

7004
A17916



**PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL ECUADOR
SEDE AMBATO**
SERÉIS MIS TESTIGOS

DEPARTAMENTO DE INVESTIGACION, POSTGRADOS Y
AUTOEVALUACION

Tema:

“BUSINESS INTELLIGENCE APLICADO A LOS CLIENTES DE
MADERVAS, PARA UNA OPORTUNA TOMA DE DECISIONES”

Tesis de grado previo a la obtención del título de Magister en
Gerencia Informática con mención en mención de desarrollo de
software y redes

Autor:

ING. MARCO VINICIO ALTAMIRANO RUIZ

Director:

ING. MSC. PATRICIO MEDINA

Ambato – Ecuador

Julio 2010



Nº de ingreso: 005526

Precio: \$80.00

canje: Donación: Compra:

Fecha de factura:

Fecha de ingreso: 08102010

PONTIFICIA UNIVERSIDAD CATOLICA DEL ECUADOR

SEDE AMBATO

DEPARTAMENTO DE INVESTIGACION, POSTGRADOS Y

AUTOEVALUACION

HOJA DE APROBACION

Tema:

“BUSINESS INTELLIGENCE APLICADO A LOS CLIENTES DE
MADERVAS, PARA UNA OPORTUNA TOMA DE DECISIONES”

Autor:

MARCO VINICIO ALTAMIRANO RUIZ

Ing. M.Sc. Patricio Medina

f.

DIRECTOR DE TESIS

Ing. M.Sc. Andrés López

f.

CALIFICADOR

Ing. M.Sc. Marco Polo Silva

f.

CALIFICADOR

Ing. Telmo Viteri

f.

JEFE DEL D.I.P.A.

Dr. Pablo Poveda

f.

SECRETARIO GENERAL PUCESA



DECLARACION DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, Marco Vinicio Altamirano Ruiz portador de la cédula de ciudadanía N° 1802916385 declaro que los resultados obtenidos en la investigación que presento como informe final, previo a la obtención del título de Magister en Gerencia Informática con mención Desarrollo de Software y Redes, son absolutamente originales auténticos y personales.

En tal virtud, declaro que el contenido, las conclusiones y los efectos legales y académicos que se desprenden del trabajo propuesto de investigación y luego de la redacción de este documento son y serán de mi sola y exclusiva responsabilidad legal y académica.



Marco Vinicio Altamirano Ruiz.

C.I. 1802916385

AGRADECIMIENTO

Mi agradecimiento a Dios que es mi luz y mi camino.

A mi amada esposa por su comprensión, y apoyo incondicional.

A mis padres Sra. María Ruiz Reyes y Sr. Marco Altamirano V., que han sabido guiarme con amor, comprensión, honradez y dedicación.

A la compañía Madervas fuente del presente estudio y en especial a su Gerente General Sr. Santiago Vásconez C., por permitir desarrollar este trabajo tan importante para el desarrollo y crecimiento de la organización.

A todas aquellas personas y amigos, que aportaron en la consecución de este trabajo investigativo.

A la Prestigiosa Pontificia Universidad Católica del Ecuador sede Ambato, por ser la fuente de nuevas oportunidades para difundir el conocimiento.

A los Maestros que con un lenguaje claro difundieron sus excelentes conocimientos.

DEDICATORIA

Al culminar el presente trabajo de investigación quiero dedicar con inmenso amor a mi esposa Mercy Elizabeth que con su apoyo incondicional he podido conseguir las metas propuestas, y a mis amados hijos que son fuente de mi inspiración y fuerza para seguir adelante Andreita Michelle y Dieguito Andrés.

RESUMEN

La empresa Madervas o Placacetro Masisa Arboriente comercializa productos de la franquicia otorgada por MASISA de Chile, además de la distribución exclusiva de tableros contrachapados de la marca Arboriente, y de toda la gama de productos que un carpintero necesita.

Siendo Madervas una empresa exitosa y en constante crecimiento, la gestión del negocio también va incrementando los requerimientos de Información, siendo cada vez mayor cantidad de datos que tienen que ser procesados manualmente, dedicando más tiempo a la generación de dicha información, quitando tiempo para su análisis y posterior toma de decisiones.

Madervas s.a. con el objetivo de ser cada vez más competitivos, y atender mejor a sus clientes, ha decidido implementar la solución de negocios inteligentes o Business Intelligence, la misma que nos permite analizar mejor a los clientes clasificados o agrupados por varios criterios: por períodos de tiempo, por vendedor, grupo de inventarios, etc., establecer los productos de baja rotación y también los de mayor movimiento.

Business Intelligence es una tecnología que para la empresa significa un apoyo importante para la toma de buenas decisiones, establecer metas, porcentajes de crecimiento, siendo de gran aporte para que los indicadores macros lleguen a las metas establecidas por la Gerencia general.

ABSTRACT

Madervas or Placacentro Masisa Arboriente Enterprise commercializes products from Masisa franchise of Chile, as well as the exclusive distribution of plywood boards from Arboriente trademark, and all the range products that a carpenter needs.

As Madervas is a successful enterprise and having a constantly growing, the business gestion is increasing the information requirements, having to process handly more amount of data, spending more time on generate that information, resting time for its analysis and making posterior decisions.

Madervas S.A. with the objective to be more competitive and attend better its customers, it has decided to implement the Business Intelligence. It allows to analyze better the customers who are classified for various criterions: time periods, salespersons, inventory groups and others. And to establish which are the low rotation and the more movement products.

Business Intelligence is a technology that represents an important support for the enterprise to make good decisions, establish goals, growth percentages, helping the macro-indicators to get the goal established by the General Management.

TABLA DE CONTENIDOS

CAPITULO I

1. PROYECTO DE INVESTIGACION.....	1
1.1. Antecedentes	1
1.2. Significado del Problema	1
1.3. Problematización.....	2
1.4. Delimitación.....	3
1.5. Justificación.....	3
1.6. Objetivos	4
1.6.1. Objetivo General	4
1.6.2. Objetivos Específicos	4
1.7. Hipótesis	5

CAPITULO II

2. Business Intelligence.....	6
2.1. Datos, Información Conocimiento.....	8
2.2. Componentes del Business Intelligence.....	11
2.2.1. Orígenes de Datos.....	12
2.2.2. Proceso ETL.....	12
2.2.3. Data WareHouse.....	13
2.2.3.1.Data Mart.....	15
2.2.4. Cubos OLAP.....	16
2.2.4.1. Sistemas de Almacenes de los Cubos OLAP.....	19
2.2.4.2. Dimensiones de los Cubos OLAP.....	22
2.2.5. Data Mining.....	23

2.2.5. Tipos de Datos.....	26
2.2.6 Extracción del conocimiento.....	27
2.3. Plataformas de Business Intelligence.....	30
2.3.1. Ejemplos de Implementación de Business Intelligence.....	30
2.4. SQL Server Business Intelligence Development Studio.....	32
2.4.1. SQL Server Integración de Servicios (SSIS).....	34
2.4.2. SQL Server Analysis Services (SSAS).....	40
2.4.2.1. Arquitectura de Analysis Services.....	41
2.4.2.2. Minería de Datos en Analysis Services.....	44
2.4.3. SQL Server Reporting Services.....	68
CAPITULO III	
3. POR QUÉ IMPLEMENTAR BUSINESS INTELLIGENCE EN MADERVAS?.....	70
3.1. Encuesta.....	70
3.1.1. Objetivo de la Investigación.....	70
3.1.2. Objetivos Específicos.....	70
3.1.3. Método de Investigación.....	72
3.1.4 Muestra de la Investigación.....	72
3.1.5. Instrumento de Recolección.....	72
3.2. Tabulación de la Encuesta.....	83
3.3. Análisis de la Encuesta.....	83
CAPITULO IV	
4. BUSINESS INTELLIGENCE EN MADERVAS.....	85
4.1. Solución Propuesta.....	85
4.2. SQL Server Business Intelligence Development Studio.....	85



4.3. Origen de Datos.....	86
4.4. Servicio de Integración de Datos.....	86
4.5. Creación de la Solución BI	90
4.6. Data Mining y Extracción del Conocimiento.....	99
4.7. Solución BI.....	108
CAPITULO V	
5. VALIDACION.....	111
5.1. Demostración de la Hipótesis.....	111
CAPITULO VI	
6. Conclusiones y Recomendaciones.....	119
6.1. Conclusiones.....	119
6.2. Recomendaciones.....	120
BIBLIOGRAFIA.....	121
GLOSARIO.....	125
ANEXOS.....	128

TABLA DE GRÁFICOS

Gráfico 2.1. Datos, Información conocimiento.....	6
Gráfico 2.2. Pirámide Datos, información Conocimiento.....	8
Gráfico 2.3. Componentes de una solución BI.....	11
Gráfico 2.4. Data Mart.....	16
Gráfico 2.5. Representación de un Cubo.....	16
Gráfico 2.6. Estructura de las tablas de un cubo OLAP.....	17
Gráfico 2.7. Almacén de datos OLAP	19
Gráfico 2.8. Almacén de datos ROLAP.....	21
Gráfico 2.9. Exploración de una Dimensión	23
Gráfico 2.10. Extracción de Conocimiento	29
Gráfico 2.11. Microsoft Business Intelligence Development Studio.....	33
Gráfico 2.12. Plataforma SQL Server 2005.....	33
Gráfico 2.13. Plataforma de la Tecnología BI.....	34
Gráfico 2.14. Microsoft Integración de Servicios.....	35
Gráfico 2.15. Arquitectura del Microsoft Integration Services.....	37
Gráfico 2.16. Ejemplo de la Arquitectura de Integration Services.....	40
Gráfico: 2.17. Proceso Data Mining.....	45
Gráfico 2.18. Algoritmo de clústeres.....	51
Gráfico 2.19 Reglas de Asociación (Algoritmo de Asociación).....	54
Gráfico 2.20. Algoritmo Serie Temporal.....	57
Gráfico 2.21. Algoritmo de regresión.....	60
Gráfico 2.22. Algoritmo de regresión	61
Gráfico 2.23. Algoritmo de regresión lineal.....	62

Gráfico 2.24. Algoritmos de minería y su aplicación.....	66
Gráfico 2.25. Algoritmos de minería y su aplicación.....	67
Gráfico 2.26. Representación Gráfica de los Algoritmo de Data Mining.	68
Gráfico 3.1. Tabulación Encuesta Pregunta N° 1.....	73
Gráfico 3.2. Tabulación Encuesta Pregunta N° 2.....	74
Gráfico 3.3. Tabulación Encuesta Pregunta N° 3.....	75
Gráfico 3.4. Tabulación Encuesta Pregunta N° 4.....	76
Gráfico 3.5. Tabulación Encuesta Pregunta N° 5.....	77
Gráfico 3.6. Tabulación Encuesta Pregunta N° 6.....	78
Gráfico 3.7. Tabulación Encuesta Pregunta N° 7.....	79
Gráfico 3.8. Tabulación Encuesta Pregunta N° 8.....	80
Gráfico 3.9. Tabulación Encuesta Pregunta N° 9.....	81
Gráfico 3.10. Tabulación Encuesta Pregunta N° 10.....	82
Gráfico 3.11. Tabulación resultado encuesta	83
Gráfico 4.1. Paquete de Integration Services para Madervas.....	87
Gráfico 4.2. Conexión a Origen de datos.....	88
Gráfico 4.3. Conexión a Destino de datos.....	88
Gráfico 4.4. Selección de Tablas.....	89
Gráfico 4.5. Guardar el paquete de exportación de datos.....	89
Gráfico 4.6. Creación Proyecto de Analysis Services.....	90
Gráfico 4.7. Conexión al Origen de Datos.....	91
Gráfico 4.8. Vista de las Tablas de Origen de Datos.....	92
Gráfico 4.9. Creación de un nuevo Cubo.....	92
Gráfico 4.10. Método de Generación del Cubo.....	93

Gráfico 4.11. Selección vista de origen de datos	93
Gráfico 4.12. Detección de tablas de hechos y dimensiones.....	94
Gráfico 4.13. Identificación de tablas de hechos y dimensiones.....	95
Gráfico 4.14. Selección de medidas.....	95
Gráfico 4.15. Medidas a generarse.....	96
Gráfico 4.16. Estructura del cubo a crearse.....	96
Gráfico 4.17. Implementación del proyecto.....	97
Gráfico 4.18. Generación del proyecto y creación de la base de datos..	98
Gráfico 4.19. Exploración de cubo generado.....	99
Gráfico 4.20. Nueva estructura de minería de datos.....	100
Gráfico 4.21. Definir la estructura de minería de datos.....	101
Gráfico 4.22. Definir técnica de de minería de datos.....	102
Gráfico 4.23. Selección de dimensión de cubo de origen.....	103
Gráfico 4.24. Selección de Clave de escenario.....	104
Gráfico 4.25. Selección de columnas de nivel escenario.....	105
Gráfico 4.26. Especificar columnas de entrada y de predicción.....	106
Gráfico 4.27. Contenido y tipo de datos de columnas	107
Gráfico 4.28. Estructura de minería de datos.....	108
Gráfico 4.29. Navegación de la Solución BI en Microsoft Excel.....	109
Gráfico 4.30. Complemento de minería de datos para Microsoft Excel..	109
Gráfico 4.31. Visor de minería de datos con el algoritmo árboles de decisión	110
Gráfico 5.1. Cubo de ventas	112
Gráfico 5.2. Cubo de Inventarios.....	113

Gráfico 5.3. Cubo de Inventarios con productos de baja rotación.....	114
Gráfico 5.4. Probabilidades de ventas por tipo de precio.....	115
Gráfico 5.5. Visor árboles de decisiones por tipo de precio.....	116
Gráfico 5.6. Probabilidad de venta por forma de pago.....	116
Gráfico 5.7. Visor árboles de decisión por forma de pago.....	117
Gráfico 5.8. Probabilidad de ventas por Zonas.....	118

TABLAS

Tabla 2.1. Aplicación de Algoritmos de Minería de datos Microsoft	66
Tabla 3.1. Tabla de objetivos e Hipótesis de la encuesta.....	70

CAPITULO I

1. PROYECTO DE INVESTIGACION

1.1. Antecedentes

Madervas es una sociedad anónima legalmente constituida, siendo su nombre comercial PLACENTRO MASISA ARBORIENTE, en el año 2002 obtiene la franquicia de la empresa Chilena Masisa para comercializar los productos MASISA tales como: MDF, Melaminas, Melamínicos, OSB, Herrajes y Accesorios, etc.; es decir todos los materiales que un maestro carpintero necesita. Como valor agregado al servicio de los clientes también ofrece: el diseño de muebles, el dimensionado de tableros con la optimización de corte, corte de tableros, y servicio de tapacanto.

El Almacén de comercialización se encuentra ubicado en la ciudad de Ambato, Provincia de Tungurahua. La empresa inicia sus actividades en Junio de 2003.

Actualmente la empresa cuenta con una amplia gama de productos que dan una excelente alternativa a los clientes para crear sus soluciones en el diseño de muebles.

1.2. Significado del problema

La gran cantidad de de información que se mantiene en la base de datos y sumada la información que ingresa diariamente en el Sistema Financiero, es imposible realizar un análisis manual de esta información. Con la ausencia de Business Intelligence la gerencia puede tomar decisiones sin analizar los riesgos y oportunidades.

1.3. Problematización

El desconocimiento de la tendencia de compra de los clientes puede ser una desventaja comercial, ya que no sabemos cuál es la frecuencia de compra, que tipo de productos adquiere, el valor promedio de los montos de compra, etc. Toda esta información es casi imposible analizarla manualmente por la gran cantidad de datos.

Adicionalmente la gerencia de ventas realiza compras sin un debido análisis de rotación de productos, creando un alto inventario de productos que se convierten en Hueso, así lo define en análisis matriz de la posición en el según BCG el Boston Consulting Group.

En este proyecto se propone analizar la información de de los clientes utilizando Business Intelligence y poder establecer los siguientes parámetros:

1. Qué tipo de producto y cantidad que adquiere.
2. Frecuencia (en periodos de tiempo).

3. Montos de compra.

Este análisis nos proporcionará información para predecir la cantidad de producto debemos tener en inventarios para cumplir con las expectativas de los clientes, o también poderle hacer conocer al cliente que tenemos en stock el producto que consume e impulsarle a comprar. Para lo cual se desarrollará e implementará un cubo de información de Clientes y de Inventarios.

El propósito principal de Business Intelligence es el proceso descubrir la información empresarial que de cualquier otra forma permanecía oculta.

1.4. Delimitación

Este proyecto de investigación se implementará en la empresa MADERVAS S.A., ubicada en la Provincia de Tungurahua, en la ciudad de Ambato. El análisis e implementación de la solución se desarrollará para realizar un mejor seguimiento y análisis a los clientes y poder tener un inventario adecuado de productos, para satisfacer a los clientes.

1.5. Justificación

Las grandes empresas generan más información en una semana que la que cualquier persona podría leer en toda su vida, e incluso las pequeñas empresas generan un volumen de datos que no son capaces de manejar.

De modo que actualmente nos enfrentamos a la paradoja de que, cuantos más datos están disponibles, menos información tenemos.

Para superar este problema, en los últimos años han surgido una serie de técnicas que facilitan el procesamiento avanzado de los datos y permiten realizar un análisis con profundidad de los mismos de forma automática. La idea clave es que los datos contienen más información oculta de la que se ve a simple vista.

Madervas s.a. debe ir mejorando tecnológicamente implementando las nuevas herramientas de inteligencia empresarial o Business Intelligence para que la Gerencia pueda tomar las decisiones a tiempo y sea una empresa más rentable y competitiva.

1.6. Objetivos

1.6.1. Objetivo General

Desarrollar e Implementar Business Intelligence para que la gerencia pueda analizar y tomar decisiones eficientes y a tiempo.

1.6.2. Objetivos Específicos

a) Implementar un cubo de información de Inventarios para poder establecer los stocks mínimos y máximos.

- b) Analizar la rotación de productos de baja rotación, y tomar una decisión en razón de los mismos.
- c) Mejorar las negociaciones con los proveedores.- al momento de realizar compras a Masisa ya que se va a tener la información necesaria, de tal forma que se adquiera los productos necesarios y en las cantidades requeridas.
- d) Predecir comportamientos de ventas utilizando mediante la extracción del conocimiento utilizando el algoritmo de arboles de decisión

1.7. Hipótesis

Aplicando Business Intelligence a los clientes de Madervas se logrará tomar decisiones para incrementar las ventas a través del análisis continuo del conocimiento extraído tomando decisiones adecuadas, para lo cual se debe mantener un inventario adecuado de productos, también logrando mejorar las negociaciones con los proveedores.

CAPITULO II

2. BUSINESS INTELLIGENCE

Conocido también como Inteligencia de Negocios, Inteligencia Empresarial o Inteligencia para la empresa; es conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales en una información estructurada, para su análisis y conversión en conocimiento, así dando un soporte a la toma de decisiones gerenciales en la empresa o negocio, ya que el futuro de una empresa u organización depende de la calidad de sus decisiones.



Gráfico 2.1. Datos, información Conocimiento.

BI apoya a la toma de decisiones basada en un análisis avanzado de datos y estimaciones que permiten afrontar una decisión minimizando el riesgo de error o aleatoriedad de la misma.

Geoffrey A. Moore Director de Chasm Group dice que: En este nuevo mundo, la información reina. Vivimos en una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios. Para mantenerse competitiva una empresa, los gerentes y tomadores de 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.

La tecnología de BI no es nueva, ha estado presente de varias formas por lo menos en los últimos 20 años, comenzando por generadores de reportes y sistemas de información ejecutiva en los 80's Afirma Candice Goodwin. Entiéndase como sinónimos de tecnología de BI los términos aplicaciones, soluciones o software de inteligencia de negocios.

BI es factor estratégico para una empresa, generando una potencial ventaja competitiva, proporcionando información privilegiada y de cierta forma oculta ya que por la gran cantidad de información no se puede analizar manualmente; para responder efectivamente a los problemas de negocio y minimizando el riesgo de una mala decisión.

Estos son algunos ejemplos de productos que se han desarrollado utilizando Business Intelligence son:

- Sistemas de Información Ejecutiva. (EIS).
- Sistemas de Soporte de a la Decisión (DSS).

- Cuadro de mando Integrales (CMI) conocido también como Balanced Score Card (BSC).

2.1. Datos, Información, Conocimiento

Según Davenport y Prusak (1999), lo define de la siguiente manera:



Gráfico 2.2. Pirámide Datos, información Conocimiento.

Datos

Un dato es un conjunto discreto, de factores objetivos sobre un hecho real. Dentro de un contexto empresarial, el concepto de dato es definido como un registro de transacciones. Un dato no dice nada sobre el porqué de las cosas, y por sí mismo tiene poca o ninguna relevancia o propósito.

Las organizaciones actuales normalmente almacenan datos mediante el uso de tecnologías. Desde un punto de vista cuantitativo, las empresas evalúan la gestión de los datos en términos de coste, velocidad y capacidad.


Todas las organizaciones necesitan datos y algunos sectores son totalmente dependientes de ellos. Bancos, compañías de seguros, agencias


gubernamentales y la Seguridad Social son ejemplos obvios. En este tipo de organizaciones la buena gestión de los datos es esencial para su funcionamiento, ya que operan con millones de transacciones diarias. Pero en general, para la mayoría de las empresas tener muchos datos no siempre es bueno. Las organizaciones almacenan datos sin sentido. Realmente esta actitud no tiene sentido por dos razones. La primera es que demasiados datos hacen más complicado identificar aquellos que son relevantes. Segundo, y todavía más importante, es que los datos no tienen significado en sí mismos.


Los datos describen únicamente una parte de lo que pasa en la realidad y no proporcionan juicios de valor o interpretaciones, y por lo tanto no son orientativos para la acción. La toma de decisiones se basará en datos, pero estos nunca dirán lo que hacer. Los datos no dicen nada acerca de lo que es importante o no. A pesar de todo, los datos son importantes para las organizaciones, ya que son la base para la creación de información.

Información


La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor:

 **Contextualizando:** se sabe en qué contexto y para qué propósito se generaron.

 **Categorizando:** se conocen las unidades de medida que ayudan a interpretarlos.

 **Calculando:** los datos pueden haber sido procesados matemática o estadísticamente.

 **Corrigiendo:** se han eliminado errores e inconsistencias de los datos.

 **Condensando:** los datos se han podido resumir de forma más concisa (agregación).

Por lo tanto, la información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre).

Conocimiento

El conocimiento es una mezcla de experiencia, valores, información y know-how que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también está en rutinas organizativas, procesos, prácticas, y normas.

2.2.1. Orígenes de datos

Estas pueden ser muy diversas, de las cuales se parte para cargar la información al data warehouse.

Sistemas Operacionales OLTP (On lineTransaction Processing): Los sistemas transacciones son la principal fuente de los datos.

Fuentes Externas: los datos externos a la empresa y que provienen de consultoras, del gobierno, u otras fuentes.

ERP: De los Sistemas de Planificación de recursos Empresariales (Enterprise Resource Planning).

2.2.2. Proceso ETL


Se denomina ETL porque es el Proceso que extrae, transforma y carga los datos en el data warehouse. Pero antes de almacenar los datos en el data warehouse éstos deben ser transformados, limpiados, filtrados y redefinidos. Por lo general la información que tenemos en los sistemas transaccionales no está preparada para la toma de decisiones, por eso es necesario ejecutar este proceso.

2.2.3. Data warehouse

Es una estructura única, y está integrada por uno o más cubos. El Data warehouse es una base de datos corporativa contiene varios componentes que permite la transferencia de datos y que se caracteriza por integrar y depurar información de uno o más orígenes de datos, para luego procesarla permitiendo su análisis y con grandes velocidades de respuesta. La creación de un data warehouse es el primer paso, para implementar una solución Business Intelligence.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales). Este tipo de información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma.

El término Data warehouse dice Bill Inmon, y se traduce literalmente como almacén de datos, y se caracteriza por ser:

 **Integrado:** los datos almacenados en el data warehouse se integran en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales se eliminan. La información suele

estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

- **Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del data warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:** el tiempo es parte implícita de la información contenida en un data warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el data warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el data warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

- **No volátil:** el almacén de información de un data warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del data warehouse la incorporación de los

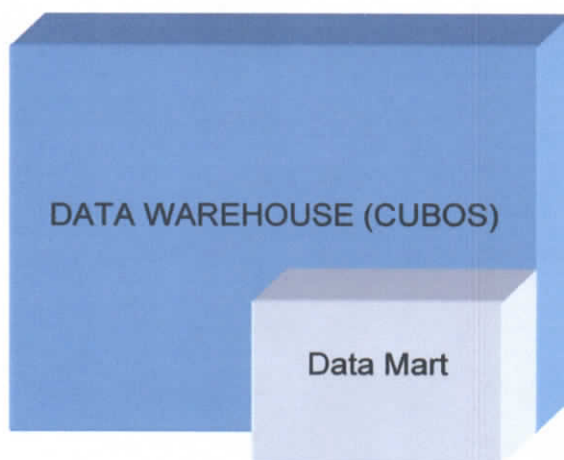
últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Otra característica del data warehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo.

2.2.3.1 Datamart

Data mart es un subconjunto de data warehouse puede ser una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica (por ejemplo: cubo de ventas, inventarios). Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.

Un data mart puede ser alimentado desde los datos de un data warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.



Fuente: Designing and Implementing OLAP Solutions. Microsoft course 2074A.

2.2.4. Cubos OLAP

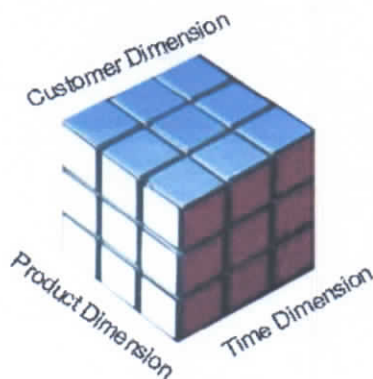


Gráfico 2.5. Representación de un Cubo

Los cubos OLAP (Online Analytical Processing) o modelos multidimensionales, se organizan los datos en torno a los datos históricos

(hechos), que tienen atributos o medidas que pueden verse en mayor o menor detalle según las dimensiones creadas.

Ejemplo: En la siguiente figura se representa parte de un almacén de datos con estructura multidimensional en la cual la tabla de hechos es la de ventas.

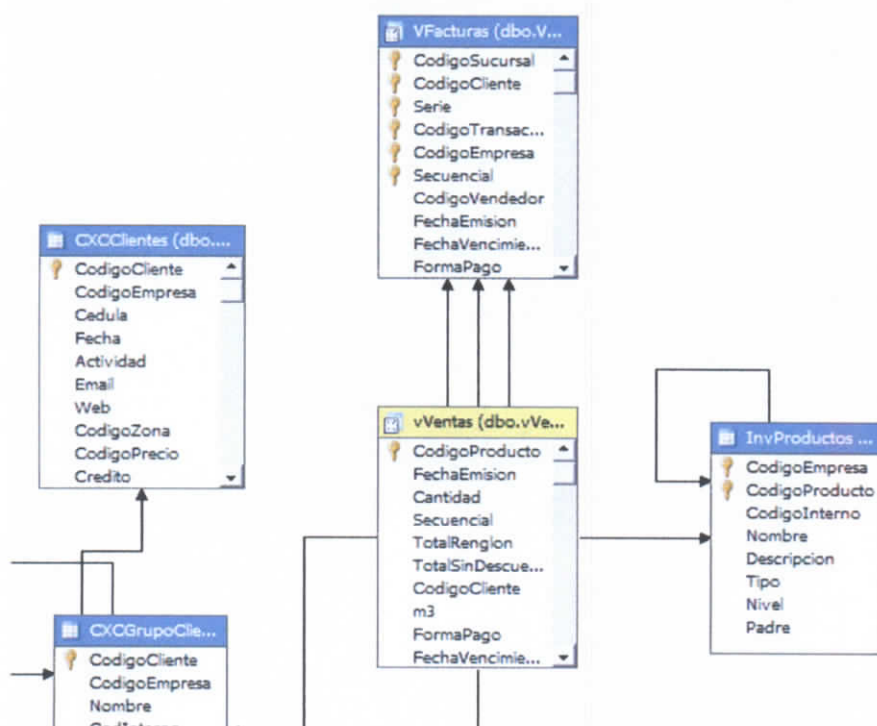


Gráfico 2.6. Estructura de las tablas de un cubo OLAP.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

La tecnología OLAP proporciona una alternativa a la tecnología de bases de datos relacionales y ofrece fácil y flexible visualización de la información,

análisis, capacidad de cálculo, consultas, funciones de planeamiento, y análisis de escenarios de grandes volúmenes de información.

Las bases de datos OLAP se adaptan a una amplia gama de aplicaciones para la inteligencia de negocios.

Las principales características de la tecnología OLAP son:

1. Las bases de datos OLAP tienen un esquema optimizado para consultas rápidas.

Las consultas OLAP son muy rápidas, y permiten una mayor interactividad a los usuarios que los típicos reportes obtenidos desde las bases de datos relacionales.

2. Los Cubos OLAP almacenan distintos niveles de datos resumidos en las estructuras de datos optimizadas para las consultas rápidas.

3. Las bases de datos OLAP tienen un robusto motor de cálculo para el análisis numérico. EL mismo que nos permite obtener informes de producción ventas simples y también puede realizar complejos algoritmos de asignación. Muchos cálculos avanzados realizados por los motores de cálculo de OLAP; no puede ser realizada por bases de datos relacionales, debido a la limitación de análisis en el RDBMS.

4. Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL).

2.2.4.1. Sistemas de Almacenes de los cubos OLAP

Los sistemas de almacenes de bases de datos pueden implementarse utilizando dos tipos de esquemas físicos:


 **MOLAP.**- Multidimensional OLAP, físicamente el almacén de datos se construye sobre estructuras basadas en matrices multidimensionales.



Gráfico 2.7. Almacén de datos OLAP

Fuente: Designing and Implementing OLAP Solutions. Microsoft course 2074A.

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la

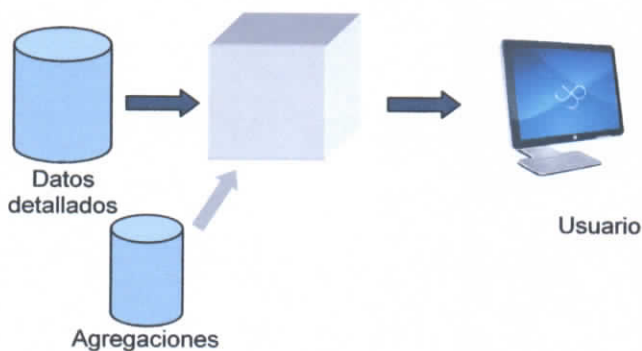


Gráfico 2.8. Almacén de datos ROLAP

Fuente: Designing and Implementing OLAP Solutions. Microsoft course 2074A.

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

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

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP. Después de que el modelo de datos para el data warehouse se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por los modelos de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos pre calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del data warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

2.2.4.2. Dimensiones de los Cubos OLAP

Las dimensiones son un punto fundamental para la construcción de un proceso analítico en línea (Cubo OLAP).

Los cubos contienen muchas dimensiones, que pueden ser públicas o privadas.

Una dimensión contiene niveles y miembros organizados jerárquicamente, que clasifica el número de combinaciones e intersecciones, con la que se analiza los datos. Cada dimensión describe un aspecto del negocio a los usuarios y facilita el acceso intuitivo y simple a la información.

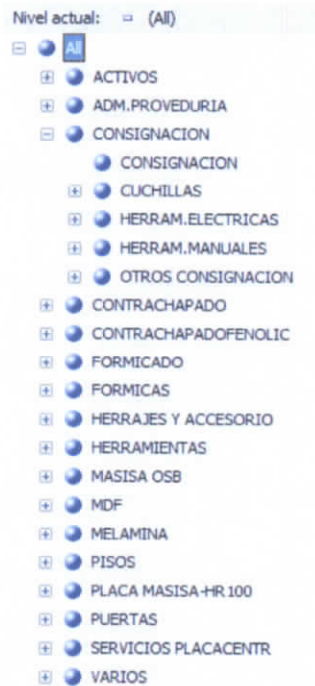


Gráfico 2.9. Exploración de una Dimensión.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

2.2.5. Data Mining

El data mining (minería de datos), es proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. La tarea fundamental es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semi-automático (asistido) y el uso de los patrones descubiertos debería ayudar a tomar decisiones más seguras que reporten beneficio a la empresa u organización.

La minería de datos tiene dos objetivos principales.

1. Trabajar con grandes volúmenes de datos, procedente la mayor parte de sistemas de información, con los problemas que ello conlleva (ruido, datos ausentes, intratabilidad, volatilidad de los datos).
2. Utilizar técnicas adecuadas para analizar estos grandes volúmenes de datos y extraer conocimiento novedoso y útil. En muchos casos la utilidad del conocimiento minado está íntimamente relacionado con la comprensibilidad del modelo inferido. No se debe olvidar que el usuario final no tiene por qué ser un experto en las técnicas de minería de datos, ni tampoco puede perder mucho tiempo interpretando los resultados. Por lo tanto es importante que la aplicación proporcione la información descubierta sea más comprensible por los humanos.

El data mining no es nuevo. Ya desde los años sesenta los estadísticos manejaban términos como data fishing, data mining o data archaeology con

la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido. A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, empezaron a consolidar los términos de data mining y Knowledge Discovery in Databases (KDD) A finales de los años ochenta sólo existían un par de empresas dedicadas a esta tecnología; en 2002 existen más de 100 empresas en el mundo que ofrecen alrededor de 300 soluciones. Las listas de discusión sobre este tema las forman investigadores de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios.

El data mining es una tecnología compuesta por etapas que integra varias áreas y que no se debe confundir con un gran software. Durante el desarrollo de un proyecto de este tipo se usan diferentes aplicaciones software en cada etapa que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente. Actualmente existen aplicaciones o herramientas comerciales de data mining muy poderosas que contienen un sinfín de utilerías que facilitan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta.

La tecnología de Data Mining puede generar nuevas oportunidades de negocios al proveer estas capacidades:

• Predicción automatizada de tendencias y comportamientos. Data Mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que tradicionalmente requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos.

• Descubrimiento automatizado de modelos previamente desconocidos. Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un sólo paso.

Cuando las herramientas de Data Mining son implementadas en sistemas de procesamiento paralelo de alta performance, pueden analizar bases de datos masivas en minutos. Procesamiento más rápido significa que los usuarios pueden automáticamente experimentar con más modelos para entender datos complejos. Alta velocidad hace que sea práctico para los usuarios analizar inmensas cantidades de datos. Grandes bases de datos, a su vez, producen mejores predicciones.

2.2.5.1 Tipos de datos

El data Mining se puede aplicar a cualquier tipo de información, siendo la técnica de minería diferentes para cada una de ellas.

Entre los tipos de datos están:

- Datos estructurados provenientes de bases de datos relacionales.
- Otros tipos de datos estructurados en bases de datos como son: espaciales, temporales, textuales y multimedia.
- Y datos no estructurados provenientes de la web o de otros tipos de repositorios de documentos.

2.2.6. Extracción del Conocimiento

Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

El proceso de extracción del conocimiento o KDD, se organiza en cinco fases:

- 1. Fase de Integración y recopilación de datos.-** Se determinan las fuentes de información que pueden ser útiles y donde conseguirlas. Luego se transforman todos los datos a un formato común, frecuentemente mediante un almacén de datos que consiga unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias.
- 2. Fase de Selección, limpieza y transformación.-** En esta fase se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos.

3. Fase de Minería de datos.- Se decide cuál es la tarea a realizar como es clasificar, agrupar, etc., y se elige el método que se va a utilizar.

4. Fase de Evaluación e interpretación.- Se evalúan los patrones y se analizan por los expertos y si es necesario se vuelve a las fases anteriores para una nueva iteración.

5. Fase de Difusión.- Se hace uso del nuevo conocimiento, y se hace comparte y difunde la información al personal involucrado.

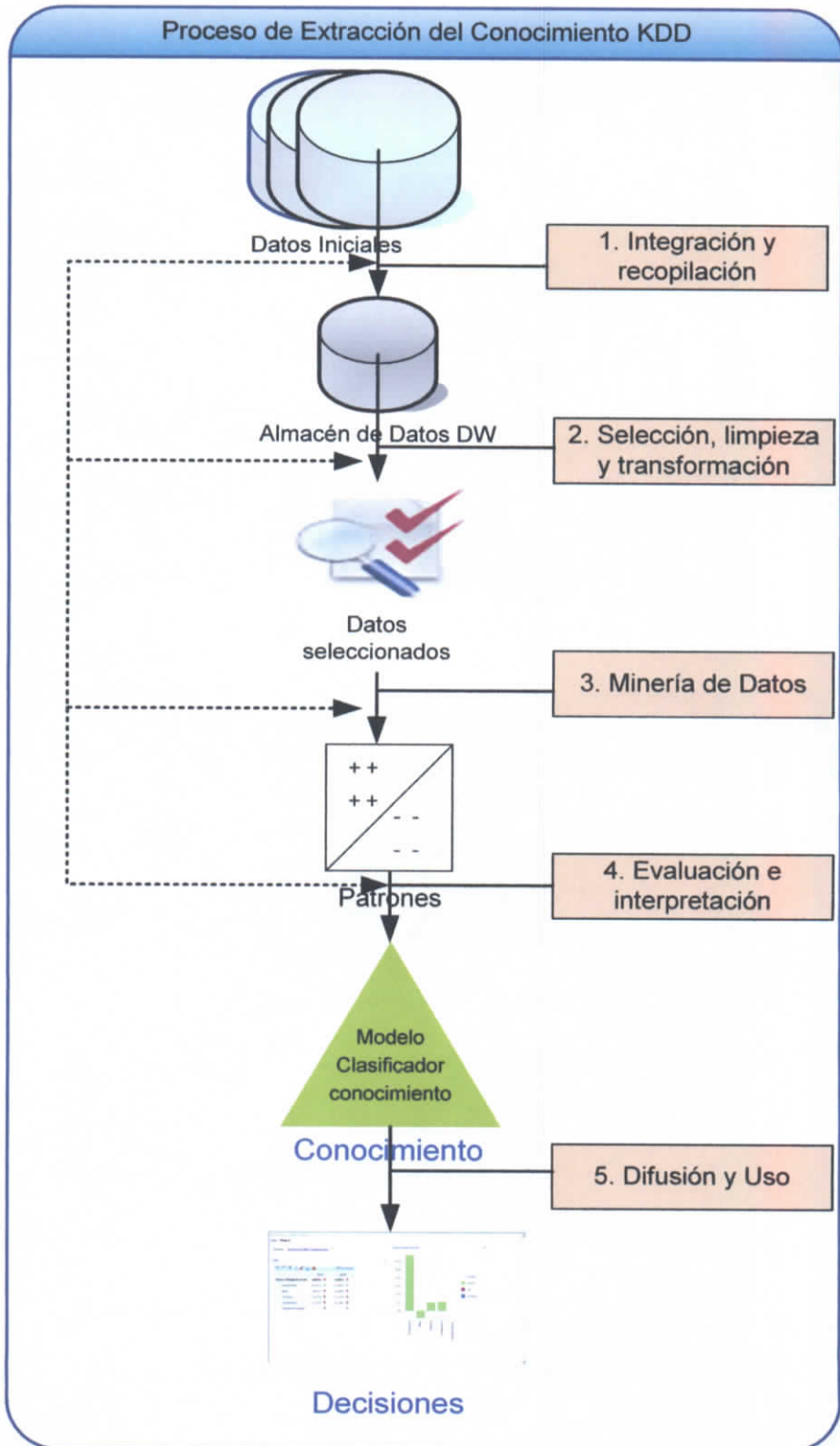


Gráfico 2.10. Extracción de Conocimiento.

Fuente: Introducción a la Minería de Datos. PEARSON EDUCACION S.A., Madrid, 2004.

2.3. Plataformas de Business Intelligence

Existen varias plataformas que se utilizan para business Intelligence y estas son:



2.3.1. Ejemplos de Implementación de Business Intelligence

A nivel mundial existen muchos ejemplos de business Intelligence que se puede citar.

AC. Milan Gana más juegos con Microsoft

AC Milán el equipo de fútbol más laureado del mundo. Una parte clave de su éxito es el laboratorio de Milán, que ayuda a mejorar el rendimiento de los jugadores basados en el análisis de sus datos de rendimiento que recopila. Milan Lab. utiliza software de Microsoft para reunir y analizar esos datos. Microsoft se ha convertido en una parte integral del éxito de este equipo ya que ayuda a AC Milan aportar mayor significado a sus datos

Departamento de Transporte de Mississippi

Este departamento es responsable de mantener las carreteras del estado y otras infraestructuras de transporte utiliza Business Intelligence para resolver los problemas de transporte, analizando la movilidad de las personas, y dar una solución efectiva.

Proyecto Sky Image Cataloguing y Analysis Tool - SKYCAT.

En astronomía y ciencias del espacio, el problema es manejar los grandes volúmenes de información acumulados, correlacionar los datos, extraer y visualizar las tendencias importantes, este problema aumentó con la llegada de nuevos telescopios, detectores, misiones espaciales con flujo de datos en terabytes, enfrentándose a una crítica necesidad de una tecnología de procesamiento de información y metodología para gestionar estos datos.

En el lapso de 6 años, el "Second Palomar Observatory Sky Survey" (POSS-II) recopiló tres terabytes de imágenes que contenían aproximadamente dos millones de objetos en el cielo, 3000 fotografías fueron digitalizadas a una resolución de 16 bits por píxel con 23.040 x 23.040 píxeles por imagen. El propósito de SKICAT es permitir y maximizar la extracción de información significativa de una gran base de datos en forma oportuna. El sistema Sky Image Cataloguing and Analysis Tool (SKYCAT) se basa en técnicas de agrupación (clustering) y árboles de decisión para poder clasificar los objetos en estrellas, planetas, sistemas, galaxias, etc., con una alta confiabilidad (Fayyad y otros, 1996). Los resultados han ayudado a los astrónomos a descubrir dieciséis nuevos quásars con corrimiento hacia el rojo que los

incluye entre los objetos más lejanos del universo y, por consiguiente, más antiguos. Estos quásars son difíciles de encontrar y permiten saber más acerca de los orígenes del universo.

2.4. SQL Server Business Intelligence Development Studio



La visión de Microsoft para Business Intelligence es el de mejorar la velocidad y la calidad de la toma de decisiones a través de Microsoft SQL Server 2005 y Microsoft .NET. Estos permiten la entrega rápida de la información inteligente y relevante a través de servicios integrados y la prestación de abrir interfaces para acceder y compartir información.

Microsoft proporciona el poder de convertir, los datos en información, y de esta información extraer el conocimiento. Para lograr este objetivo Microsoft nos proporciona la herramienta “**SQL Server Business Intelligence Development Studio**”, la misma que viene integrada con el paquete de SQL server.

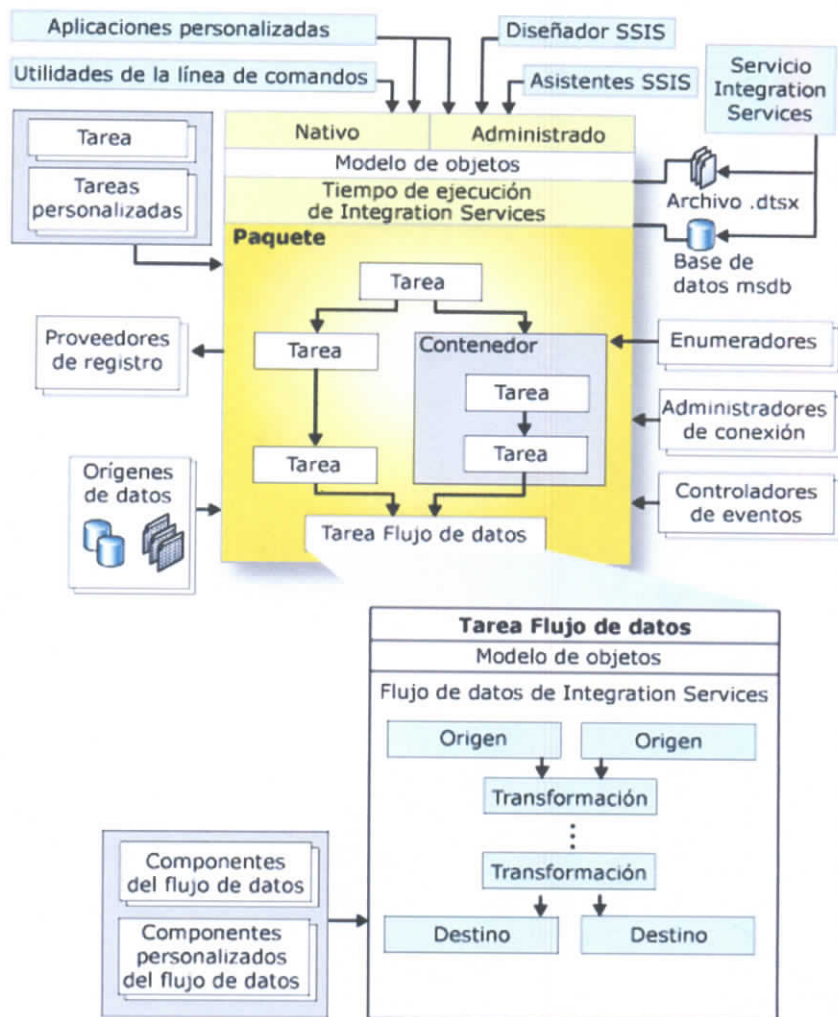


Gráfico 2.16. Arquitectura del Microsoft Integration Services

Fuente: Documentación de Microsoft Visual Studio.

Los siguientes componentes son importantes para utilizar Integration Services adecuadamente:

Diseñador SSIS

El Diseñador SSIS es una herramienta gráfica que se puede usar para crear y mantener paquetes Integration Services. El Diseñador SSIS está

disponible en Business Intelligence Development Studio como parte de un proyecto de Integration Services.

Motor en tiempo de ejecución

El tiempo de ejecución de Integration Services guarda el diseño de paquetes, ejecuta paquetes y admite registros, puntos de interrupción, configuración, conexiones y transacciones.

Tareas y otros ejecutables

Los ejecutables de tiempo de ejecución de Integration Services son el paquete, los contenedores, las tareas y los controladores de eventos que incluye Integration Services. Los ejecutables de tiempo de ejecución también incluyen tareas personalizadas que el usuario puede desarrollar.

Motor de flujo de datos (también conocido como canalización) y componentes de flujo de datos

La tarea Flujo de datos encapsula el motor de flujo de datos. El motor de flujo de datos proporciona los búferes en memoria que mueven datos desde el origen hasta el destino y llama los orígenes que extraen datos de archivos

y bases de datos relacionales. El motor de flujo de datos también administra las transformaciones que modifican datos y los destinos que cargan datos o los ponen a disposición de otros procesos. Integration Services Los componentes de flujo de datos son los orígenes, transformaciones y destinos que se utilizan diferentes tipos de elementos de flujo de datos: orígenes que extraen datos, transformaciones que Integration Services incluye. También puede incluir componentes personalizados en un flujo de datos.

Integración de Servicios

El Integration Services (Integración de Servicios) permite usar SQL Server Management Studio para supervisar paquetes Integration Services en ejecución y para administrar el almacenamiento de los paquetes.

Asistente para importación y exportación de SQL Server

El Asistente para importación y exportación de SQL Server puede copiar datos entre orígenes de datos para los que esté disponible un proveedor de datos de .NET Framework administrado o un proveedor OLE DB nativo. El Asistente ofrece también el método más simple para crear un paquete Integration Services que copia datos de un origen en un destino.

Analysis Services combina los mejores aspectos del análisis tradicional basado en OLAP y la elaboración de informes basada en relaciones al permitir a los programadores definir un único modelo de datos, denominado Unified Dimensional Model (UDM), a partir de uno o más orígenes de datos físicos. Todas las consultas de usuario final desde aplicaciones OLAP, de elaboración de informes y de BI personalizadas obtienen acceso a los orígenes de datos subyacentes a través del modelo UDM, que proporciona una única vista empresarial de estos datos relacionales.

Analysis Services proporciona un amplio conjunto de algoritmos de minería de datos para permitir a los usuarios empresariales recopilar los datos mediante la búsqueda de patrones y tendencias específicos. Estos algoritmos de minería de datos se pueden utilizar para analizar los datos a través de un modelo UDM o directamente a partir de un almacén de datos físico.

2.4.2.1. Arquitectura de Analysis Services

Microsoft SQL Server Analysis Services (SSAS) utiliza componentes de servidor y de cliente para proporcionar la funcionalidad de procesamiento analítico en línea (OLAP) y de minería de datos para aplicaciones de Business Intelligence:

- El componente de servidor de Analysis Services se implementa como servicio de Microsoft Windows. SQL Server Analysis Services admite varias instancias en el mismo equipo, con cada instancia de Analysis Services implementada como instancia independiente del servicio de Windows.
- Pueden emitirse comandos de consulta mediante los siguientes lenguajes: SQL; MDX (Expresiones multidimensionales), un lenguaje de consulta estándar para el análisis; o Extensiones de minería de datos (DMX), un lenguaje de consulta estándar orientado a la minería de datos. También se puede utilizar el lenguaje ASSL (Analysis Services Scripting Language) para administrar objetos de base de datos de Analysis Services.
- Analysis Services también admite un motor de cubo local que permite a las aplicaciones en clientes desconectados examinar datos multidimensionales almacenados localmente.

Arquitectura de servidor (Analysis Services)

El componente de servidor de Microsoft SQL Server 2005 Analysis Services (SSAS). Esta aplicación está formada por componentes de seguridad, un componente de escucha XML for Analysis (XMLA), un componente de

procesador de consultas y otros componentes internos que realizan las siguientes funciones:

- Analizar instrucciones recibidas de clientes.
- Administrar metadatos.
- Controlar transacciones.
- Procesar cálculos.
- Almacenar datos de celdas y dimensiones.
- Crear agregaciones.
- Programar consultas.
- Almacenar objetos en la caché.
- Administrar recursos del servidor.

Arquitectura de cliente (Analysis Services)

Microsoft SQL Server Analysis Services (SSAS) admite una arquitectura de cliente ligero. El motor de cálculo de Analysis Services depende totalmente del servidor, por lo que todas las consultas se resolverán en él. En consecuencia, para cada consulta sólo se necesita realizar un viaje de ida y vuelta entre el cliente y el servidor, lo que produce un rendimiento escalable a medida que las consultas aumenten en complejidad.

2.4.2.2. Minería de datos en Analysis Services



Luego de haber creado el proyecto de Analysis Services, puede agregar estructuras de minería de datos y uno o más modelos de minería de datos basados en cada estructura. Una estructura de minería de datos, incluyendo las tablas y columnas, se deriva de una vista de origen de datos o un cubo OLAP existentes en el proyecto. Al agregar una estructura de minería de datos, que se puede utilizar para definir la estructura y para especificar un algoritmo y los datos de entrenamiento que se utilizarán en la creación de un modelo inicial basado en esa estructura.

Un modelo de minería de datos forma parte de un proceso mayor que incluye desde la definición del problema básico que resolverá el modelo hasta la implementación del modelo en un entorno de trabajo. El proceso que utiliza Analysis Services se puede definir mediante los seis pasos básicos siguientes:

1. Definir el problema.
2. Preparar los datos.
3. Explorar los datos.
4. Generar modelos.
5. Explorar y validar los modelos.
6. Implementar y actualizar los modelos.

El siguiente es un diagrama que describe las relaciones entre cada paso del proceso y las tecnologías de Microsoft SQL Server 2005 que se pueden utilizar para completar cada paso.

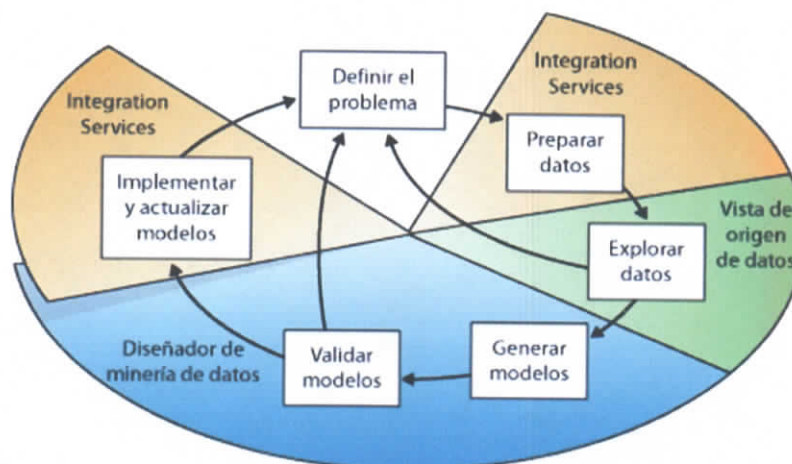


Gráfico 2.17. Proceso Data Mining.

Fuente: Documentación de Microsoft Visual Studio.

La creación de un modelo de minería de datos es un proceso dinámico e iterativo. Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de minería de datos adecuados y que, por tanto, debe buscar más datos. Puede generar varios modelos y descubrir que no responden al problema planteado cuando lo definió y que, por tanto, debe volver a definir el problema. Es posible que deba actualizar los modelos una vez implementados debido a que haya más datos disponibles. Por esto, es importante comprender que la creación de un modelo de minería de datos es un proceso, y que cada paso del proceso

puede repetirse tantas veces como sea necesario para crear un modelo válido.

Algoritmos de Minería de Datos (Analysis Services)

El algoritmo de minería de datos es el mecanismo que crea un modelo de minería de datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos y luego busca patrones y tendencias específicos. El algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de minería de datos. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas.

El modelo de minería de datos que crea un algoritmo puede tomar diversas formas, incluyendo:

- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción.
- Un árbol de decisión que predice si un cliente determinado comprará un producto.
- Un modelo matemático que predice las ventas.
- Un conjunto de clústeres que describe cómo se relacionan los casos de un conjunto de datos.
- Cada uno produce un resultado diferente.
- Algunos pueden producir más de un resultado.

- Utilizar diferentes algoritmos para diferentes tareas.
 - No es necesario usar los algoritmos independientemente. Se pueden usar múltiples algoritmos para solucionar un problema particular.
 - Hay algoritmos para explorar datos.
 - Hay algoritmos para predecir una salida.
-
- Árboles de Regresión para proporcionar un forecasting financiero.
 - Algoritmos Basados en Reglas (CART) para market basket analysis.
 - Árboles de Decisión (algoritmos de clasificación) tanto para predicción como para reducir el número de columnas en un data set que no afectan al modelo final.
 - Algoritmos Clustering para profiling.

Microsoft SQL Server Analysis Services (SSAS) incluye los siguientes algoritmos:

- Algoritmo de árboles de decisión de Microsoft.
- Algoritmo de clústeres de Microsoft.
- Algoritmo Bayes naive de Microsoft.
- Algoritmo de asociación de Microsoft.
- Algoritmo de clústeres de secuencia de Microsoft.
- Algoritmo de serie temporal de Microsoft.
- Algoritmo de red neuronal de Microsoft (SSAS).

- Algoritmo de regresión logística de Microsoft.
- Algoritmo de regresión lineal de Microsoft

Algoritmo de árboles de decisión de Microsoft

El algoritmo de árboles de decisión 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.

Por ejemplo: en un escenario para predecir qué clientes van a adquirir probablemente una bicicleta, si nueve de diez clientes jóvenes compran una bicicleta, pero sólo lo hacen dos de diez clientes de edad mayor, el algoritmo infiere que la edad es un buen elemento de predicción en la compra de bicicletas. El árbol de decisión realiza predicciones basándose en la tendencia hacia un resultado concreto.

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

El algoritmo de árboles de decisión de Microsoft 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.

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. La selección de características es importante para evitar que los atributos irrelevantes utilicen tiempo de procesador. Si utiliza demasiados atributos de predicción o de entrada al diseñar un modelo de minería de datos, el modelo puede tardar mucho tiempo en procesarse o incluso quedarse sin memoria.

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

- **Una única 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.

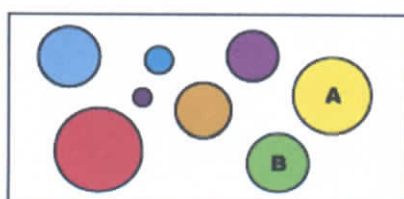
Algoritmo de clústeres de Microsoft

El algoritmo de clústeres de Microsoft es un algoritmo de segmentación. El algoritmo utiliza técnicas iterativas para agrupar los escenarios de un

conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

Los modelos de clústeres identifican las relaciones en un conjunto de datos que no se podrían derivar lógicamente a través de la observación casual. Por ejemplo, puede discernir lógicamente que las personas que se desplazan a sus trabajos en bicicleta no viven, por lo general, a gran distancia de sus centros de trabajo. Sin embargo, el algoritmo puede encontrar otras características que no son evidentes acerca de los trabajadores que se desplazan en bicicleta.

En el siguiente diagrama, el clúster A representa los datos sobre las personas que suelen conducir hasta el trabajo, en tanto que el clúster B representa los datos sobre las personas que van hasta allí en bicicleta.



A = Trabajadores que conducen para ir al trabajo
B = Trabajadores que van en bicicleta al trabajo

Gráfico 2.18. Algoritmo de clústeres.

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

Algoritmo Bayes naive de Microsoft

El algoritmo Bayes naive de Microsoft es un algoritmo de clasificación para el modelado de predicción. Este algoritmo calcula la probabilidad condicional entre columnas de entrada y de predicción y supone que las columnas son independientes. Esta suposición de independencia implica, de manera un tanto ingenua en ocasiones, que este algoritmo no tiene en cuenta las dependencias que puedan existir.

Desde el punto de vista computacional, el algoritmo es menos complejo que otros algoritmos de Microsoft y, por tanto, resulta útil para generar rápidamente modelos de minería de datos para descubrir relaciones entre columnas de entrada y columnas de predicción. Puede utilizar este algoritmo para realizar exploraciones iniciales de datos y, más adelante, aplicar los

Regla
Road Bottle Cage = Existing, Cycling Cap = Existing -> Water Bottle = Existing
Mountain-200 = Existing, Mountain Tire Tube = Existing -> HL Mountain Tire = Existing
Mountain-200 = Existing, Water Bottle = Existing -> Mountain Bottle Cage = Existing
Touring-1000 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing
Road-750 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing
Touring Tire = Existing, Sport-100 = Existing -> Touring Tire Tube = Existing

Gráfico 2.19 Reglas de Asociación (Algoritmo de Asociación).

Fuente: Documentación Microsoft Technet.

Como se muestra en el diagrama, el algoritmo de asociación puede encontrar potencialmente muchas reglas dentro de un conjunto de datos. El algoritmo usa dos parámetros, compatibilidad y probabilidad, para describir los conjuntos de elementos y las reglas que genera. Por ejemplo, si X e Y representan dos elementos que pueden formar parte de una cesta de compra, el parámetro de compatibilidad es el número de casos del conjunto de datos que contienen la combinación de elementos, X e Y. Mediante el uso del parámetro de compatibilidad en combinación con los parámetros MINIMUM_SUPPORT y MAXIMUM_SUPPORT definidos por el usuario, el algoritmo controla el número de conjuntos de elementos que se generan. El parámetro de probabilidad, denominado también confianza, representa la fracción de casos del conjunto de datos que contiene X y que también contiene Y. Mediante el uso del parámetro de probabilidad en combinación con el parámetro MINIMUM_PROBABILITY, el algoritmo controla el número de reglas que se generan.

Algoritmo de clústeres de secuencia de Microsoft

El algoritmo de clústeres de secuencia es un algoritmo de análisis de secuencia. Puede utilizar este algoritmo para explorar los datos que contienen eventos que pueden vincularse mediante rutas o secuencias. El algoritmo encuentra las secuencias más comunes mediante la agrupación de las secuencias idénticas. Estas secuencias pueden tomar muchas formas, incluyendo:

- Los datos que describen las rutas de clics que los usuarios siguen a través de un sitio Web.
- Los datos que describen el orden en el que un cliente agrega elementos en una cesta de compra de un comerciante electrónico.

Este algoritmo es similar al Algoritmo de clústeres de Microsoft. Sin embargo, en lugar de encontrar clústeres de escenarios que contienen atributos similares, el algoritmo de clústeres de secuencia de Microsoft encuentra clústeres de escenarios que contienen rutas similares en una secuencia.

El modelo de minería de datos que crea este algoritmo contiene descripciones de las secuencias más comunes en los datos. Puede usar las descripciones para predecir el siguiente paso probable de una nueva secuencia. Cuando el algoritmo agrupa registros, también puede tener en cuenta las columnas de datos que no están relacionadas directamente con las secuencias. Debido a que el algoritmo incluye las columnas no relacionadas, puede usar el modelo resultante para identificar las relaciones entre los datos de las secuencias y los datos que no se producen en una secuencia.

Algoritmo de serie temporal de Microsoft

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

El siguiente diagrama muestra un modelo típico de previsión en el tiempo de las ventas de un producto en cuatro regiones de ventas diferentes. El modelo presentado en el diagrama de ventas muestra las ventas de cada región como líneas de color rojo, amarillo, púrpura y azul. La línea de cada región consta de dos partes:

- La información histórica aparece a la izquierda de la línea vertical y representa los datos que el algoritmo utiliza para crear el modelo.
- La información de la predicción aparece a la derecha de la línea vertical y representa la previsión realizada por el modelo.

A la combinación de los datos de origen y los datos de la predicción se le denomina serie.

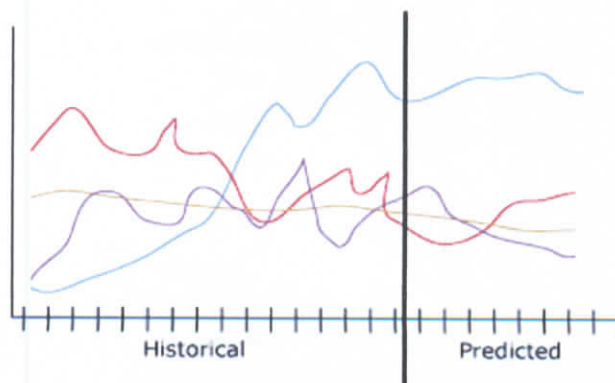


Gráfico 2.20. Algoritmo Serie Temporal.

Fuente: Documentación Microsoft Technet.

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

Algoritmo de red neuronal de Microsoft

El algoritmo de red neuronal Microsoft crea modelos de minería de datos de clasificación y regresión mediante la generación de una red de neuronas de tipo perceptrón multicapa. De forma similar al algoritmo de árboles de decisión de Microsoft, el algoritmo de red neuronal de Microsoft calcula las probabilidades para cada posible estado del atributo de entrada cuando se da cada estado del atributo de predicción. Posteriormente, puede utilizar

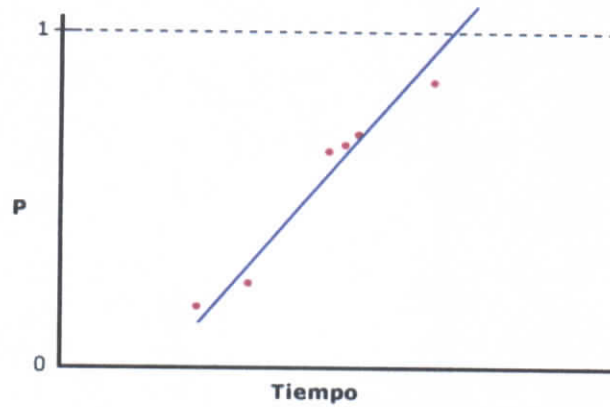


Gráfico 2.21. Algoritmo de regresión.

Fuente: Documentación Microsoft Technet.

El eje x contiene los valores de una columna de entrada. El eje y contiene las probabilidades de que la columna de predicción tenga un estado o el otro. El problema que puede surgir es que la regresión lineal no limite la columna a los valores 0 y 1, a pesar de que son los valores máximo y mínimo de la columna. Una forma de resolver el problema es llevar a cabo una regresión logística. En vez de crear una línea recta, el análisis de regresión logística crea una curva con forma de "S" que contiene las restricciones máxima y mínima.

Por ejemplo, el siguiente diagrama muestra los resultados que se obtienen si lleva a cabo una regresión logística con los mismos datos utilizados en el ejemplo anterior.

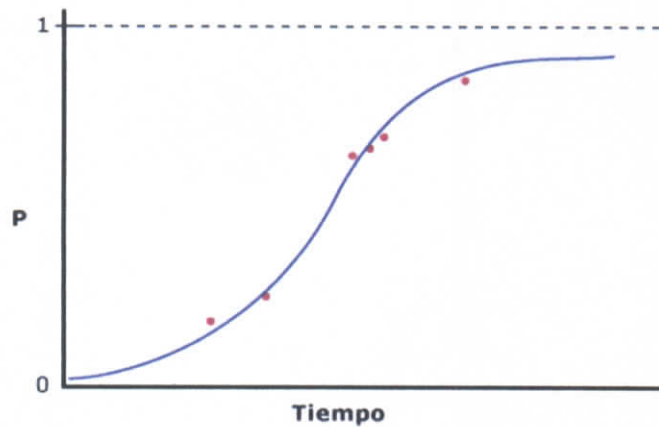


Gráfico 2.22. Algoritmo de regresión.

Fuente: Documentación Microsoft Technet.

Observe cómo la curva nunca va por encima del 1 ni por debajo del 0. Puede utilizar la regresión logística para describir qué columnas de entrada son importantes a la hora de determinar el estado de la columna de predicción.

Algoritmo de regresión lineal de Microsoft

El algoritmo Regresión lineal de Microsoft es una variación del algoritmo de árboles de decisión de Microsoft, el algoritmo no crea nunca una división y, por tanto, lleva a cabo una regresión lineal.

Puede utilizar la regresión lineal para determinar una relación entre dos columnas continuas. La relación toma la forma de una ecuación para la línea

que mejor represente una serie de datos. Por ejemplo, la línea del siguiente diagrama muestra la mejor representación lineal de los datos.

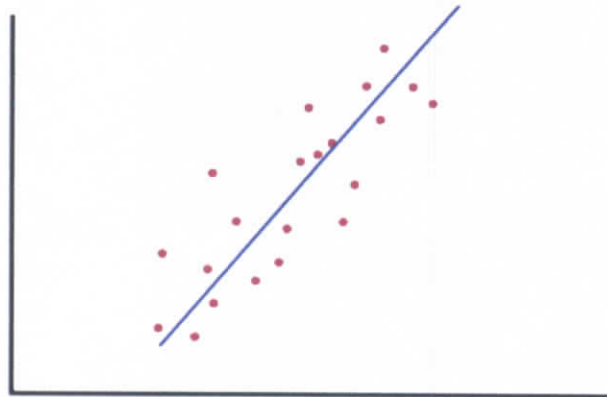


Gráfico 2.23. Algoritmo de regresión Lineal.

Fuente: Documentación Microsoft Technet.

La ecuación que representa la línea en el diagrama toma la forma general de $y = ax + b$ y es conocida como la ecuación de regresión. La variable Y representa la variable de salida, X representa la variable de entrada y a y b son coeficientes ajustables. Cada punto de datos del diagrama tiene un error asociado con su distancia con respecto a la línea de regresión. Los coeficientes a y b de la ecuación de regresión ajustan el ángulo y la ubicación de la línea de regresión. Puede obtener la ecuación de regresión ajustando a y b hasta que la suma de los errores asociados con los puntos alcance su cifra más baja.

Tipos de algoritmos de minería de datos

Analysis Services incluye los siguientes tipos de algoritmos:

- **Algoritmos de clasificación.-** que predicen una o más variables discretas, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de clasificación es el Algoritmo de árboles de decisión de Microsoft.
- **Algoritmos de regresión.-** que predicen una o más variables continuas, como las pérdidas o los beneficios, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de regresión es el Algoritmo de serie temporal de Microsoft (Analysis Services - Minería de datos).
- **Algoritmos de segmentación.-** que dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares. Un ejemplo de algoritmo de segmentación es el Algoritmo de clústeres de Microsoft (Analysis Services - Minería de datos).

- **Algoritmos de asociación.-** que buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación, que pueden utilizarse en un análisis de la cesta de compra. Un ejemplo de algoritmo de asociación es el Algoritmo de asociación de Microsoft.

- **Algoritmos de análisis de secuencias.-** que resumen secuencias o episodios frecuentes en los datos, como un flujo de rutas Web. Un ejemplo de algoritmo de análisis de secuencias es el Algoritmo de agrupación en clústeres de secuencia de Microsoft.

Aplicar los algoritmos

La elección del mejor algoritmo para una tarea empresarial específica puede ser un desafío. Aunque puede utilizar diferentes algoritmos para realizar la misma tarea, cada uno de ellos genera un resultado diferente, y algunos pueden generar más de un tipo de resultado. Por ejemplo, puede usar el algoritmo Árboles de decisión de Microsoft no sólo para la predicción, sino también como una forma de reducir el número de columnas de un conjunto de datos, ya que el árbol de decisión puede identificar las columnas que no afectan al modelo de minería de datos final.

Tampoco es necesario usar los algoritmos de modo independiente. En una única solución de minería de datos se pueden usar algunos algoritmos para explorar datos y, posteriormente, usar otros algoritmos para predecir un resultado específico a partir de esos datos. Por ejemplo, puede utilizar un algoritmo de agrupación en clústeres, que reconoce patrones, para dividir los datos en grupos que sean más o menos homogéneos, y luego usar los resultados para crear un mejor modelo de árbol de decisión. Puede utilizar varios algoritmos dentro de una solución para realizar tareas independientes, por ejemplo, usar un algoritmo de árbol de regresión para obtener información de previsiones financieras y un algoritmo basado en reglas para llevar a cabo un análisis de la cesta de compra.

Los modelos de minería de datos pueden predecir valores, generar resúmenes de datos y buscar correlaciones ocultas. Para ayudarle a seleccionar los algoritmos para su solución de minería de datos, la siguiente tabla proporciona sugerencias sobre qué algoritmos usar en tareas específicas.

Tarea	Algoritmos de Microsoft que se pueden usar
<p>Predecir un atributo discreto.</p> <p>Por ejemplo, predecir si el destinatario de una campaña de envío de correo directo adquirirá un producto.</p>	<p>Algoritmo de árboles de decisión de Microsoft</p> <p>Algoritmo Bayes naive de Microsoft</p> <p>Algoritmo de clústeres de Microsoft (Analysis Services - Minería de datos)</p> <p>Algoritmo de red neuronal de Microsoft (Analysis Services - Minería de datos)</p>

<p>Predecir un atributo continuo.</p> <p>Por ejemplo, preveer las ventas del año próximo.</p>	<p>Algoritmo de árboles de decisión de Microsoft</p> <p>Algoritmo de serie temporal de Microsoft (Analysis Services - Minería de datos)</p>
<p>Predecir una secuencia.</p> <p>Por ejemplo, realizar un análisis clickstream del sitio web de una empresa.</p>	<p>Algoritmo de agrupación en clústeres de secuencia de Microsoft</p>
<p>Buscar grupos de elementos comunes en las transacciones.</p> <p>Por ejemplo, utilizar el análisis de la cesta de la compra para sugerir a un cliente la compra de productos adicionales.</p>	<p>Algoritmo de asociación de Microsoft</p> <p>Algoritmo de árboles de decisión de Microsoft</p>
<p>Buscar grupos de elementos similares.</p> <p>Por ejemplo, segmentar datos demográficos en grupos para comprender mejor las relaciones entre atributos.</p>	<p>Algoritmo de clústeres de Microsoft (Analysis Services - Minería de datos)</p> <p>Algoritmo de agrupación en clústeres de secuencia de Microsoft</p>

Tabla 2.1. Aplicación de Algoritmos de Minería de datos Microsoft.

Fuente: Documentación Microsoft Technet.

En el gráfico siguiente se resumen los algoritmos de minería y su aplicación

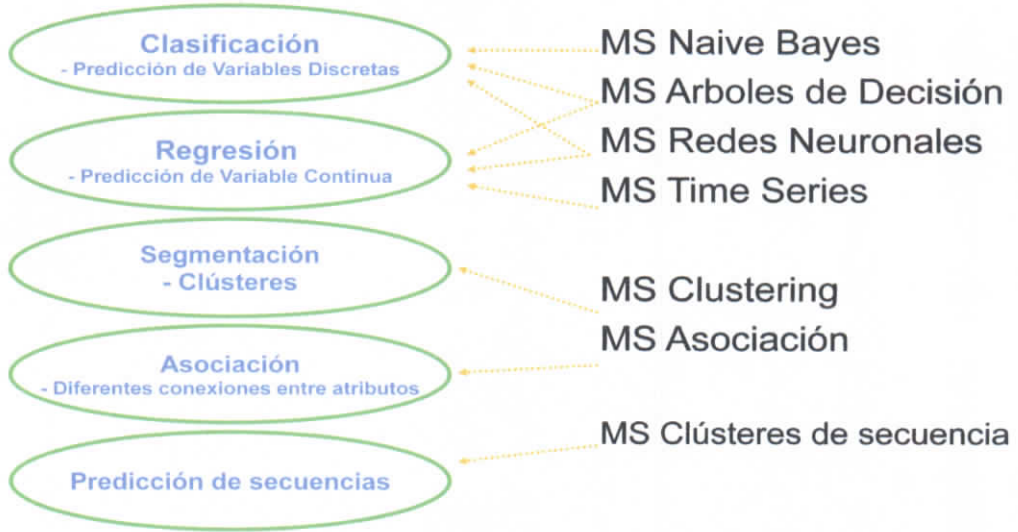


Gráfico 2.24. Algoritmos de minería y su aplicación.

Fuente: Visión General de Data Mining, Edgar Iqira. Microsoft Services.

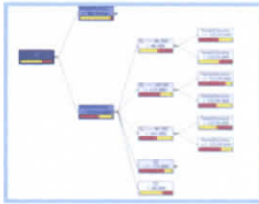
	Arboles decisión	Bayes Naïve	Clustering	Clústeres de secuencia	Serie Temporal	Asociación	Redes Neuronales	
Clasificación	✓	✓	✓	✓		✓	✓	
Regresión	✓	✓	✓	✓			✓	
Segmentación			✓	✓			✓	
Asociación	✓	✓	✓	✓		✓	✓	
Detección Anomalía			✓	✓			✓	
Análisis de Secuencias				✓				
Serie temporal					✓			

✓ - Primera Opción
 ✓ - Segunda Opción

Gráfico 2.25. Algoritmos de minería y su aplicación.

Fuente: Visión General de Data Mining, Edgar Iqira. Microsoft Services.

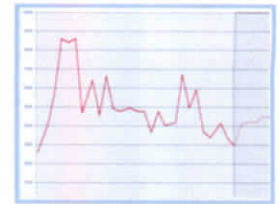
Gráficos de los algoritmos



Arboles de decisión



Clústeres



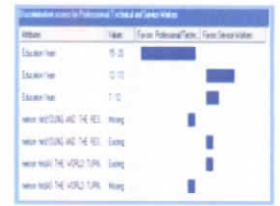
Series Temporales



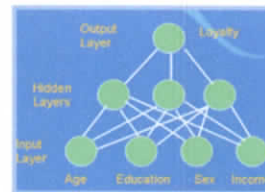
Clústeres de secuencia



Asociación



Bayes Naïve



Redes Neuronales

Gráfico 2.26. Representación Gráfica de los Algoritmo de Data Mining.

Fuente: Visión General de Data Mining, Edgar Iquira. Microsoft Services.

2.4.3. SQL Server Reporting Services. (SSRS)



SQL Server Reporting Services incluye el Diseñador de informes, una completa herramienta para la creación de informes que se aloja en el entorno de Microsoft Visual Studio. Para generar un informe en el Diseñador de informes, cree el informe, agregue los datos y organice la distribución de los datos y los elementos gráficos. También puede agregar características interactivas al informe y manipular el resultado utilizando expresiones. Una

vez terminado el informe, utilice el Diseñador de informes para mostrar una vista previa del mismo y publicarlo directamente en el servidor de informes.

Cuando se crea un informe con el Diseñador de informes u otra herramienta, en realidad se está creando una definición de informe. Una definición de informe contiene información acerca del origen, la estructura y la distribución de los datos y objetos del informe. La definición de informe se guarda como archivo RDL (Report Definition Language) en un proyecto de servidor de informes, que se incluye en una solución de Visual Studio 2005.

CAPITULO III

3. POR QUÉ IMPLEMENTAR BUSINESS INTELLIGENCE EN MADERVAS?

Para Madervas conocer y predecir el la ventas, analizar la rotación de inventarios, mejorar las negociaciones con los proveedores, se convierte en una ventaja competitiva para lo cual implementando una solución Business Intelligence ayuda y soporta a la toma de decisiones eficaces.

3.1. Encuesta.

3.1.1. Objetivo de la Investigación.

Se establece como objetivo determinar si la empresa Madervas requiere de una solución BI para mejorar la gestión de gerencia de ventas.

3.1.2. Objetivos Específicos.

Objetivo	Hipótesis
1. Conocer si se ha establecido cuál es el producto estrella de la empresa.	Si tienen establecido.
2. Conocer si se ha establecido	Si tienen establecido que productos

que productos son importantes para la empresa.	son los más importantes.
3. Conocer si se ha establecido que clientes son los más importantes.	Si se tiene establecido que clientes son los más importantes para la empresa.
4. Determinar si se ha perdido cuota en el mercado con respecto a la competencia	No se ha establecido, no existe la información suficiente
5. Conocer si se recibe a tiempo la información.	No se recibe oportunamente la información.
6. Establecer si tienen el tiempo suficiente para analizar la información.	No tienen el suficiente tiempo, para el análisis.
7. Saber si tienen alguna ventaja competitiva con respecto a las empresas del sector.	Si tienen establecido las ventajas.
8. Conocer si el personal de ventas cumple con los objetivos.	No se conoce, si el personal de ventas cumple los objetivos planificados.
9. Determinar si se conoce el comportamiento de compras de los clientes.	No se conoce que comportamiento de compras tienen los clientes.
10. Conocer si se puede procesar	Si se conoce cuales son los más

<p>rápidamente la información para saber que producto fueron los más rentables dentro de un periodo.</p>	<p>rentables, pero no con la oportunidad necesaria.</p>
--	---

Tabla 3.1. Tabla de objetivos e Hipótesis de la encuesta.

3.1.3. Método de Investigación

Se utiliza el método de investigación descriptiva – cuantitativa; para determinar los requerimientos de información de la empresa y desarrollar la solución BI para mejorar la gestión y cumplir los objetivos propuestos en este estudio.

3.1.4. Muestra de la investigación

Para establecer la necesidad e importancia de implementar business Intelligence se realiza y ejecuta la encuesta al gerente de ventas de la empresa.

3.1.5. Instrumento de recolección

El instrumento de recolección de la información es mediante una encuesta.

1. Sabe Ud.Cuál es su producto estrella o que le da la mayor rentabilidad en la empresa?

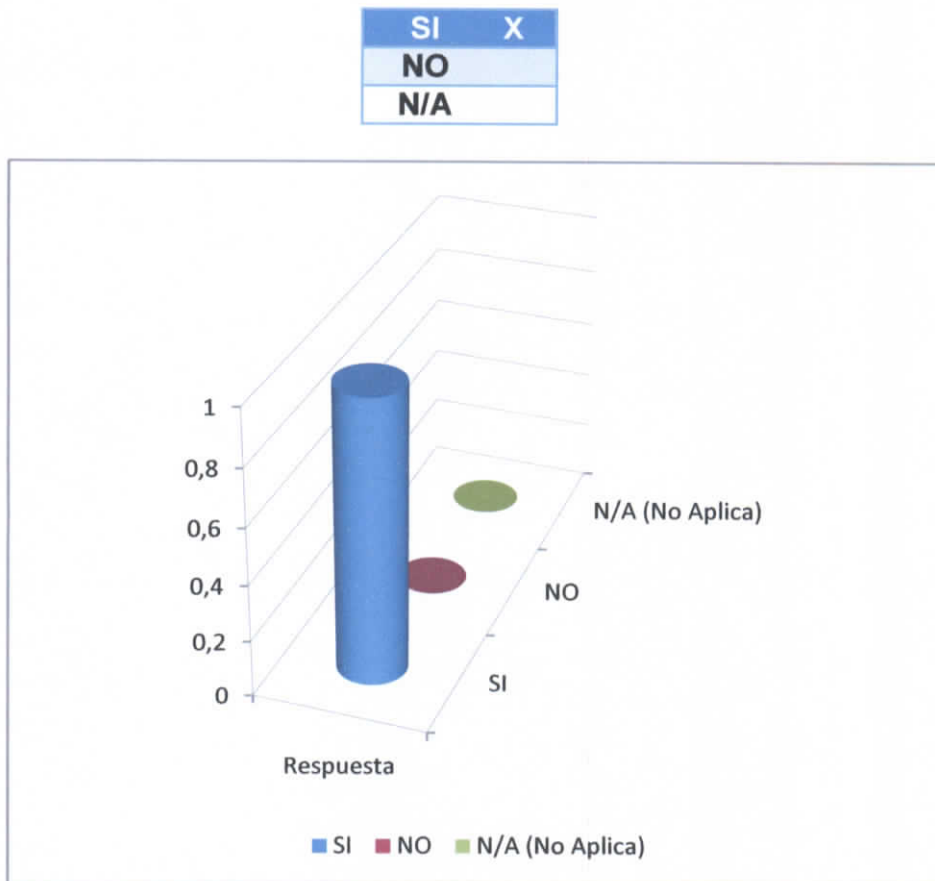


Gráfico 3.1. Tabulación Encuesta Pregunta N° 1.

Fuente: El investigador.

Interpretación: SI se conoce y se ha definido cuál es el producto estrella de la empresa, es decir el que más se vende y es rentable.

Observación.- Se procesa los datos manualmente para obtener los informes.

2. ¿Está seguro de qué productos son los más importantes para su empresa?

SI	X
NO	
N/A	

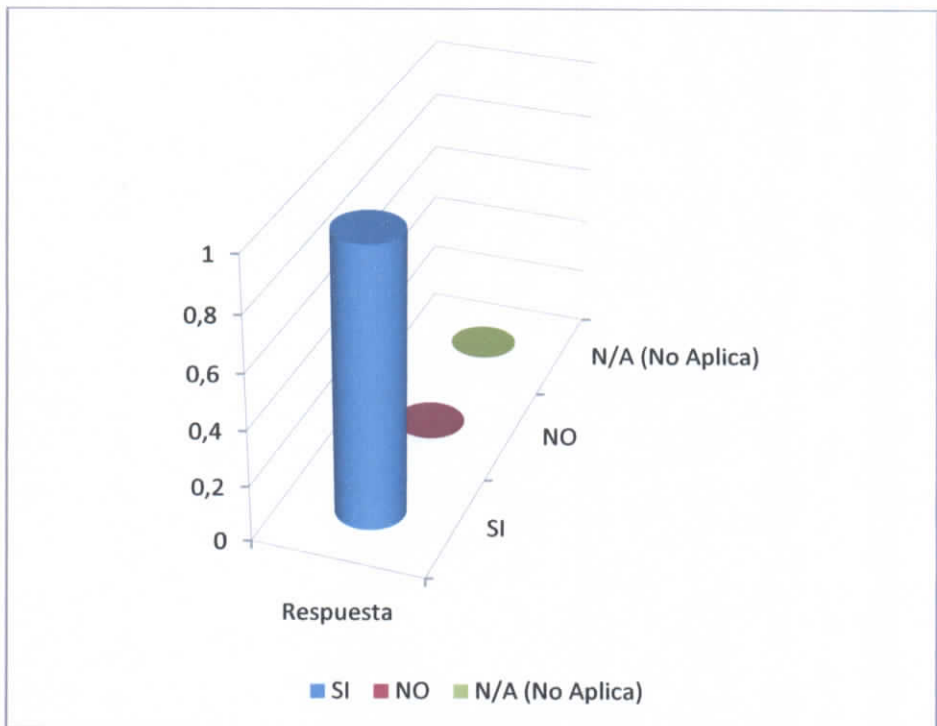


Gráfico 3.2. Tabulación Encuesta Pregunta N° 2.

Fuente: El investigador.

Interpretación: Si se conoce que productos son los más importantes de la empresa, se tiene definido que son los productos otorgados por la franquicia como son MDF, Melaminas, OSB, etc.

Observación.- Se procesa los datos manualmente para realizar los informes.

3. ¿Está seguro de qué clientes son los más importantes para su empresa?

SI	X
NO	
N/A	

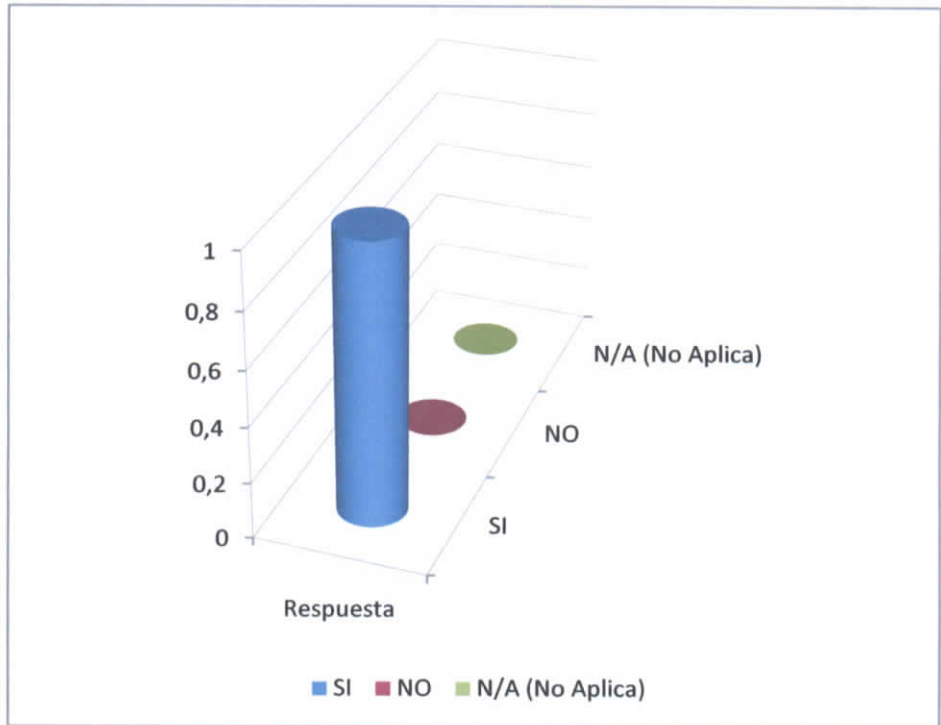


Gráfico 3.3. Tabulación Encuesta Pregunta N° 3.

Fuente: El investigador.

Interpretación: SI se ha determinado que clientes son los más importantes para la empresa.

Observación.- Se procesa los datos manualmente para realizar los informes.

4. ¿Sabe si está perdiendo cuota de mercado con respecto a su competencia?

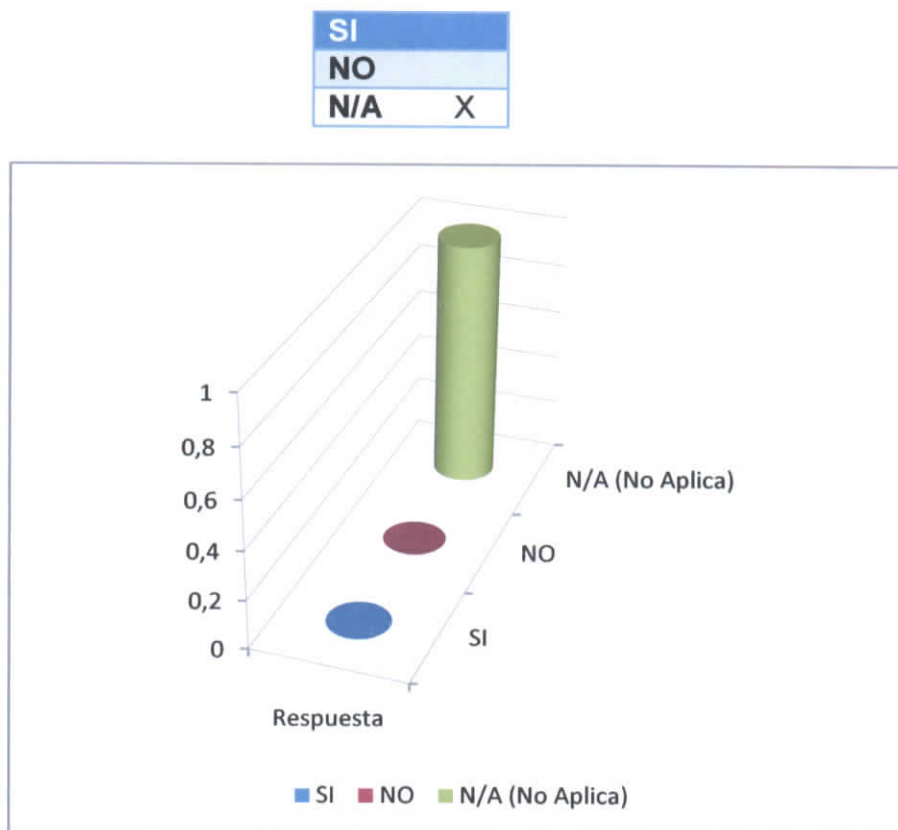


Gráfico 3.4. Tabulación Encuesta Pregunta N° 4.

Fuente: El investigador.

Interpretación.- No Aplica, se requiere de mayor información para poder establecer si se está perdiendo en el mercado, además se necesita información de cantidades que comercializa la competencia en este caso Edimca, lo cual es muy difícil obtener.

Observación.- Se conoce informalmente que existen varias formas de comercialización por parte de la competencia.

5. ¿La información requerida la recibe oportunamente para realizar los negocios y/o tomar decisiones?

SI	
NO	X
N/A	

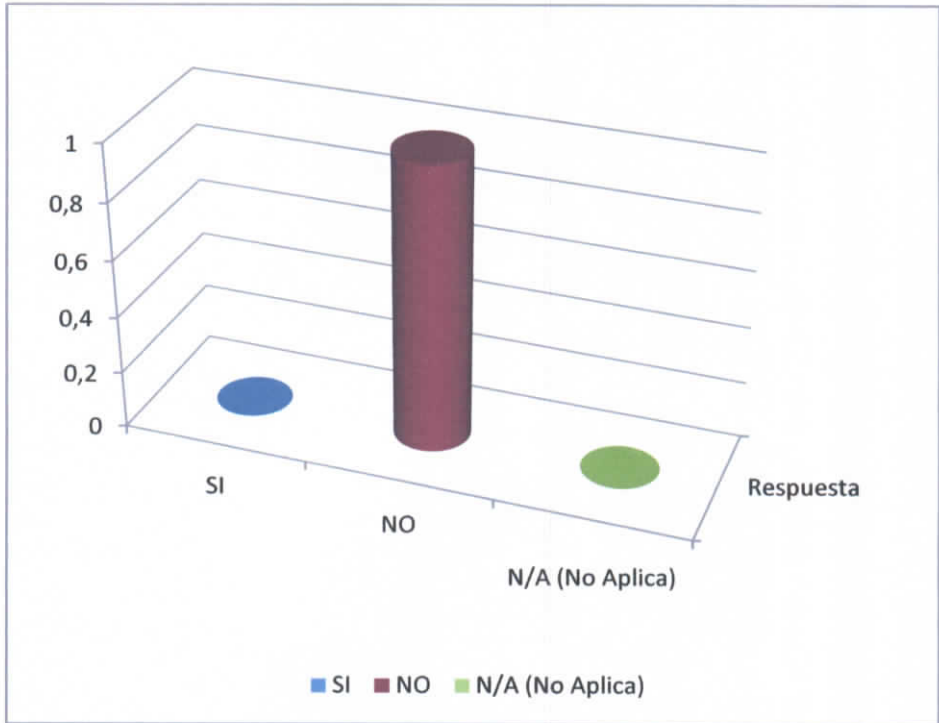


Gráfico 3.5. Tabulación Encuesta Pregunta N° 5.

Fuente: El investigador.

Interpretación: No se recibe oportunamente, se requiere de tiempo para procesar la información.

Observación.- Se requiere de mayor tiempo para procesar los datos, y además digitar valores.

6. ¿Le queda el tiempo suficiente para analizar documentos e informes?

SI	
NO	X
N/A	

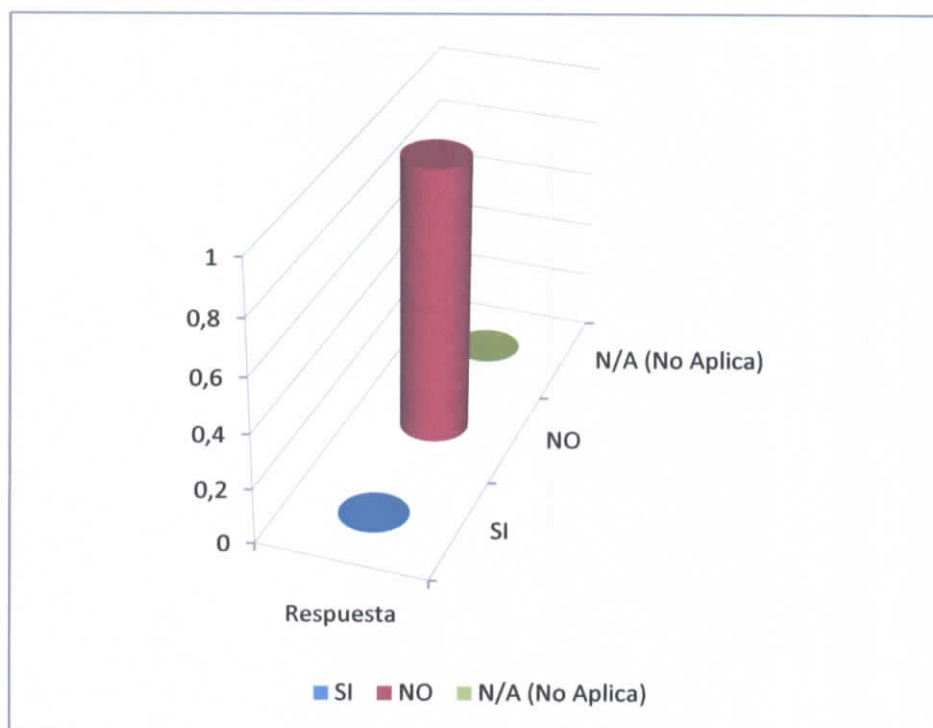


Gráfico 3.6. Tabulación Encuesta Pregunta N° 6.

Fuente: El investigador.

Interpretación.- No, reciben oportunamente la información ya que le dedican más tiempo a proceso manual para el informe, quedando poco tiempo para su análisis.

Observación.- Se procesa los datos manualmente para realizar los informes.

7. ¿Dispone de alguna ventaja competitiva clara con respecto a las demás empresas de su sector?

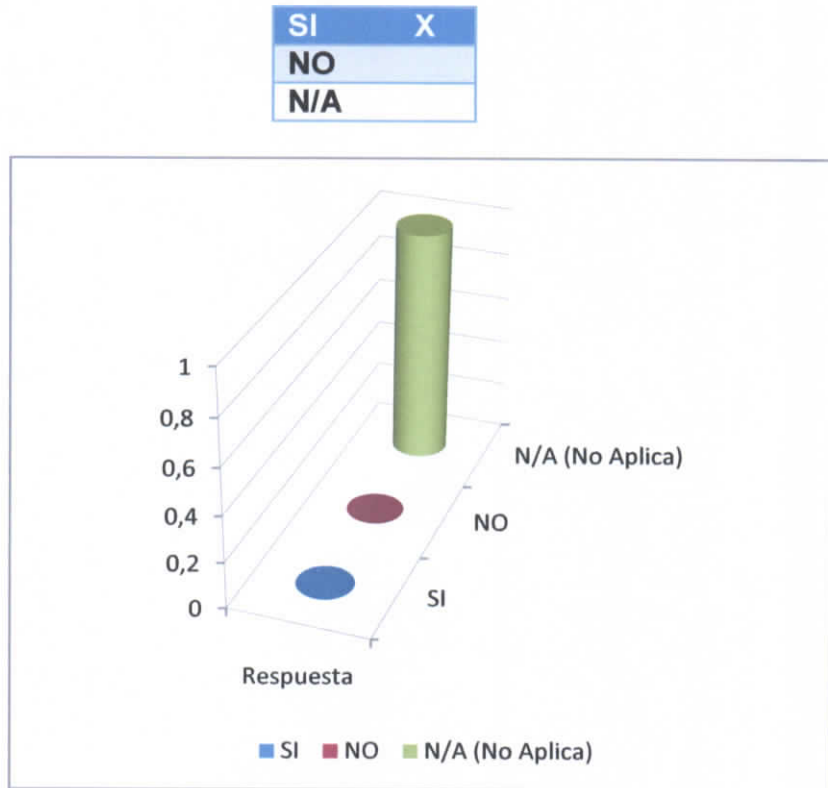


Gráfico 3.7. Tabulación Encuesta Pregunta N° 7.

Fuente: El investigador.

Interpretación.- Si, se tiene definida la ventaja competitiva por el tipo de producto que comercializan por medio de la franquicia de MASISA, y además por los productos adicionales que forman parte del servicio a los clientes.

Observación.- Se puede decir que la ventaja competitiva se puede definir también como saber las tendencias de compras de los clientes.

8. ¿Sabe con certeza si su personal de ventas cumple con los objetivos planificados?

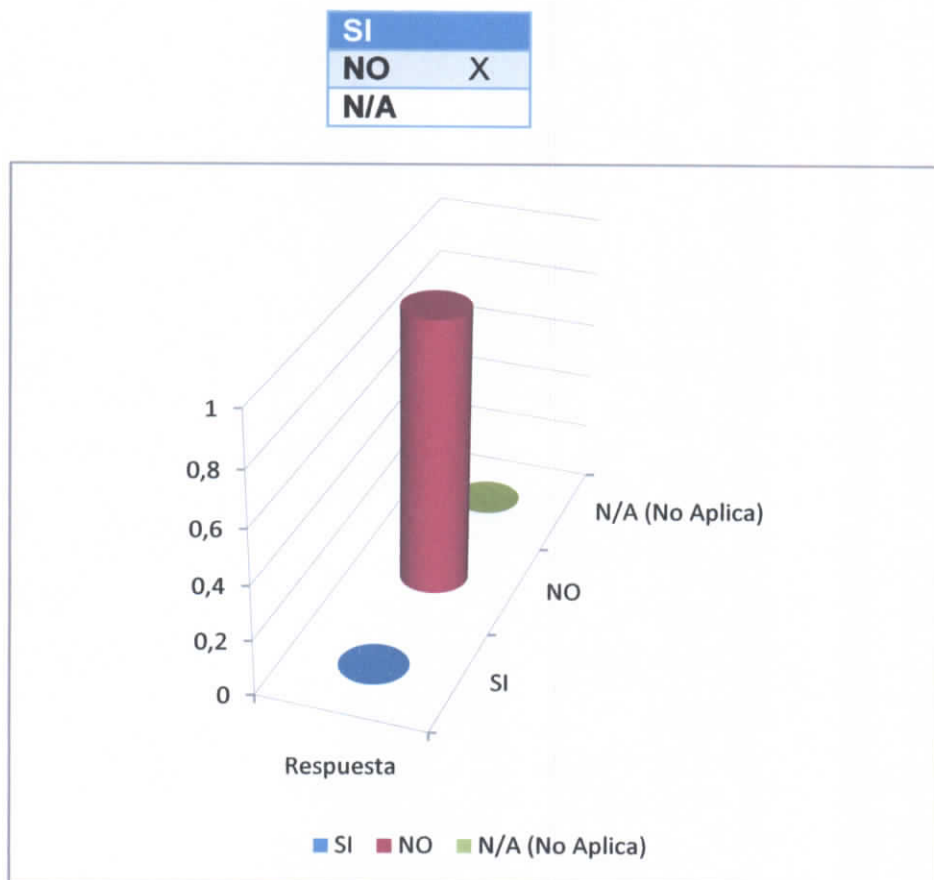


Gráfico 3.8. Tabulación Encuesta Pregunta N° 8.

Fuente: El investigador.

Interpretación.- No se conoce si el personal de ventas cumple con los objetivos planificados.

Observación.- Este tipo de información no se ha analizado, porque no se tiene el reporte por vendedor.

9. ¿Analizado cuáles son las tendencias de compra de sus clientes, por período, por producto, etc.?

SI	
NO	X
N/A	

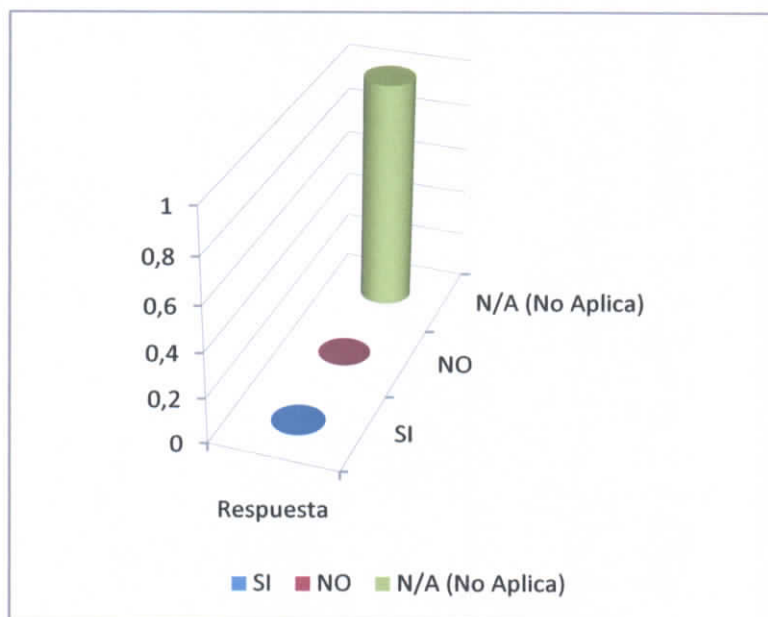


Gráfico 3.9. Tabulación Encuesta Pregunta N° 9.

Fuente: El investigador.

Interpretación.- No se analizado la tendencia de compras de los clientes.

Observación.- Procesar manualmente esta información es casi imposible, para esto se requiere implementar una solución de BI.

10. ¿Puede establecer de forma rápida que productos fueron los más rentables durante un periodo determinado?

SI	
NO	X
N/A	

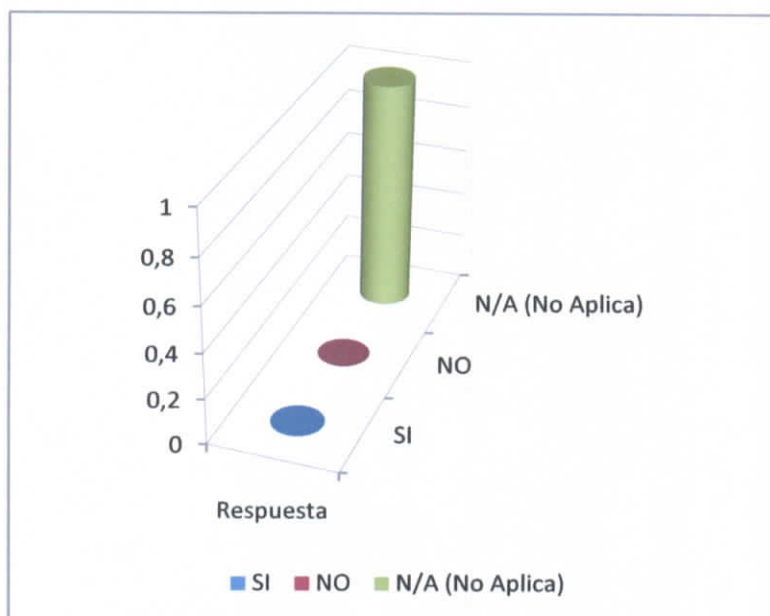


Gráfico 3.10. Tabulación Encuesta Pregunta N° 10.

Fuente: El investigador.

Interpretación.- Si se ha establecido cuales son los productos más rentables, pero no a la oportunidad necesaria.

Observación.- Para este tipo de análisis se requiere de información detallada y clasificada o agrupada por periodos de tiempo.

3.2. Tabulación de la Encuesta



Gráfico 3.11. Tabulación resultado encuesta.

Fuente: El investigador.

3.3. Análisis de la Encuesta.

Se puede concluir que con este diagnóstico ejecutado en la empresa, que:

El 40% de la información si te tiene o procesa, pero manualmente desde los reportes generados por el sistema administrativo financiero y desde tablas dinámicas diseñadas.

Con el 50% de las respuestas negativas se puede analizar que se requiere implementar una solución de BI para que se obtenga información importante,

CAPITULO IV

4. BUSINESS INTELLIGENCE EN MADERVAS

Business Intelligence se ha convertido en una tecnología "Imprescindible", impulsada por la importancia de obtener una ventaja competitiva mediante el desarrollo de un conocimiento profundo del cliente.

4.1. Solución propuesta

Se desarrolla un sistema de BI como soporte a la toma de decisiones (DSS), para lo cual debemos implementar y desarrollar un cubo de información de ventas y otro de inventarios.

Para la extracción del conocimiento KDD es decir la predicción de tendencias de compras de los clientes (Business Intelligence) utilizaremos el Algoritmo de árboles de decisión de Microsoft.

4.2. SQL Server Business Intelligence Development Studio

La solución Business Intelligence en la empresa Madervas, se desarrolla en SQL Server Business Intelligence Development Studio ya que este es un paquete que viene incluido en SQL Server 2005, del mismo que se tiene las

analice la información y por supuesto para tomar decisiones que ayuden a mejorar la gestión de ventas.

licencias cumpliendo las normales legales establecidas por la empresa Microsoft.

4.3. Origen de datos

La fuente principal de información son los datos históricos y actuales del sistema administrativo financiero denominado Ecuamat.

La base de datos está alojada en SQL SERVER 2005.

Las tablas que se utilizadas para el desarrollo de la solución BI son:

Inventarios – INV.

Facturación - FAC.

Cuentas por Cobrar – CXC.

Vistas, desarrolladas para las necesidades de la solución BI.

4.4. Servicio de Integración de Datos SSIS

Utilizando el SSIS, esta herramienta que nos permite realizar el proceso de ETL extracción, transformación y Carga de datos.

Este asistente nos permite cargar los datos desde la base de datos en producción en este caso llamada MaderVas, hasta la nueva base MadervasDW.

Para la solución BI, se utilizamos las tablas de:

Inventarios.- que se identifican en la base con las Iniciales INV.

Facturación.- que se identifican en la base con las Iniciales FAC.

Cuentas Por Cobrar.- que se identifican en la base con las Iniciales CXC.

Estas tablas deben ser seleccionadas y creando un proyecto de Integration services, para pasar los datos a la base MadervasDW.

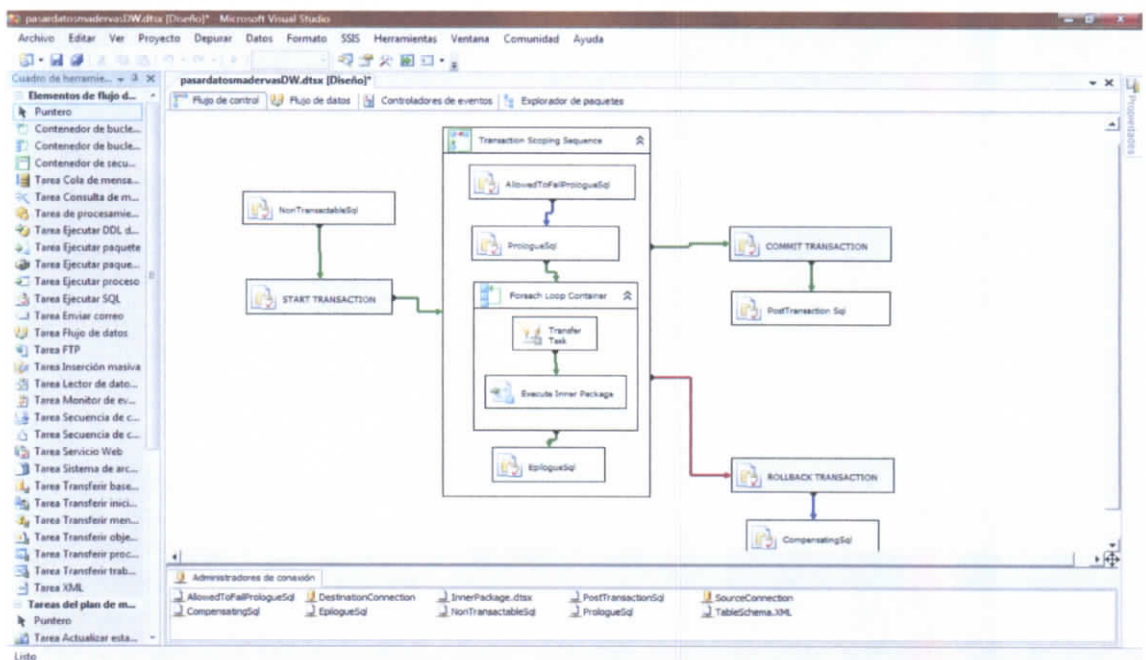


Gráfico 4.1. Paquete de Integration Services para Madervas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

También se puede realizar este proceso de forma más sencilla utilizando el asistente de importación/exportación de datos con el SQL Server Management Studio:

a.) Seleccionando la base origen.

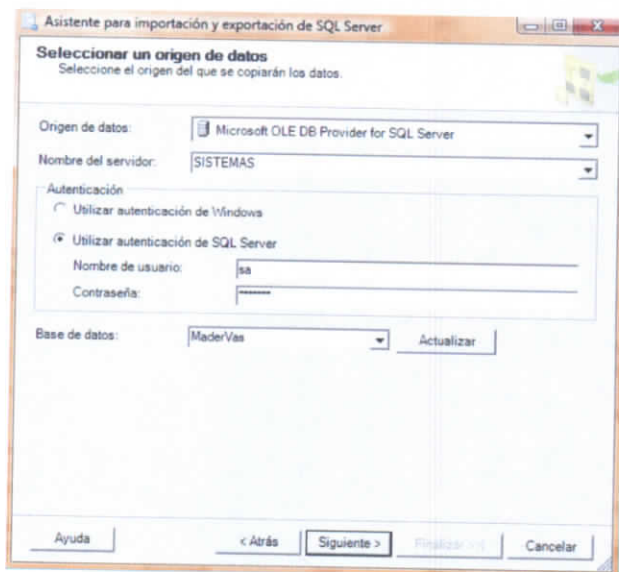


Gráfico 4.2. Conexión a Origen de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

b.) Seleccionar la base destino.

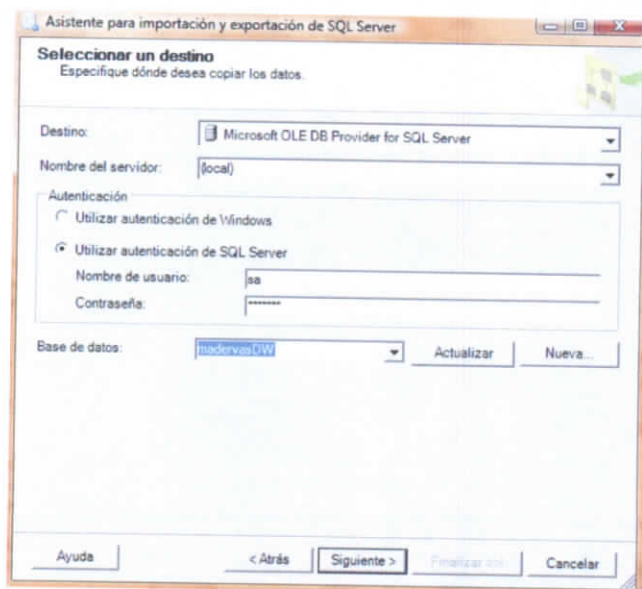


Gráfico 4.3. Conexión a Destino de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

c.) Posteriormente seleccionamos las tablas requeridas para la solución.

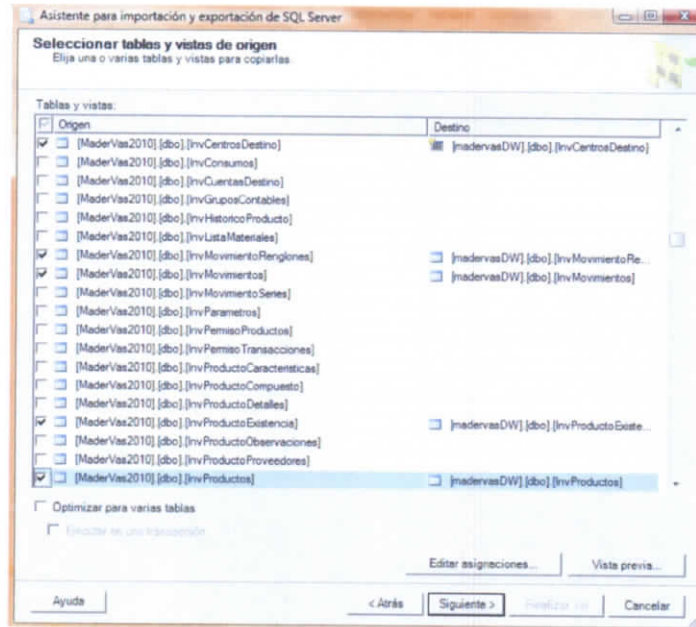


Gráfico 4.4. Selección de Tablas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

d.) Guardamos el paquete creado, ya que este le se realiza la programación como un trabajo programado en el SQL Server Management Studio.

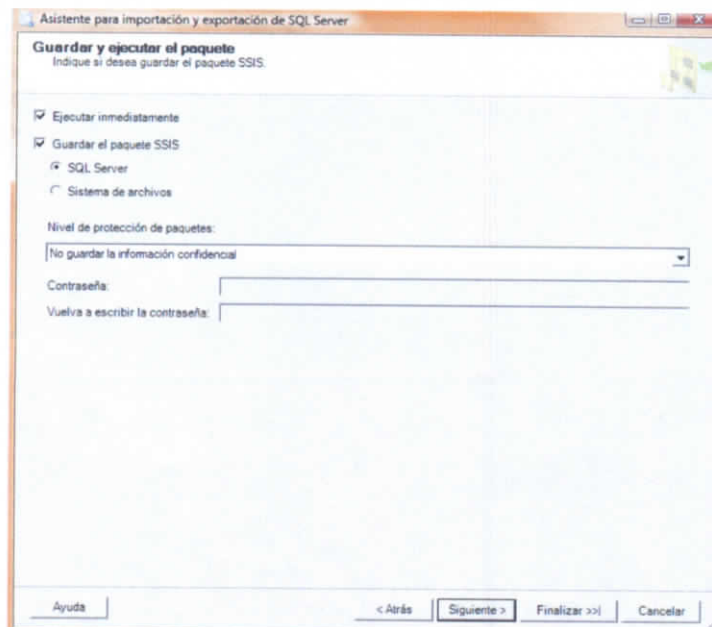


Gráfico 4.5. Guardar el paquete de exportación de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

4.5. Creación de la Solución BI

Con los datos cargados a la base MadervasDW, utilizando "SQL Server Business Intelligence Development Studio" seguimos los siguientes pasos:

1. Creamos el proyecto de Analysis Services.

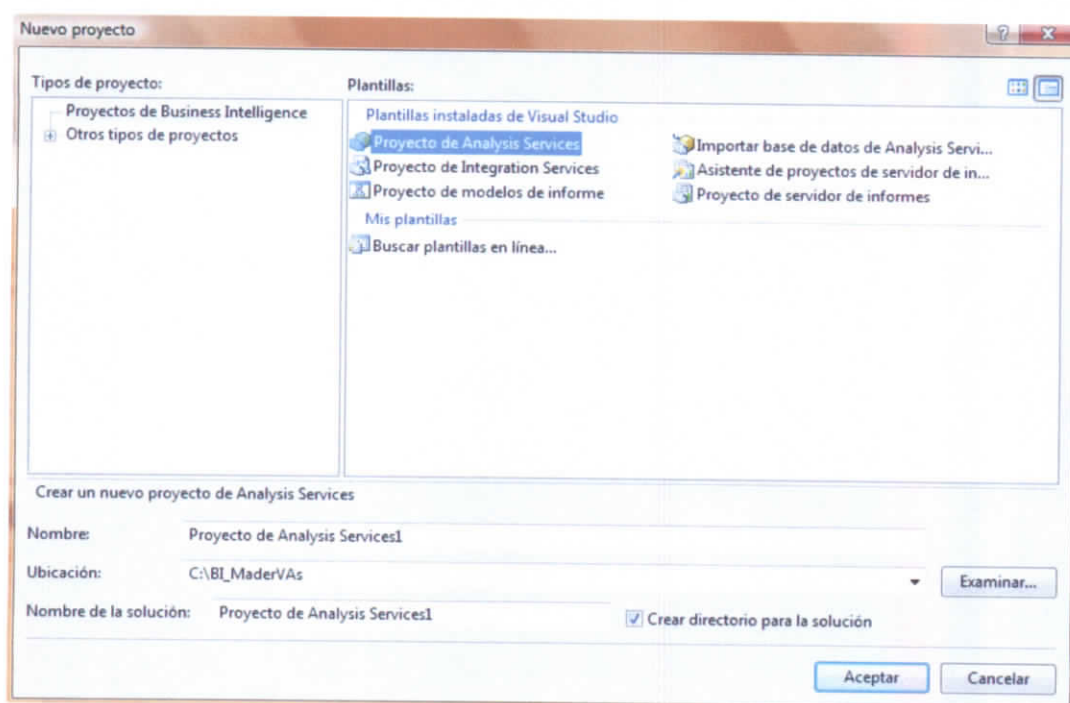


Gráfico 4.6. Creación Proyecto de Analysis Services.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

2. Creado el proyecto, se debe establecer la conexión a la base de datos utilizando el OLE DB Nativo\Microsoft OLE DB Provider for SQL Server.

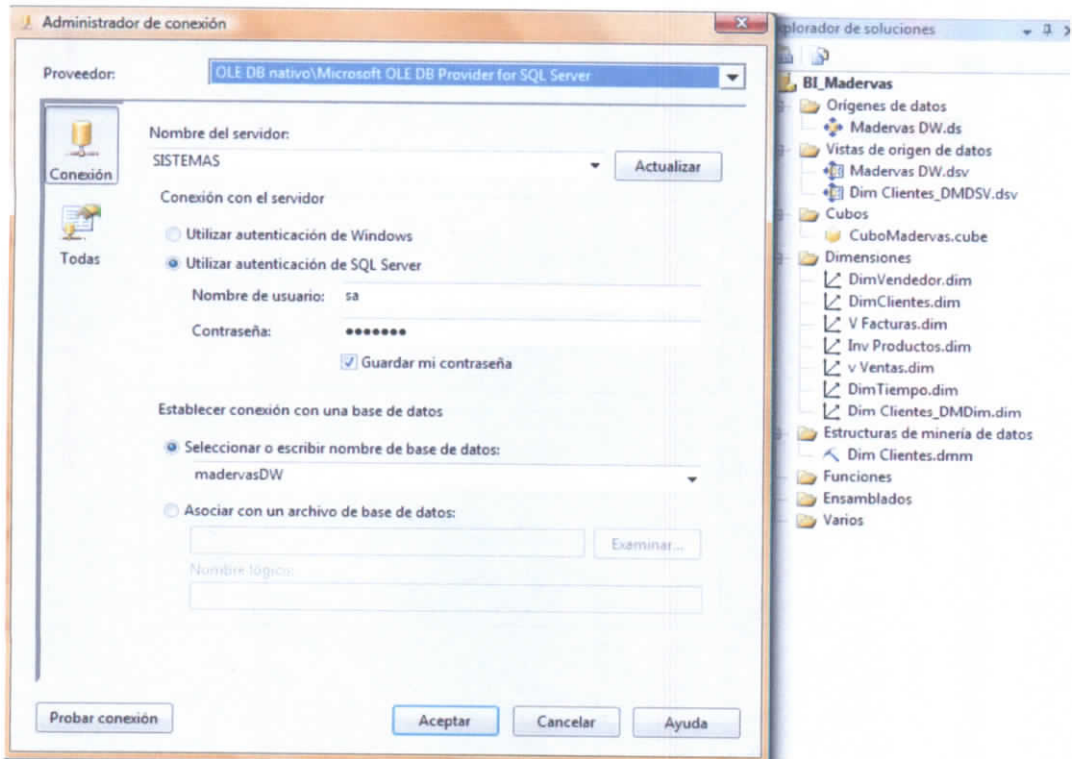


Gráfico 4.7. Conexión al Origen de Datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

3. Definir Vista de Origen de Datos.- se define las tablas y las relaciones que se va utilizar, (tabla de hechos) que es la tabla central del esquema dimensional para la creación de un cubo) y las dimensiones.

Las tablas que se muestran a la vista de origen de datos deben estar ya depuradas, filtradas, dependiendo de los requerimientos.

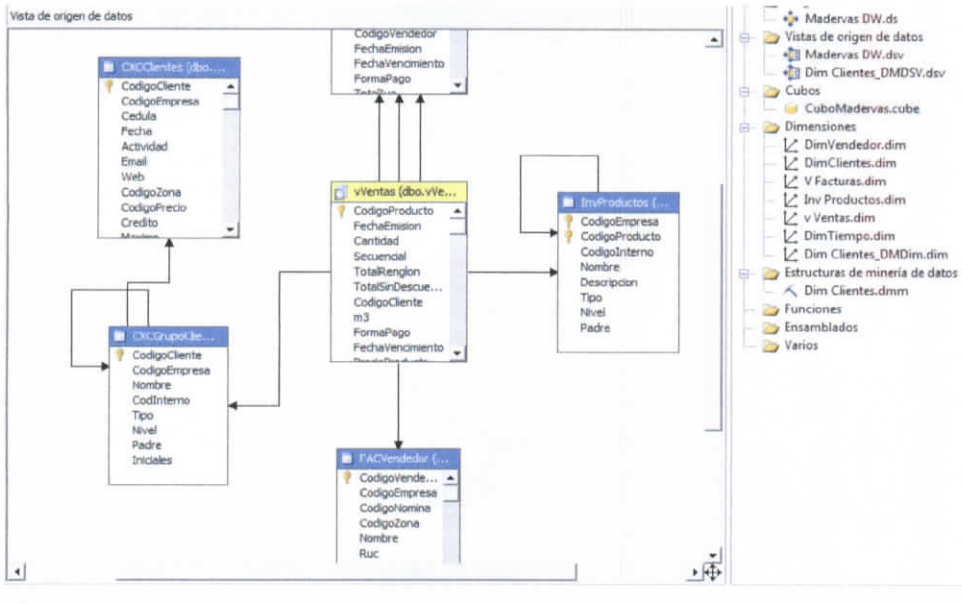


Gráfico 4.8. Vista de las Tablas de Origen de Datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

4. Creación del CUBO OLAP.

Dando un click derecho en CUBOS en el árbol de explorador de soluciones y escogemos crear cubo.

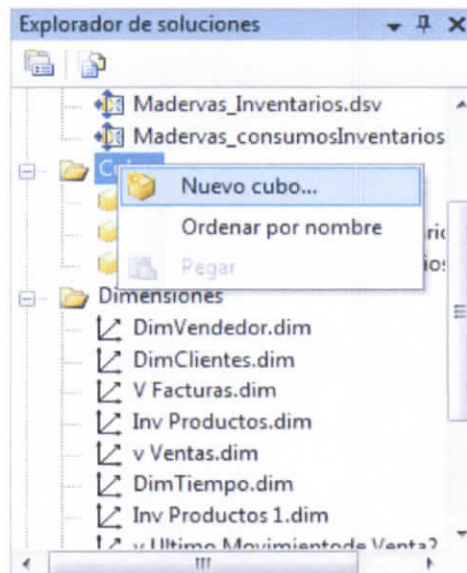


Gráfico 4.9. Creación de un nuevo Cubo.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Automáticamente se inicia el asistente para la creación de cubos, debemos indicar que método de generación deseamos, escoger la generación automática.

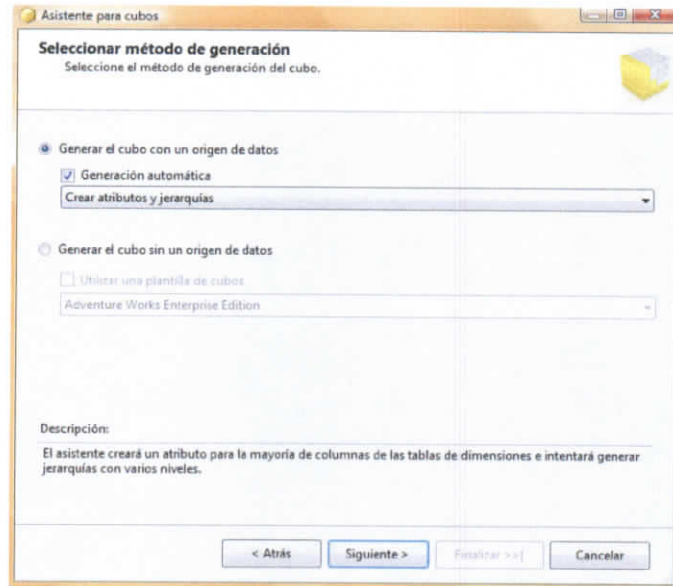


Gráfico 4.10. Método de Generación del Cubo.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Posteriormente escogemos que cual vista de datos vamos a crear el cubo

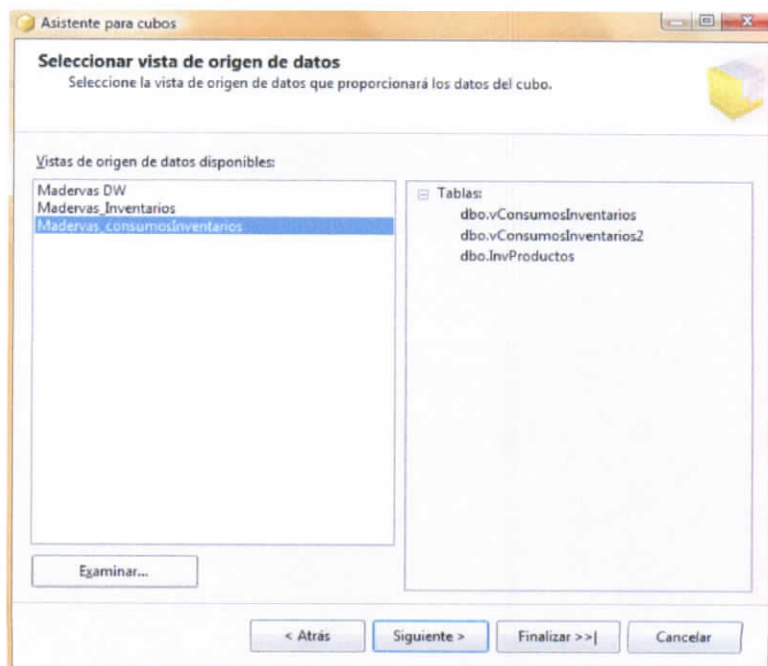


Gráfico 4.11. Selección vista de origen de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Al escoger esta opción de generación este proceso detecta automáticamente las tablas de movimientos (hechos) y las dimensiones.

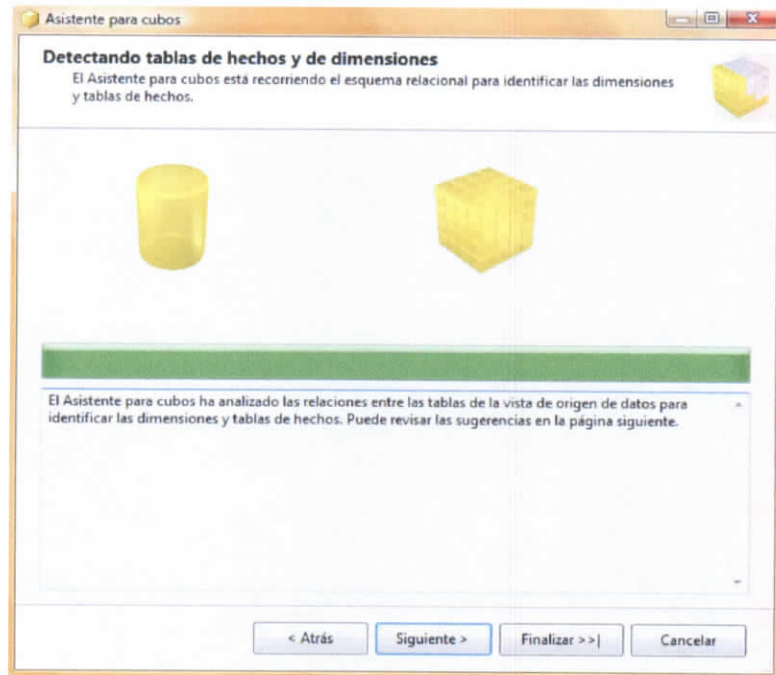


Gráfico 4.12. Detección de tablas de hechos y dimensiones.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Obteniendo como resultado la creación del cubo con la respectiva tabla de hechos y sus dimensiones

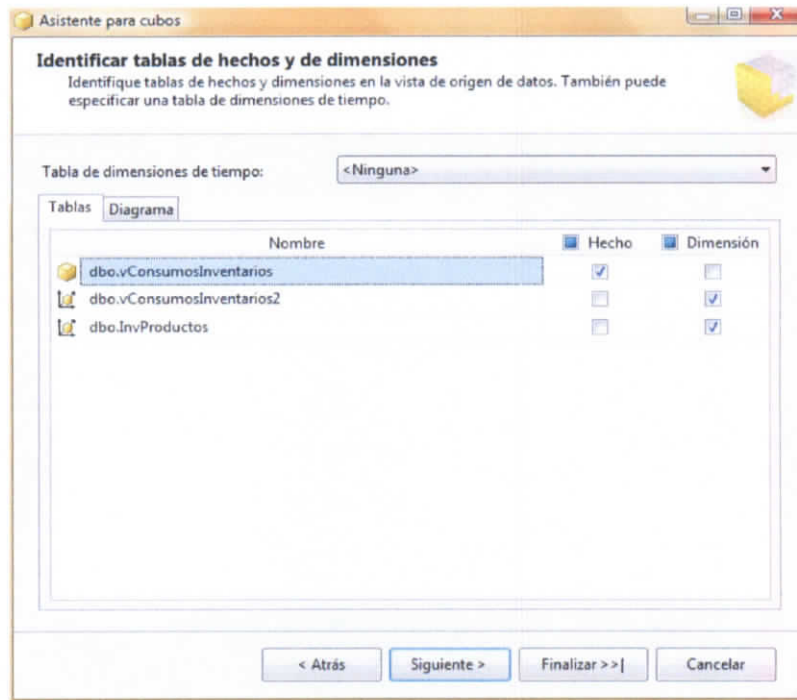


Gráfico 4.13. Identificación de tablas de hechos y dimensiones.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Luego el asistente nos muestra las Medidas que se van a generar

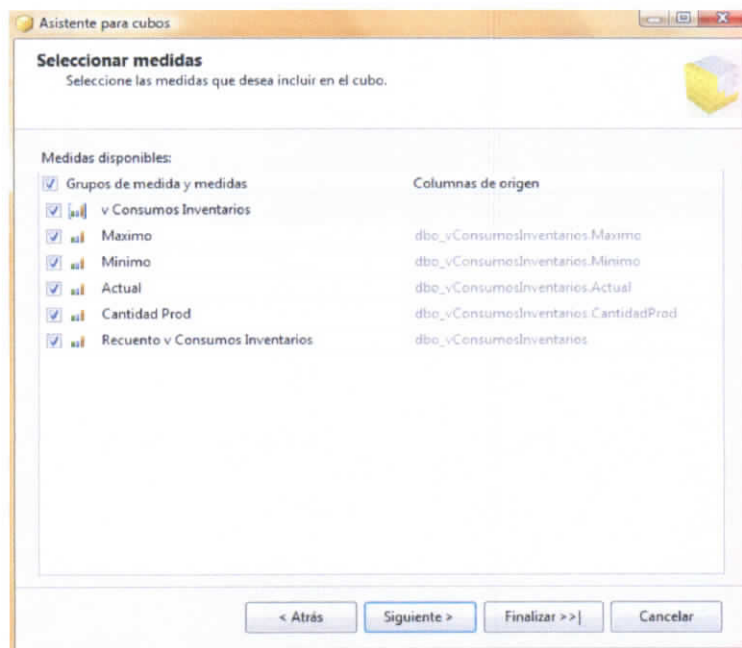


Gráfico 4.14. Selección de medidas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

En el siguiente paso tenemos el detalle de las dimensiones a crearse

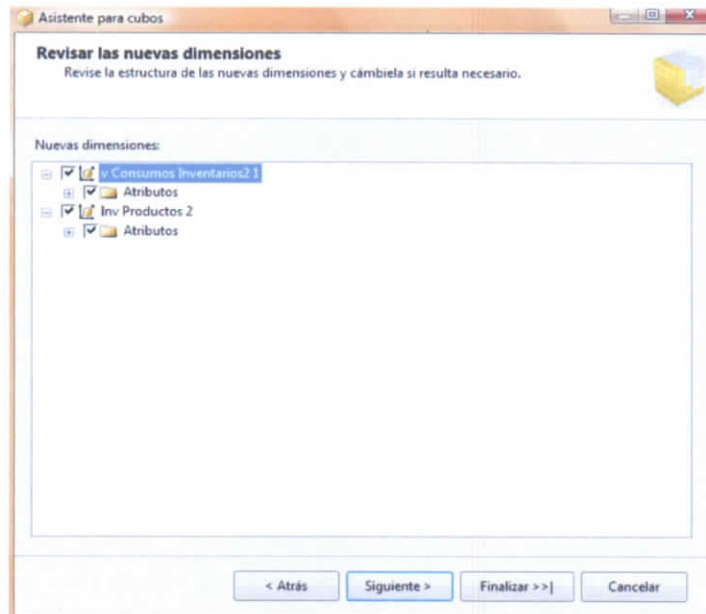


Gráfico 4.15. Medidas a generarse.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Y como último paso es la creación del cubo de datos con sus respectivas dimensiones y medidas.

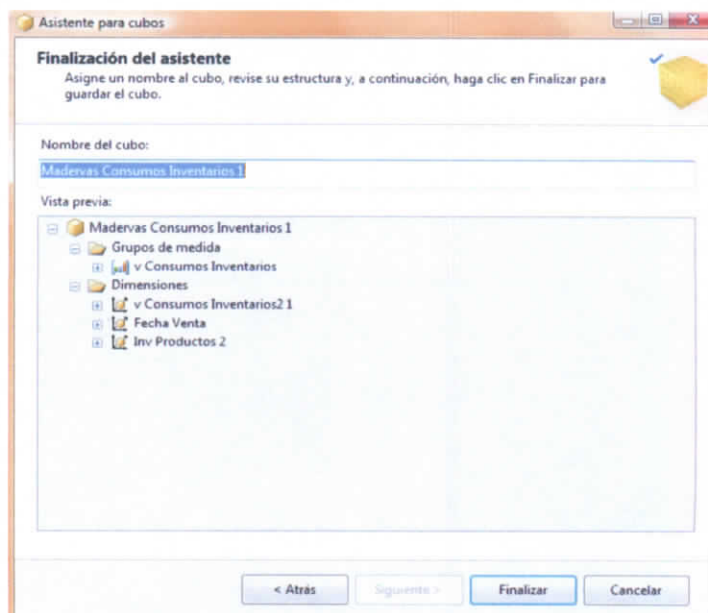


Gráfico 4.16. Estructura del cubo a crearse.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Luego de creado el cubo en el explorador de soluciones se generan automáticamente el cubo, las dimensiones.

5. Implementación del proyecto generado - Data Mining

Para implementar el proyecto debemos dar un click en el menú de generar o dar un click en Implementar que se muestra junto con el nombre del proyecto de analysis services.

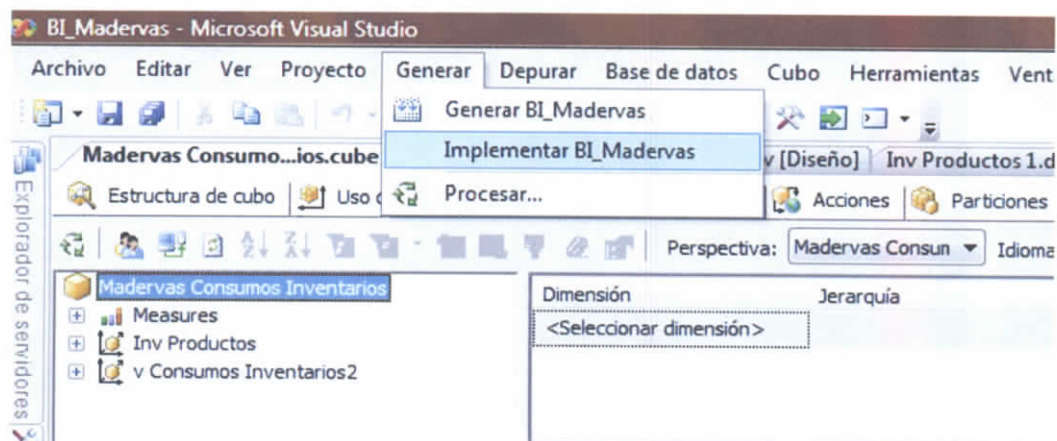


Gráfico 4.17. Implementación del proyecto.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Al iniciar la implementación de proyecto se genera:

a.) En el servidor de analysis services se crea la base de datos con el nombre del proyecto, es decir el data warehouse con los datos procesados y convertidos en información útil para poder ser analizada.

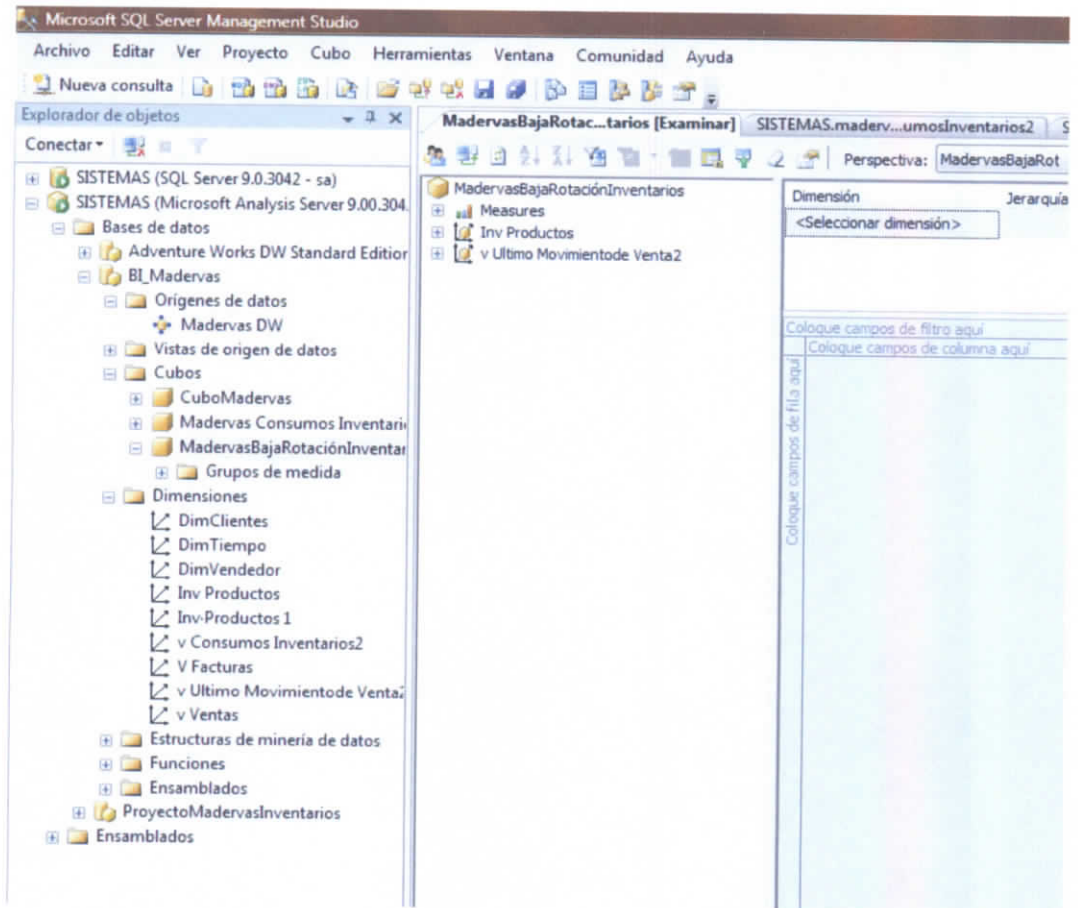


Gráfico 4.18. Generación del proyecto y creación de la base de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

b.) En la ventana de examinar el cubo de información podemos iniciar la revisión de la información generada en el cubo de información.

Dimension	Jerarquía	Operador	Expresión de filtro		
-Seleccionar dimensión-					
Categorías de campo de filtro aquí					
Categorías de campo de columna aquí					
Nivel 02	Máximo	Mínimo	Actual	Cantidad Prod	Recuento v Consumos Inventarios
CONSIGNACION	0,00	0,00	208,00	321,00	281,00
CONTRACHAPADO	0,00	0,00	3.247.273,00	313.154,32	12.880,00
CONTRACHAPADOPEHOLE	0,00	0,00	8.959,00	2.171,00	256,00
FORMICADO	0,00	0,00	1.512,00	207,00	63,00
FORMICAS	0,00	0,00	13.724,00	3.453,00	670,00
herrajes y accesorio	20.250,00	4.050,00	9.108.670,00	799.824,47	46.813,00
herramientas	1,00	1,00	28.083,00	2.980,99	2.056,00
MASSA DSB	0,00	0,00	48.153,00	22.484,00	1.835,00
MOF	3.025.448,00	676.214,00	596.215,00	232.961,50	26.434,00
MELEMINA	0,00	0,00	116.905,00	18.394,00	5.716,00
PCISO	0,00	0,00	15.781,00	30.088,80	1.994,00
PLACA MASSA-HILLO	0,00	0,00	189.430,00	16.760,00	2.882,00
PUERTAS	0,00	0,00	61.635,00	13.837,00	2.040,00
SERVICIOS PLACACENTR	0,00	0,00	-1.234.027,000,00	1.064.423,25	3.464,00
VARIOS	0,00	0,00	-2.286,00	71,00	54,00
Total general	3.045.599,00	680.265,00	-1.220.616,000,00	2.521.031,33	107.428,00

Gráfico 4.19. Exploración de cubo generado.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

De esta forma se generan los cubos de:

Un Cubo de Información de Ventas.- para poder analizar la información agrupada por cliente, por productos, por grupo de inventarios, por cliente, por tipo de precio.

Dos Cubos de Información de Inventarios.- Para analizar la baja rotación de productos, y también para analizar el movimiento de los productos y establecer los stocks máximos y Mínimos

4.6. Data Mining y Extracción del Conocimiento

Para el proceso de predecir el comportamiento de ventas, debemos extraer la información desde el cubo de ventas el mismo que se generó en el proceso anterior.

Para el proceso de extracción de conocimiento en la solución de BI, utilizamos el algoritmo árboles de decisión, ya que este es uno de los utilizados para predicciones.

Para implementar la minería de datos debemos seguir los siguientes pasos:

a.) En el explorador de soluciones crear una nueva Estructura de minería de datos.

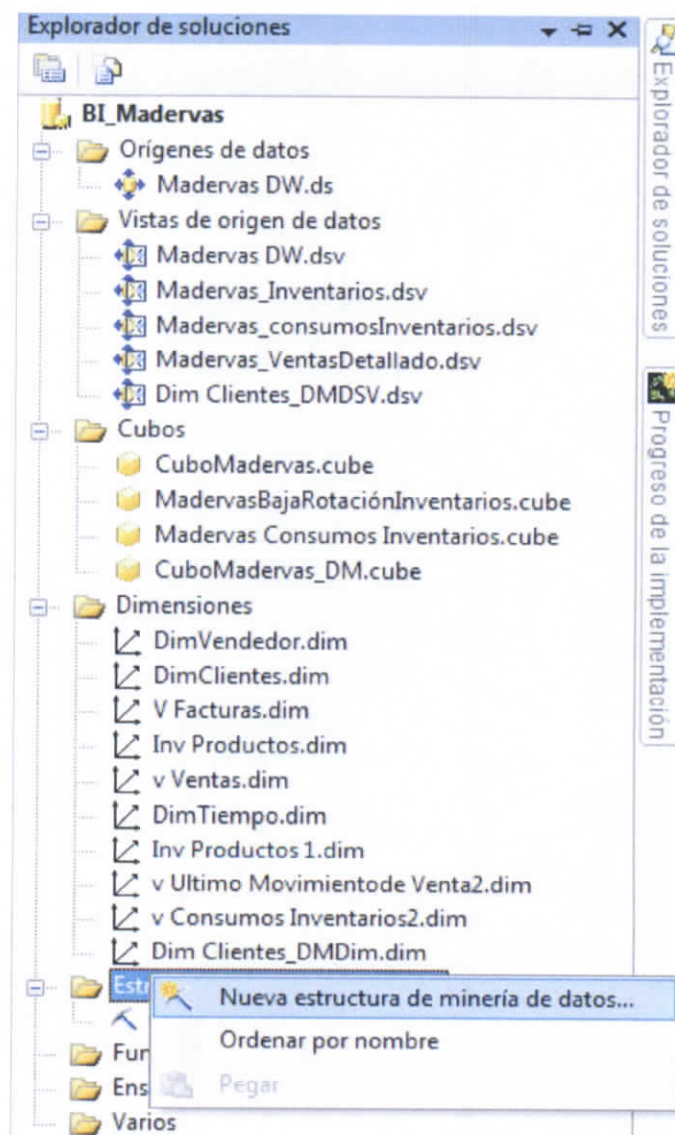


Gráfico 4.20. Nueva estructura de minería de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

b.) Seleccionamos que vamos a utilizar el método de definición: **a partir de un cubo existente.**

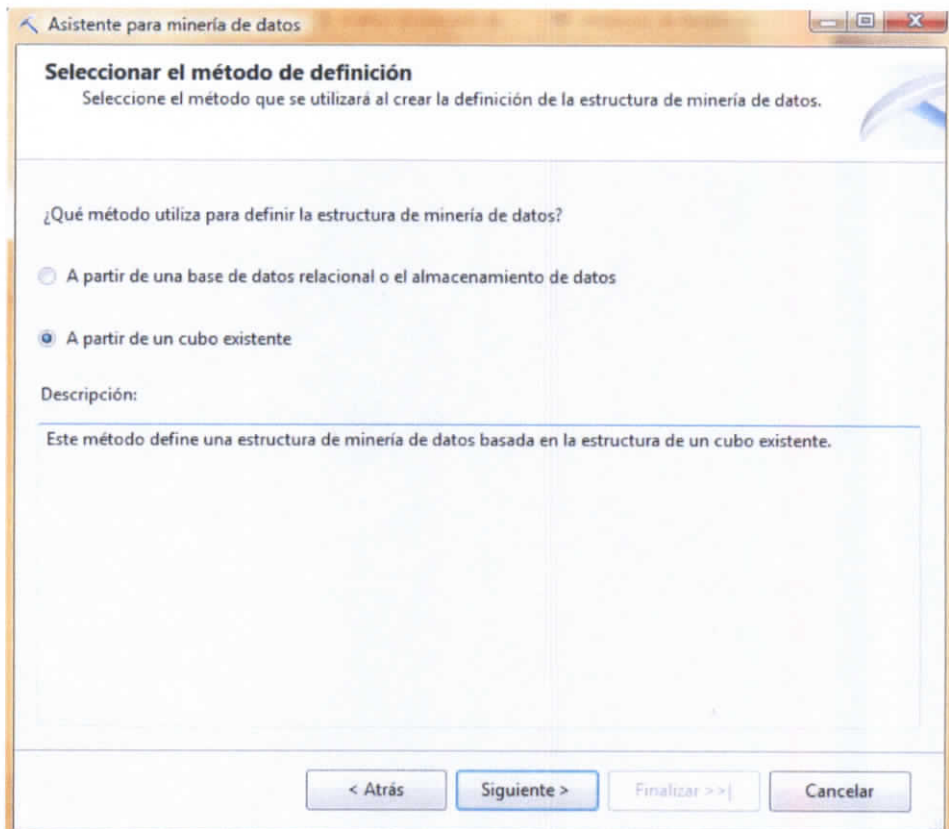


Gráfico 4.21. Definir la estructura de minería de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

c.) Para la técnica de minería de datos escogemos el algoritmo Microsoft Decision Trees.

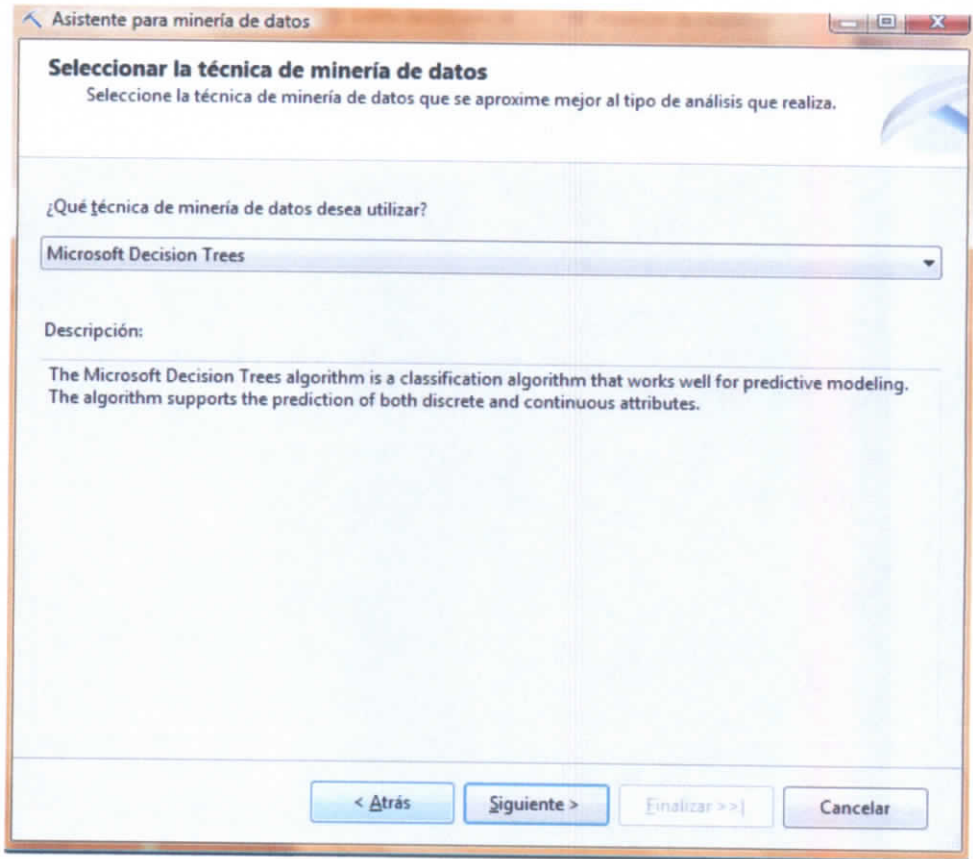


Gráfico 4.22. Definir técnica de de minería de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

d.) La dimensión del cubo de origen que vamos utilizar en este caso es CuboMadervas, y seleccionamos la dimensión DimClientes.

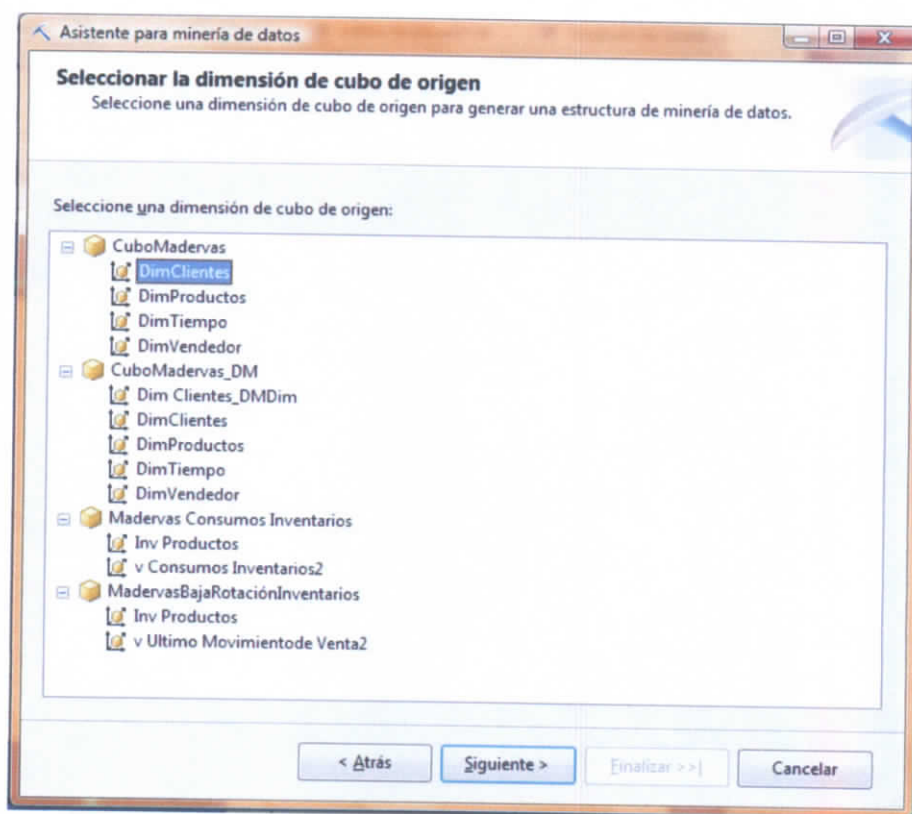


Gráfico 4.23. Selección de dimensión de cubo de origen.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

e.) La clave del escenario es decir por la que vamos analizar o agrupar es CXCGrupoSclientes.

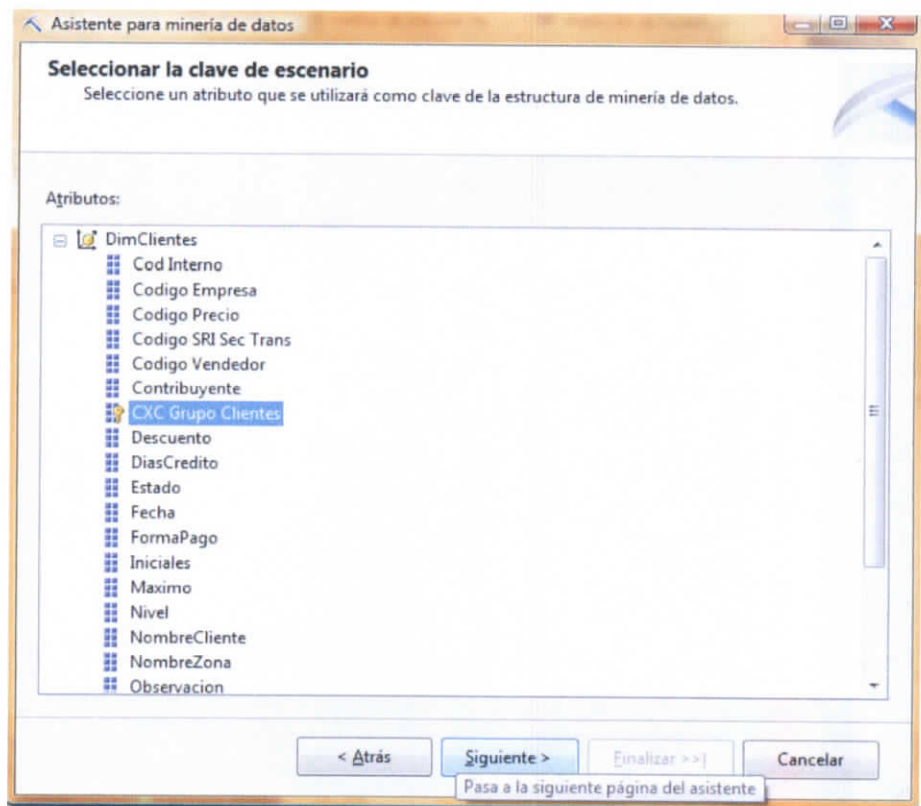


Gráfico 4.24. Selección de Clave de escenario.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

f.) Escogemos las columnas del nivel de escenario: Atributos:

CódigoPrecio, CódigoVendedor, FormaPago, NombreCliente, Zona, y las

Medidas: Cantidad, m3.

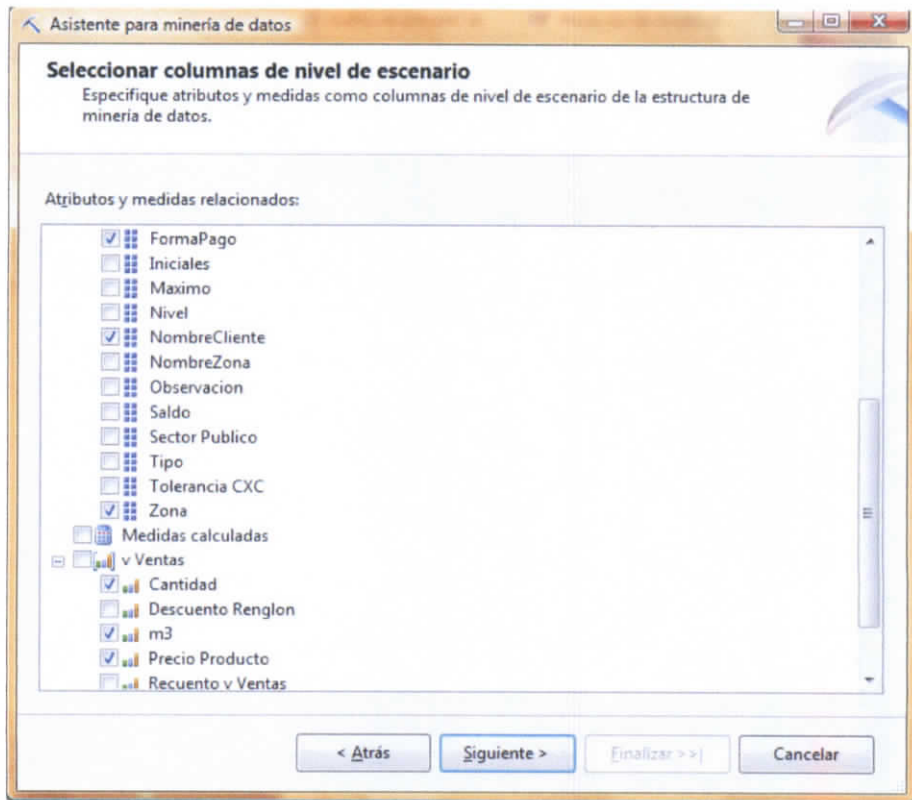


Gráfico 4.25. Selección de columnas de nivel escenario.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

g.) Seleccionamos las columnas de entrada y de predicción tal como se muestra en la gráfica.

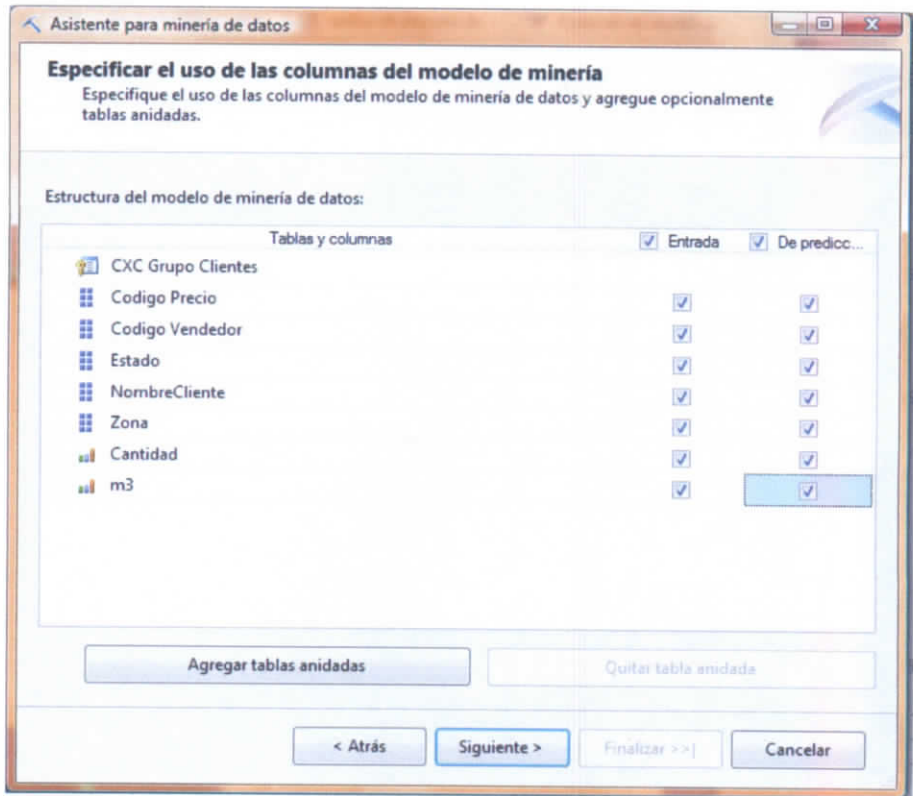


Gráfico 4.26. Especificar columnas de entrada y de predicción.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

h.) En el siguiente paso automáticamente se especifican que tipo de datos de las columnas son las que se van a generar.

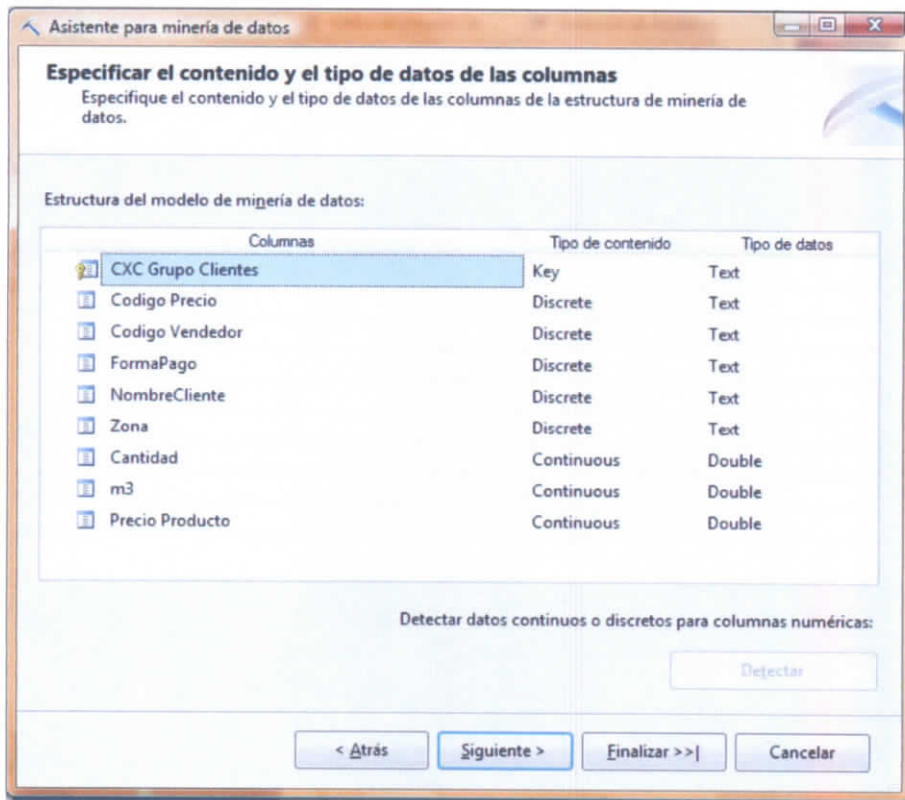


Gráfico 4.27. Contenido y tipo de datos de columnas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

i.) No vamos utilizar la opción de segmentar el cubo.

j.) Finalmente, tenemos la gráfica de cómo está estructura de la minería de datos.

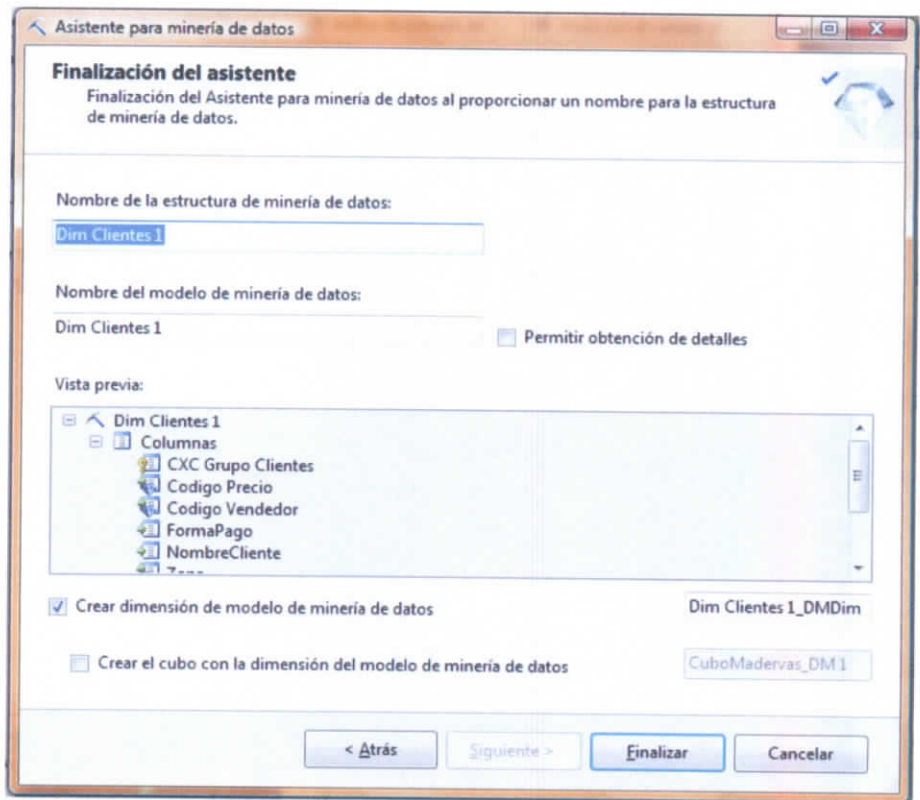


Gráfico 4.28. Estructura de minería de datos.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

4.7. Solución BI

Para explorar e interactuar con la solución BI, utilizaremos el Microsoft Excel con las tablas pivote para la revisión del cubo de información generado, siendo una de las aplicaciones de escritorio más utilizadas en el mundo.

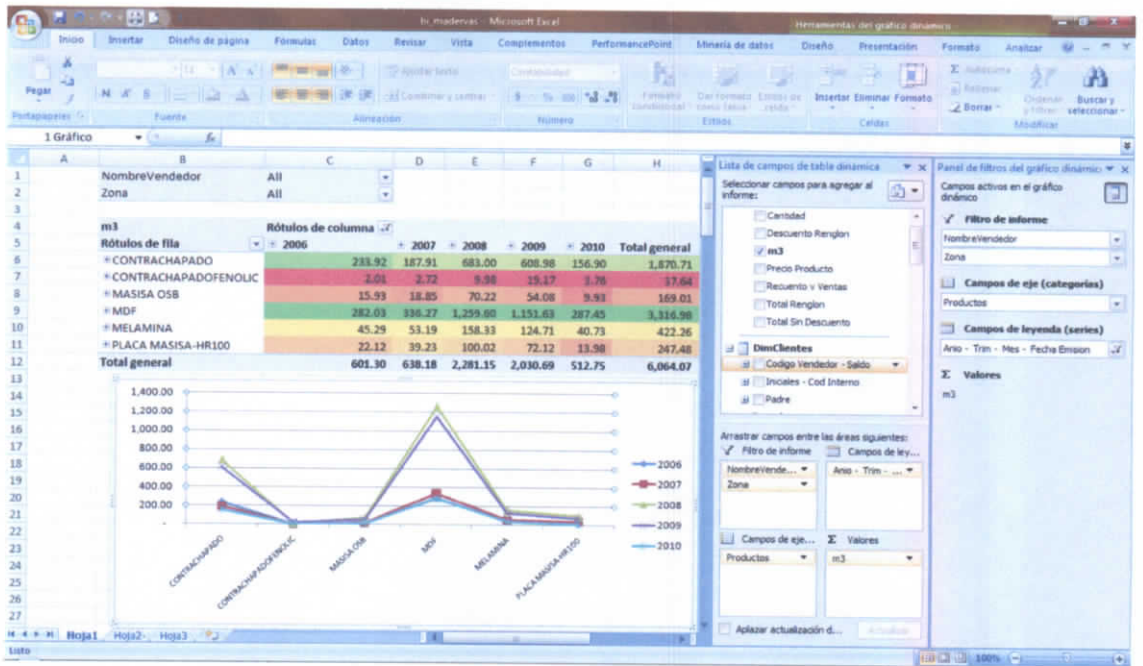


Gráfico 4.28. Navegación de la Solución BI en Microsoft Excel.

Fuente: Interface Microsoft Excel.

Para la generación y extracción del conocimiento utilizamos el complemento de minería de datos para Excel.

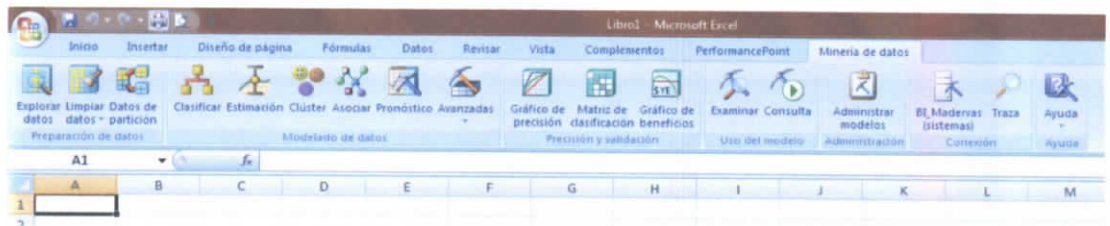


Gráfico 4.30. Complemento de minería de datos para Microsoft Excel.

Fuente: Interface Microsoft Excel.

También podemos analizar la información por medio del visor de modelos de minería de datos incluido en el SQL Server Business Intelligence Development Studio.

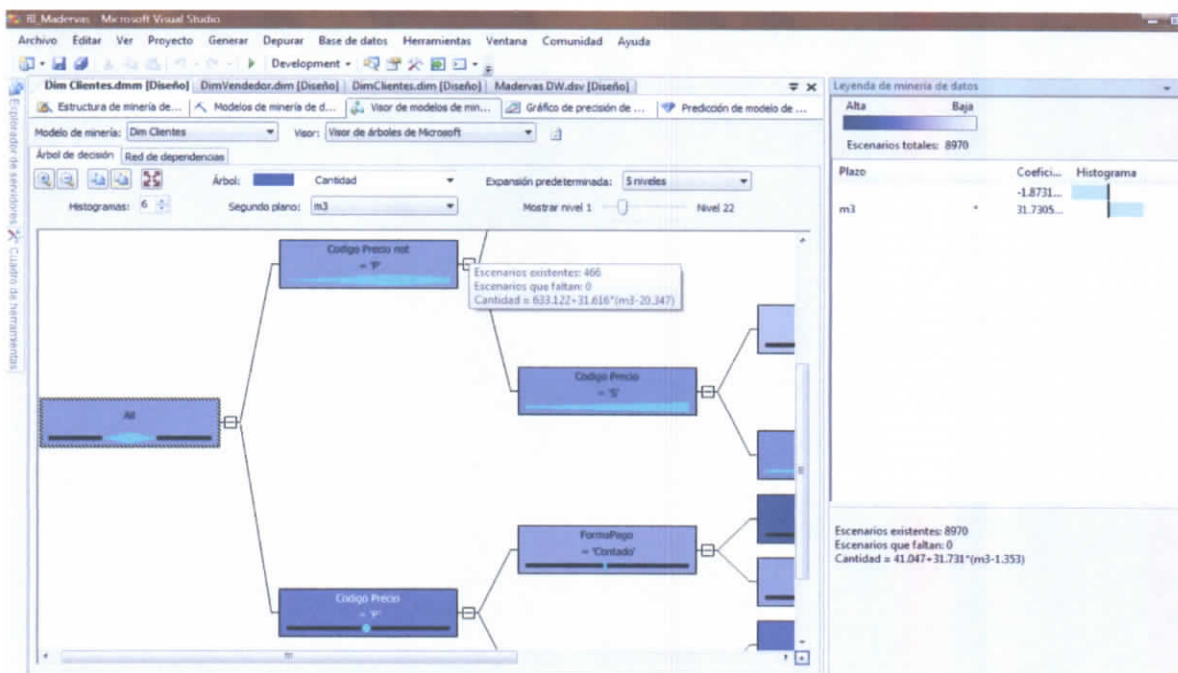


Gráfico 4.31. Visor de minería de datos con el algoritmo árboles de decisión.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

CAPITULO V

5. VALIDACION.

5.1. Demostración de la hipótesis.

La Hipótesis planteada fue: Aplicando Business Intelligence a los clientes de Madervas se logrará tomar decisiones para incrementar las ventas a través del análisis continuo del conocimiento extraído tomando decisiones adecuadas, para lo cual se debe mantener un inventario adecuado de productos, también logrando mejorar las negociaciones con los proveedores.

Para demostrar la hipótesis tenemos que cumplir los objetivos específicos:

a) Implementar un cubo de información de ventas que nos permita revisar y analizar la información utilizando los diferentes tipos de dimensiones como son:

DimClientes.- Con esta dimensión podemos analizar la información agrupada por clientes, y también analizar con un segundo nivel de grupos de inventarios.

DimProductos.- La dimensión inventarios nos permite utilizar para agrupar, como también para utilizar filtros si se arrastra la dimensión al área de filtro.

DimTiempo.- Esta dimensión nos permite agrupar por períodos de tiempo como: Año, Mes, Trimestre.

Dim Vendedor.- La dimensión de vendedores nos permite agrupar por vendedores.

		Año			
		2005	2006	2007	2008
		Cantidad	Cantidad	Cantidad	Cantidad
		m3	m3	m3	m3
DimProductos	Total	190,00	288,00	300,00	10,72
DimProductos	CONTRACHAPADO	190,00	288,00	300,00	10,72
DimProductos	CONTRACHAPADO 12mm	100,00	4,47	223,00	9,96
DimProductos	CONTRACHAPADO 18mm	45,00	2,41	31,00	1,66
DimProductos	CONTRACHAPADO 28mm	2,713,00	22,30	3,670,00	43,70
DimProductos	CONTRACHAPADO 6mm	250,00	4,47	189,00	3,35
DimProductos	CONTRACHAPADO 9mm	930,00	24,92	1,055,00	28,36
DimProductos	CONTRACHAPADO TP-6mm	30,00	0,24	25,00	0,20
DimProductos	Total	4,258,00	75,59	5,471,00	96,91
DimCuentas	MEX	1,572,00	92,09	1,835,00	123,60
DimCuentas	PLACA MASSA HR 100	243,00	18,62	323,00	20,21
DimCuentas	MASSA OSB	1,20	0,03	6,00	0,20
DimCuentas	MELAMINA	6,00	0,42	6,00	0,41
DimCuentas	Total	6,080,00	183,75	7,751,00	241,33
DimVendedor	CHASQUIZA FUERTES VICTOR HUGO	3,138,00	114,53	7,156,00	155,74
DimVendedor	SANCHEZ ALDAZ JOSE VICENTE	1,870,00	21,80	3,735,00	38,38
DimVendedor	ROBALINO PEREZ GLORIA BENE	1,211,00	86,71	3,484,00	85,16
DimVendedor	MORALES FUNDAS GUIDO ENRIQUE	64,00	2,90	1,393,00	87,49
DimVendedor	SARZON NUÑEZ LARSON ROBERT	2,778,00	112,21	1,340,00	69,21
DimVendedor	ECOTRIPLIX	408,00	6,09	3,694,00	57,97
DimVendedor	GUDMAN JACOME WITMANE	386,00	12,00	1,015,00	28,61
DimVendedor	CABRERA ANDA PABLO HERNAN			89,00	5,78
DimVendedor	MORALES ANDRADE ROSA LEONOR	1,138,00	39,03	1,656,00	53,64
DimVendedor	VACA ZUÑIGA SILVIA VERONICA (REMASI FERRERIA)			722,00	19,06
DimVendedor	SORIA VILLACEZ DAVID ALFREDO			1,407,00	48,19
DimVendedor	ORLIZ VILLAGAS LUIS ROBERTO			1,234,00	38,58
DimVendedor	VALDIVIEZO VALDIVIEZO CLARA SUSANA	180,00	6,14	30,00	0,36
DimVendedor	SILVA FLORES DAVID ALFONSO			200,00	7,01
DimVendedor	Total				

Gráfico 5.1. Cubo de ventas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

b) Implementar un cubo de información de Inventarios para poder establecer los stocks mínimos y máximos.- en el cubo de información podemos analizar los inventarios agrupados por productos, o por grupo de inventarios.

Dimensión	Jerarquía	Operador	Expresión de filtro	Prad	Mesmo	Actual	Cantidad Prod	Mesmo	Actual
Level 04	ANGULO DE TABLERO DE 33 BLANCO			42,00	10,00	63,00	188,00		
Level 05	MOP MELC. CEDRO			42,00	10,00	63,00	188,00		
Level 06	MOP MELC. CEDRO 15m			35,00	0,00	18,00	30,00		
Level 07	MOP MELC. CEDRO AMAZONICO 15x1830x2500 JC Total			42,00	0,00	63,00	218,00		
	MOP MELC. CEDRO 18mm			42,00	0,00	4,00	17,00		
	MOP MELC. CEDRO 18mm Total			35,00	0,00	8,00	17,00		
	MOP MELC. CEDRO 18mm			0,00	0,00	1,00	1,00	0,00	0,00
	MOP MELC. LAUREL			9,00	0,00	3,00	1,00		
	MOP MELC. JODAL			0,00	0,00	2,00	1,00		
	TMBER ADAPTADOR MAGNETICO			0,00	0,00	1,00	1,00		
	TMBER PRESA BISSAGA PRESION 26mm			0,00	0,00	1,00	1,00		
	TMBER PRESA REGULAR 7/8			0,00	0,00	2,00	1,00		
	TMBER PRESA MEDIA CAÑA 1/8			0,00	0,00	1,00	2,00		
	TMBER PRESA MEDIA CAÑA 3/16			0,00	0,00	1,00	1,00		
	TMBER PRESA RANURADO 1/8			0,00	0,00	1,00	1,00		
	TMBER PRESA RANURADO V 90			0,00	0,00	1,00	1,00		
	TMBER PRESA RANURADO 1/4			0,00	0,00	1,00	1,00		
	TMBER PRESA REBASES 1 1/4			0,00	0,00	1,00	2,00		
	TMBER PRESA RECTA 1/2 X 1/2"			0,00	0,00	1,00	2,00		
	TMBER PRESA RECTA 1/2 X 1/2-7/8"			0,00	0,00	1,00	4,00		
	TMBER PRESA RECTA 1/2X3/8			0,00	0,00	1,00	3,00		
	TMBER PRESA RECTA 1/2X5/8			0,00	0,00	1,00	3,00		
	TMBER PRESA RECTA 1/4 X 2-1/4"			0,00	0,00	1,00	2,00		
	TMBER PRESA RECTA 1/4 X 2-7/8"			0,00	0,00	1,00	3,00		
	TMBER PRESA RECTA 3/16			0,00	0,00	1,00	2,00		
	TMBER PRESA RECTA 5/16			0,00	0,00	1,00	1,00		
	TMBER PRESA RECTA 9/32			0,00	0,00	1,00	2,00		
	TMBER PRESA RECTA ROD. SUP. 1/2 X 2-1/2"			0,00	0,00	1,00	2,00		

Gráfico 5.2. Cubo de Inventarios.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

c) Analizar la rotación de productos de baja rotación, y tomar una decisión en razón de los mismos.- implementando el cubo de información de inventarios con productos de baja rotación, o que no se han vendido, de un solo vistazo podemos definir los productos hueso, ante lo cual gerencia debe tomar decisiones de cómo disponer ese producto.

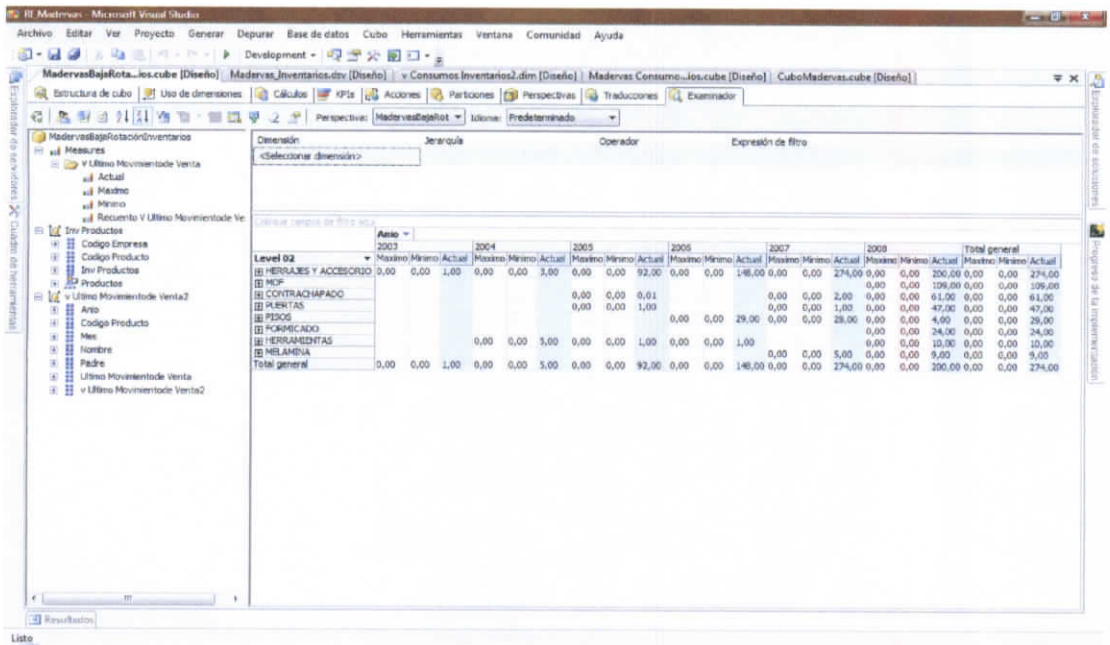


Gráfico 5.3. Cubo de Inventarios con productos de baja rotación.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

En el cuadro podemos analizar los productos hueso desde el año 2008 al 2003, y que no han tenido movimiento.

- d) Mejorar las negociaciones con los proveedores.- al momento de realizar compras a Masisa ya que se va a tener la información necesaria, de tal forma que se adquiera los productos necesarios y en las cantidades requeridas.- al definir los stocks máximos y mínimos como también el análisis de consumo no se solicitará productos que no han rotado y al mismo tiempo se solicitará los productos que si tienen movimiento.
- e) Predecir comportamientos de ventas utilizando mediante la extracción del conocimiento utilizando el algoritmo de arboles de decisión.-

implementado la estructura de minería de datos, podemos analizar a los clientes clasificados, por:

- **Código Precio.-** además esta agrupado por la forma de pago contado y no contado, y en la ventana de leyenda tenemos las probabilidades de distribución de compras de los clientes, esto en razón del análisis del algoritmo de árboles de decisión.

Interpretando la leyenda.- tenemos una probabilidad de seguir vendiendo al precio P (Público) del 94,67%, seguido de un 4.19% para F (Frecuentes), a subdistribuidores (S) 1.05%, y como precio de distribución (D) el 0.06%.

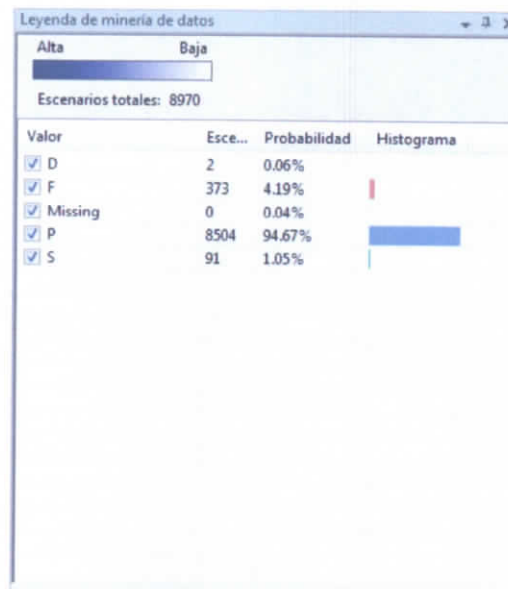


Gráfico 5.4. Probabilidades de ventas por tipo de precio.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

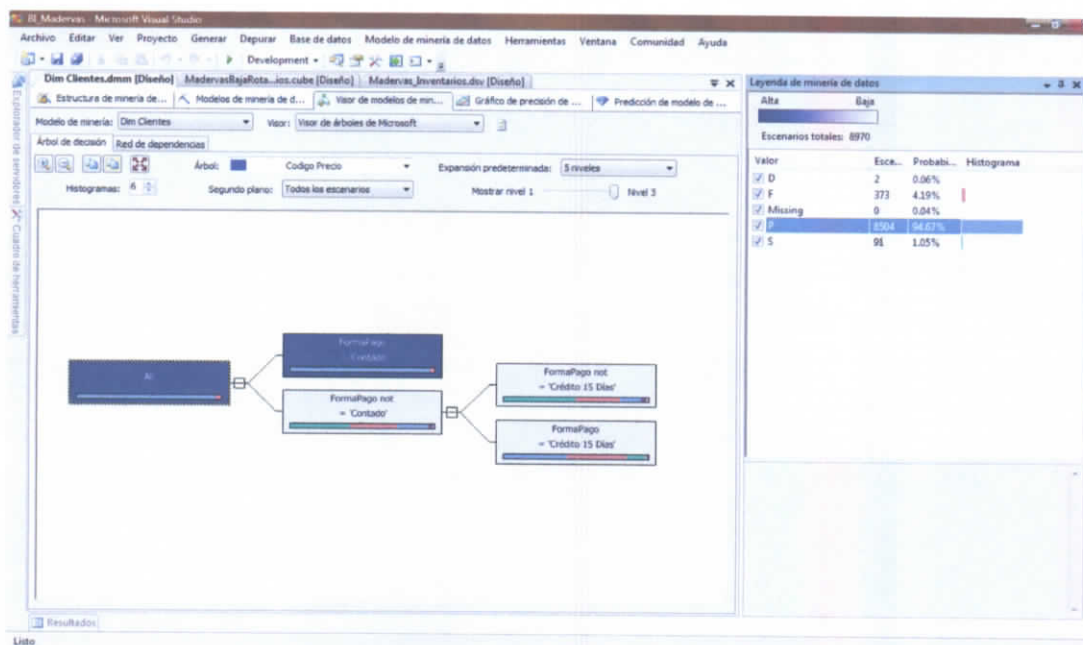


Gráfico 5.5. Visor árboles de decisiones por tipo de precio.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

- **Forma de Pago.-** en este análisis se puede analizar que la forma la tendencia de tipos de pago que la empresa va a mantener se muestra en la ventana de leyenda, Interpretando la información, o conocimiento extraído podemos decir que la probabilidad de vender de contado es del 98.51%, es decir se debe mantener las políticas de crédito como hasta hoy.

Leyenda de minería de datos			
Alta		Baja	
Escenarios totales: 8970			
Valor	Esce...	Probabi...	Histograma
<input checked="" type="checkbox"/> Contado	8850	98.51%	
<input checked="" type="checkbox"/> Crédito 10 Días	3	0.06%	
<input checked="" type="checkbox"/> Crédito 15 Días	31	0.37%	
<input checked="" type="checkbox"/> Crédito 30 Días	71	0.81%	
<input checked="" type="checkbox"/> Crédito 45 Días	4	0.07%	
<input checked="" type="checkbox"/> Crédito 60 Días	6	0.09%	
<input checked="" type="checkbox"/> Crédito 8 Días	5	0.08%	
<input checked="" type="checkbox"/> Missing	0	0.02%	

Gráfico 5.6. Probabilidad de venta por forma de pago.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

Explorar el modelo de arboles de decisión:

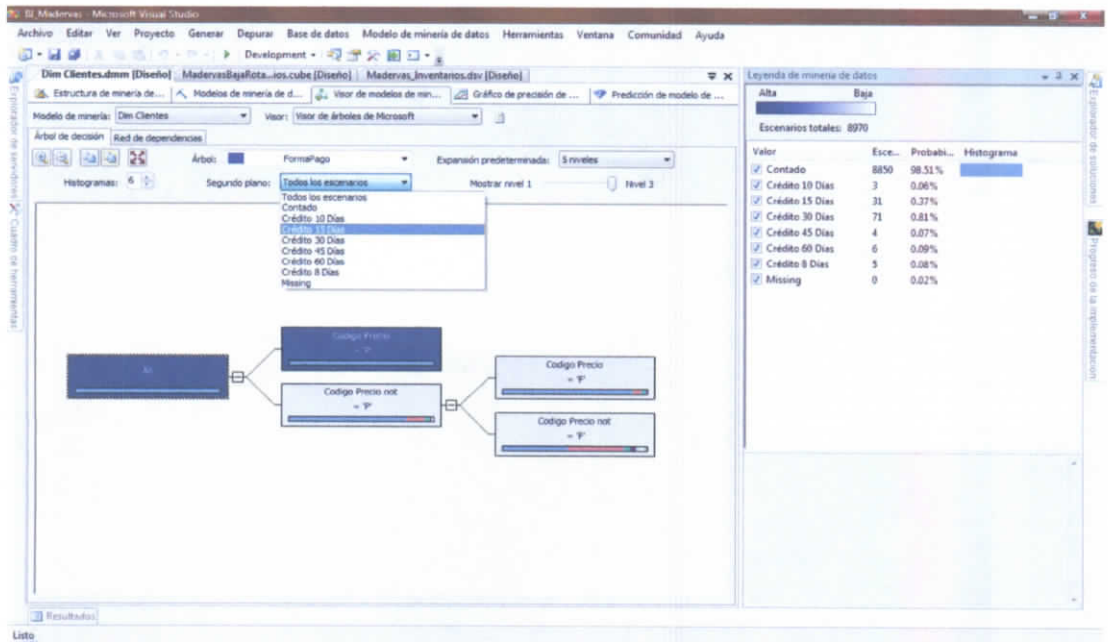


Gráfico 5.7. Visor árboles de decisión por forma de pago.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

- **Zona.-** al analizar por zona podemos deducir que el 84% de ventas del Placacentro se realizarán en Ambato, tal como se ha venido haciendo hasta hoy, Seguido de un 3.27% en Pelileo.

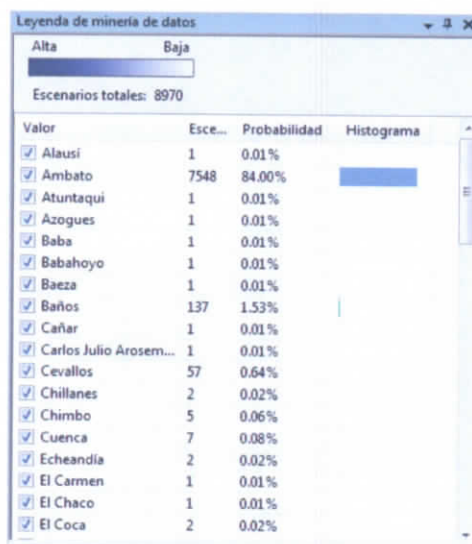


Gráfico 5.8. Probabilidad de ventas por Zonas.

Fuente: Interface de SQL Server Business Intelligence Development Studio.

CAPITULO VI

6. Conclusiones y Recomendaciones

6.1. Conclusiones

Con la tecnología Business Intelligence en la empresa Madervas, se facilita y apoya a la gerencia en la toma de decisiones, considerando que el acceso y análisis a la información se puede hacer en cualquier momento. La información está disponible y actualizada al momento requerido.

Al contar con BI el personal ya no dedica tiempo adicional a procesar manualmente la información, y por su puesto minimizando errores de digitación. El tiempo ahora se va a dedicar a analizar la información y tomar decisiones; o ratificar las ya tomadas.

Al hacer el análisis de los cubos OLAP de ventas e inventarios se tiene una visión más amplia del negocio ya que se puede utilizar varios escenarios y agrupamientos.

Con la extracción del conocimiento (KDD), podemos concluir que esta tecnología sirve de apoyo a los tomadores de decisiones, ya que se pueden analizar las probabilidades en el modelo de arboles de decisión.

La información va estar centralizada para poder acceder a la misma en cualquier momento. No es necesario que el usuario ejecute alguna actividad para procesar y obtener los reportes requeridos.

6.2. Recomendaciones

La tecnología Business Intelligence es una herramienta muy útil para los gerentes, ahora que se ha implementado en la empresa Madervas se recomienda seguirla utilizando, y mejorando con los nuevos requerimientos de análisis.

Las probabilidades o tendencias generadas por los algoritmos de Analysis Services realizan cálculos matemáticos complejos, los mismos que siempre tienen que ser validados por el personal que utiliza dicha información.

Si se realiza algún cambio tanto en los cubos OLAP como en la estructura de minería de datos se debe validar los resultados, ya que puede haber varios factores que afecten al modelo, y puede tornarse sensible ante cualquier cambio.

BIBLIOGRAFIA:

- OLAP Train, LLC for Microsoft Corporation. Designing and Implementing OLAP Solutions with Microsoft SQL Server 2000. Colombia. Cargraphics S.A. Microsoft Training and Certification. Course Number: 2074a. 2001

- José Hernández Orallo, Ma.José Ramirez Quintana, César Ferri Ramirez. Introducción a la Minería de Datos. Madrid. Pearson Educación S.A. 2004

Artículos Electrónicos.

- Max Chickering and Ramon Iyer. A Tutorial for constructing a Plug-in Algorith.. Microsoft Corporation. Agosto 2004.

<http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnsql90/html/pluginalg2.asp>

- BI-Spain.com Portal en Español sobre Business Intelligence
<http://www.bi-spain.com/>

- Business Intelligence de Microsoft
<http://www.microsoft.com/bi/>

- Business Intelligence + Informática Estratégica
http://www.sinnexus.com/business_intelligence/

- Business Intelligence basado en Excel. DataCycle Reporting
<http://www.apesoft.es/>

- Extracción del conocimiento

<http://kdd.org/kdd>

- Gestión del conocimiento

<http://www.gestiondelconocimiento.com/>

- Luis Carlos Molina Felix. Data Mining: Torturando a los datos hasta que confiesen.

<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>

- Lynn Langit. Foundations of SQL Server 2005 Business Intelligence. James Huddleston. New York.

http://books.google.com.ec/books?id=N7tzbajkNccC&printsec=frontcover&dq=Foundations+of+SQL+Server+2005+Business+Intelligence&source=bl&ots=NDfHXz8Mgl&sig=K8GmaukSEPow-vYuDW0ELN7U I&hl=es&ei=47nDS8WpDMKB8qb37YT9CA&sa=X&oi=book_result&ct=result#v=onepage&q&f=false

- Microsoft Technet.

<http://technet.microsoft.com>

- Open Source Business Intelligence

<http://www.pentaho.com/>

- Raman Iyer and Bogdan Crivat. SQL Server Data Mining: Plug-In Algorithms. Microsoft Corporation.

<http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnsq190/html/ssdmpia.asp>

- Robert Wrembel, Robert Wrembel y Christian Koncilia. Data Warehouses and OLAP Concepts, Architectures and Solutions. April Schmidt. United States Of America. Editorial: IGI Global. 11 Diciembre 2006

http://books.google.com/books?id=XFivorxZDm8C&dq=Data+Warehouses+and+OLAP+Concepts,+Architectures+and+Solutions&pg=PP1&ots=Ss44ExDxWv&source=bn&sig=Isck8bEkRQdIVE70kDDyD9L9Z90&hl=en&sa=X&oi=book_result&resnum=4&ct=result#v=onepage&q&f=false

- **Sql Server Tutorial**

<http://msdn.microsoft.com/en-US/library/ms170486%28v=SQL.90%29.aspx>

- **SQL Reporting Services**

[http://msdn.microsoft.com/es-es/library/ms166358\(SQL.90\).aspx](http://msdn.microsoft.com/es-es/library/ms166358(SQL.90).aspx)

- **SQL Server Analysis Services**

[http://msdn.microsoft.com/es-es/library/ms166364\(SQL.90\).aspx](http://msdn.microsoft.com/es-es/library/ms166364(SQL.90).aspx)

- **SQL Server Book on Line**

<http://msdn.microsoft.com/en-US/library/ms167593%28v=SQL.90%29.aspx>

- Wikipedia

<http://es.wikipedia.org/>

- Wordreference.

<http://ww.wordreference.com>

GLOSARIO

D

DTS.- (Data Transformation Services).- servicio de transformación de datos.

M

Metadatos.- literalmente «sobre datos», son datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso.

MDX.- (Expresiones multidimensionales), un lenguaje de consulta estándar para el análisis; o Extensiones de minería de datos (DMX)

MDDB.- Base de Datos multidimensional

O

OLAP .- Online Analytical Processing Las bases de datos OLAP están organizadas para facilitar la recuperación y análisis de grandes cantidades de datos.

P

Probabilidad.- mide la frecuencia con la que se obtiene un resultado (o conjunto de resultados) al llevar a cabo un experimento aleatorio, del que se

conocen todos los resultados posibles, bajo condiciones *suficientemente* estables. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemática, la ciencia y la filosofía para sacar conclusiones sobre la probabilidad de sucesos potenciales y la mecánica subyacente de sistemas complejos.

Predicción.- Anuncio o aviso previo de un hecho que va a suceder

R

RDBMS.- Relational Database Management System

S

Sistemas legacy.- sistema heredado (o sistema *legacy*) es un sistema informático (equipos informáticos o aplicaciones) que ha quedado anticuado pero continúa siendo utilizado por el usuario (típicamente una organización o empresa) y no se quiere o no se puede reemplazar o actualizar de forma sencilla.

T

Tendencia.- El concepto de tendencia es absolutamente esencial para el enfoque técnico del análisis de mercados. Todas las herramientas usadas por el analista técnico tienen un solo propósito: detectar y medir las

tendencias del precio para establecer y manejar operaciones de compra-venta dentro de un cierto mercado.

SI
NO
N/A

6. ¿Le queda el tiempo suficiente para analizar documentos e informes?

SI
NO
N/A

7. ¿Dispone de alguna ventaja competitiva clara con respecto a las demás empresas de su sector?

SI
NO
N/A

8. ¿Sabe con certeza si su personal de ventas cumple con los objetivos planificados?

SI
NO
N/A

9. ¿Analizado cuáles son las tendencias de compra de sus clientes, por período, por producto, etc.?

SI
NO
N/A

10. ¿Puede establecer de forma rápida qué productos fueron los más rentables durante un periodo determinado?

SI	
NO	X
N/A	