FC_MARKETING_WORK	importe_iva
dboletav,dfactura,dnotcre	importe	FC_MARKETING_WORK	importe_total
dboletav,dfactura,dnotcre	Pretot/tipocambio	FC_MARKETING_WORK	valor_venta_me
dboletav,dfactura,dnotcre	Importe-pretot/tipocambio	FC_MARKETING_WORK	Importe_iva_me
dboletav,dfactura,dnotcre	Importe/tipocambio	FC_MARKETING_WORK	Importe_total_me
Función obtienecosto		FC_MARKETING_WORK	costo_venta
dboletav,dfactura,dnotcre	Calculo del descuento aplicado en la transacción de venta	FC_MARKETING_WORK	importe_descuento
Función obtienecosto_me		FC_MARKETING_WORK	Costo_venta_me
dboletav,dfactura,dnotcre	Calculo del descuento aplicado en la transacción de venta	FC_MARKETING_WORK	Importe_descuento_me

Tabla 7: Especificación Técnica detallada del CUS: Transformar Información (continuación)

## DIAGRAMA DE CLASES: TRANSFORMAR INFORMACION (TRABAJO)

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de trabajo (tablas temporales) del Datawarehouse. A continuación se muestra el Diagrama de Clases: Transformar Información. Ver Gráfico 41.

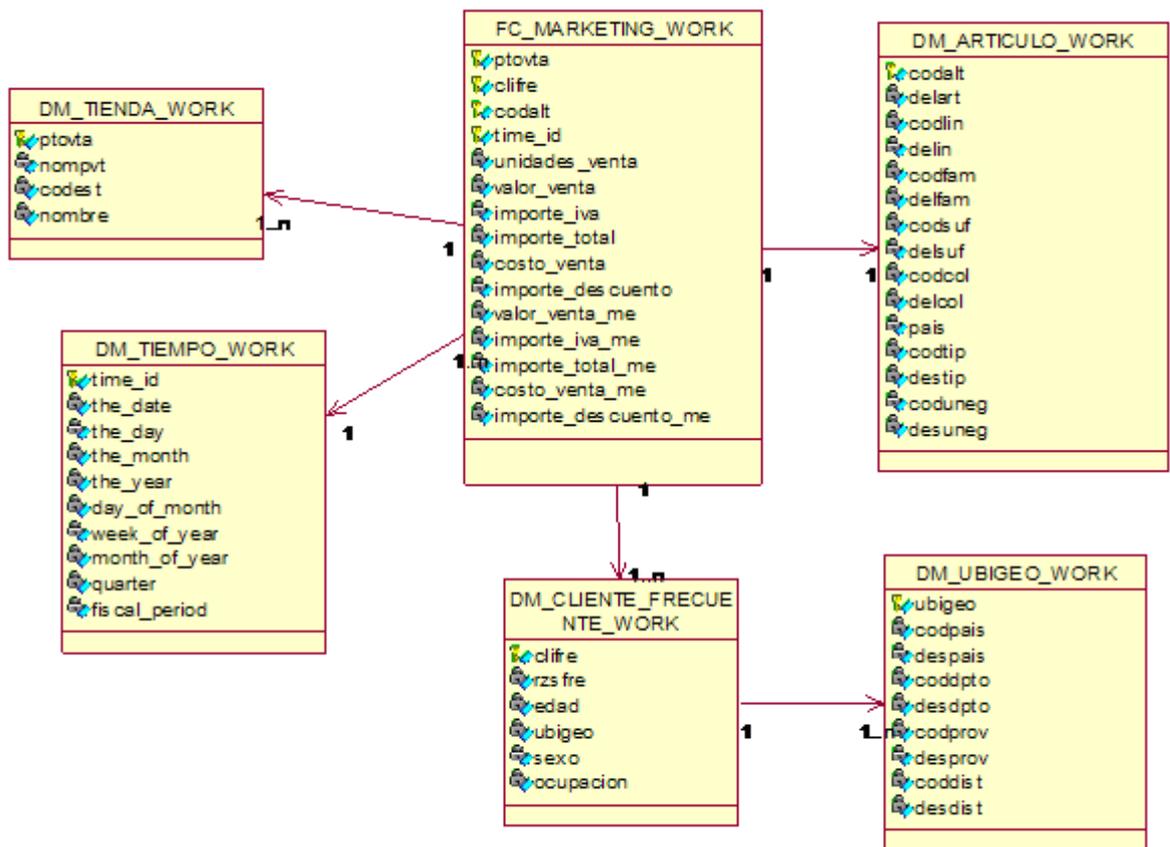


Gráfico 41: Diagrama de Clases: Transformar Información

## **DIAGRAMA DE SECUENCIA CUS: TRANSFORMAR INFORMACION**

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Transformar Información. Ver gráfico 42.



## Construcción del CUS: Transformar Información

1. Se procede a crear la conexión de la Base de Datos en el “Administrador de Conexiones”, ver gráfico 43.

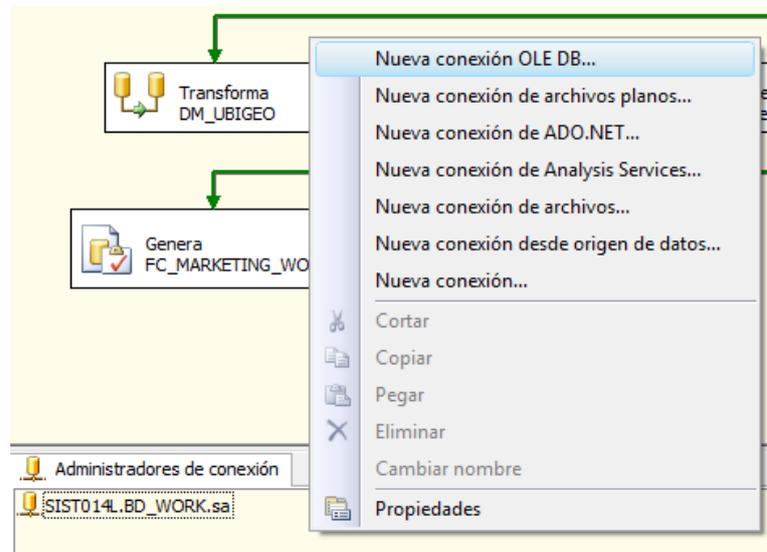
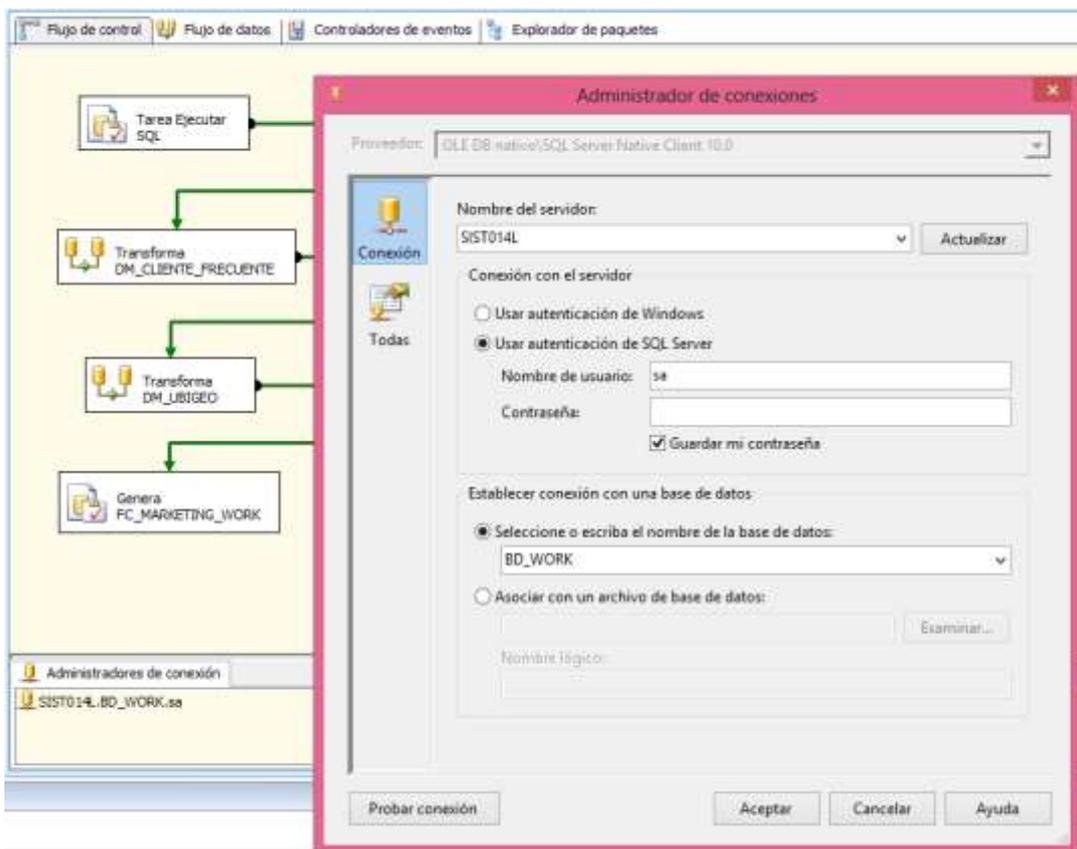


Gráfico 43: Administrador de Conexiones



2. Se procede a generar el flujo de trabajo para la transformación de las tablas maestras y de ventas.

3. Se genera la tarea de ejecución SQL que realiza la acción de truncar todas las tablas de trabajo dimensional y de hechos, indicadas en Especificación Técnica Detallada – numeral 2) a. Ver gráfico 44.

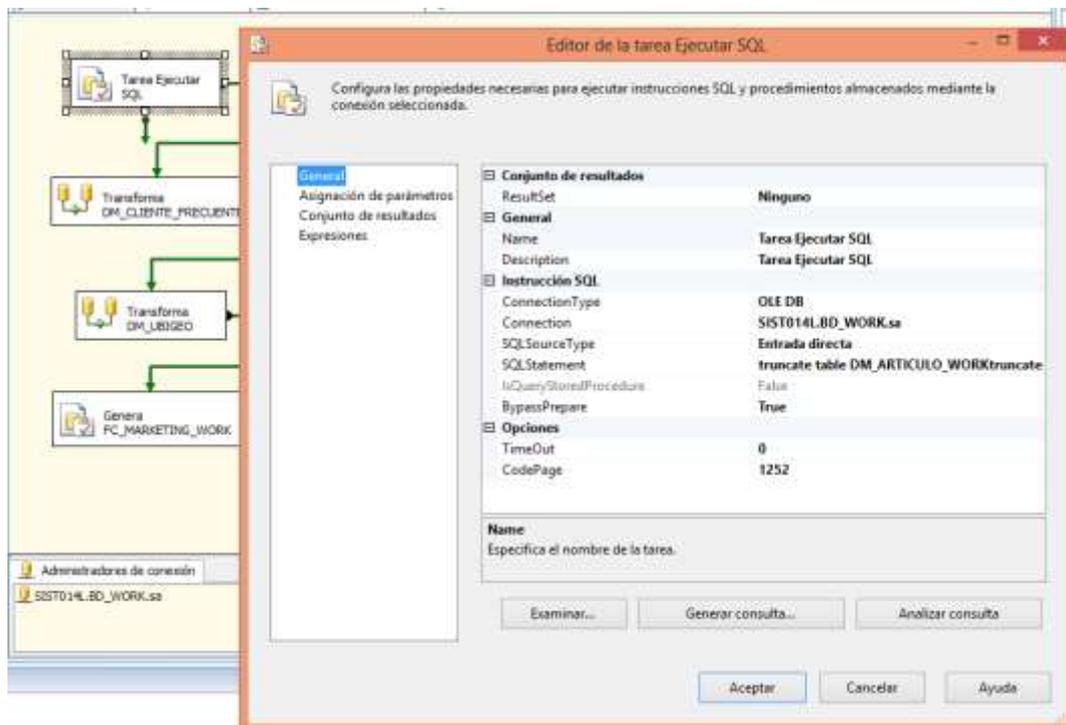


Gráfico 44: Truncar las tablas de trabajo dimensional y de hechos

4. Se actualizan los datos de tabla clifrecuente para evitar caracteres especiales, validar que campos no contengan valores como #,\$,& y otros que puedan ocasionar errores de inserción de datos. Ver gráfico 45.

```

Escribir consulta SQL

-- Evita conflictos por tipo de datos en cliente frecuente
update clifrecuente
set codpais='00'
where ISNUMERIC(codpais)=0

update clifrecuente
set coddpto='00'
where ISNUMERIC(coddpto)=0

update clifrecuente
set codprov='00'
where ISNUMERIC(codprov)=0

update clifrecuente
set coddist='00'
where ISNUMERIC(coddist)=0

update clifrecuente set codpais=
case when LEN(isnull(trim(trim(clifrecuente.codpais)), '00'))=1 then '0'+isnull(trim(trim(clifrecuente.codpais)), '00') else isnull(trim(trim(clifrecuente.codpais)), '00') end
where LEN(isnull(trim(trim(clifrecuente.codpais)), '00'))=1

```

Gráfico 45: Actualización de datos Tabla Cliente

5. Se procede a generar el flujo de control del DTS con un nuevo flujo de trabajo para cada entidad a seleccionar y generar las de trabajo dimensionales y de hecho, según Especificación Técnica Detallada – numeral 2 de la letra b a la letra g.

a. En el siguiente ejemplo se muestra el flujo de trabajo para la tabla de trabajo DM\_ARTICULO\_WORK. Ver gráfico 46.

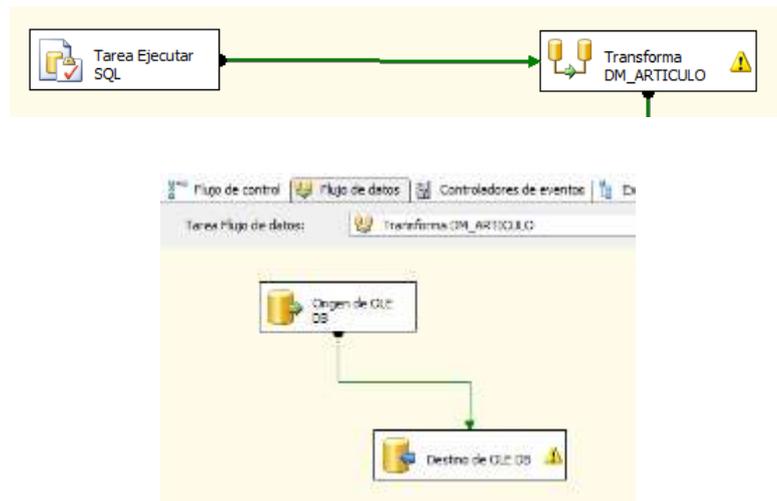


Gráfico 46: Flujo de trabajo para la tabla de trabajo dm\_articulo\_work

b. Se genera el origen de la conexión OLEDB, desde la base de datos de trabajo ( BD\_WORK) de la tabla articulos1. Ver gráfico 47.

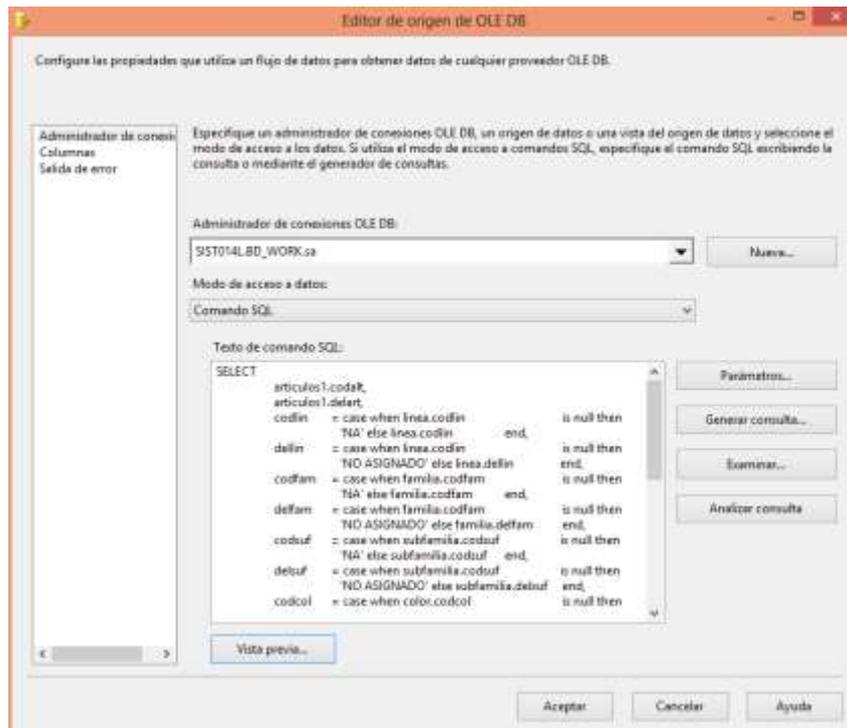


Gráfico 47: Conexión desde la base de datos de trabajo, de la tabla articulos1

- c. Se genera el destino de la conexión OLEDB, hacia la base de datos de trabajo (BD\_WORK) en la tabla de dimensión DM\_ARTICULO\_WORK. Ver gráfico 48.

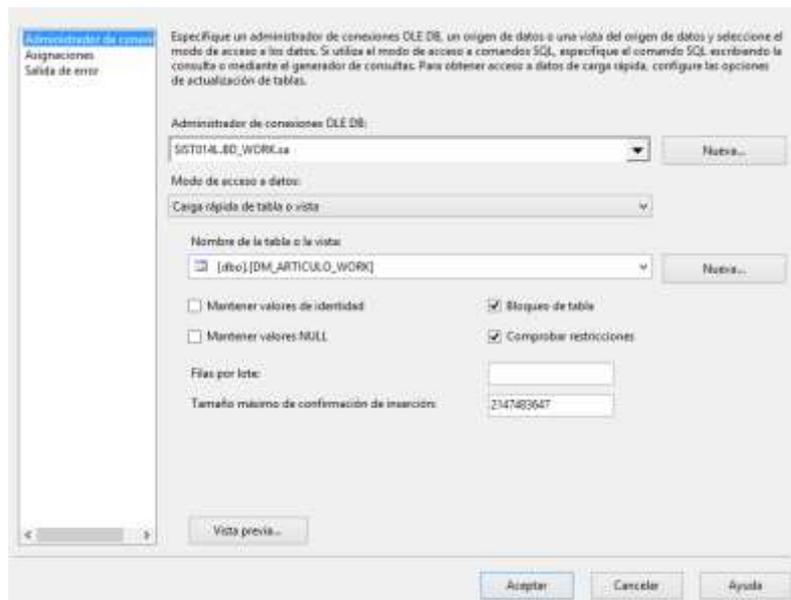


Gráfico 48: Conexión hacia la BD de Trabajo en la tabla de dimensión dm\_articulo\_work

d. Se realiza el mismo procedimiento para el resto de tablas de trabajo dimensionales y de hecho. Ver gráfico 49.

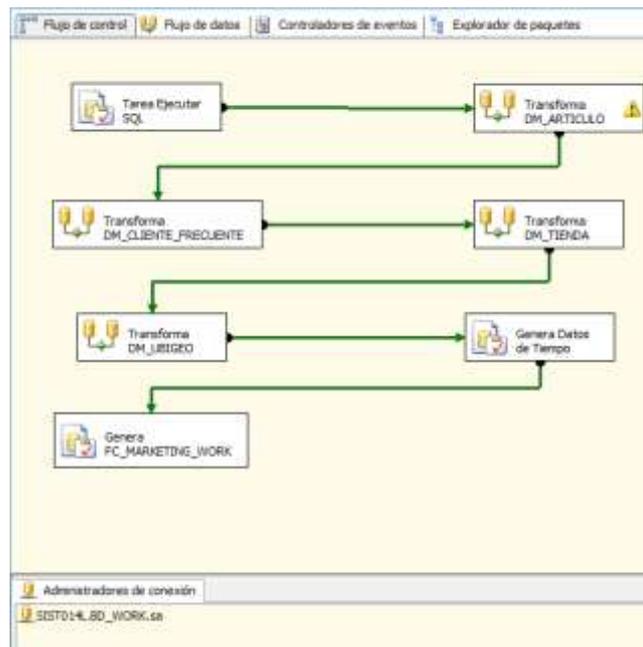


Gráfico 49 Flujo de trabajo de las tablas trabajo dimensionales y de hecho:

A continuación se detalla la descripción del caso de Uso del Sistema: **Cargar Información.**

## DESCRIPCION DEL CASO DE USO DEL SISTEMA CARGAR INFORMACION

### 1. DESCRIPCION CORTA

La fase de carga es el momento en el cual los datos de la fase de transformación son cargados en el Datawarehouse.

### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Servicio DTS

### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Ejecutar tarea de conexión al origen de datos de trabajo
2	S	Establece la conexión solicitada
3	A	Ejecutar tarea de conexión al origen de datos del Datawarehouse
4	S	Establece la conexión solicitada.
5	A	Eliminar los registros de las tablas dimensiones del datawarehouse.
6	S	Elimina los registros seleccionados.
7	A	Eliminar los registros de la tabla de hechos del datawarehouse
8	S	Eliminar los registros seleccionados.
9	A	Ejecutar tarea de almacenar dimensiones al Datawarehouse
10	S	Realiza almacenamiento de dimensiones.
11	A	Ejecutar tarea de almacenar tablas de hechos al Datawarehouse.
12	S	Realizar almacenamiento de tabla de hechos.

### 4. FLUJO ALTERNATIVO DE EVENTOS

**4.1. Paso 1.1:** Si el origen de datos no está disponible

Paso	Tipo	Descripción del paso
1	S	Registrar Log de Errores de conexión

### 5. PRECONDICIONES

Ítem	Precondiciones
1	Haber ejecutado el Caso de Uso : Transformar Información

## 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para el Analysis Services.

## 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Cargar Información tendrá una frecuencia diaria.

A continuación se detalla la Especificación Técnica detallada del Caso de Uso: Cargar Información, tal como se visualiza en la Tabla 8.

## Especificación Técnica detallada del CUS: Cargar Información

1. El proceso de carga se inicia con establecer la conexión a la base de datos del datawarehouse que permitirá almacenar la información final de las dimensiones y la tabla de hechos.

2. Una vez establecida la conexión se ejecuta la tarea de carga de los datos, esta tarea realiza las siguientes acciones:

a. Eliminar los registros de las dimensiones y la tabla de hechos

Eliminar registros de tabla DM\_ARTICULO\_

Eliminar registros de tabla DM\_CLIENTE\_FRECUENTE\_WORK

Eliminar registros de tabla DM\_TIENDA\_WORK

Eliminar registros de tabla DM\_UBIGEO\_WORK

Eliminar registros de tabla DM\_TIEMPO\_WORK

Eliminar registros de tabla FC\_MARKETING\_WORK

b. Se inicia la selección de la información de la dimensión DM\_ARTICULO\_WORK para generar la tabla de tipo dimensión para los artículos en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo		Tabla Destino	Campo
DM_ARTICULO_WORK	codalt		DM_ARTICULO	codalt
DM_ARTICULO_WORK	delart		DM_ARTICULO	delart
DM_ARTICULO_WORK	pais		DM_ARTICULO	pais
DM_ARTICULO_WORK	codlin		DM_ARTICULO	codlin
DM_ARTICULO_WORK	dellin		DM_ARTICULO	dellin
DM_ARTICULO_WORK	codfam		DM_ARTICULO	codfam
DM_ARTICULO_WORK	delfam		DM_ARTICULO	delfam
DM_ARTICULO_WORK	codsuf		DM_ARTICULO	codsuf
DM_ARTICULO_WORK	delsuf		DM_ARTICULO	delsuf
DM_ARTICULO_WORK	codcol		DM_ARTICULO	codcol
DM_ARTICULO_WORK	delcol		DM_ARTICULO	delcol
DM_ARTICULO_WORK	codtip		DM_ARTICULO	codtip
DM_ARTICULO_WORK	destip		DM_ARTICULO	destip
DM_ARTICULO_WORK	coduneg		DM_ARTICULO	coduneg
DM_ARTICULO_WORK	desuneg		DM_ARTICULO	desuneg

Tabla 8 Especificación Técnica detallada del CUS: Cargar Información

## Especificación Técnica detallada del CUS: Cargar Información

- c. Se inicia la selección de la información de la dimensión DM\_TIENDA\_WORK para generar la tabla de tipo dimensión para las tiendas en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo	Tabla Destino	Campo
DM_TIENDA_WORK	ptovta	DM_TIENDA	ptovta
DM_TIENDA_WORK	nompvt	DM_TIENDA	nompvt
DM_TIENDA_WORK	codest	DM_TIENDA	codest
DM_TIENDA_WORK	nombre	DM_TIENDA	nombre

- d. Se inicia la selección de la información de la dimensión DM\_TIEMPO\_WORK para generar la tabla de tipo dimensión para los datos de fechas en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo	Tabla Destino	Campo
DM_TIEMPO_WORK	time_id	DM_TIEMPO	time_id
DM_TIEMPO_WORK	the_date	DM_TIEMPO	the_date
DM_TIEMPO_WORK	the_day	DM_TIEMPO	the_day
DM_TIEMPO_WORK	The_month	DM_TIEMPO	The_month
DM_TIEMPO_WORK	the_year	DM_TIEMPO	the_year
DM_TIEMPO_WORK	day_of_month	DM_TIEMPO	day_of_month
DM_TIEMPO_WORK	week_of_year	DM_TIEMPO	week_of_year
DM_TIEMPO_WORK	Month_of_year	DM_TIEMPO	Month_of_year
DM_TIEMPO_WORK	Quarter	DM_TIEMPO	Quarter
DM_TIEMPO_WORK	Fiscal_period	DM_TIEMPO	Fiscal_period

- e. Se inicia la selección de la información de la dimensión DM\_CLIENTE\_FRECUENTE\_WORK para generar la tabla de tipo dimensión para los clientes frecuentes en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla 8: Especificación Técnica detallada del CUS: Cargar Información (continuación)

## Especificación Técnica detallada del CUS: Cargar Información

Tabla Origen	Campo	Tabla Destino	Campo
DM_CLIENTE_FRECUENTE_WORK	clifre	DM_CLIENTE_FRECUENTE	clifre
DM_CLIENTE_FRECUENTE_WORK	rzsfre	DM_CLIENTE_FRECUENTE	rzsfre
DM_CLIENTE_FRECUENTE_WORK	Edad	DM_CLIENTE_FRECUENTE	Edad
DM_CLIENTE_FRECUENTE_WORK	ubigeo	DM_CLIENTE_FRECUENTE	ubigeo
DM_CLIENTE_FRECUENTE_WORK	sexo	DM_CLIENTE_FRECUENTE	sexo
DM_CLIENTE_FRECUENTE_WORK	ocupación	DM_CLIENTE_FRECUENTE	ocupación

- c. Se inicia la selección de la información de la dimensión DM\_UBIGEO\_WORK para generar la tabla de tipo dimensión para el ubigeo en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla Origen	Campo	Tabla Destino	Campo
DM_UBIGEO_WORK	ubigeo	DM_UBIGEO	ubigeo
DM_UBIGEO_WORK	codpais	DM_UBIGEO	codpais
DM_UBIGEO_WORK	coddpto	DM_UBIGEO	coddpto
DM_UBIGEO_WORK	codprov	DM_UBIGEO	codprov
DM_UBIGEO_WORK	coddist	DM_UBIGEO	coddist
DM_UBIGEO_WORK	desdist	DM_UBIGEO	desdist
DM_UBIGEO_WORK	despais	DM_UBIGEO	despais
DM_UBIGEO_WORK	desdpto	DM_UBIGEO	desdpto
DM_UBIGEO_WORK	desprov	DM_UBIGEO	desprov

- d. Se inicia la selección de la información de la tabla de hechos FC\_MARKETING\_WORK para generar la tabla de hechos para las ventas en el datawarehouse. La carga se realiza con el siguiente origen de datos y destino de datos:

Tabla 8: Especificación Técnica detallada del CUS: Cargar Información (continuación)

## Especificación Técnica detallada del CUS: Cargar Información

Tabla Origen	Campo	Tabla Destino	Campo
FC_MARKETING_WORK	Ptovta	FC_MARKETING	Ptovta
FC_MARKETING_WORK	time_id	FC_MARKETING	time_id
FC_MARKETING_WORK	Clifre	FC_MARKETING	Clifre
FC_MARKETING_WORK	Codalt	FC_MARKETING	Codalt
FC_MARKETING_WORK	unidades_ventas	FC_MARKETING	unidades_ventas
FC_MARKETING_WORK	valor_venta	FC_MARKETING	valor_venta
FC_MARKETING_WORK	importe_iva	FC_MARKETING	importe_iva
FC_MARKETING_WORK	importe_total	FC_MARKETING	importe_total
FC_MARKETING_WORK	valor_venta_me	FC_MARKETING	valor_venta_me
FC_MARKETING_WORK	Importe_iva_me	FC_MARKETING	Importe_iva_me
FC_MARKETING_WORK	Importe_total_me	FC_MARKETING	Importe_total_me
FC_MARKETING_WORK	costo_venta	FC_MARKETING	costo_venta
FC_MARKETING_WORK	importe_descuento	FC_MARKETING	importe_descuento
FC_MARKETING_WORK	Costo_venta_me	FC_MARKETING	Costo_venta_me
FC_MARKETING_WORK	Importe_descuento_me	FC_MARKETING	Importe_descuento_me

Tabla 8: Especificación Técnica detallada del CUS: Cargar Información (continuación)

## DIAGRAMA DE CLASES: CARGAR INFORMACION (DESTINO)

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de trabajo finales del Datawarehouse. A continuación se muestra el Diagrama de Clases: Cargar Información. Ver Gráfico 50.

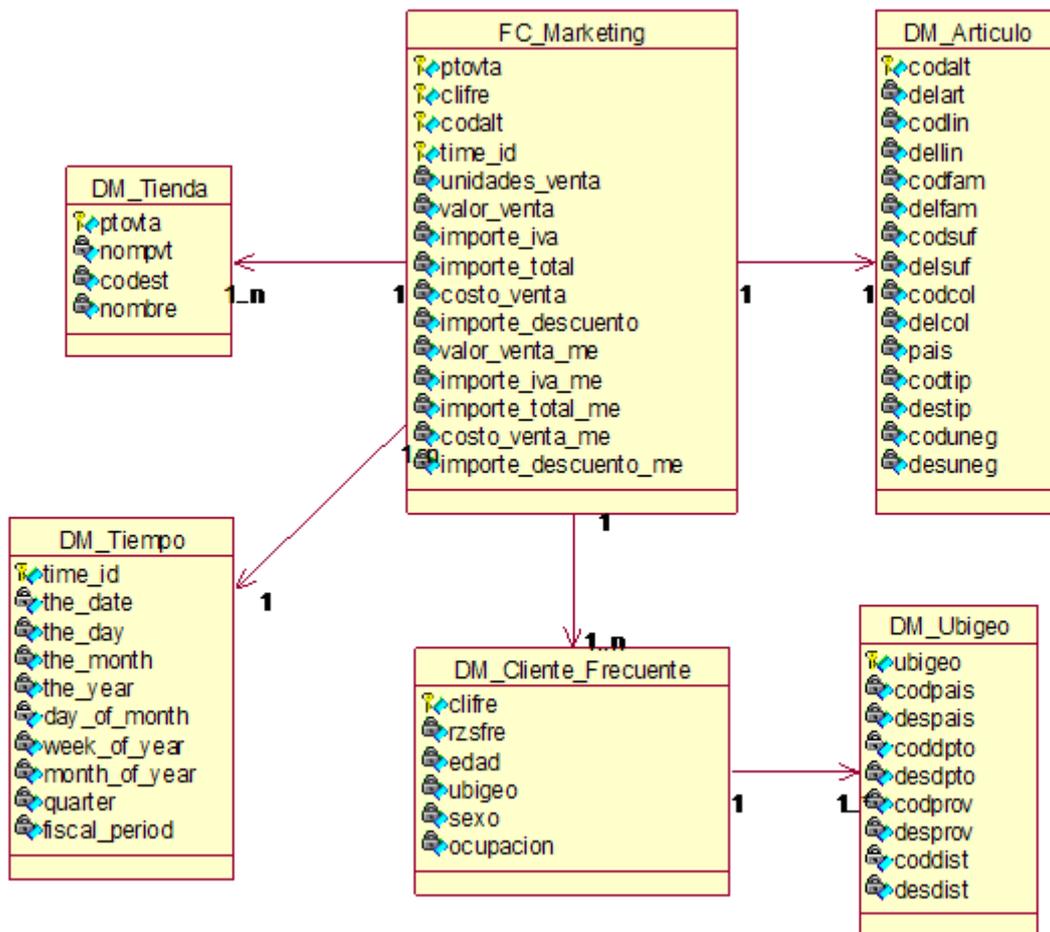


Gráfico 50: Diagrama de Clases: Cargar Información

## **DIAGRAMA DE SECUENCIA CUS: CARGAR INFORMACION**

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Cargar Información. Ver gráfico 51.

.

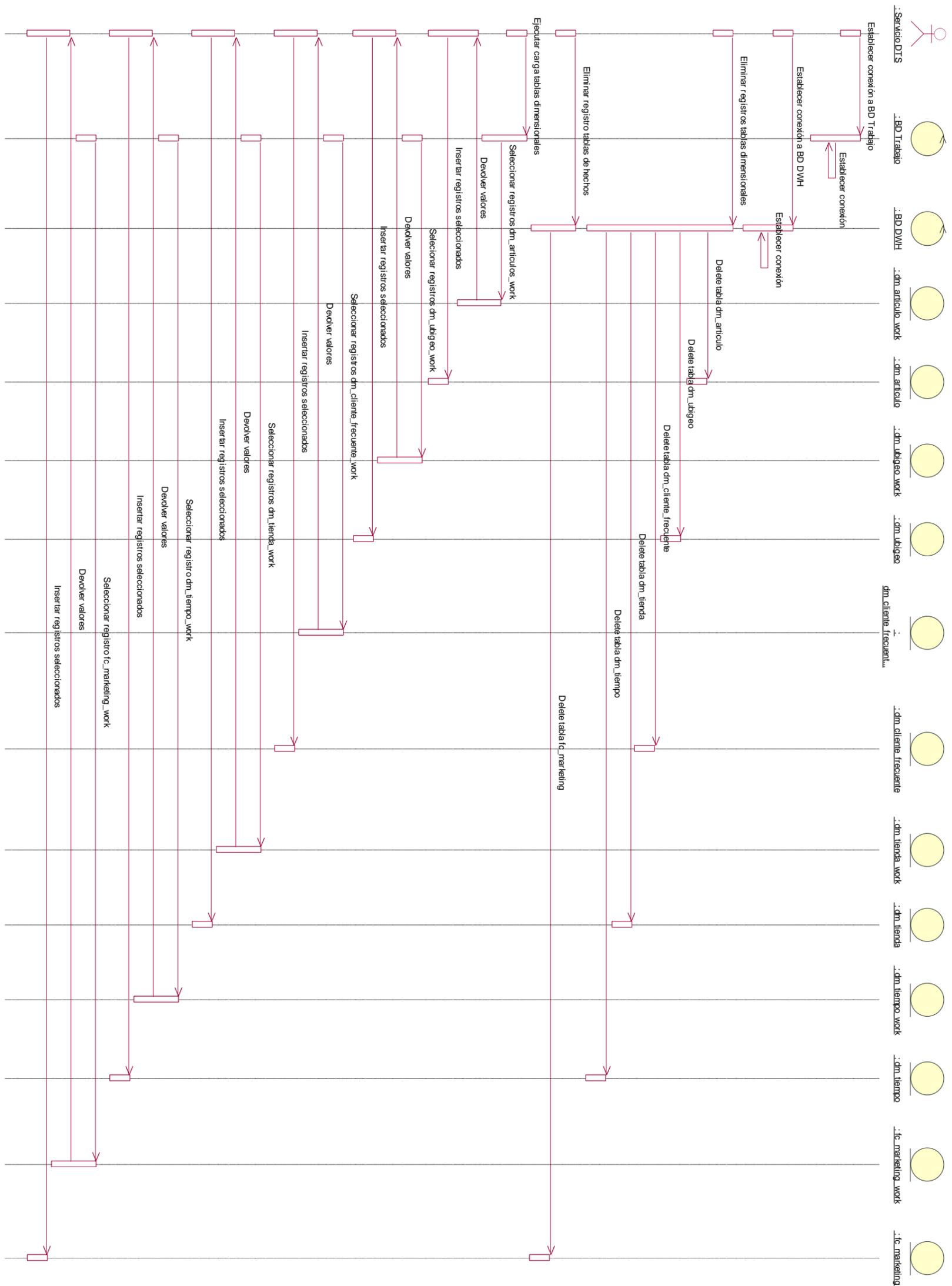


Gráfico 51: Diagrama de Secuencia: Cargar Información

## Construcción del CUS: Cargar Información

1. Se procede a crear la conexión de la BD en el “Administrador de Conexiones”. Ver gráfico 52.

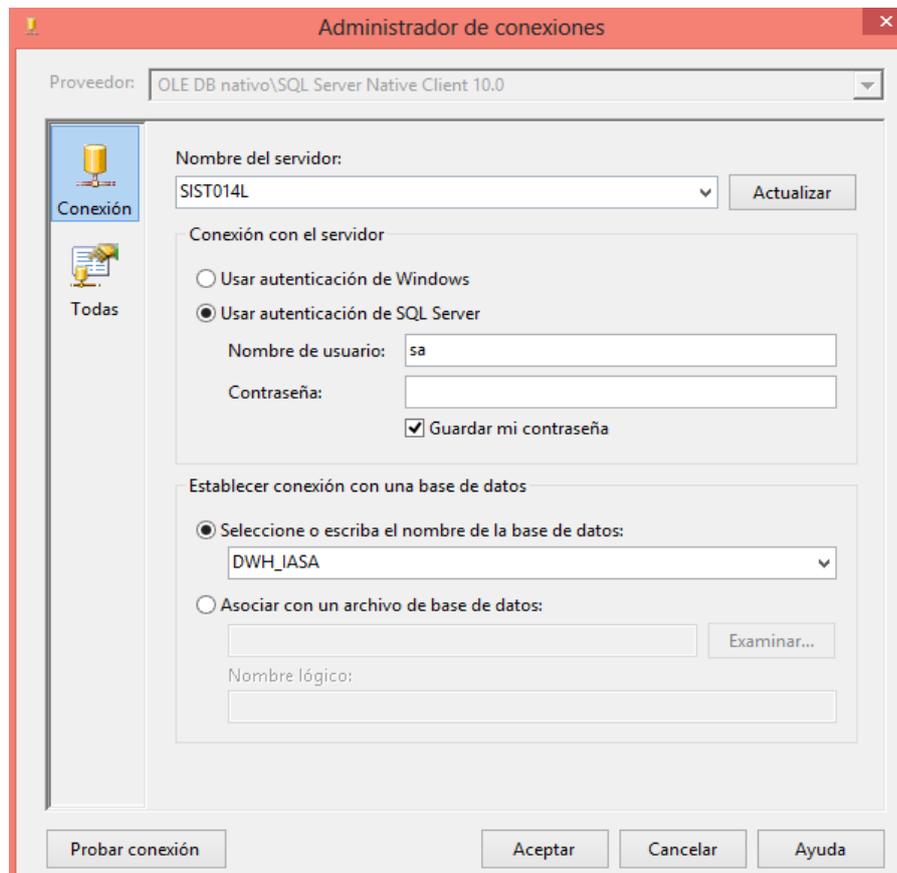
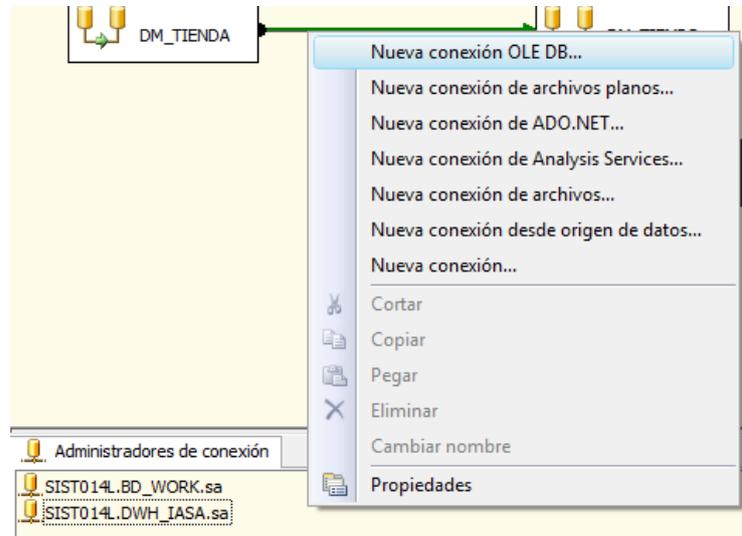


Gráfico 52: Administrador de Conexiones

2. Se procede a generar el flujo de trabajo para la carga de información.
3. Se genera la tarea de ejecución SQL que realiza la acción de eliminar los registros de las tablas de trabajo dimensional y de hechos, indicadas en la Especificación Técnica Detallada – numeral 2) a. Ver gráfico 53.

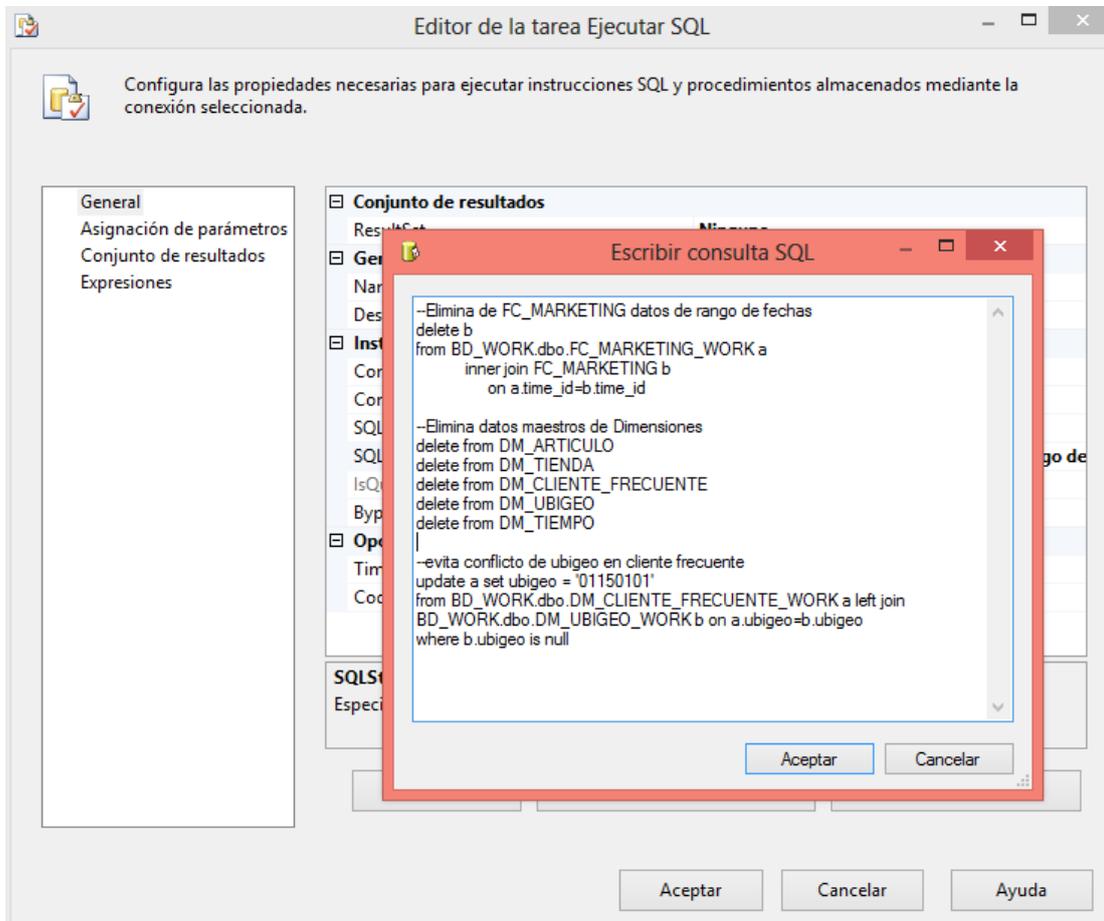


Gráfico 53: Eliminación de registros de las tablas de trabajo dimensional y de hechos

4. Se procede a generar el flujo de control del DTS con un nuevo flujo de trabajo para cada entidad a seleccionar y generar las tablas dimensionales y de hecho, según la Especificación Técnica Detallada numeral 2 de la letra b a la letra g.

- a. En el siguiente ejemplo se muestra el flujo de trabajo para la tabla dimensional DM\_ARTICULO. Ver gráfico 54.

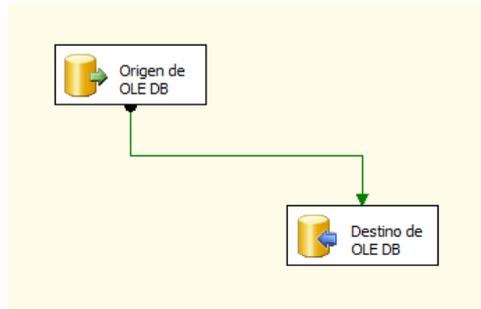


Gráfico 54: Flujo de trabajo para la tabla dimensional dm\_artículo

- b. Se genera el origen de la conexión OLEDB, desde la base de datos de trabajo (BD\_WORK) de la tabla DM\_ARTICULO\_WORK. Ver gráfico 55.

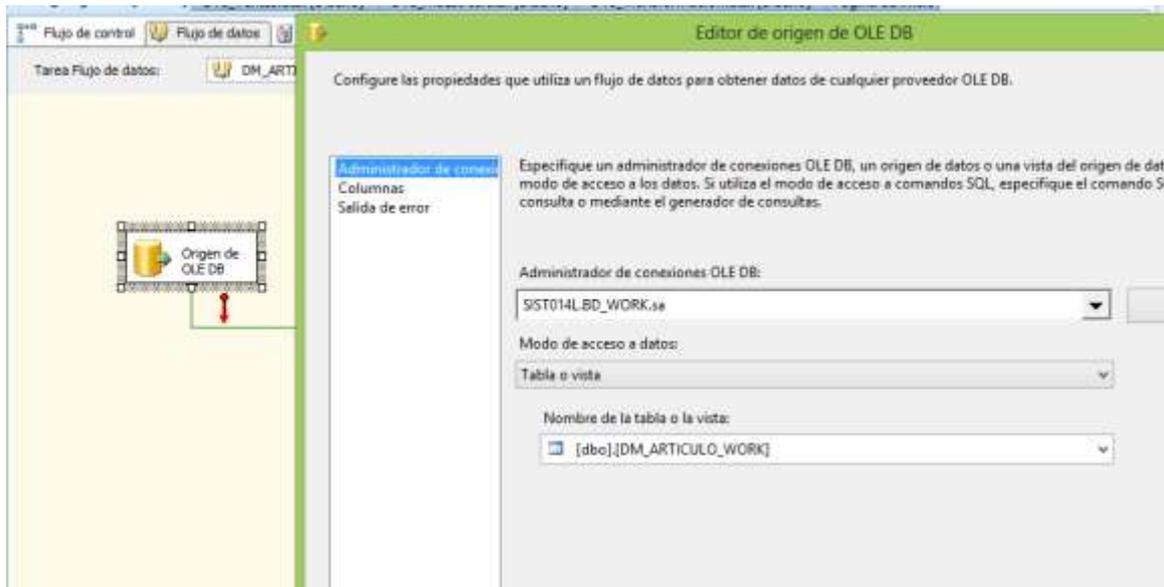


Gráfico 55: Conexión desde la base de datos de trabajo, de la tabla dm\_articulo\_work

- c. Se genera el destino de la conexión OLEDB, hacia el Datawarehouse (DWH\_IASA) en la tabla dimensión DM\_ARTICULO. Ver gráfico 56.

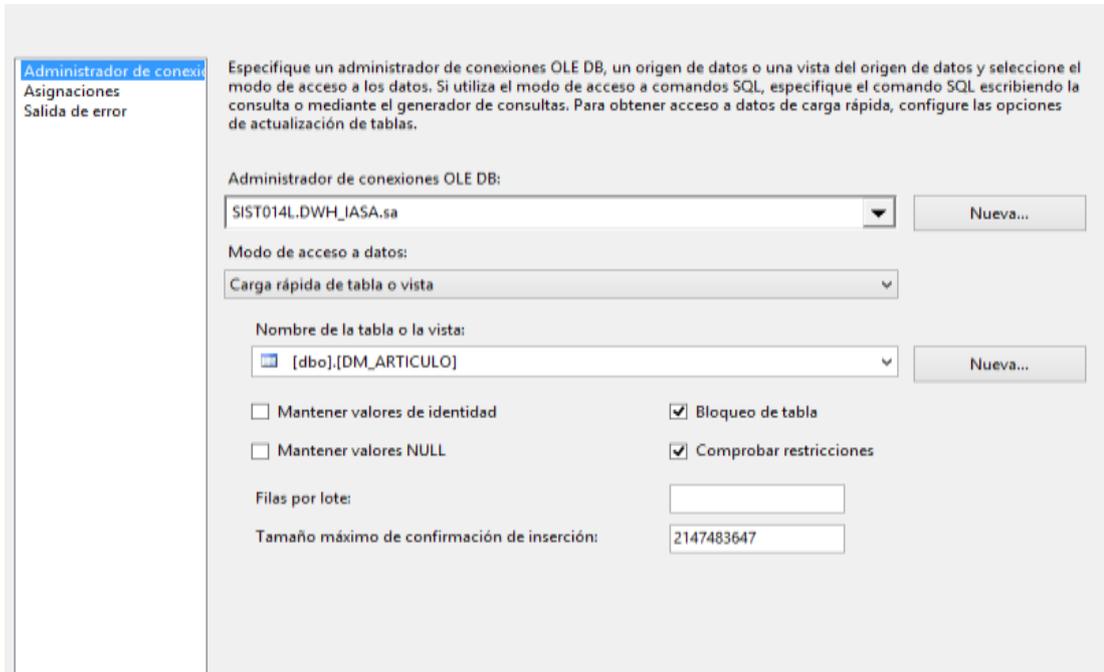


Gráfico 56: Conexión hacia el Datawarehouse en la tabla dimensión dm\_articulo

- d. Se realiza el mismo procedimiento para el resto de tablas de trabajo dimensionales y de hecho. Ver gráfico 57.

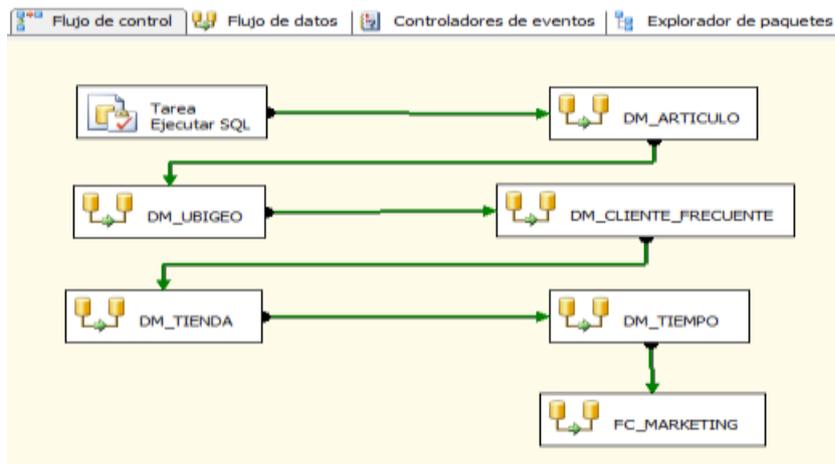


Gráfico 57: Flujo de trabajo para las tablas de trabajo dimensionales y de hecho

A continuación se detalla la descripción del caso de Uso del Sistema: Generar Clasificación de Clientes.

## DESCRIPCION DEL CASO DE USO DEL SISTEMA

### GENERAR CLASIFICACION DE CLIENTES

#### 1. DESCRIPCION CORTA

Este proceso prepara, sondea y explora los datos para obtener la información de clasificación de clientes a través del algoritmo de Árbol de Decisión.

#### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Analysis Services

#### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Realizar selección de información de clientes y ventas al origen de datos.
2	S	Realizar la selección de datos solicitada.
3	A	Generar la estructura de minería de datos la cual está basada en las tablas y columnas del datawarehouse.
4	S	Realizar la generación de estructura de minería de datos.
5	A	Utilizar las variables de predicción y las variables de entrada en el modelo de minería de datos.
6	S	Realizar la selección de las variables de predicción y entrada.
7	A	Ejecutar la técnica de minería de datos: árbol de decisión.
8	S	Realizar la ejecución de la técnica solicitada.

#### 4. FLUJO ALTERNATIVO DE EVENTOS

Paso	Tipo	Descripción del paso
1	S	

#### 5. PRECONDICIONES

Ítem	Precondiciones
1	Haber ejecutado el proceso de ETL

#### 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para la consulta del reporte.

#### 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Generar Clasificación de clientes será diaria.

## DIAGRAMA DE CLASES: GENERAR CLASIFICACION DE CLIENTES

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de Datawarehouse. A continuación se muestra el Diagrama de Clases: Generar Clasificación de Clientes. Ver Gráfico 58.

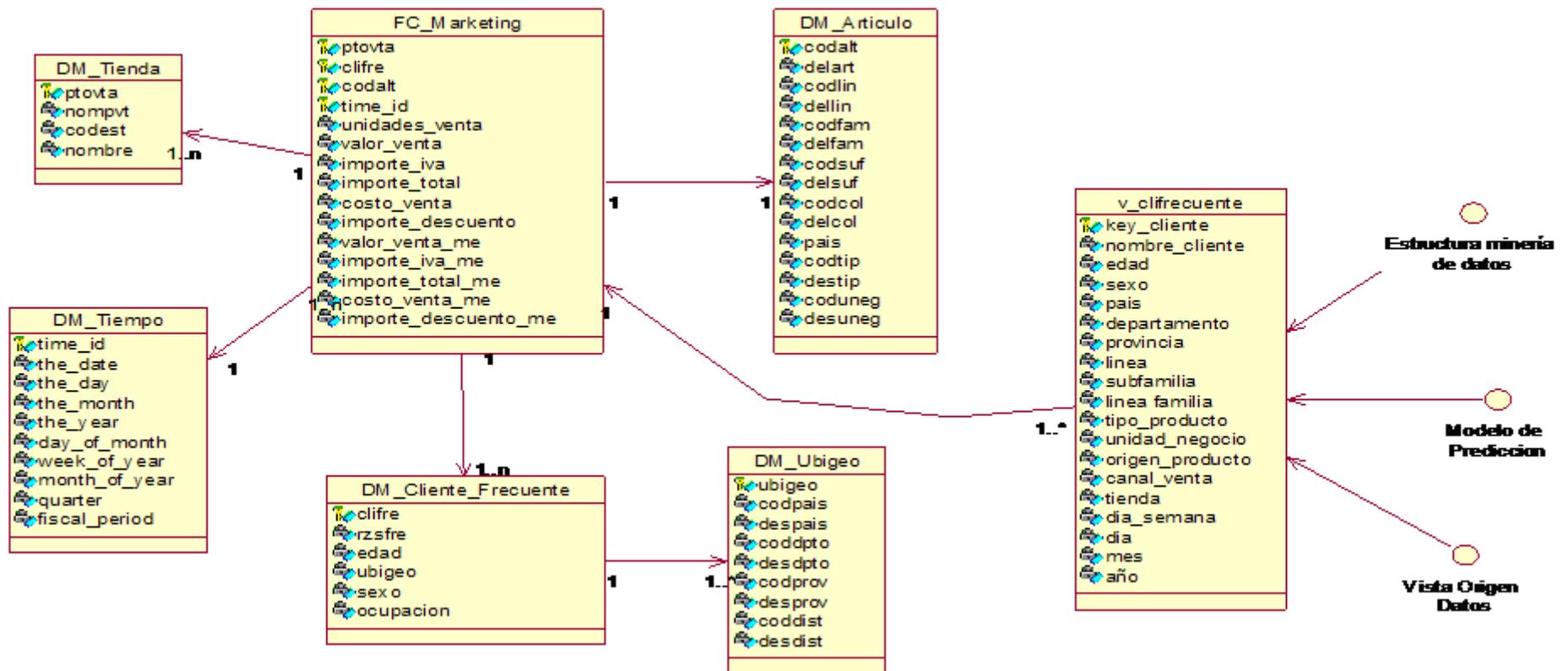


Gráfico 58: Diagrama de Clases: Generar Clasificación de Clientes

## DIAGRAMA DE SECUENCIA CUS: GENERAR CLASIFICACION DE CLIENTES

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Generar Clasificación de Clientes. Ver gráfico 59.

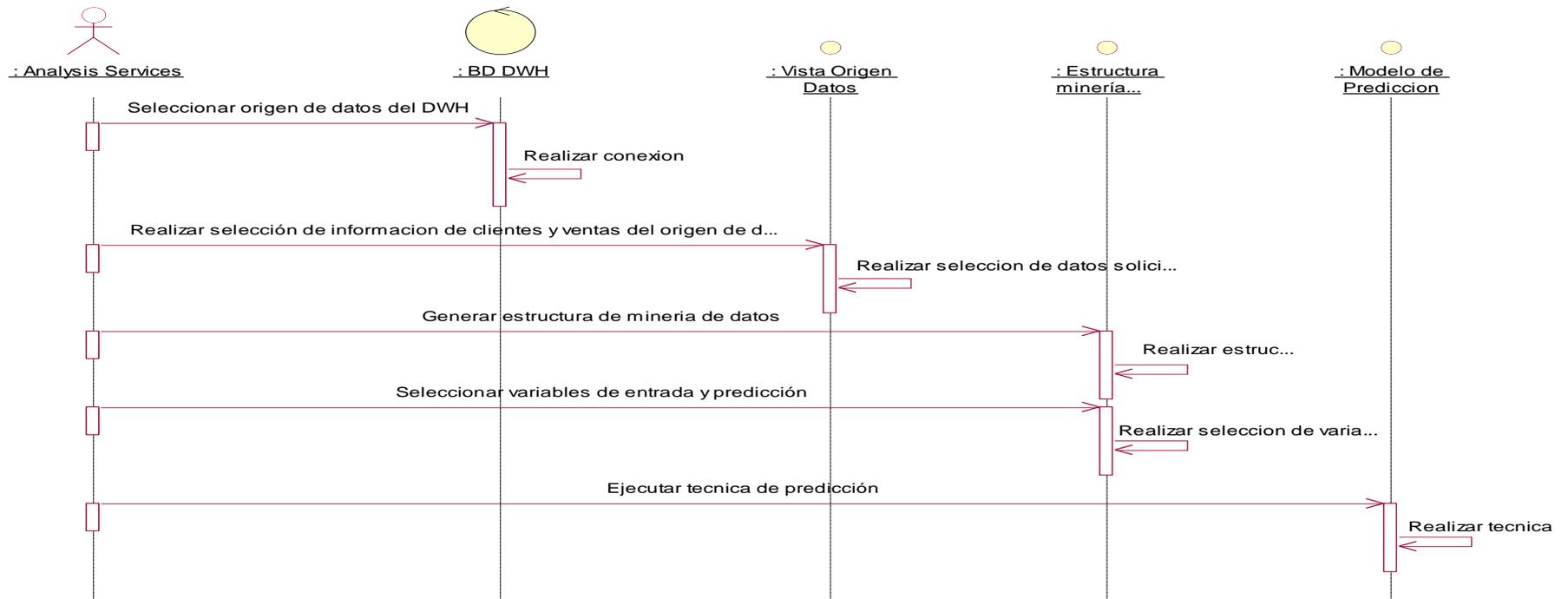


Gráfico 59: Diagrama de Secuencia: Generar Clasificación de Clientes

## Construcción del CUS: Generar Clasificación de Clientes

Se usará como herramienta para ejecutar las tareas descritas, SQL Server Business Intelligence Development Studio, en el cual se creará un nuevo proyecto. Ver gráfico 60.

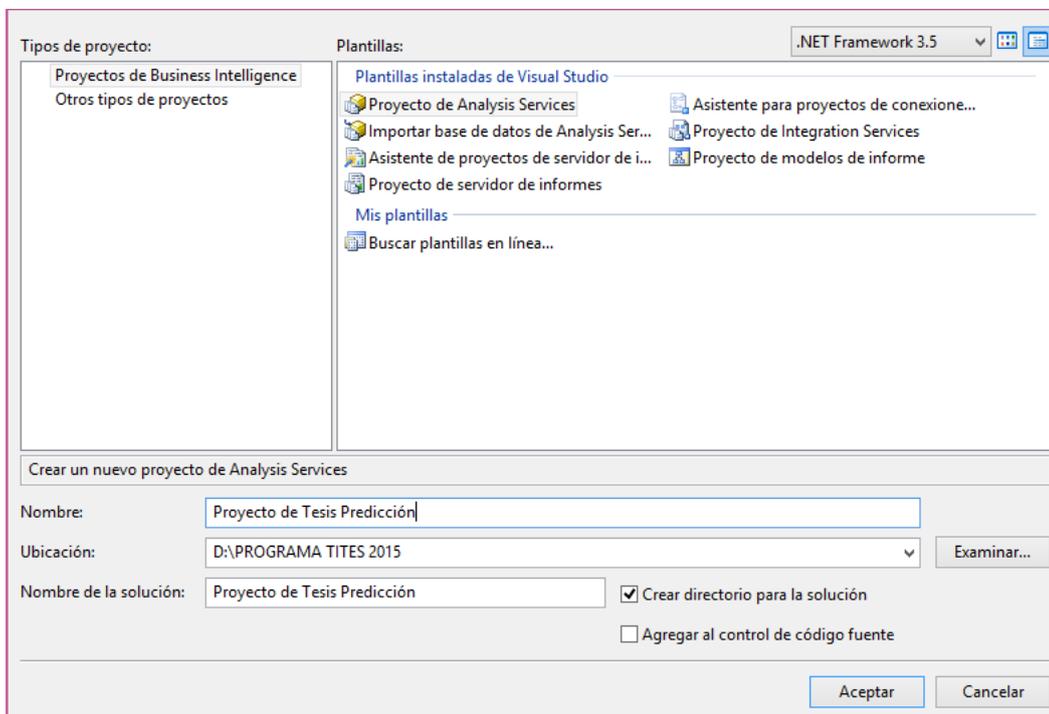


Gráfico 60: Creación de Proyecto

1. Crearemos la conexión con el origen de datos ( datawarehouse ) para nuestro modelo de predicción. Ver gráfico 61.

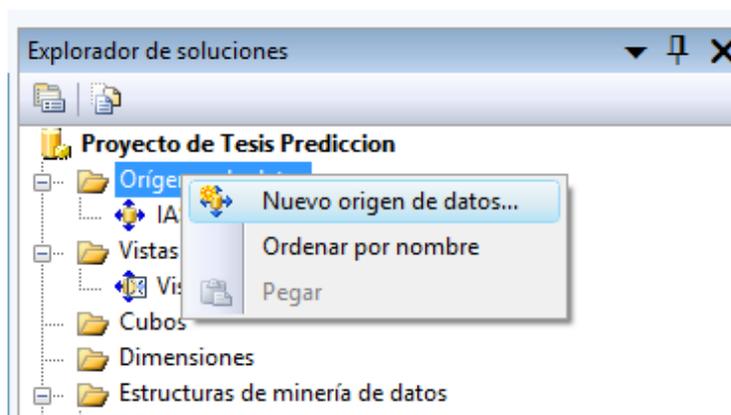


Gráfico 61: Creación al nuevo origen de datos

a. Iniciamos el asistente para orígenes de datos. Ver gráfico 62.

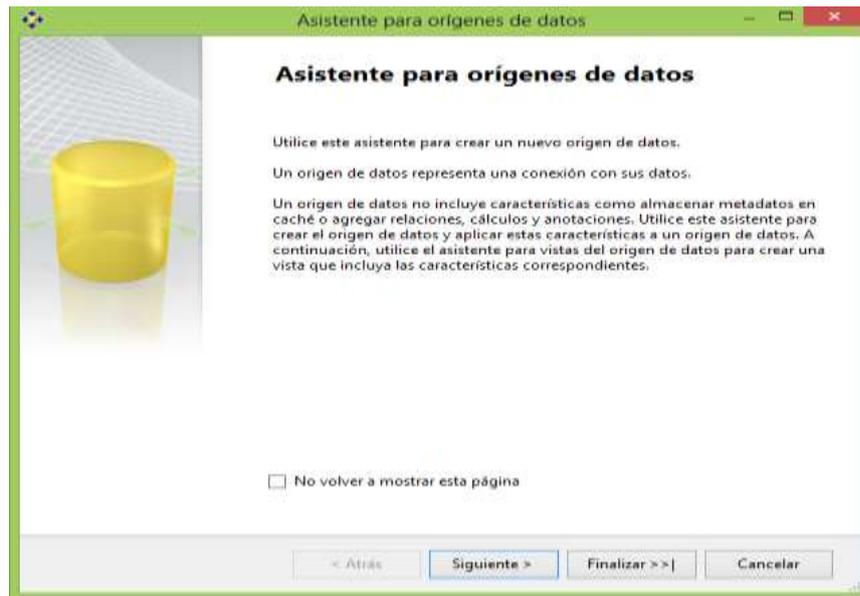


Gráfico 62: Asistente para la creación del origen de datos

b. Seleccionamos la base de datos origen, en este caso DWH\_IASA( datawarehouse). Ver gráfico 63.

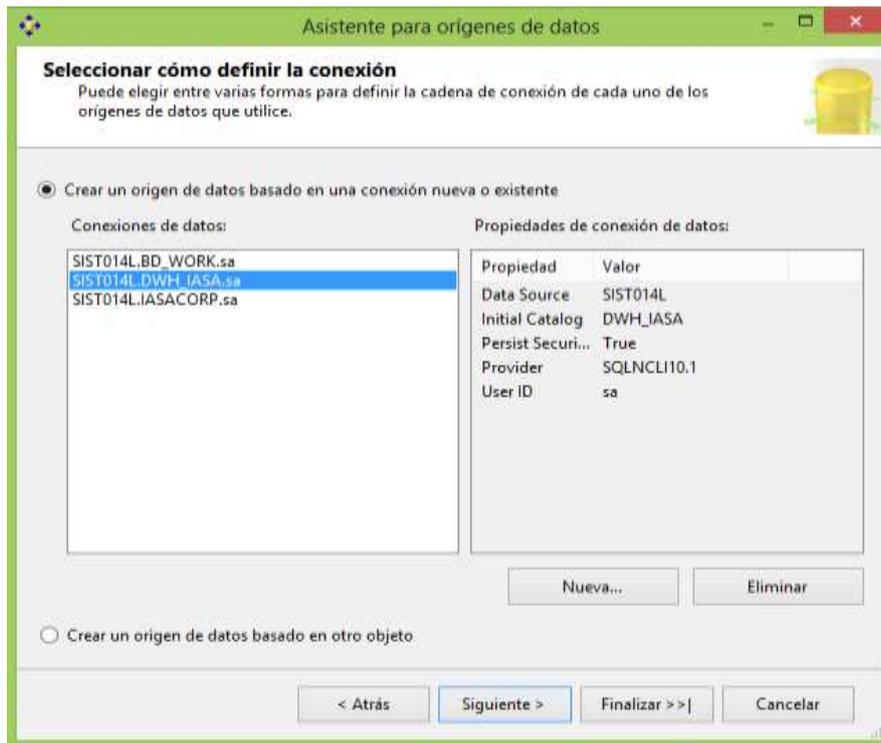


Gráfico 63: Selección de la base de datos origen

c. Ingresamos el nombre para nuestro origen de datos. Ver gráfico 64.

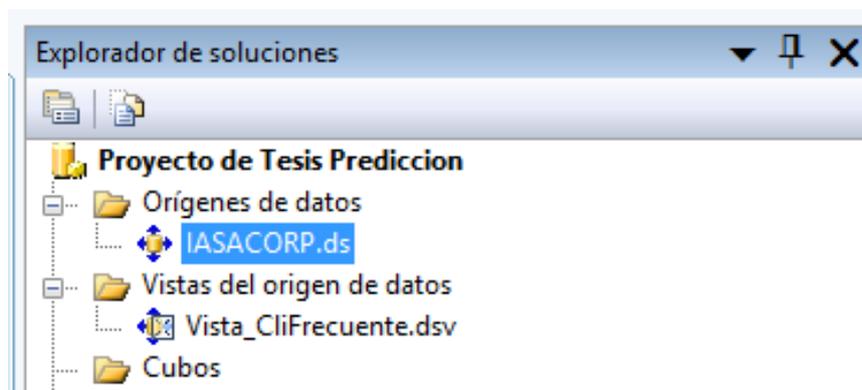
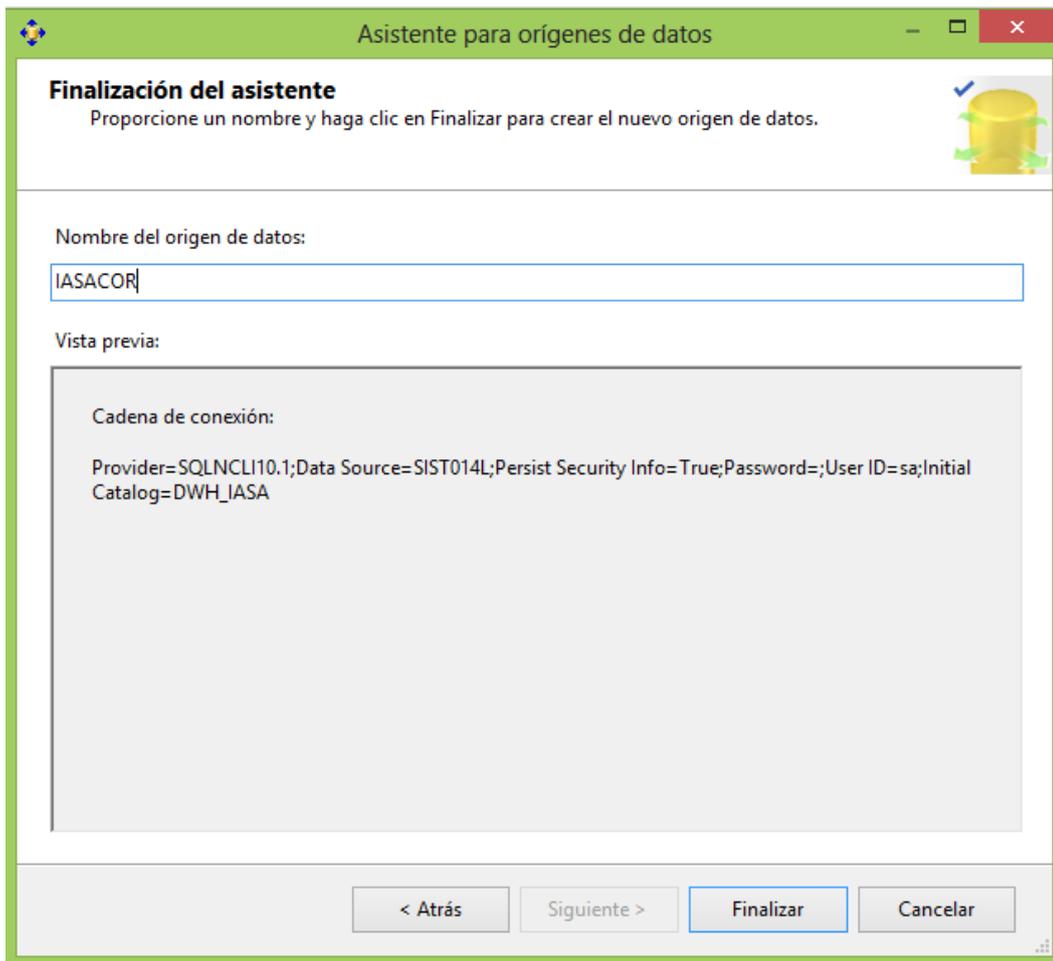


Gráfico 64: Generación de la conexión del origen de datos

2. Creamos la vista del origen de datos que contendrá el modelo lógico de nuestra tabla dimensional con la que trabajaremos las variables para armar nuestra estructura de minería de datos. Ver gráfico 65.

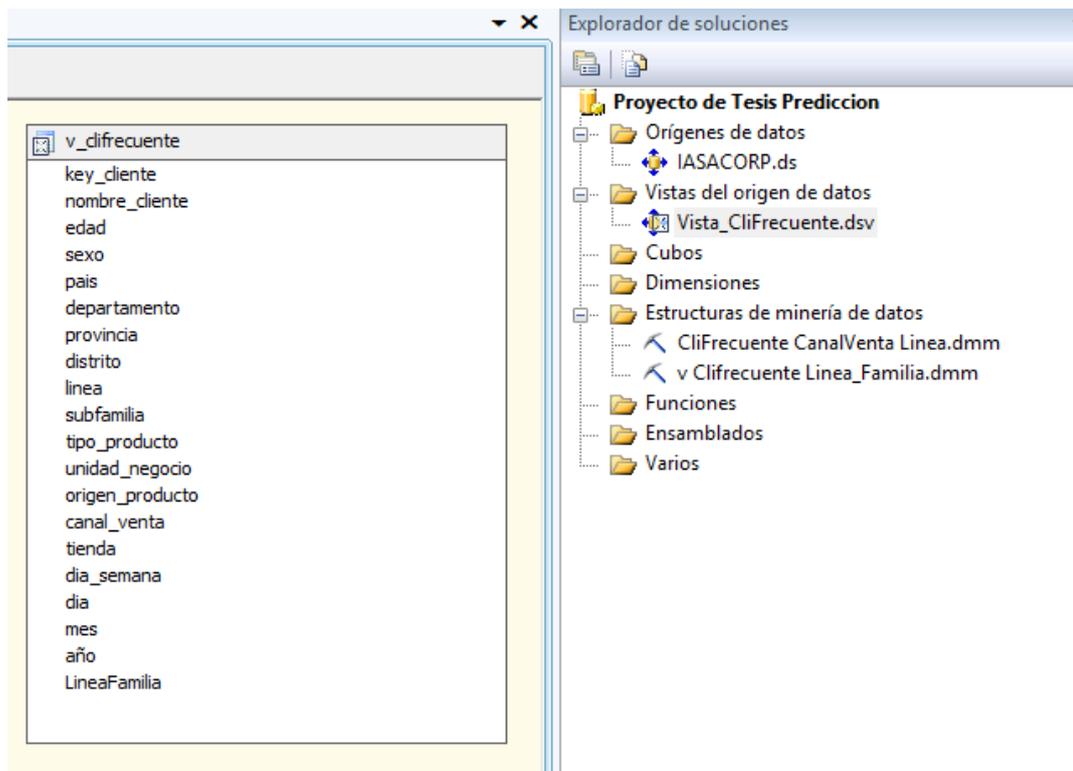


Gráfico 65: Creación de la Vista de Origen de Datos

- a. Esta vista la creamos a partir del siguiente script en el analizador de SQL Server generandose como objeto la vista, en el script realizamos el filtro de todos los clientes frecuentes registrados en la base de datos transaccional.

```

alter view v_clifrecuente
As
select distinct a.clifre as key_cliente,b.rzsfre as nombre_cliente,b.edad as edad,b.sexo as
sexo,c.despais as pais,c.desdpto as departamento,
c.desprov as provincia,c.desdist as distrito,
d.dellin as linea,d.delfam as familia,d.delsuf as subfamilia,d.destip as
tipo_producto,
d.desuneg as unidad_negocio,d.pais as origen_producto,
e.nombre as canal_venta,e.nompvt as tienda,f.the_day as
dia_semana,f.day_of_month as dia,f.the_month as mes,f.the_year as año
from FC_MARKETING a
inner join DM_CLIENTE_FRECUENTE b on b.clifre = a.clifre
inner join DM_UBIGEO c on b.ubigeo = c.ubigeo
inner join DM_ARTICULO d on d.codalt = a.codalt
inner join DM_TIENDA e on e.ptovta = a.ptovta
inner join DM_TIEMPO f on f.time_id = a.time_id
where a.clifre <> '0000001'

-- select * from v_clifrecuente

```

1. Creamos la estructura de minería de datos en base a las consultas o reportes con las variables de entrada y predicción que deseamos analizar.
  - a. Como primera estructura creamos la consulta de Clientes Frecuentes por Línea y Canal de Venta, seleccionamos la opción Nueva estructura de minería de datos. Ver gráfico 66.

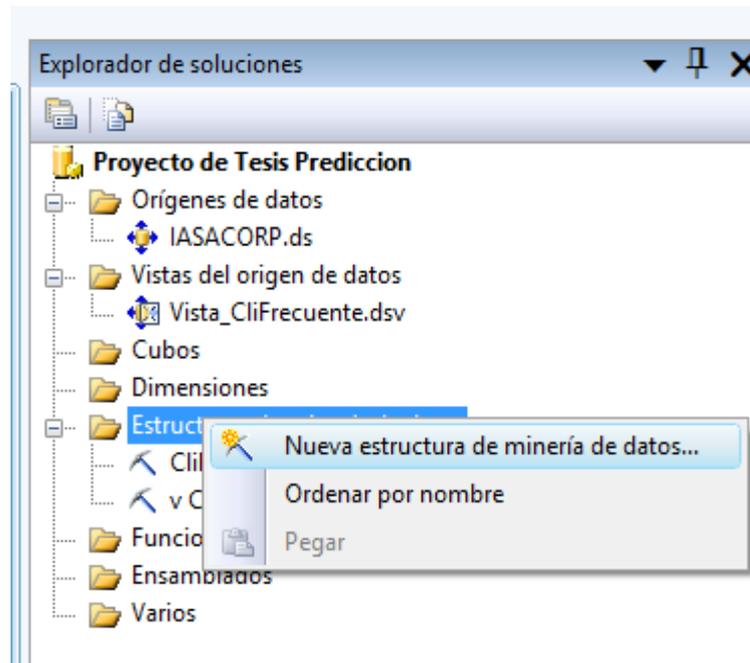


Gráfico 66: Creación de la estructura de minería de datos: Por Canal de Venta y Línea

- b. Iniciamos el asistente para minería de datos dando click en Siguiente. Ver gráfico 67.

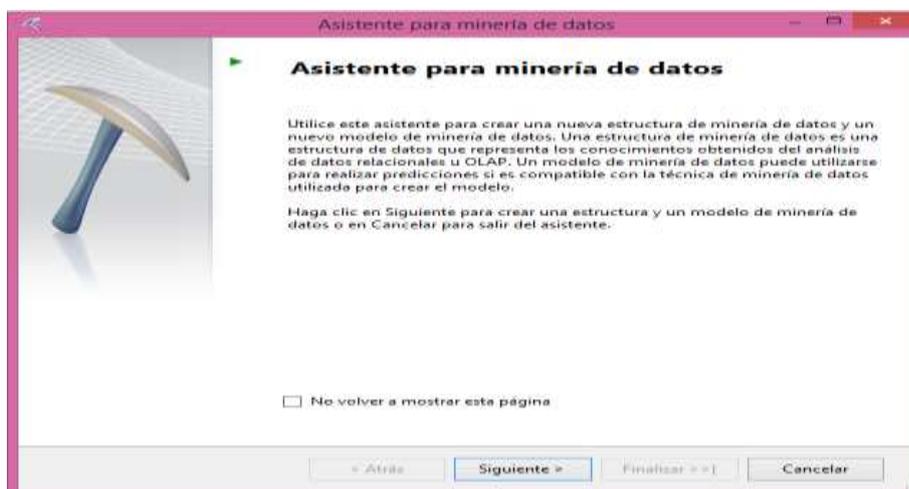


Gráfico 67: Asistente de minería de datos

- c. Seleccionamos el método de definición : A partir de una base de datos relacional o de almacenamiento de datos. Ver gráfico 68.

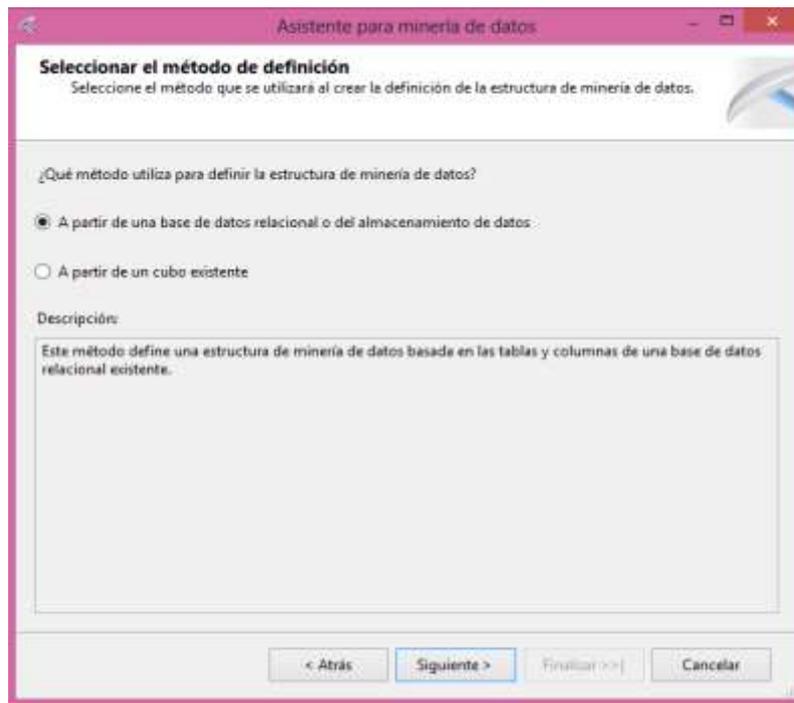


Gráfico 68: Selección Base de Datos relacional

- d. Seleccionamos la técnica de minería de datos : Arbol de Decisión. Ver gráfico 69.

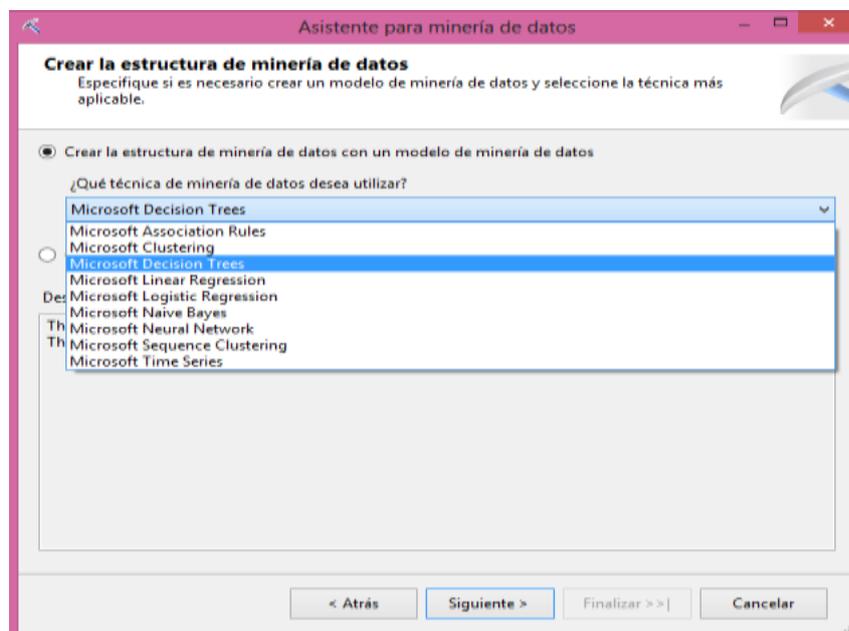


Gráfico 69: Selección de la técnica de minería de datos

e. Seleccionamos la vista de origen de datos creada. Ver gráfico 70.

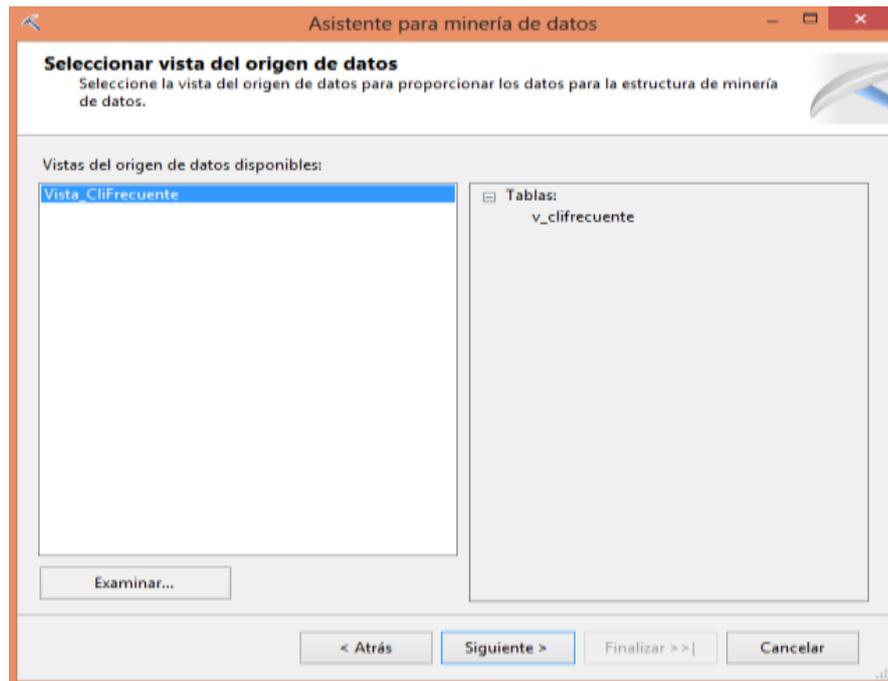


Gráfico 70: Selección la vista de origen de datos creada

f. Especificamos el tipo de tablas. Ver gráfico 71.

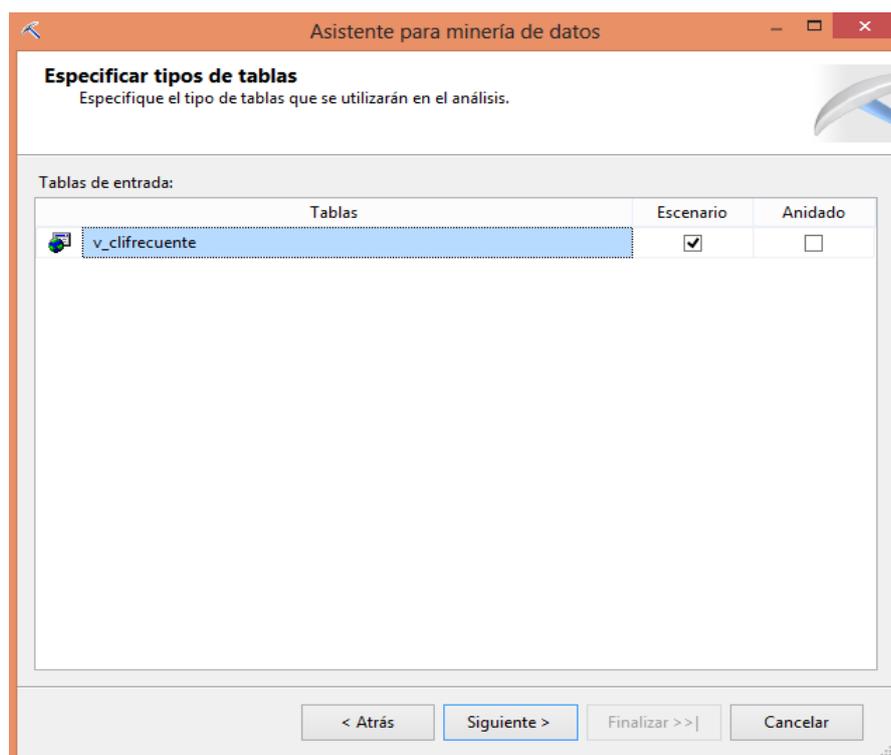


Gráfico 71: Selección del tipo de tabla v\_clifrecuente

- g. Especificamos las columnas que utilizaremos para el análisis de predicción : **Clave** : key\_cliente, **Entrada** : Canal de Venta , **Predicción**: Línea. Ver gráfico 72.

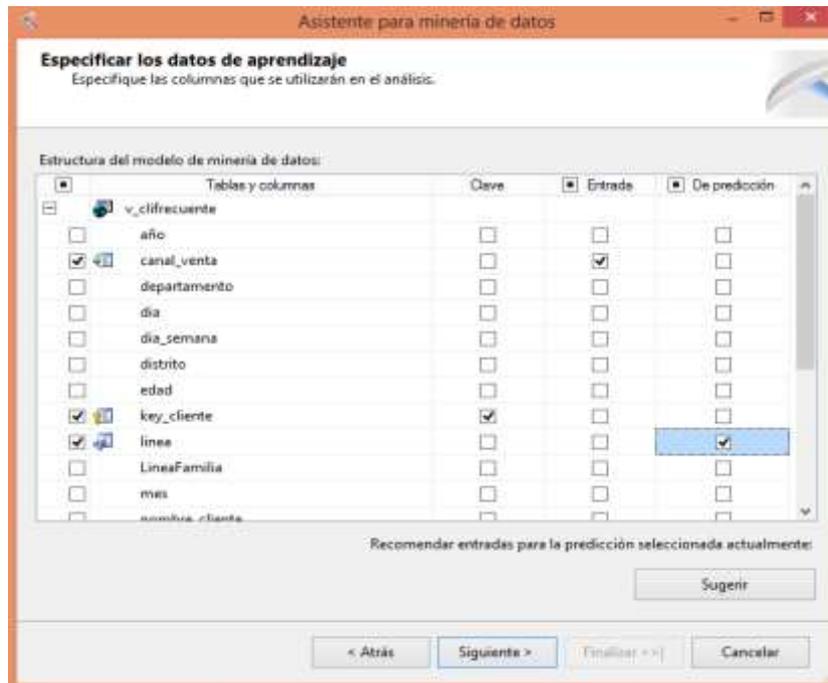


Gráfico 72: Selección de variables para el análisis de predicción

- h. Especificamos el contenido y tipos de datos. Ver gráfico 73.

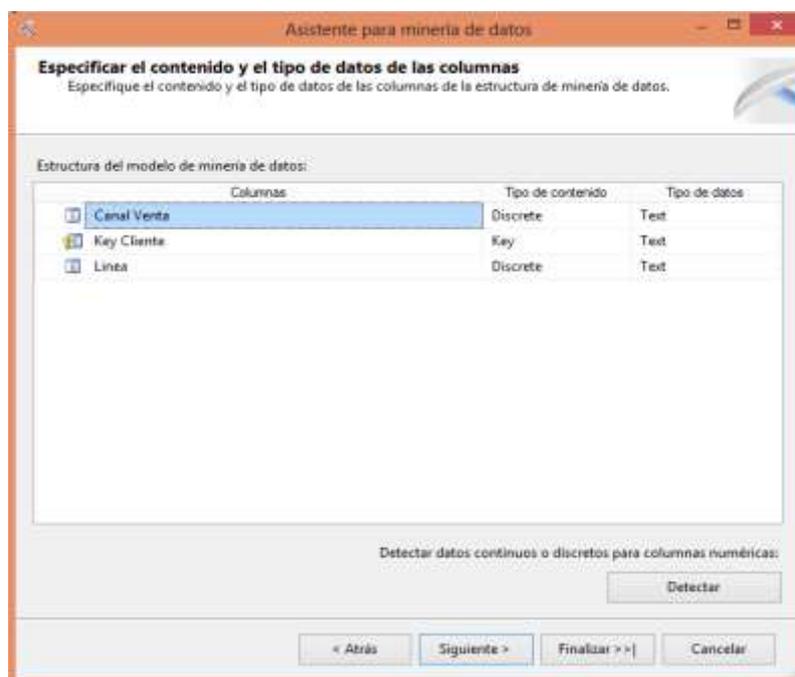


Gráfico 73: Visualización del contenido y tipos de datos

- i. Nombramos la estructura de minería de datos y el modelo de minería de datos. Ver gráfico 74.

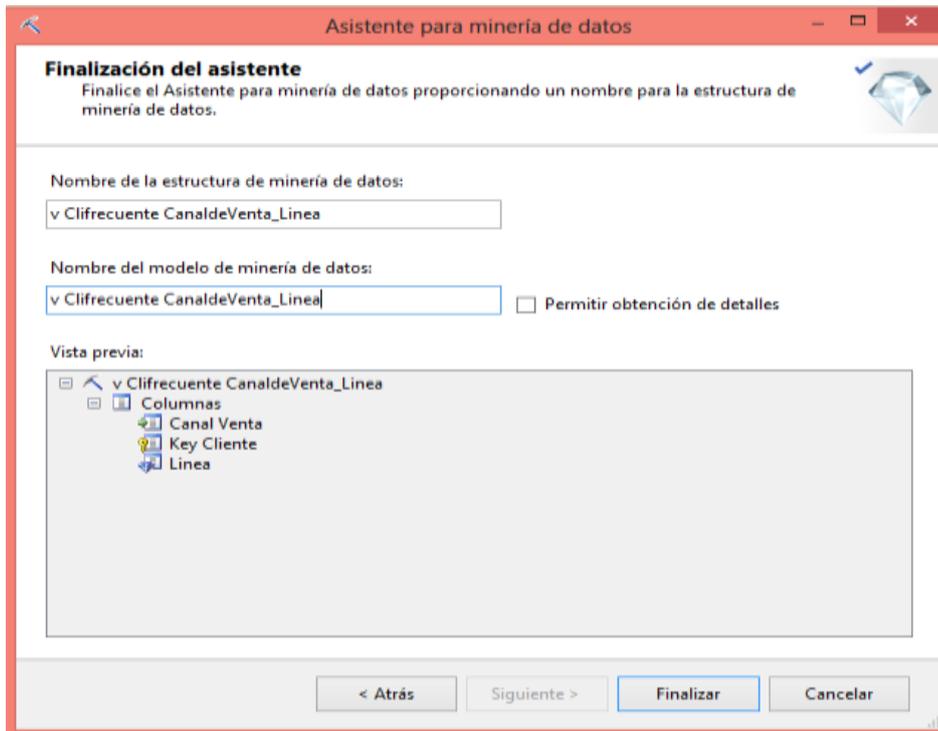


Gráfico 74: Asignación del nombre de la estructura y modelo de minería de datos

2. Verificamos la estructura de minería de datos creada. Ver gráfico 75.

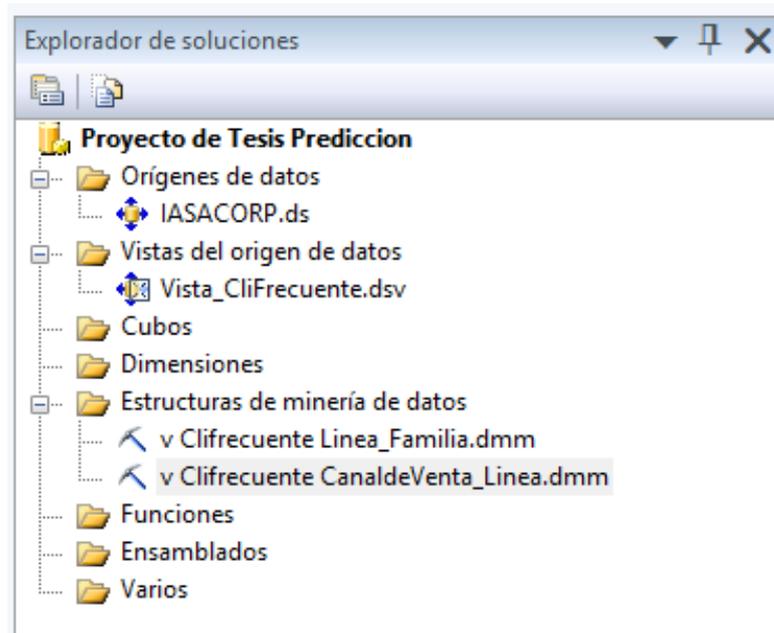


Gráfico 75: Vista de la estructura de minería de datos creada

a. Estructura de minería de datos. Ver gráfico 76.

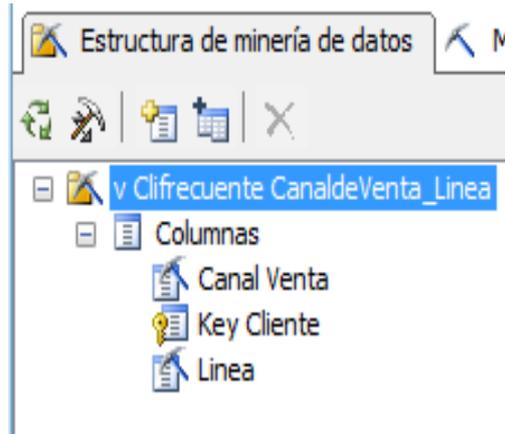


Gráfico 76: Estructura de minería de datos

b. Modelo de minería de datos. Ver gráfico 77.

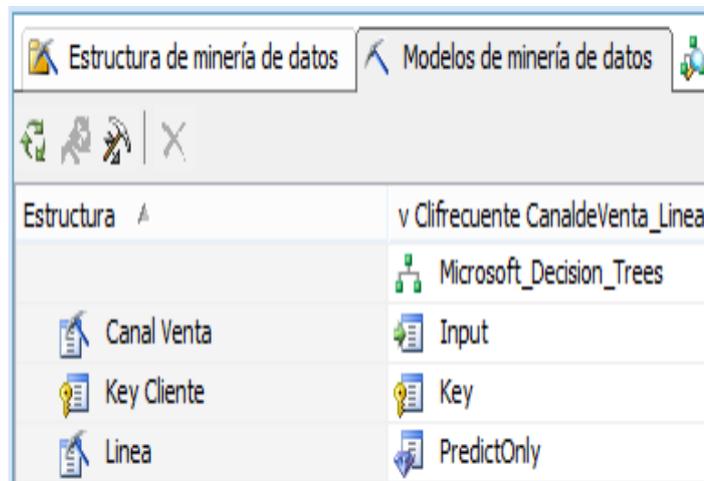
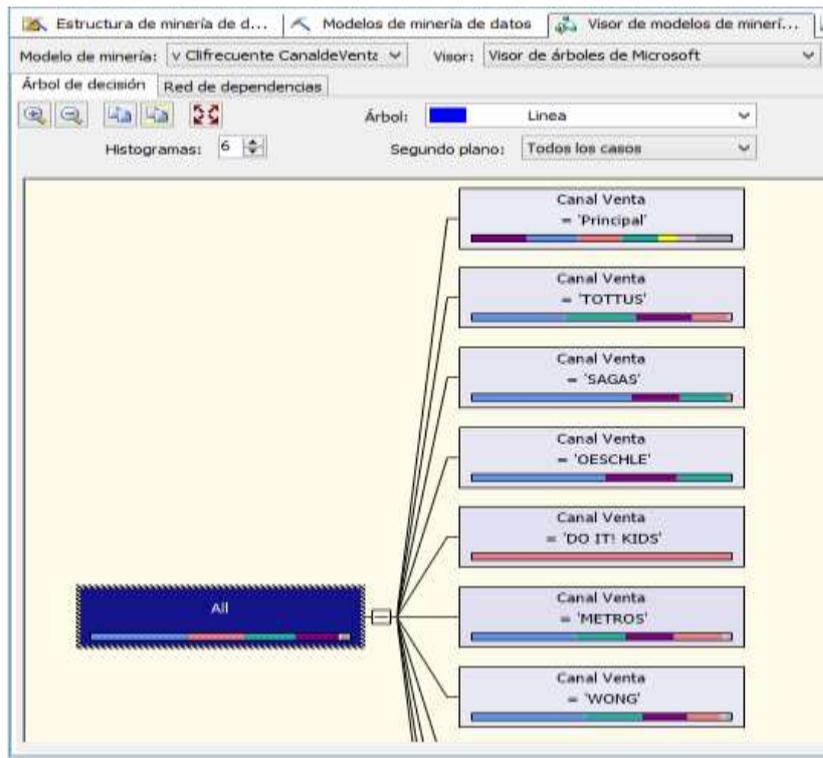


Gráfico 77: Modelo de minería de datos

c. Visor de minería de datos/Leyendo de minería de datos. Ver gráfico 78.



Leyenda de minería de datos

Alta Baja

Escenarios totales: 34024

Valor	Esce...	Probabi...	Histograma
<input checked="" type="checkbox"/> ACCES. DE PELO FF	40	0.12%	
<input checked="" type="checkbox"/> ACCESORIOS DE CA...	6793	19.96%	
<input checked="" type="checkbox"/> ANTEOJOS DE SOL	329	0.97%	
<input checked="" type="checkbox"/> ARTICULOS PROM...	4	0.01%	
<input checked="" type="checkbox"/> BISUTERIA	11708	34.40%	
<input checked="" type="checkbox"/> BISUTERIA FF	52	0.15%	
<input checked="" type="checkbox"/> BOLSAS DE REGALO	65	0.19%	
<input checked="" type="checkbox"/> CALZADO	231	0.68%	
<input checked="" type="checkbox"/> CALZADO FF	1	0.01%	
<input checked="" type="checkbox"/> COMPLEMENTOS D...	5773	16.96%	
<input checked="" type="checkbox"/> COMPLEMENTOS FF	38	0.11%	
<input checked="" type="checkbox"/> COSMETICOS	523	1.54%	
<input checked="" type="checkbox"/> ESCOLAR	9	0.03%	
<input checked="" type="checkbox"/> HOGAR	358	1.05%	
<input checked="" type="checkbox"/> INCENTIVOS	0	0.00%	
<input checked="" type="checkbox"/> LENTES DE SOL FF	0	0.00%	
<input checked="" type="checkbox"/> Missing	0	0.00%	

Gráfico 78: Visor de minería de datos-Leyenda de minería de datos

A continuación se detalla la descripción del caso de Uso del Sistema: **Analizar información por Canal de Venta y Línea.**

## DESCRIPCION DEL CASO DE USO DEL SISTEMA

### ANALIZAR INFORMACION POR CANAL DE VENTA Y LINEA

#### 1. DESCRIPCION CORTA

Este proceso muestra el reporte de análisis predictivo por Línea y Canal de Venta en base a la información solicitada.

#### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Gerente de MKT/Jefe de Línea

#### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Ejecutar estructura de minería de datos por Línea y Canal de Venta
2	S	Mostrar vista gráfica de Árbol de Decisión
3	A	Seleccionar y analizar vistas gráficas y leyendas de minería de datos.
4	S	Mostrar leyenda de minería de datos.

#### 4. FLUJO ALTERNATIVO DE EVENTOS

Paso	Tipo	Descripción del paso
1	S	

#### 5. PRECONDICIONES

Ítem	Precondiciones
1	<ul style="list-style-type: none"> <li>- Haber conectado el origen de datos en la herramienta de Analysis Services</li> <li>- Haber creado la vista Clifrecuente</li> <li>- Haber creado la estructura de minería de datos.</li> </ul>

#### 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para la consulta del reporte.

#### 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Analizar Información por Línea y Canal de Venta será diaria.

## DIAGRAMA DE CLASES: ANALIZAR INFORMACION POR CANAL DE VENTA Y LINEA

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de Datawarehouse. A continuación se muestra el Diagrama de Clases: Analizar información por Canal de Venta y Línea. Ver Gráfico 79.

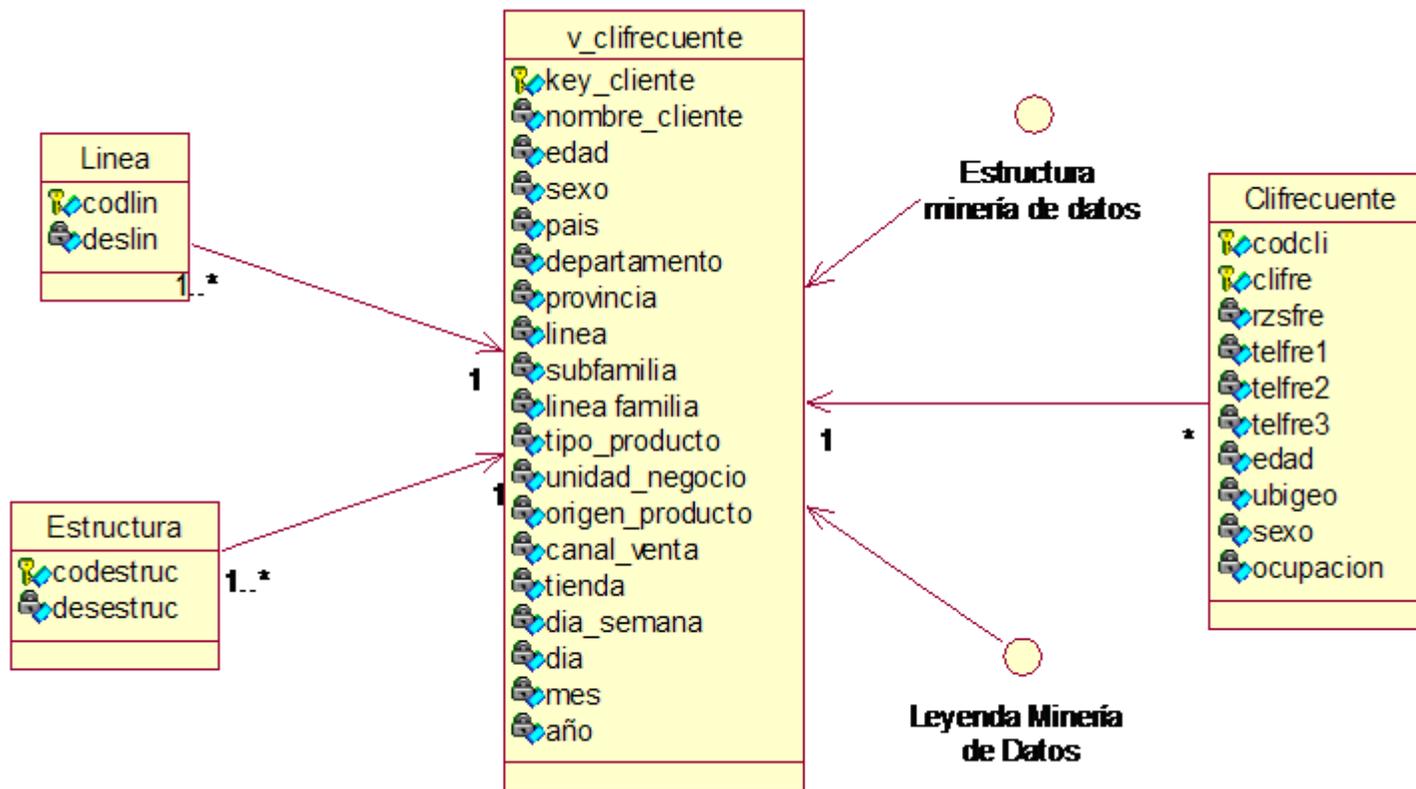


Gráfico 79: Diagrama de Clases: Analizar información por Canal de Venta y Línea

## DIAGRAMA DE SECUENCIA CUS: ANALIZAR INFORMACION POR CANAL DE VENTA Y LINEA

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Analizar información por Canal de Venta y Línea. Ver gráfico 80.

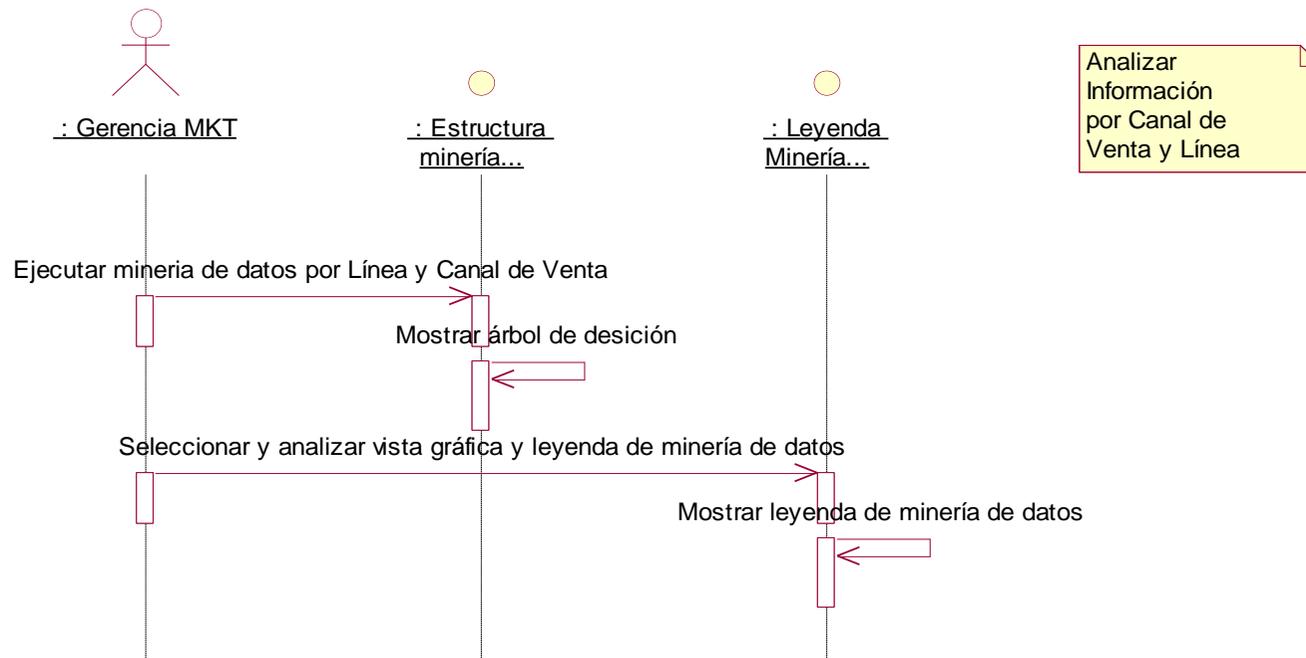


Gráfico 80: Diagrama de Secuencia: Analizar información por Canal de Venta y Línea

## Construcción del CUS: Analizar información por Canal de Venta y Línea

1. Ejecutar la estructura de minería de datos Clifrecuente CanaldeVenta\_Linea. Ver gráfico 81.

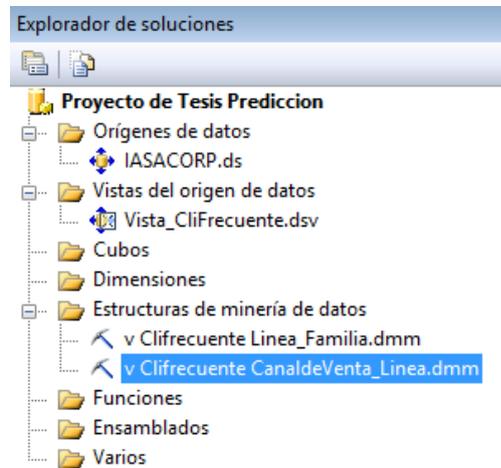


Gráfico 81: Ejecución estructura de minería de datos Clifrecuente CanaldeVenta\_Linea

2. Ingresamos a la opción Visor de modelos de minería de datos y visualizamos el árbol de decisión. Ver gráfico 82.

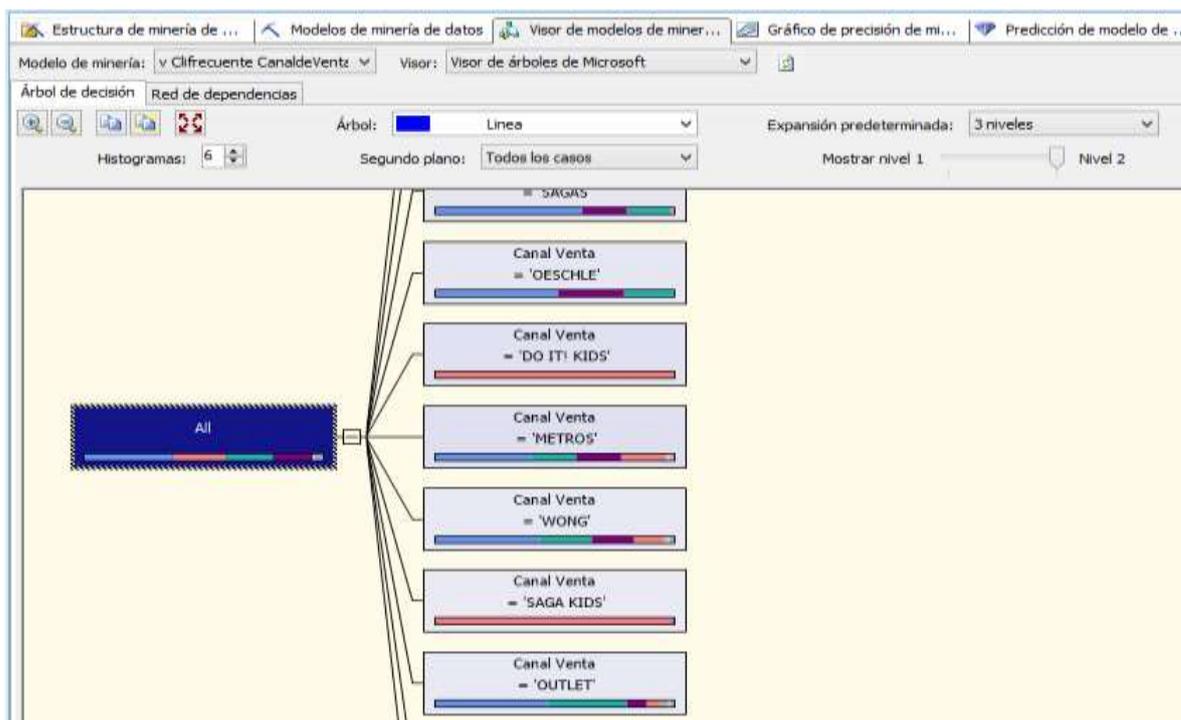


Gráfico 82: Visor de modelos de minería de datos y visualizamos el árbol de decisión

3. En una primera instancia analizamos el primer nodo o primer nivel con la siguiente información: Ver gráfico 83.

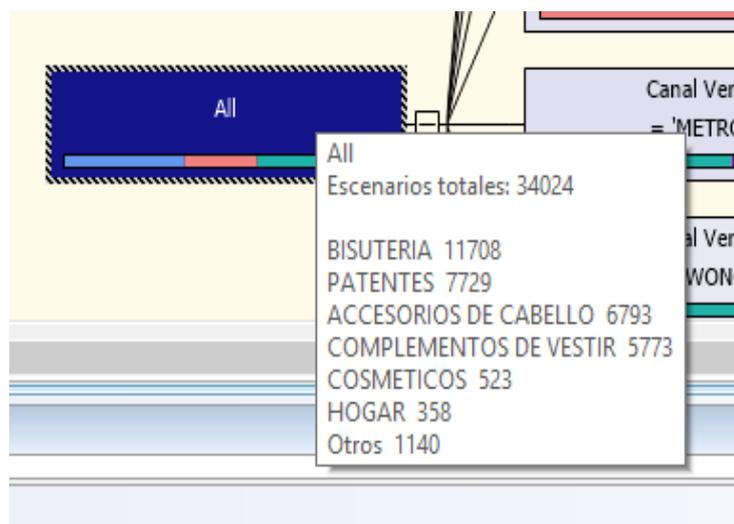


Gráfico 83: Vista de Escenarios del análisis predictivo

Del 34,024 total de escenarios este se distribuye en las siguientes probabilidades de ventas por línea de Clientes Frecuentes, tal como se muestra en la tabla 9.

Línea	Escenarios	% Probabilidad
Bisutería	11708	34.40%
Patentes	7729	22.71%
Accesorios de Cabello	6793	19.96%
Complementos de Vestir	5773	16.96%
Cosméticos	523	1.54%
Hogar	358	1.05%
Otros	1140	3.38%

Tabla 9 Escenarios y probabilidades de Ventas por Línea de Cliente Frecuentes

Esta información también se puede analizar en la Leyenda de Minería de datos. Ver gráfico 84.

Leyenda de minería de datos			
Alta		Baja	
Escenarios totales: 34024			
Valor	Escenarios	Probabilidad	Histograma
<input checked="" type="checkbox"/> ACCES. DE PELO FF	40	0.12%	
<input checked="" type="checkbox"/> ACCESORIOS DE CA...	6793	19.96%	
<input checked="" type="checkbox"/> ANTEOJOS DE SOL	329	0.97%	
<input checked="" type="checkbox"/> ARTICULOS PROM...	4	0.01%	
<input checked="" type="checkbox"/> BISUTERIA	11708	34.40%	
<input checked="" type="checkbox"/> BISUTERIA FF	52	0.15%	
<input checked="" type="checkbox"/> BOLSAS DE REGALO	65	0.19%	
<input checked="" type="checkbox"/> CALZADO	231	0.68%	
<input checked="" type="checkbox"/> CALZADO FF	1	0.01%	
<input checked="" type="checkbox"/> COMPLEMENTOS D...	5773	16.96%	
<input checked="" type="checkbox"/> COMPLEMENTOS FF	38	0.11%	
<input checked="" type="checkbox"/> COSMETICOS	523	1.54%	
<input checked="" type="checkbox"/> ESCOLAR	9	0.03%	
<input checked="" type="checkbox"/> HOGAR	358	1.05%	
<input checked="" type="checkbox"/> INCENTIVOS	0	0.00%	
<input checked="" type="checkbox"/> LENTES DE SOL FF	0	0.00%	
<input checked="" type="checkbox"/> Missing	0	0.00%	
<input checked="" type="checkbox"/> NO ASIGNADO	3	0.01%	
<input checked="" type="checkbox"/> PATENTES	7729	22.71%	
<input checked="" type="checkbox"/> PROMOCIONES	13	0.04%	
<input checked="" type="checkbox"/> REGALERIA	206	0.61%	
<input checked="" type="checkbox"/> REGALERIA FF	14	0.04%	
<input checked="" type="checkbox"/> RELOJES	51	0.15%	
<input checked="" type="checkbox"/> TEXTIL	84	0.25%	
<input checked="" type="checkbox"/> TEXTIL FF	0	0.00%	

Gráfico 84: Leyenda de Minería de datos

4. Analizamos a un 2do Nivel por Canal de Venta. Ver gráfico 85.

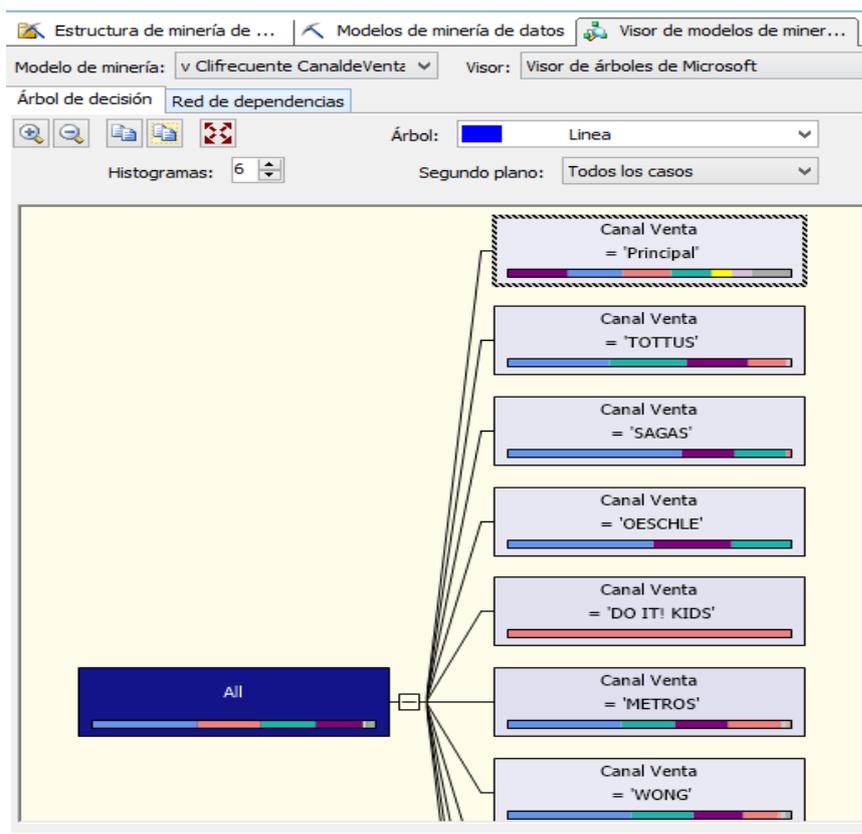


Gráfico 85: Análisis por Canal de Venta

5. Tomamos como ejemplo el Canal de Venta Tottus, del total de 1187 escenarios este se distribuye en las siguientes probabilidades de venta de Cliente Frecuentes por Línea. Ver gráfico 86 y Tabla 10.

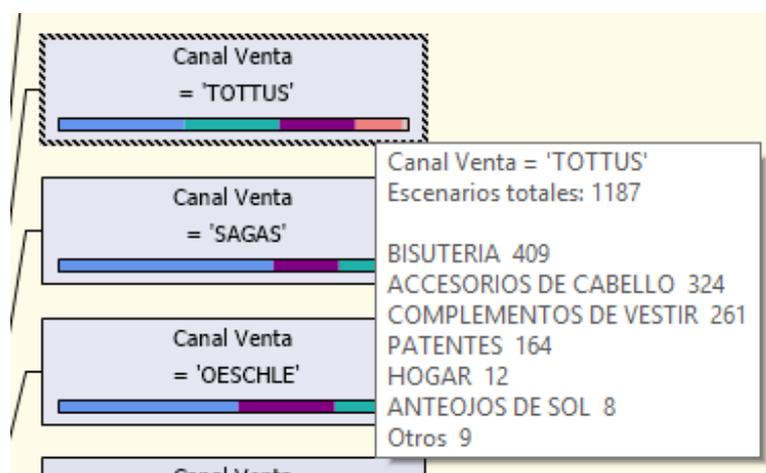


Gráfico 86: Probabilidades de venta de Cliente Frecuentes por Línea

Línea	Escenarios	% Probabilidad
Bisutería	409	34.42%
Accesorios de Cabello	324	27.27%
Complementos de Vestir	261	21.97%
Patentes	164	13.81%
Hogar	12	1.02%
Anteojos de Sol	8	0.68%
Otros	9	3.38%

Tabla 10 Escenarios y Probabilidades de Ventas de Clientes Frecuentes

Esta información también se puede analizar en la Leyenda de Minería de datos. Ver gráfico 87.

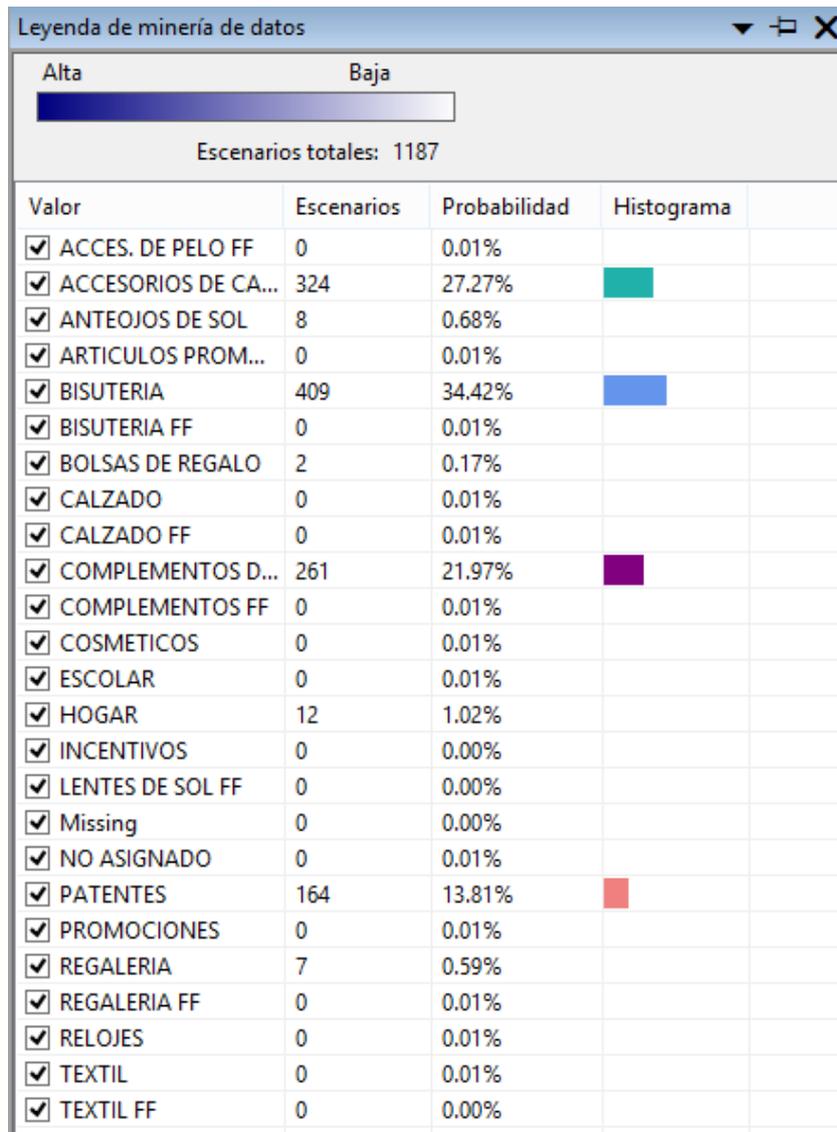


Gráfico 87: Leyenda de Minería de datos

A continuación se detalla la descripción del caso de Uso del Sistema: **Analizar información por Línea y Familia.**

## DESCRIPCION DEL CASO DE USO DEL SISTEMA ANALIZAR INFORMACION POR LINEA Y FAMILIA

### 1. DESCRIPCION CORTA

Este proceso muestra el reporte de análisis predictivo por Línea y Familia en base a la información solicitada.

### 2. ACTORES ASOCIADOS

Actor	Tipo	Descripción
1	Principal	Gerente de MKT/Jefe de Línea

### 3. FLUJO BASICO DE EVENTOS

Paso	Tipo	Descripción del paso
1	A	Ejecutar estructura de minería de datos por Línea y Familia
2	S	Mostrar vista gráf86a de Árbol de Decisión
3	A	Seleccionar y analizar vistas gráficas y leyendas de minería de datos.
4	S	Mostrar leyenda de minería de datos.

### 4. FLUJO ALTERNATIVO DE EVENTOS

Paso	Tipo	Descripción del paso
1	S	

### 5. PRECONDICIONES

Ítem	Precondiciones
1	<ul style="list-style-type: none"> <li>- Haber conectado el origen de datos en la herramienta de Analysis Services</li> <li>- Haber creado la vista Clifrecuente</li> <li>- Haber creado la estructura de minería de datos.</li> </ul>

### 6. POSTCONDICIONES

Ítem	Postcondiciones
1	Se genera la información requerida para la consulta del reporte.

### 7. FRECUENCIA

La frecuencia de ejecución del caso de uso Analizar Información por Línea y Familia será diaria.

## DIAGRAMA DE CLASES: ANALIZAR INFORMACION POR LINEA Y FAMILIA

El siguiente diagrama de clases, muestra las relaciones entre las clases de las tablas de Datawarehouse. A continuación se muestra el Diagrama de Clases: Analizar información por Línea y Familia. Ver Gráfico 88.

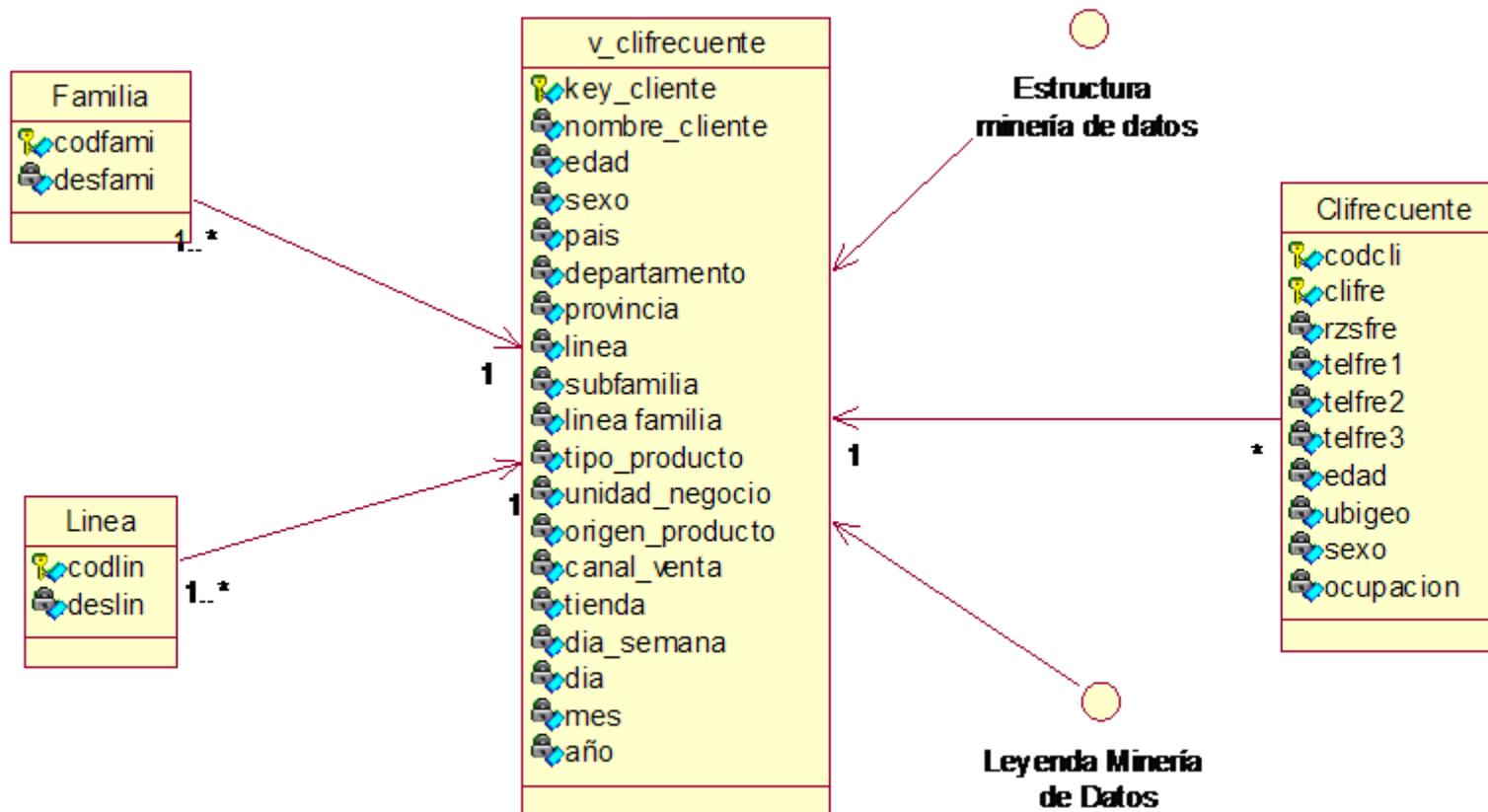


Gráfico 88: Diagrama de Clases Analizar información por Línea y Familia

## DIAGRAMA DE SECUENCIA CUS: ANALIZAR INFORMACION POR LINEA Y FAMILIA

El siguiente diagrama de secuencia, muestra la interacción entre los objetos y clases del sistema que son usados para implementar los mensajes intercambiados entre los objetos. A continuación se muestra el Diagrama de Secuencia: Analizar información por Línea y Familia. Ver gráfico 89.

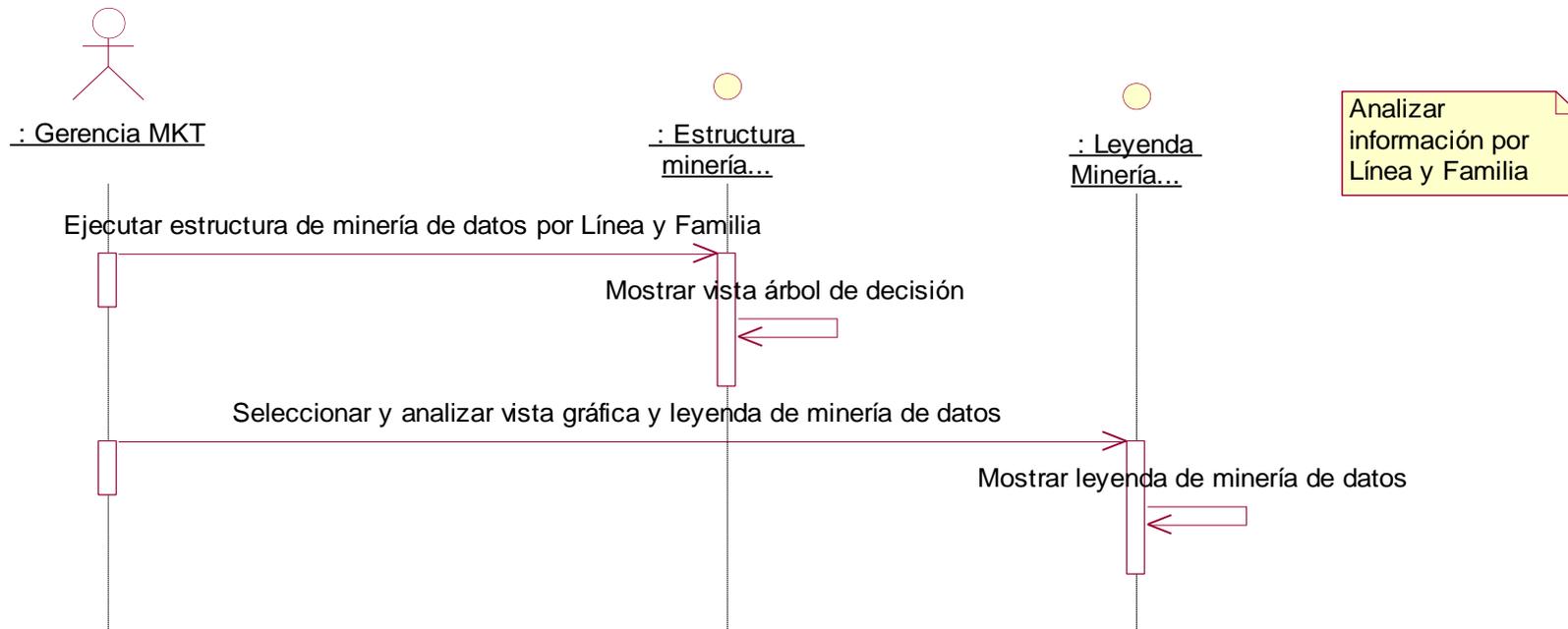


Gráfico 89: Diagrama de Secuencia CUS Analizar información por línea y familia

## Construcción del CUS: Analizar información por Línea y Familia

1. Ejecutar la estructura de minería de datos Clifrecuente Linea\_Familia. Ver Gráfico 90.

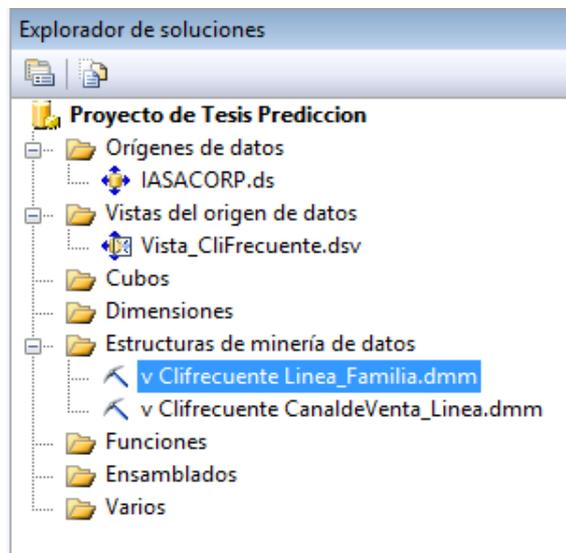


Gráfico 90: Ejecución estructura de minería de datos Clifrecuente Linea\_Familia

2. Ingresamos a la opción Visor de modelos de minería de datos y visualizamos el árbol de decisión. Ver gráfico 91.

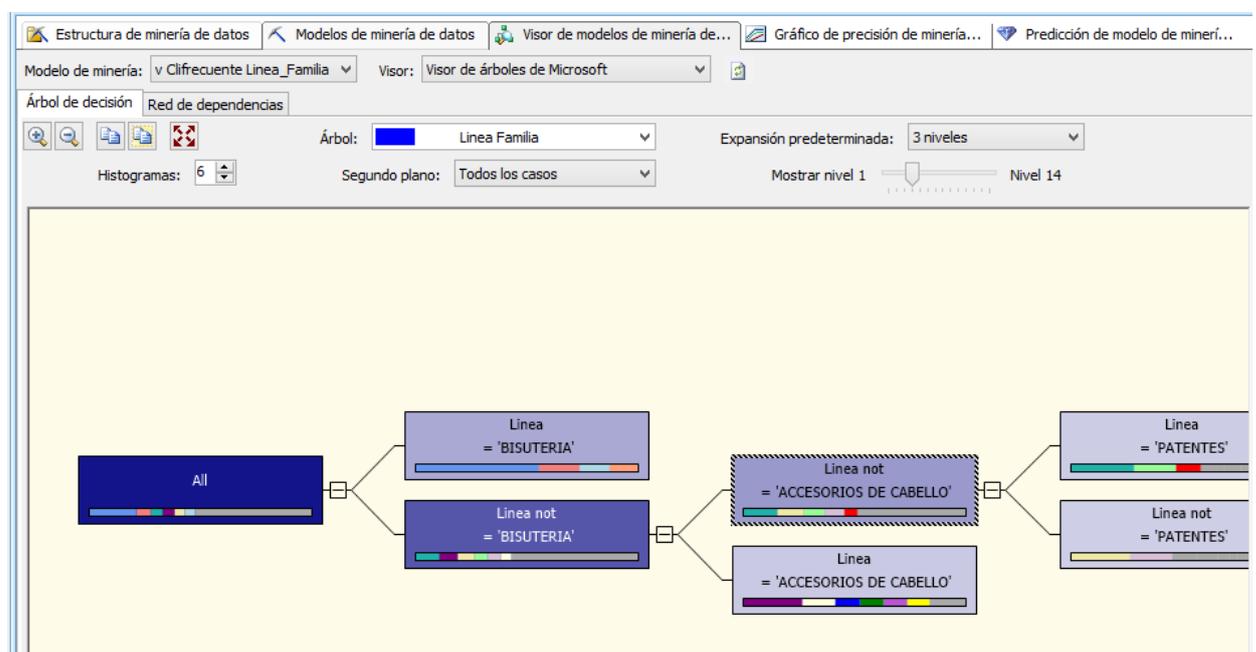


Gráfico 91: Visor de modelos de minería de datos y visualizamos el árbol de decisión

3. En una primera instancia analizamos el primer nodo o primer nivel con la siguiente información. Ver gráfico 92.

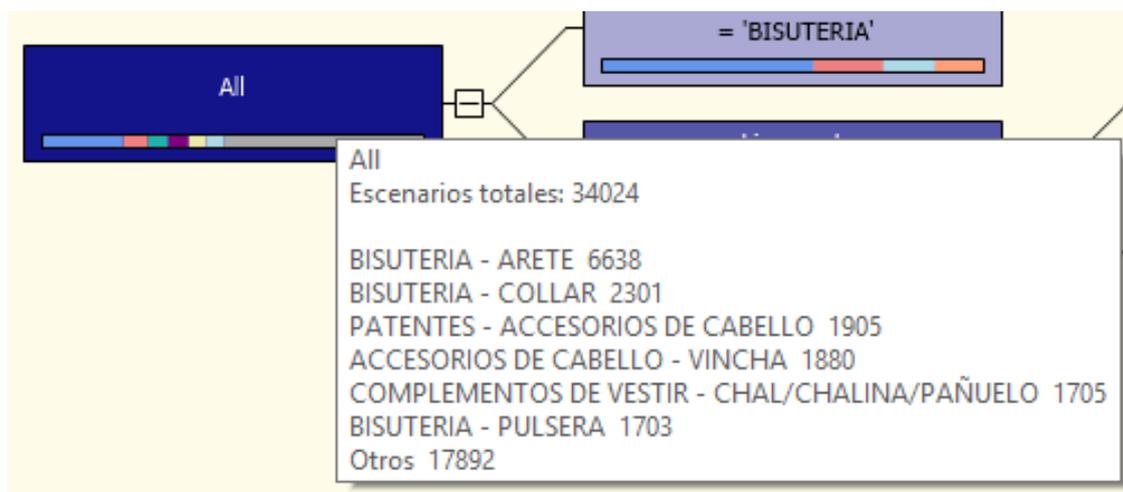


Gráfico 92: Visor de Escenarios del análisis predictivo

Del 34,024 total de escenarios este se distribuye en las siguientes probabilidades de ventas por Familia de Clientes Frecuentes.

Estos datos se visualizan en la tabla 11.

Familia	Escenarios	% Probabilidad
Bisutería/Arete	6638	19.50%
Bisutería/Collar	2301	6.76%
Patentes/Accesorios de Cabello	1905	5.60%
Accesorios de Cabello/Vincha	1880	5.52%
Complemento de Vestir/Chal..	1705	5.01%
Bisuteria/Pulsera	1703	5.00%
Otros	17892	52.61%

Tabla 11 Escenarios y Probabilidades de ventas por Familia de Clientes Frecuentes

Esta información también se puede analizar en la Leyenda de Minería de datos.  
Ver gráfico 93.

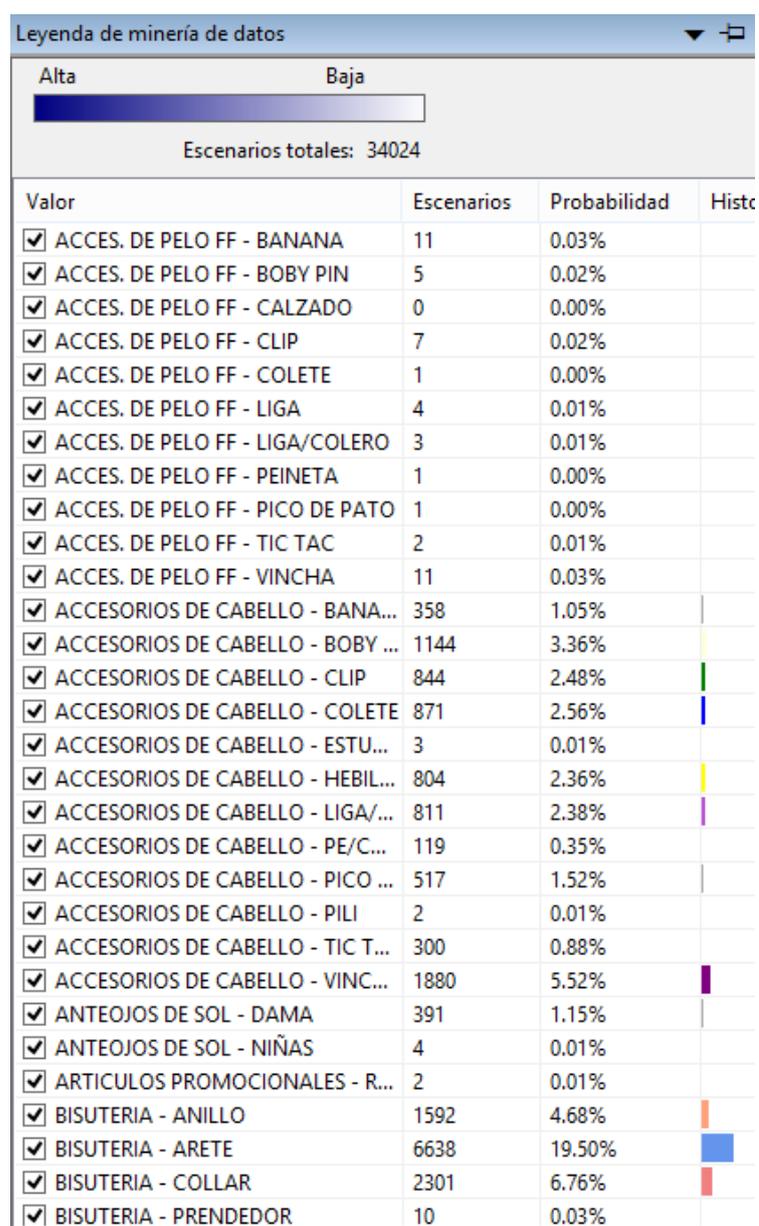


Gráfico 93: Leyenda de Minería de datos

4. Analizamos a un 2do Nivel por Línea Bisutería. Ver gráfico 94.

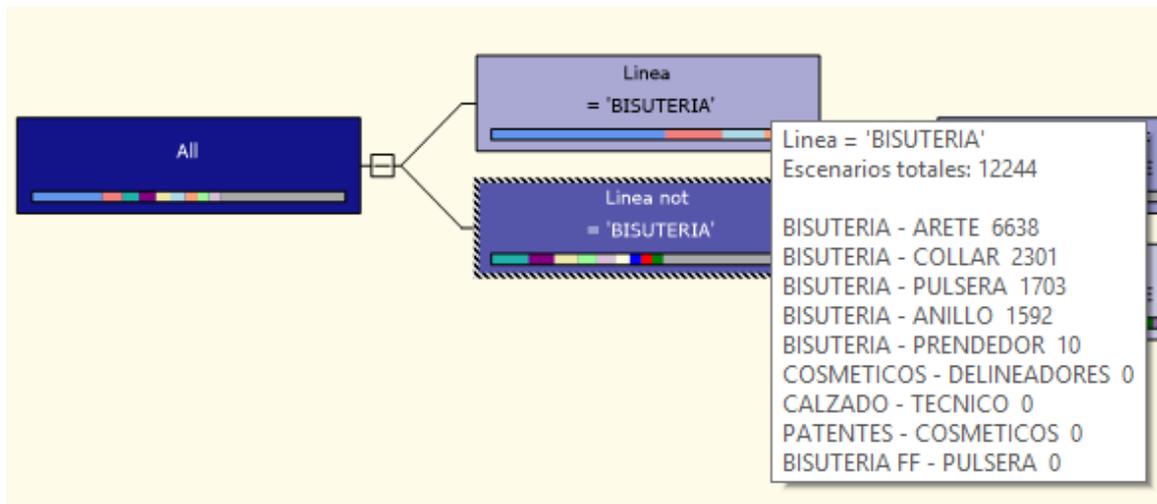


Gráfico 94: Análisis por Línea Bisutería

Del total de 12244 escenarios este se distribuye en las siguientes probabilidades de venta de Cliente Frecuentes por la línea Bisutería y Familia, ver tabla 12.

Línea Bisutería/Familia	Escenarios	% Probabilidad
Bisutería/Arete	6638	54.21%
Bisutería/Collar	2301	18.79%
Bisutería/Pulsera	1703	13.91%
Bisutería/Anillo	1592	13.00%
Bisutería/Prendedor	10	0.08%

Tabla 12 Escenarios y Probabilidades de venta por Lines Bisutería y Familia

Esta información también se puede analizar en la Leyenda de Minería de datos. Ver gráfico 95.

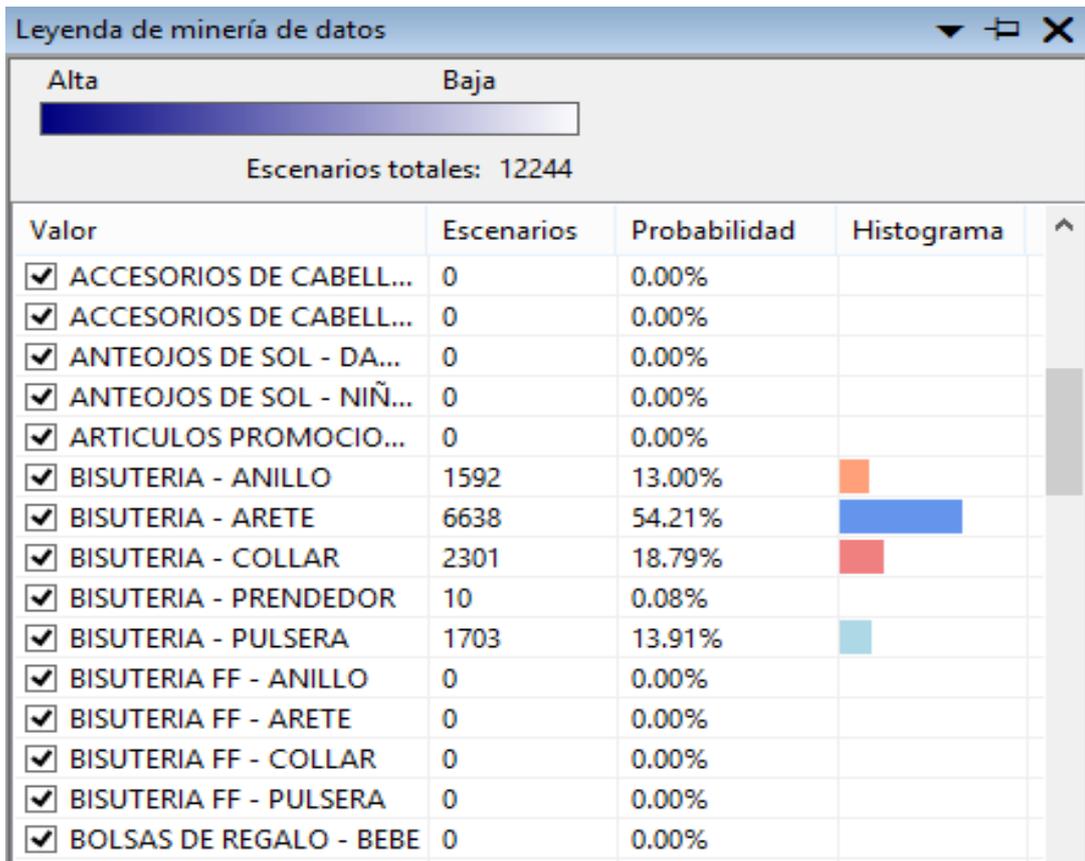


Gráfico 95: Leyenda de Minería de datos

3. Para la línea que no es Bisutería del total de 21,780 escenarios este se distribuye en las siguientes probabilidades de venta de Cliente Frecuentes para las siguientes líneas y familias. Ver gráfico 96 y Tabla 13.

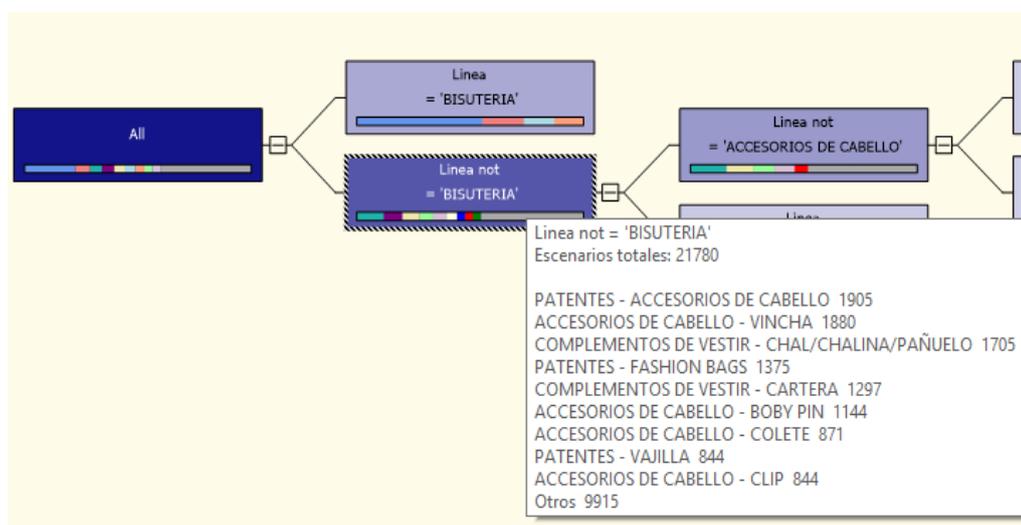


Gráfico 96: Probabilidades de venta de Cliente Frecuentes por Línea y Familia

Línea / Familia	Escenarios	% Probabilidad
Patentes/Accesorios de Cabello	1905	8.74%
Accesorios de Cabello/Vincha	1880	8.63%
Complementos de Vestir/Chal...	1705	7.82%
Patentes/Fashion Bags	1375	6.31%
Complementos de Vestir/Cartera	1297	5.95%
Accesorios de Cabello/Boby Pin	1144	5.25%
Accesorios de Cabello/Colete	871	4.00%
Patentes/Vajilla	844	3.87%
Accesorios de Cabello/Clip	844	3.87%
Otros	9915	45.56%

Tabla 13 Escenarios y Probabilidades de Venta de Clientes Frecuentes

Esta información también se puede analizar en la Leyenda de Minería de datos. Ver gráfico 97.

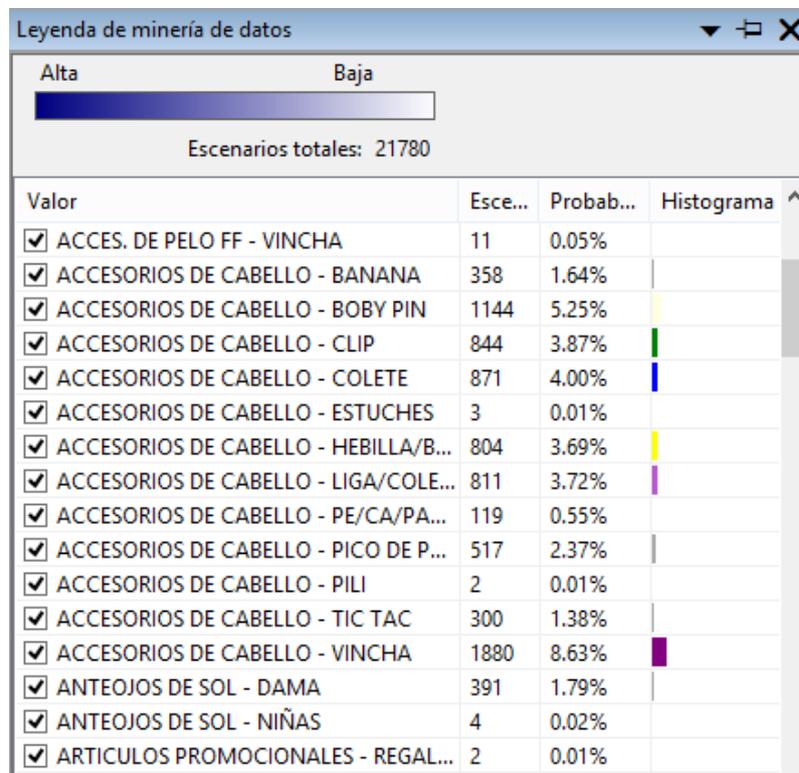


Gráfico 97: Leyenda de Minería de datos

## 5. Evaluación de Resultados

Se evalúan los logros y cambios expresados por el proyecto a través de las múltiples acciones de implementación. Estos logros se basan en los objetivos planteados inicialmente en el proyecto y los cuales se detallan a continuación:

Se realizó al 100% la identificación dentro del Sistema Transaccional de Ventas, la información relevante de los clientes preferenciales, así como toda la información de las transacciones realizadas, detalle de los productos adquiridos, etc.

Se construyó al 100% el modelo de datos multidimensional que mediante un proceso de ETL (Extracción, Transformación y Carga) contiene la información comercial para su posterior análisis.

Se implementó el algoritmo de árboles de decisión dentro de la herramienta de análisis predictivo, que permitió generar información de probabilidad sobre el comportamiento del consumo de los clientes frecuentes, basada en las líneas de producto que comercializa la empresa, la cual pueda ser **usada para la proyección de ventas, estimaciones de compras y lanzamiento de estrategias comerciales.**

Cabe señalar, que los resultados se muestran en cada especificación de los casos de uso Analizar Información por Canal de Venta y Línea, el caso de uso Analizar Información por Línea y Familia, mostrando las probabilidades y posibles escenarios.

## CONCLUSIONES

1. La tecnología de BI actualmente contribuye con las empresas a nivel estratégico y operacional ya que permite tomar decisiones en las diferentes áreas de una organización a través de los indicadores de gestión, así de cómo el análisis de tendencias y predicción del futuro a través de la minería de datos.
2. Es así que en la presente tesis se ha desarrollado la minería de datos para obtener información valiosa que contribuya a mejorar los procesos de presupuestos de ventas y compras de la organización IASACORP INTERNATIONAL S.A., además de poder obtener tendencias del comportamiento de los clientes para poder ofrecer mejores estrategias de promociones, ofertas y descuentos.
3. En cuanto a la mejora del proceso de presupuestos de ventas, la herramienta de predicción, ha permitido que genere una proyección del cálculo de probabilidades de ventas de las líneas y familias de productos con una aproximación del 95% de confiabilidad, y de cómo resultado un proceso de presupuesto de ventas mejor del que se venía usando. Esto principalmente por que el uso de la información estadística en la herramienta de predicción, nos da un respaldo en base al algoritmo usado (árbol de decisión).
4. Respecto al proceso de poder determinar el comportamiento de clientes de la organización, la minería de datos aplicada da como resultado estar un paso delante de los competidores y mejorar los resultados de ventas. La información que la herramienta de minería de datos brinda en este aspecto es muy valiosa ya que permite tomar decisiones efectivas en las estrategias de promociones, ofertas y descuentos ya que se cuenta con información estadística de cómo los clientes consumirán las líneas de productos de la empresa y así poder definir las estrategias mencionadas que finalmente tendrá un impacto en el aumento del volumen de ventas.

5. En cuanto a la mejora en los tiempos de preparar los reportes de estimación de venta y plan de compras en base a los resultados del pronóstico, la herramienta de predicción ha permitido que este proceso se realice en 1 día, versus el proceso anterior que tomaba entre 3 a 4 días.

## RECOMENDACIONES

1. Cuando se desarrolla un proyecto de BI, se debe asegurar la calidad de los datos del sistema que se va procesar, evitando así la no información no veraz o errónea. No es posible lograr un buen resultado final en un proyecto, acorde a los objetos marcados, sino se realiza previamente una buena limpieza de los datos. Esto da una idea de la enorme necesidad de tomarse muy en serio esta etapa, realizándola acorde a unos parámetros correctos y teniendo en cuenta las recomendaciones de los expertos.
2. Esta solución puede servir de base para cualquier empresa en el rubro retail.
3. El sistema de información desarrollado debería ampliarse a las diferentes áreas de la empresa aprovechando al máximo la información resultante para pronosticar o bien para encontrar las agrupaciones o relaciones entre datos, que a simple vista no se logra ver. Logrando hacer realizad beneficios importantes para la empresa.
4. Así mismo, existen diferentes técnicas de minería de datos que se pueden utilizar en otros rubros, por ejemplo:
  - ✓ Se puede segmentar conjunto de clientes, el conjunto de valores e índices financieros, el espectro de observaciones astronómicas, el conjunto de empleados y de sucursales u oficinas, etc. Utilizando el algoritmo de Clustering o “agrupamiento”.
  - ✓ Por ejemplo si se quiere determinar si se puede jugar o no un partido de futbol, tenis, etc. de acuerdo al clima, a partir de los atributos que modelan el tiempo (vista, temperatura, humedad y viento) se puede usar el algoritmo de clasificación.
  - ✓ Las redes neuronales están siendo utilizadas en distintos sectores como la industria, el gobierno, el ejército, las comunicaciones, banca y las finanzas, los seguros, la medicina, la distribución, la robótica, el marketing, etc.

## REFERENCIAS BIBLIOGRAFICAS

1. Inteligencia de Negocios  
<http://www.usmp.edu.pe/publicaciones/boletin/fia/info46/sistemas/articulo6.htm>
2. Blog sobre Business Intelligence  
<http://www.lantares.com/blog/algoritmos-de-predictive-analytics-y-su-aplicación-practica>
3. Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones  
<http://www.monografias.com/trabajos75/inteligencia-negocios/inteligencia-negocios3.shtml#conclusioa#ixzz3ifbsmM7J>
4. Inteligencia a Medida de tu Empresa: Dafne Rosso Pelayo-2012.  
<http://www.sistemasinteligentespyme.com/>
5. Datawarehouse  
<http://www.monografias.com/trabajos24/bodega-de-datos/bodega-de-datos.shtml>  
<http://www.definicionabc.com/tecnologia/datawarehouse.php>
6. Elaboración de Algoritmos  
<http://es.slideshare.net/ballbreakerhouse/elaboracin-de-algoritmos>
7. Fundamentos de Programación: Algoritmos y Estructuras. Luis Jayanes Aguilar. Editorial Mc Graw Hill.
8. Tecnologías de Información y Estrategia: Aníbal Goicochea 2009  
<http://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>
9. Koneges  
<http://www.konoges.com/blogs/54-tecnologia/424-algoritmos-analisis-predictivo>
10. Guía de Predictive Analytics  
<http://www.lantares.com/analisis-predictivo>
11. Implantación en una empresa de un sistema Business Intelligence SaaS  
<https://riunet.upv.es/bitstream/handle/10251/8591/Proyecto%20II%20-%20C1%20-%20DMA%20-%2056-09.pdf>
12. Minería de Datos  
[http://www.sinnexus.com/business\\_intelligence/datamining.aspx](http://www.sinnexus.com/business_intelligence/datamining.aspx)  
<http://www.prodinet.es/datamining/>  
<http://www.universoabierto.com/13494/monografico-mineria-de-datos/>

### 13. Metodología RUP

<http://procesosdesoftware.wikispaces.com/METODOLOGIA+RUP>

## ANEXOS

### Anexo 1: Script para probar el DTS de extracción de las tablas maestros y ventas

```
USE BD_WORK
```

```
-- Script para probar el DTS de extracción de las tablas maestros y tablas de  
ventas
```

```
-- Si accion es 'S' hace select a las tablas maestros o si acción es 'V' hace select  
a las tablas de ventas, en caso contrario elimina registros para volver a probar  
DTS
```

```
declare @accion char(1)
```

```
set @accion='V'
```

```
IF @accion = 'S'
```

```
begin
```

```
    select * from linea  
    select * from familia  
    select * from subfamilia  
    select * from tiparticulo  
    select * from uninegocio  
    select * from color  
    select * from puntovta  
    select * from estructura  
    select * from pais  
    select * from departamento  
    select * from provincia  
    select * from distrito  
    select * from articulos1  
    select * from clifrecuente
```

```
end
```

```
else
```

```
begin
```

```
    truncate table linea  
    truncate table familia  
    truncate table subfamilia  
    truncate table tiparticulo  
    truncate table uninegocio  
    truncate table color  
    truncate table puntovta  
    truncate table estructura  
    truncate table pais  
    truncate table departamento  
    truncate table provincia  
    truncate table distrito  
    truncate table articulos1  
    truncate table clifrecuente
```

```
end
```

```
IF @accion = 'V'  
begin  
    select * from cboletav  
    select * from dboletav  
    select * from cfactura  
    select * from dfactura  
    select * from cnotcre  
    select * from dnotcre  
    select * from cgdespacho  
    select * from dgdespacho  
    select * from cguiadevolucion  
    select * from dguiadevolucion  
end  
else  
begin  
    truncate table cboletav  
    truncate table dboletav  
    truncate table cfactura  
    truncate table dfactura  
    truncate table cnotcre  
    truncate table dnotcre  
    truncate table cgdespacho  
    truncate table dgdespacho  
    truncate table cguiadevolucion  
    truncate table dguiadevolucion
```

**Anexo 2** : Script para comprobar la carga y eliminación de los registros en las tablas de trabajo dimensionales y de hechos.

```
declare @valor char(1)
set @valor='S'
if @valor='S'
begin
    select * from DM_ARTICULO_WORK
    select * from DM_CLIENTE_FRECUENTE_WORK
    select * from DM_TIENDA_WORK
    select * from DM_TIEMPO_WORK
    select * from DM_UBIGEO_WORK
    select * from FC_MARKETING_WORK
end
else
begin
    truncate table DM_ARTICULO_WORK
    truncate table DM_CLIENTE_FRECUENTE_WORK
    truncate table DM_TIENDA_WORK
    truncate table DM_TIEMPO_WORK
    truncate table DM_UBIGEO_WORK
    truncate table FC_MARKETING_WORK
end
```

**Anexo 3** : Script para comprobar la carga o eliminación de los registros en las tablas dimensionales y de hechos en el datawarehouse.

```
USE DWH_IASA
declare @valor char(1)
set @valor='S'

if @valor='S'
begin
    SELECT * FROM DM_ARTICULO
    SELECT * FROM DM_CLIENTE_FRECUENTE
    SELECT * FROM DM_TIENDA
    SELECT * FROM DM_TIEMPO
    SELECT * FROM DM_UBIGEO
    SELECT * FROM FC_MARKETING
end
else
begin
    DELETE FROM FC_MARKETING
    DELETE FROM DM_ARTICULO
    DELETE FROM DM_CLIENTE_FRECUENTE
    DELETE FROM DM_TIENDA
    DELETE FROM DM_TIEMPO
    DELETE FROM DM_UBIGEO

end
```