

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE
MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**



**IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE
NEGOCIOS BASADO EN EL ALGORITMO DE SERIE TEMPORAL
PARA LA MEJORA DEL PROCESO DE TOMA DE DECISIONES
GERENCIALES EN UNA EMPRESA COMERCIAL**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

ANALI DEL ROSARIO LÓPEZ PALACIOS

Chiclayo, 16 de diciembre de 2015

**“IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA
DE NEGOCIOS BASADO EN EL ALGORITMO DE SERIE
TEMPORAL PARA LA MEJORA DEL PROCESO DE TOMA DE
DECISIONES GERENCIALES EN UNA EMPRESA COMERCIAL”**

POR:

ANALI DEL ROSARIO LÓPEZ PALACIOS

**Presentada a la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR EL JURADO INTEGRADO POR

**Mgr. Juan Antonio Torres Benavides
PRESIDENTE**

**Ing. José Adalberto Espinoza Delgado
SECRETARIO**

**Ing. Segundo José Castillo Zumarán
ASESOR**

INDICE

I. INTRODUCCIÓN.....	10
II. MARCO TEÓRICO	14
2.1 Antecedentes	14
2.1.1. Antecedentes de Investigación	14
2.1.1.1.- Locales	14
2.1.1.2.- Nacionales	15
2.1.1.2.- Internacional.....	16
2.1.2. Antecedentes de la aplicación	17
2.2 Bases Teórico Científicas	18
2.2.1.- Inteligencia de Negocios	18
2.2.1.1. Definiciones	18
2.2.1.2. Características de la Inteligencia de Negocios.....	18
2.2.1.3. Beneficios de la Inteligencia de Negocio para la empresa.....	19
2.2.1.4. OLTP	19
2.2.1.5. OLAP	20
2.2.1.6. DATA WAREHOUSE.....	20
2.2.1.7. DATA MART	21
2.2.1.8. Proceso ETL.....	22
2.2.2.- Minería de datos	23
2.2.2.1. Definiciones	23
2.2.2.2. Métodos de minería de datos	23
2.2.2.3. Algoritmos de Serie de Tiempo.....	24
2.2.2.4.- Comparación de la Minería de Datos con otra solución alternativa – La estadística	25
2.2.3.- Metodologías De Desarrollo	26
2.2.3.1. Metodología de desarrollo de Data Warehouse.....	26
2.2.3.2. Metodología de desarrollo de minería.....	32
2.2.4.- Herramientas de desarrollo.....	35
2.2.4.1. Plataforma SQL	36
2.2.4.2.- Plataforma Pentaho.....	36
III. MATERIALES Y MÉTODOS	37
3.1 Diseño de investigación.....	37
3.2. Metodología	42
3.2.1 Descripción de la Metodología – Ralph Kimball.....	42
3.2.2.- Descripción de la Metodología de CRISP – DM	44

IV. RESULTADOS	46
4.1.-Planificación del proyecto	46
4.2.-Definición de los Requerimientos del Negocio	46
4.2.1. Dimensiones VS Hechos – Matriz Bus	47
4.2.2.- Start Mart	48
4.3. Modelado Dimensional	48
4.3.1.- Análisis de Procesos	49
4.3.1.1- Procesos Ventas	49
4.3.1.2.- Procesos Compras	49
4.3.2.- Dimensiones	50
4.3.2.1.- Dimensión Producto	50
4.3.2.2.- Dimensión Tiempo	51
4.3.2.3.- Dimensión Proveedor	52
4.3.2.4.- Dimensión Cliente	53
4.3.2.5.- Dimensión Empleado	53
4.3.3.- Tabla de Hechos	54
4.4. Diseño y Desarrollo de Presentación de Datos	54
4.5. ETL Diseño y Desarrollo	68
4.6.Diseño de la Arquitectura Técnica	70
4.7. Modelado de la minería	71
4.8. Evaluación de la minería	72
4.9. Despliegue	75
4.10. Mantenimiento y crecimiento	89
V.- DISCUSIÓN	90
VI. CONCLUSIONES	94
VII. REFERENCIAS BIBLIOGRÁFICAS	95
VIII.- ANEXOS	98

ÍNDICE DE TABLAS

Tabla N° 01: Comparación Bill Inmon - Ralph Kimbal.....	32
Tabla N° 02: Operacionalización de las variables.....	38
Tabla N° 03: Técnica de recolección de datos – 40	39
Tabla N°04: Matriz de congruencia.....	40
Tabla N°05: Cronograma de desarrollo.....	46
Tabla N°06: Matriz Bus.....	47
Tabla N° 07: Consideraciones y generales.....	49
Tabla N° 08: Dimensiones.....	50
Tabla N°09: Dimensión Producto.....	51
Tabla N°10: Descripción Producto.....	51
Tabla N° 11: Dimensión Tiempo.....	51
Tabla N°12: Dimensión Producto.....	51
Tabla N° 13: Dimensión Proveedor.....	52
Tabla N° 14: Descripción Proveedor.....	52
Tabla N° 15: Dimensión Cliente.....	53
Tabla N° 16: Descripción Cliente.....	53
Tabla N° 17: Dimensión Empleado.....	54
Tabla N° 18: Lista de Hechos.....	54
Tabla N° 19: Producto Descripción.....	55
Tabla N° 20: Categoría Descripción.....	55
Tabla N° 21: SubCategoría Descripción.....	55
Tabla N° 22: Producto Limpieza de Datos.....	55
Tabla N°23: Producto Fuente de Datos.....	55
Tabla N°24: Categoría Fuente de Datos.....	56
Tabla N°25: SubCategoría Fuente de Datos.....	56
Tabla N°26: DimensiónProducto Destino.....	56
Tabla N° 27: Tiempo Venta Descripción.....	57
Tabla N° 28: Tiempo Compra Descripción.....	57
Tabla N° 29: Tiempo Limpieza de Datos.....	57
Tabla N° 30: Tiempo Venta Fuente de Datos.....	57
Tabla N° 31: Tiempo Compra Fuente de Datos.....	57
Tabla N° 32: DimensionTiempo Tabla Destino.....	59
Tabla N° 33: Proveedor Descripción.....	59
Tabla N°34: Proveedor Limpieza de Datos.....	59
Tabla N°35: Proveedor CliPro Fuentes de Datos.....	59
Tabla N°36: Proveedor Depa Fuentes de Datos.....	59
Tabla N°37: DimensionProveedor Tabla Destino.....	60
Tabla N°38: Cliente Descripción.....	60
Tabla N° 39: Cliente Limpieza de Datos.....	60
Tabla N° 40: Clientes CliPro Fuente de Datos.....	60
Tabla N° 41: DimensionClientes Tabla Destino.....	60
Tabla N° 42: Empleado Descripción.....	61

Tabla N° 43: Empleado Limpieza de Datos.....	61
Tabla N° 44: Empleado Fuente de Datos.....	62
Tabla N° 45: DimensionEmpleado Tabla Destino.....	62
Tabla N° 46: Venta Descripción.....	62
Tabla N° 47: Venta Limpieza de Datos.....	63
Tabla N° 48: Venta Fuente de Datos.....	63
Tabla N° 49: Venta Dventa Fuente de Datos.....	63
Tabla N° 50: HechoVenta Tabla Destino.....	64
Tabla N° 51: Compras Tablas Fuente.....	64
Tabla N° 52: Compras Limpieza de Datos.....	64
Tabla N° 53: Compras Fuentes de Datos.....	66
Tabla N° 54: Compras Dcompras Fuentes de Datos.....	66
Tabla N° 55: HechoCompras Tabla Destino.....	66
Tabla N° 56: MAPE con el Algoritmo MIXED (ARTXP y ARIMA).....	73
Tabla N° 57: MAPE con el Algoritmo ARTXP.....	73
Tabla N° 58: MAPE con el Algoritmo ARIMA.....	74
Tabla N° 59: Mantenimiento y crecimiento.....	89
Tabla N° 60 Tiempo de elaboración de informes solicitados.....	90
Tabla N°61: Escala de medición para indicadores cualitativos.....	91
Tabla N° 62: Nivel de calidad de la información solicitada.....	91
Tabla N° 63: Prueba de hipótesis del segundo indicador.....	91
Tabla N° 64: Nivel de satisfacción de la gerencia sobre la información solicitada.....	92
Tabla N° 65: Prueba de hipótesis del tercer indicador.....	92

ÍNDICE DE FIGURAS

Figura N°01: Metodología Inmon.....	27
Figura N° 02.- Fases de Metodología de Kimball.....	28
Figura N°03: Fases de SEMMA.....	33
Figura N°04.- Fases de la metodología CRISP-DM.....	34
Figura N°05: Comparación CRISP DM – SEMMA.....	35
Figura N° 06: Fases de Metodología de Kimball.....	43
Figura N° 07: Fases de la metodología CRISP.....	45
Figura N° 08: Start Mart Ventas.....	65
Figura N° 09: Start Mart Compras.....	65
Figura N° 10: Modelo dimensional.....	66
Figura N° 11: Flujo de control.....	68
Figura N° 12: Flujo de datos Cliente.....	68
Figura N° 13: Flujo de Datos Empleado.....	69
Figura N° 14 : Flujo de Datos HechoVentas.....	69
Figura N° 15: Cubo OLAP.....	69
Figura N° 16: Arquitectura del Proyecto.....	70
Figura N° 17: Selección de datos de aprendizaje.....	71
Figura N° 18: Proyección de productos.....	72
Figura N° 19: Consulta MDX en Pentaho Data Ingration.....	75
Figura N°20: Dashboard 1.....	76
Figura N°21: Dashboard 2 – Evaluación por Mes.....	76
Figura N°22: Dashboard 2 – Evaluación por Campaña.....	78
Figura N°23: Dashboard 2 – Evaluación por Año.....	79
Figura N°24: Dashboard 3 – Evaluación por Mes.....	80
Figura N°25: Dashboard 3 – Proyección de Ventas.....	81
Figura N°26: Dashboard 3 – Evaluación por Campaña.....	82
Figura N°27: Dashboard 3 – Evaluación por Año.....	83
Figura N°28: Dashboard 4 – Evaluación por Mes.....	85
Figura N°29: Dashboard 4 – Evaluación por Campaña.....	86
Figura N°30: Dashboard 4 – Evaluación por Anio.....	88

RESUMEN

El presente trabajo de investigación tuvo como objetivo mejorar el proceso de toma de decisiones de la empresa “El Ofertón SAC ”. Dentro de los problemas identificados se encontró que los reportes solicitados por la gerencia eran entregados con demoras debido a que el personal encargado de realizar dichos reportes tomaba mucho tiempo en la culminación de los mismos, generando adicionalmente, un costo para la empresa. Los reportes generados presentaban un nivel de calidad de la información bajo, por lo que varias veces los clientes solicitaban productos que no tenían en stock mientras existían otros productos que durante grandes periodos de tiempo no tenían movimientos de ventas, generando que la gerencia tuviera un nivel de satisfacción bajo por la información brindada en los reportes.

La implementación de la Solución de Inteligencia de Negocios basado en el algoritmo de Serie Temporal ha mejorado significativamente el proceso de toma de decisiones en la empresa “El Ofertón SAC”.

La solución redujo en un 96,25% el tiempo promedio en la elaboración de informes solicitados por la gerencia, se incrementó el nivel de calidad de la información solicitada de 1,8667 a 4,600 y se incrementó el nivel de satisfacción sobre la información solicitada de 2,2000 a 4,5333 , ambos basados en una escala de 5 puntos del estilo Likert.

PALABRAS CLAVE: Inteligencia de Negocios, Serie Temporal, Toma de decisiones, Minería de Datos, ETL, Dashboards.

ABSTRACT

This research aims to improve the decision-making process of the company “El Ofertón SAC ”. Among the problems identified it was found that the reports requested by management were delivered with delays due to staff responsible for carrying out these took a long time in the completion thereof, further generating a cost to the company reports. The reports generated showed a level of information quality low, so many times customers had not requested products in stock as there were other products for long periods of time did not have sales movements, causing management to have a level Low satisfaction with the information provided in the reports.

The implementation of the Business Intelligence solution based on the Time Series algorithm has significantly improved the process of decision making in the company “El Ofertón SAC ”.

The solution reduced by 96.25% the average time in preparing reports requested by management, the quality of the information requested from 1.8667 to 4.600 and increased the level of satisfaction increased over the requested information from 2,2000 to 4,5333, both based on a 5-point scale of Likert style.

KEYWORDS: Business Intelligence, Time Series, Decision Making, Data Mining, ETL, Dashboards.

I. INTRODUCCIÓN

Hoy en día la información es uno de los activos potencialmente más valiosos en una empresa, por lo que Méndez (2006) indica que el valor real de esa información depende de cómo es gestionada, del tiempo que se emplea en procesarla y traducirla en lanzamientos de productos o servicios. Es por eso que en la última década, las organizaciones han gastado considerables sumas de dinero y dedicado recursos a construir Sistemas Transaccionales Online (OLTP) y Enterprise Resource Planning (ERP), en castellano, sistemas operacionales, donde las organizaciones pueden operar y realizar sus actividades propias de su negocio.

Sin embargo estos sistemas generan una gran cantidad de datos que son difícilmente aprovechables tal y como se generan, para un análisis y la adecuada toma de decisiones, siendo uno de los usos principales de los sistemas OLAP el almacenar información del día a día, ésta debería ser aplicada en los retos de mañana, ya que los datos del ayer hacen que la acción del mañana sea posible.

La empresa “Distribuciones y Representaciones El Ofertón S.A.C” dedicada a la venta al por mayor y menor de productos de papelería, pasamanería, abarrotos, útiles escolares y de escritorio, no es ajena al cambio en las organizaciones, por lo que tiene implementado un sistema informático llamado “Pecano ERP”, el cual cuenta con módulos especializados para diversas realidades empresariales, y dentro la empresa en estudio, el software permite realizar las siguientes funciones:

Genera el registro de compras y ventas; Permite el registro de todo tipo de productos; Diversos tratamientos del IGV: costo, reintegro tributario, crédito fiscal; Ventas y compras en moneda nacional y extranjera; Registro de ventas por vendedor logueado; Los precios de venta se pueden configurar para registrarse incluido IGV ó sin IGV; Múltiples series para los documentos de compras y ventas; Manejo de todo tipo de comprobante de venta: boletas, facturas, facturas guía, tickets, etc; Impresión de todo tipo de comprobantes de ventas: boletas, facturas, tickets, etc.; Tipos de cambio según SUNAT para la conversión a moneda nacional de ventas realizadas en moneda extranjera; Facilidad para la revisión de la digitación de los talonarios de venta; Verificación de documentos pendientes de digitar; Padrón de proveedores y clientes.

Y como parte del proceso de toma de decisiones en la empresa, el área gerencial solicita diversos reportes para en base a ello poder realizar sus respectivas acciones decisorias. El proceso para la generación de los reportes es el siguiente:

El gerente solicita 2 veces al mes un conjunto de reportes que típicamente son los siguientes:

Cantidad de ventas por cliente; Monto de ventas por cliente; Cantidad de ventas por producto; Monto de ventas por producto; Cantidad de ventas por vendedor; Monto de ventas por vendedor; Cantidad de ventas por departamento; Monto de ventas por departamento; Cantidad de compras por proveedor; Monto de compras por proveedor; todos estos reportes son presentados en función del mes en proceso, y los mismos tipos de reportes en función un comparativo de meses (2 meses anteriores).

Al recibir la solicitud del gerente, el encargado de sistemas realiza el proceso de extracción de los datos del sistema Pecano ERP, mediante la opción de exportar data a Excel que brinda el sistema Pecano ERP, se selecciona el año de la data que se quiere obtener y los datos se exportan en forma de tabla en un archivo de Excel, entonces el encargado de sistema prosigue a crear los reportes solicitados en diversas hojas de Excel.

Dichos reportes permiten la toma de 7 decisiones positivas en base a datos, lo que representa el 37% del total de 19 decisiones, las decisiones que permiten son:

Determinar los clientes que a través de los periodos anteriores generaron ingresos importantes para la empresa para bríndales beneficios para su fidelización; definir los departamentos donde tengo más acogida por los clientes para implementar estrategias de publicidad conociendo sus preferencias por periodos; definir los productos de los periodos anteriores que fueron buenos para comprarlos nuevamente; determinar los productos de los periodos anteriores que no fueron tan buenos y por tanto no debería comprar o debería impulsar su venta; determinar los vendedores que vendieron más y quienes menos en los periodos anteriores para brindarles técnicas personalizadas para su mejora; determinar los productos que actualmente no tienen salida para realizar estrategias de venta; determinar los clientes que actualmente realizan más compras y generan más ingresos para ofrecerle un beneficio.

Para la creación de los 10 reportes del último mes, le toma un tiempo de 5 horas, pero para la culminación de los 20 reportes comparativos (sobre mes anteriores) se toma un tiempo de 11 horas, por lo que el 33% de los reportes son entregados a tiempo y el 67% de los reportes solicitados son entregados con demoras, durante ese tiempo de demoras el área gerencial no puede tomar 5 de sus decisiones mientras estos no están, lo que representa el 21% de decisiones que no pueden ser tomadas cuando se solicitan.

Las 12 decisiones restantes, son tomadas en base a la intuición o experiencia del área gerencial, por lo que dichas decisiones tiene un siempre un margen de error mucho mayor porque no se basa en datos, lo que indica que el 63% de las decisiones tienen una mayor tendencia a poder ser incorrectas por la falta de datos en su elección, básicamente las decisiones que toma mediante esta forma son las siguientes:

Definir hacia dónde dirigir primordialmente la atención actualmente para alcanzar los objetivos: clientes, ventas o utilidades; determinar los tipos de beneficios: descuentos o productos adicionales; determinar los productos en los que brindo ese beneficio; brindar beneficios menores medios o grandes; definir las categorías y subcategorías que tienen mayores montos y cantidades de venta para comprar nuevos productos pertenecientes a ellas; Definir los productos que en este mes que no están teniendo margen de beneficios adecuados para reevaluarlos; definir la cantidad de productos que debo comprar para el siguiente mes; determinar mis descuentos por campañas que surgieron efecto para de esa forma continuar con esa estrategia en la próxima campaña de ese tipo; definir los proveedores con los que puedo realizar algún tipo de alianza; determinar los productos me lograron más ganancia en soles y en porcentaje para ofrecerlos nuevamente; determinar las categorías y subcategorías que generaron más soles y porcentaje de ganancias para comprar nuevos productos pertenecientes a ellas; determinar los productos que no me brindaron tantas ganancias para reducir su compra, o reevaluar su valores.

Y es en base a ello que en la empresa el Oferton SAC presente diversas dificultades como:

Que al menos 3 veces por semana los clientes soliciten productos que no tengan en almacén, teniendo que realizar nuevos pedidos a sus proveedores, cuyos productos demoran en llegar un aproximado de 1 a 4 días, dependiendo de la distancia del proveedor ocasionando así la insatisfacción o pérdida de esos clientes. Asimismo, teniendo en cuenta que el 93% (76) de sus proveedores no realizan los envíos de mercadería de manera gratuita, no es recomendable realizar pedidos con pocas cantidades ya que las empresas de transporte le cobran por los km de distancia incurridos en el envío y no por la cantidad a transportar.

Paradójicamente, el 25% (447) de los productos que están en almacén no presentan ningún tipo de movimiento de venta desde su compra, ocasionando gastos de almacén y un problema para el ordenamiento del mismo.

Tanto el tema de campañas especiales, las cuales se realizan 4 veces al año: campañas escolares (de enero a mayo y de agosto a octubre), campañas de fiestas patrias (junio a julio), campañas navideñas (noviembre a diciembre), como la fidelización y alianzas con los clientes y proveedores respectivamente, no son medidos de la forma como le gustaría al área gerencial.

Las demoras en las generación de reportes no solo genera gastos de tiempo, sino también un costo para la empresa ya que se le tiene que pagar al personal por esas horas incurridas.

De esta manera, el área gerencial indica que el nivel de calidad de la información brindada en los reportes es de 1,8667 y su nivel de satisfacción por la información brindada es de 2,200, ambos basados en una escala de 5 puntos del estilo Likert.

Analizando la situación de la empresa, se plantea el siguiente problema de investigación:

¿De qué manera se podrá mejorar el proceso de toma de decisiones en la empresa “El Ofertón SAC”?

Para la cual planteamos la siguiente hipótesis: “La implementación de una solución de Inteligencia de Negocios basado en el algoritmo de Serie Temporal ayudará a la mejora del proceso de toma de decisiones gerenciales en la empresa El Ofertón SAC”

El objetivo general de la tesis es mejorar el proceso de toma de decisiones de la empresa “El Ofertón SAC”, a través de la construcción de una solución de Inteligencia de Negocios basado en el algoritmo de Serie Temporal

Objetivos específicos

- Reducir el tiempo en la elaboración de informes solicitados.
- Incrementar el nivel de calidad de la información.

- Incrementar el nivel de satisfacción sobre la información solicitada.

La presente tesis se justifica de forma tecnológica ya que se implementa una solución de Inteligencia de Negocios encargada de integrar, procesar y analizar los datos que están almacenados en el sistema de transaccional que posee la empresa, para alinearlos hacia los objetivos de la organización, permitiendo con ello mejoras en el proceso de toma de decisiones.

La tesis se justifica desde el punto de vista científico porque se hace uso del algoritmo de Serie Temporal para la predicción del proceso de ventas, y debido a que es una empresa que cuenta con datos incrementales, el algoritmo de serie temporal es el más adecuado, puesto que a cuanta mayor dimensionalidad y robustez en la data almacenada, más optimas son las predicciones brindadas por esta técnica, contrariamente a alternativas de solución basadas meramente en la estadística. Así mismo, el proyecto servirá como referencia bibliográfica para otras investigaciones.

La justificación económica que tiene este proyecto es que le permite a la empresa reducir costes y mejorar la rentabilidad al mejorar el proceso de toma de decisiones y teniendo en cuenta que el despliegue de la información de la solución de BI es a través de una herramienta de software libre, es esto un gran punto a favor debido a su ventaja principal, el costo económico.

La justificación social de esta tesis es que el desarrollo de este trabajo contribuye a la satisfacción de la comunidad, ya que la aplicación de BI basada en el algoritmo de Serie Temporal, le permite a la empresa saber las necesidades que requiere la demanda de la sociedad para que así la empresa tome las medidas necesarias para poder cubrir a todas ellas y lograr la complacencia de todos sus clientes.

II. MARCO TEÓRICO

2.1 Antecedentes

2.1.1. Antecedentes de Investigación

2.1.1.1.- Locales

- TESIS: “DESARROLLO DE UNA SOLUCIÓN DE BUSINESS INTELLIGENCE PARA MEJORAR EL PROCESO DE TOMA DE DECISIONES ESTRATÉGICAS EN LA GESTIÓN COMERCIAL DE LA EMPRESA TRUCK AND MOTORS DEL PERÚ S.A.C”

La tesis fue desarrollada por Carlos Jhonny Galán Sánchez, de la Universidad Católica Santo Toribio de Mogrovejo - Chiclayo, año 2011, quien realizó una solución de Inteligencia de Negocios para mejorar el proceso de ventas de la empresa comercial Trucks and Motors, donde se concluye, efectivamente, que con la solución de BI realizada se redujo el tiempo promedio de un 92.44% en los que incurre el obtener información estratégica; utilizando para ello la metodología de Data Warehouse para su desarrollo y la herramienta Microsoft SQL Server 2008 como herramienta de aplicación.

Esta investigación tiene relación con mi proyecto de tesis ya que en ella realizó una solución de Inteligencia de Negocios, que le permite a la gerencia obtener reportes e indicadores de gestión para realizar un mejor análisis de la información. Además de que en ella se muestra una serie de comparativos entre las herramientas de Inteligencia de Negocio como son: Pentaho, Oracle y DTS

- TESIS: “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA WAREHOUSE QUE DE SOPORTE A LA TOMA DE DECISIONES PARA LA GESTIÓN DE COMERCIALIZACIÓN DE LA EMPRESA COMERCIAL CYNTHIA”

La tesis fue desarrollada por Manuel Alfonso De la Serna Erazo, de la Universidad Católica Santo Toribio de Mogrovejo, Chiclayo - Chiclayo, año 2010, quien diseñó e implementó un Datawarehouse para mejorar el proceso de toma de decisiones de la empresa comercial Cynthia, para lo cual se necesitó de la creación del cubo de ventas utilizando Analysis Management, a su vez de un solución OLAP para un análisis multidimensional y

de Microsoft Excel 2007 para el diseño de las tablas, gráficos y dimensiones.

Esta investigación tiene relación con mi proyecto de investigación ya que en ella se elaboró el diseño de un Datamart en proceso de expansión a Data Warehouse para la gestión de ventas de una empresa comercial para de esa forma implementar una herramienta de Inteligencia de Negocios

- TESIS: “DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS: APLICANDO LA TÉCNICA DE DATA PROFILING PARA MEJORAR LA CALIDAD DE LOS DATOS EN LA EMPRESA "BIOAGRO" S.R.L”.

La tesis fue desarrollada por Leny Karen Salazar Ahumada, de la Universidad Católica Santo Toribio de Mogrovejo - Chiclayo, año 2013, quien formuló una propuesta coherente y viable para la empresa Biagro S.R.L, que necesitó realizar un crecimiento en sus ventas; y debido a que en esta empresa se almacena la información de manera inadecuada, es decir existe inconsistencia y duplicidad en los datos, no puede realizar un análisis de los mismos. Motivo por el cual se hace uso del ETL para una adecuada implementación de la solución de BI.

Esta investigación tiene relación con mi proyecto de investigación ya que en ella se elaboró un Data Warehouse para la incrementar las ventas de una empresa comercial centrándose mucho en la limpieza de datos para incrementar la calidad de información brindada en la Solución de BI

2.1.1.2.- Nacionales

- TESIS: “ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE FINANZAS DE LA MUNICIPALIDAD METROPOLITANA DE LIMA”

La tesis fue desarrollada por Grace Isabel Núñez Soto, de la Pontificia Universidad Católica del Perú –Lima, año 2010, cuya tesis surge por la necesidad que tienen los usuarios del Área de Finanzas de la Municipalidad Metropolitana de Lima de acceder de manera rápida a información confiable relacionada con su trabajo. El procedimiento a realizar por los usuarios para acceder a esta información, era solicitar al Área de sistemas la exportación de los resultados de la Base de Datos, vaciar estos datos en un archivo en Excel, depurar, ordenar, organizar y clasificar dichos datos para obtener la información que requieren para realizar sus labores. Este procedimiento manual también generaba dependencia del Área de Sistemas, y la información generada

tenía una alta posibilidad de error, así como de gran inversión de tiempo estos procedimientos de información dispersa en archivos de Excel, por lo que se plantea como objetivo de la tesis la Implementación de una Solución de Inteligencia de Negocios para el Área de Finanzas.

Esta investigación tiene relación con mi proyecto de investigación porque nos indica todo el procedimiento y pasos que llevaron que se llevaron a cabo, desde el análisis de la problemática, hasta el desarrollo e implantación de la solución de BI, por lo que se pudo saber los resultados favorables que trabajo dicha implementación en la Municipalidad de Lima.

- TESIS: “MINERÍA DE DATOS APLICADOS A LAS VENTAS CON TARJETA DE CRÉDITO CLÁSICA REALIZADOS EN LAS TIENDAS SAGA FALABELLA EN LA CIUDAD DE LIMA”

La tesis fue desarrollada por Hober Willy Siccha Vega, de la Universidad Tecnológica del Perú–Lima, año 2012, quien a través de las técnicas de minería de datos determinó el comportamiento a futuro y la naturaleza de los datos históricos de ventas con tarjeta de crédito clásica de Saga Falabella en la ciudad de Lima, lo que también permitió la creación de diversas estrategias de mercado antes no posibles sin dicha implementación

Esta investigación tiene relación con mi proyecto de investigación porque se está aplicando la minería de datos para la predicción utilizando la herramienta de Microsoft Analysis Services haciendo una explicación de la utilización de dicha herramienta.

2.1.1.2.- Internacional

- TESIS: INTELIGENCIA DE NEGOCIOS. APLICACIÓN EN LA ADMINISTRACIÓN DEL PRESUPUESTO EN UNA EMPRESA DEL SECTOR PÚBLICO.

La tesis fue desarrollada por Gustavo David Vega Vivas, del Instituto Politécnico Nacional, México D.F, año 2010, quien diseñó e implementó una herramienta de software usando tecnología de negocios, que cuenta con las características de confiabilidad y oportunidad, para el apoyo a la toma de decisiones en los niveles estratégicos de la Secretaría de Comunicaciones y Transporte. Para ello, realizó una identificación de la información y de las fuentes de datos que utilizan las áreas involucradas para la elaboración de informes. Después diseñó una base de datos multidimensional para consolidar esa información y se constituyeron las interfaces de usuario necesarias, que sirven para mostrar los parámetros de control que los usuarios requieran evaluar.

Esta investigación tiene relación con mi proyecto de investigación porque nos indica cuáles son los factores de éxito de un Sistema de Información Ejecutivo, tal como el apoyo, compromiso de la alta dirección, apoyo operativo, tecnología apropiada, definición de los requerimientos adecuados, entre otros. Permitió conocer como mejora la Inteligencia de negocios en el proceso de toma de decisiones en las empresas. La implementación de este BI fue utilizando cubos OLAP, nos mostró las estimaciones de espacio en el disco, en memoria que esta necesita para su funcionamiento. Se permitió conocer los resultados después de la implementación del software

2.1.2. Antecedentes de la aplicación

➤ **CASO DE ÉXITO: BUSINESS INTELLIGENCE (DESARROLLO AD-HOC) – EMPRESA DEL SECTOR PETROLERO**

Este caso, cuyo autor es ICA2 Innovación y Tecnología, año 2011, consiste en el desarrollo tecnológico AD-HOC, basado en una solución Business Intelligence que permitió al cliente adquirir un mejor conocimiento del negocio mediante la gestión de KPI (Key Performance Indicators), accesible a nivel mundial mediante interfaz Web tanto para representantes de la organización como para sus socios de negocios.

La correlación con la investigación es que esta empresa se implementó de una solución de Business Intelligence, generando los reportes vía interfaz para que el gerente sepa cuándo, cuánto, que y a quién comprar para satisfacer las necesidades del mercado.

➤ **CASO DE ÉXITO: BODEGAS CHANDON S.A**

Este caso, cuyo autor es Alejandro Muño, año 2012, consiste en el desarrollo de un set de herramientas de análisis que brindan a Bodengas Chandon la oportunidad de entender mejor la dinámica del mercado de la industria, para tal proceso se realizó la implementación de una herramienta de extracción, transformación de datos para luego cargarlo en el Data Warehouse, pudiendo a partir de ello realizar un análisis multidimensional

La correlación con la investigación es que esta empresa se implementó de una solución de Business Intelligence, se logró plantear escenarios futuros, lo que permitió mejorar la productividad de las ventas en un 12%. Además nos indica todos aquellos factores que fueron la clave del éxito, los cuales son necesarios conocer para realizar una mejor implementación de la solución.

➤ CASO DE ÉXITO: ARCOR

Este caso, cuyo autor es SPPerú, año 2011, consiste en la implementación de soluciones de BI en la empresa Arcor, la principal productora de caramelos, dichas soluciones consolidaron la información de los sistemas en todos los países donde tiene operación, la aplicación de BI abarcó las áreas de ventas, compras, marketing, y administración financiera, la ventaja que lograron con el BI fue la integración con las diversas unidades de negocio y plantas industriales, que están ubicadas en diversos países de Latinoamérica.

La correlación con la investigación es que en esta empresa se implementó de una solución de Business Intelligence, logrando el objetivo base que era la integración de las 40 plantas industriales, y el manejo individual de cada una de ellas, obteniendo mayor eficiencia en la confección de reportes y mejora en el proceso de toma de decisiones

2.2 Bases Teórico Científicas

2.2.1.- Inteligencia de Negocios

2.2.1.1. Definiciones

Curto (2012), define a Inteligencia de Negocios o Business Intelligence como el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.

El Data Warehouse Institute (2013) define que Inteligencia de Negocios, reúne datos, tecnología, análisis y conocimiento humano para optimizar las decisiones de negocios y, finalmente, impulsar el éxito de una empresa. Los programas de Inteligencia de negocio suelen combinar un almacén de datos empresariales y una plataforma de BI o conjunto de herramientas para transformar los datos en información útil.

Al mismo tiempo Vit (2002) indica que la Inteligencia de Negocios es un enfoque para la gestión empresarial que le permite a una organización definir qué información es útil y relevante para la toma de decisiones corporativas. Inteligencia de Negocios es un esquema polifacético que fortalece a las organizaciones para tomar mejores decisiones rápidamente, convertir los datos en información y usar una estrategia inteligente para la gestión empresarial.

2.2.1.2. Características de la Inteligencia de Negocios

Tvrđiková (2009) expresa que para que una herramienta software sea considerada de inteligencia de negocio, debe garantizar las siguientes características:

- Accesibilidad: debe garantizar el acceso de los usuarios a los datos, siendo independiente de la procedencia de los mismos
- Orientada al usuario: buscar la independencia entre los conocimientos técnicos que puedan poseer los distintos usuarios
- Apoyo a la toma de decisiones: debe distinguirse de una simple presentación de la formulación de la información, debe poseer herramientas de análisis que permita seleccionar y manipular solamente aquellos datos de análisis que sean interesantes para el usuario

2.2.1.3. Beneficios de la Inteligencia de Negocio para la empresa

Tvrđiková (2009) manifiesta que los beneficios de la Inteligencia de Negocio son:

- Unificación de la información tratada bajo el mismo sistema gestor, proveniente de clientes, proveedores, ventas, previstas, empleados, futuros mercados, etc.
- Extracción del conocimiento que puede generar beneficios tales como la previsión de ventas y devoluciones, tendencias del mercado.
- Centralización de toda la información de la empresa, la cual nos permitirá compartir dicha información entre empleados.

Curto (2012) agrega otros beneficios que proporciona la Inteligencia de negocios para la organización:

- Crear un círculo virtuoso de la información (los datos se transforman en información que genera un conocimiento que permite tomar mejores decisiones que se traducen en mejores resultados y que generan nuevos datos).
- Permitir una visión única, histórica, tenaz y de calidad de toda la información.
- Poder diferenciar lo relevante de lo superfluo.
- Acceder más rápido a la información.
- Tener mayor agilidad en la toma de las decisiones.

2.2.1.4. OLTP

Ramos (2011), indica que los sistemas OLTP están diseñados para gestionar un gran número de peticiones concurrentes sobre sus bases de datos, y que los usuarios puedan insertar, modificar, borrar y consultar dichos datos. Están enfocados a que cada operación (transacción) trabaje con pequeñas cantidades de filas, y a que ofrezcan una respuesta rápida. Habitualmente utilizan sistemas de bases de datos relacionales para gestionar los datos, y suelen estar altamente normalizados. En ellos es muy importante la integridad de los datos, y deben cumplir las propiedades ACID (Atomicity, Consistency, Isolation, Durability):

- ✓ Atomicidad: una operación, o se realiza por completo o no se realiza, nunca debe quedar a medias.
- ✓ Consistencia: sólo se ejecutan las operaciones que cumplen las reglas de integridad de la base de datos.
- ✓ Aislamiento (Isolation): una operación no puede afectar a otras, dos transacciones sobre los mismos datos son independientes y no generan errores entre sí.
- ✓ Durabilidad: una vez realizada una operación, ésta es persistente y no se puede deshacer.

2.2.1.5. OLAP

Aranceta (2010) explica que los cubos de información o cubos OLAP (On Line Analytic Process) funcionan como los cubos de rompecabezas en los juegos, en el juego se trata de armar los colores y en el data warehouse se trata de organizar los datos por tablas o relaciones; los primeros (el juego) tiene tres dimensiones, los cubos OLAP tiene un número indefinido de dimensiones, razón por la cual también reciben el nombre de hipercubos. A la información se puede acceder mediante “tablas dinámicas” en una hoja de cálculo o a través de programas personalizados. Las tablas dinámicas permiten manipular las vistas (cruces, filtrados, organización, totales) de la información con mucha facilidad.

Por tanto, Ramos (2011) indica que el Procesamiento Analítico en Línea (OLAP) tiene como objetivo agilizar la consulta de grandes volúmenes de información. Para ello utiliza estructuras multidimensionales, cubos OLAP, que contienen datos pre-calculados y agregados. Estos sistemas tienen una velocidad de respuesta muy superior a los sistemas OLTP

Entonces, un cubo OLAP está estructurado en dimensiones, que son las diferentes perspectivas desde las que queremos analizar la información, y en medidas, que son los diferentes hechos con valores concretos que solicita el usuario.

2.2.1.6. DATA WAREHOUSE

Según Curto (2012), un Data Warehouse es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización –independientemente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios–, con las propiedades siguientes: estable, coherente, fiable y con información histórica. Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes). Las bases de datos relacionales son el soporte técnico más comúnmente usado para almacenar las estructuras de estos datos y sus grandes volúmenes.

Ramos (2011) indica que el Data Warehouse se caracteriza por ser:

- ✓ Orientado a temas: los datos están organizados por temas para facilitar el entendimiento por parte de los usuarios, de forma que todos los datos relativos a un mismo elemento de la vida real queden unidos entre sí. Por ejemplo, todos los datos de un cliente pueden estar consolidados en una misma tabla, todos los datos de los productos en otra, y así sucesivamente.
- ✓ Integrado: los datos se deben integrar en una estructura consistente, debiendo eliminarse las inconsistencias existentes entre los diversos sistemas operacionales. La información se estructura en diversos niveles de detalle para adecuarse a las necesidades de consulta de los usuarios. Algunas de las inconsistencias más comunes que nos solemos encontrar son: en nomenclatura, en unidades de medida, en formatos de fechas, múltiples tablas con información similar (por ejemplo, hay varias aplicaciones con tablas de clientes).
- ✓ Histórico (variante en el tiempo): los datos, que pueden ir variando a lo largo del tiempo, deben quedar reflejados de forma que al ser consultados reflejen estos cambios y no se altere la realidad que había en el momento en que se almacenaron, evitando así la problemática que ocurre en los sistemas operacionales, que reflejan solamente el estado de la actividad de negocio presente. Un Data Warehouse debe almacenar los diferentes valores que toma una variable a lo largo del tiempo. Por ejemplo si un cliente ha vivido en tres ciudades diferentes, debe almacenar el periodo que vivió en cada una de ellas y asociar los hechos (ventas, devoluciones, incidencias, etc.) que se produjeron en cada momento a la ciudad en la que vivía cuando se produjeron, y no asociar todos los hechos históricos a la ciudad en la que vive actualmente
- ✓ No volátil: la información de un Data Warehouse, una vez introducida, debe ser de sólo lectura, nunca se modifica ni se elimina, y ha de ser permanente y mantenerse para futuras consultas.

2.2.1.7. DATA MART

Curto (2012) define al Data Mart como subconjunto de los datos del data warehouse cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población de usuarios específica. Al igual que en un data warehouse, los datos están estructurados en modelos de estrella o copo de nieve. Curto indica también que el Data Mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento o área dentro la organización

Así mismo, Kroenke (2003) expresa que el Data Mart es una facilidad parecida a un Data Warehouse, pero con un dominio mucho más pequeño. El Data Mart se puede restringir a un tipo en particular de datos, a una determinada función del negocio, una unidad de negocio específico o a un área geográfica.

Ramos (2011), indica que la diferencia de un Data Mart con respecto a un Data Warehouse es solamente en cuanto al alcance. Mientras que un Data Warehouse es un sistema centralizado con datos globales de la empresa y de todos sus procesos operacionales, un Data Mart es un subconjunto temático de datos, orientado a un proceso o un área de negocio específica. Debe tener una estructura óptima desde todas las perspectivas que afecten a los procesos de dicha área. Es más, según Ralph Kimball, cada Data Mart debe estar orientado a un proceso determinado dentro de la organización, por ejemplo, a pedidos de clientes, a compras, a inventario de almacén, a envío de materiales, etc. Para Ralph Kimball el conjunto de Data Marts forma el Data Warehouse.

2.2.1.8. Proceso ETL

Ramos (2012), explica que un Data Warehouse, o un Data Mart, se cargan periódicamente, y en él se unifica información procedente de múltiples fuentes, creando una base de datos que cumple una serie de características descritas anteriormente. Esto implica que deben existir una serie de procesos que leen los datos de las diferentes fuentes, los transforman y adaptan al modelo que hayamos definido, los depuran y limpian, y los introducen en esta base de datos de destino. Esto es lo que se conoce como procesos ETL, procesos de Extracción, Transformación y Carga (Load).

Es muy importante diseñar un buen proceso ETL, en él se deben reconciliar todos los datos de las diferentes fuentes, realizar los cálculos necesarios, mejorar la calidad de los datos, y por supuesto, adaptarlos al nuevo modelo físico y almacenarlos en él.

Trujillo y Mazón (2011), indican que los procesos ETL (Extraction, Transformation, and Loading, extracción, transformación y carga) son de crucial relevancia en la arquitectura del almacén de datos. Ya que estos procesos son los responsables de extraer los datos de las fuentes de datos transaccionales, realizar las transformaciones necesarias, cargarlos en el almacén de datos una vez hayan sido tratados y realizar las cargas sucesivas de datos durante la vida del almacén de datos.

Por tanto los procesos ETL se pueden dividir en tres etapas:

- ✓ Extracción de los datos.- Uno de los aspectos fundamentales que debemos considerar a la hora de diseñar es si optamos por una actualización completa, que es mucho más simple, o si optamos por una actualización incremental, que es lo más conveniente. Si optamos por la primera alternativa el proceso consiste en borrar los datos y volver a cargarlos, pero si optamos por la segunda, habrá establecer una serie de controles y técnicas.
- ✓ Transformación de los datos.- En los procesos de transformación, es preciso asegurarnos de que los datos sean válidos, de su integridad y de su utilidad, lo que suele incluir realizar cálculos y generar nuevos valores. Los datos deben ser depurados para eliminar inconsistencias, discrepancias y duplicidades. Estas transformaciones suelen conllevar cambios con respecto a la estructura de origen para adaptarla al destino, cambios en el contenido de los valores de origen y creación de nuevos valores en las filas de destino.
- ✓ Carga en el Data Warehouse.- Como último paso debemos realizar el proceso de incorporar los datos al Data Warehouse y/o a los diferentes Data Marts, y a los cubos OLAP. Todo ello según la presentación, formatos definidos y periodicidad de actualización propuesta.

2.2.2.- Minería de datos

2.2.2.1. Definiciones

Pérez (2007) define a la Minería de Datos como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. Explicando que debido a la disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas se ha transformado el análisis de datos regular hacia un análisis de datos orientado técnicas especializadas englobadas bajo el nombre de minería de datos.

De manera similar, Vieira (2009) define a la Minería de Datos como un método automático para descubrir patrones de datos, sin la tendenciosidad y la limitación de un análisis sustentado en la intuición humana. Indica también que la minería de datos comprende un conjunto de técnicas para la “descripción” y “predicción” a partir de grandes masas de datos, haciendo unos de las denominadas Data Warehouse.

2.2.2.2. Métodos de minería de datos

Hernández (2004), menciona los siguientes métodos:

a. Clasificación y Predicción: clasifican datos en clases predefinidas. Son dos formas de análisis de datos que pueden ser usados para crear modelos que permiten describir clases de datos o predecir valores futuros. Existen muchos métodos de clasificación como los árboles de decisión, los métodos estadísticos, las redes neuronales y los conjuntos difusos.

b. Algoritmos de Regresión: se generan funciones predictoras. Se habla de un modelo de regresión cuando la variable de respuesta y las variables explicativas son todas cuantitativas. Es una técnica estadística que construye un modelo para determinados datos analizados y a través de este predecir datos futuros.

c. Reglas de Asociación: búsqueda de relaciones entre variables. Una regla de asociación es un criterio que implica ciertas relaciones de asociación entre distintos objetos de una base de datos, tales como “ocurren juntos” o “uno implica al otro”.

d.- Modelado de dependencias: modelos que expliquen la dependencia entre atributos.

e.- Clustering: búsqueda de conjuntos en los que se agrupan los datos cuando las clases son desconocidas. Identifica grupos de datos que son similares. La similitud se mide mediante funciones de distancia especificadas por los usuarios o por expertos.

f. Series de Tiempo: busca relaciones en datos que transcurren secuencialmente. Representan el conjunto de observaciones hechas con respecto a una variable en períodos de tiempo determinados.

2.2.2.3. Algoritmos de Serie de Tiempo

La página oficial de Microsoft SQL Server define los siguientes algoritmos:

ARTXP.- Algoritmo ARTXP es un modelo de árbol con regresión automática que permite representar los datos periódicos de una serie temporal. Este algoritmo relaciona un número variable de elementos pasados con cada elemento actual que se predice. El nombre, ARTXP, deriva del hecho de que el método de árbol de regresión automática (un algoritmo ART) se aplica a varios estados anteriores desconocidos

ARIMA.-El algoritmo ARIMA significa Modelos Autorregresivos Integrados de Medias Móviles. Permite determinar las dependencias en las observaciones tomadas secuencialmente en el tiempo y puede incorporar impactos aleatorios como parte del modelo.

Los modelos de serie temporal pueden ser estacionarios o no estacionarios. Los modelos estacionarios son aquellos que

reverten a un medio, aunque podrían tener ciclos, mientras que los no estacionarios no tienen un centro de equilibrio y están sujetos a mayores variaciones o cambios introducidos por los impactos o variables externas. Para ello se realiza una diferenciación lo cual hace que una serie temporal se estabilice y se convierta en estacionaria.

2.2.2.4.- Comparación de la Minería de Datos con otra solución alternativa – La Estadística

Son varias las soluciones que están fundamentadas en las estadísticas como tal, por ello realiza una comparación con esta técnica.

Hand (1998) hace un comparativo de las estas dos técnicas, donde indica que la Estadística y Data Mining conducen al mismo objetivo, el de efectuar “modelos” compactos y comprensibles que rindan cuenta de las relaciones establecidas entre la descripción de una situación y un resultado (o un juicio) relacionado con dicha descripción.

También apunta a mejorar la toma de decisiones mediante un conocimiento del entorno. Este entorno lo facilitan los datos almacenados en la compañía, cuantitativos o cualitativos.

Fundamentalmente, la diferencia entre ambas reside en que las técnicas del Data Mining construyen el modelo de manera automática mientras que las técnicas estadísticas necesitan ser manejadas - y orientadas - por un estadístico profesional.

Las técnicas de Data Mining permiten ganar tanto en performance como en manejabilidad e incluso en tiempo de trabajo. La posibilidad de realizar uno mismo sus propios modelos sin necesidad de sub-contratar ni ponerse de acuerdo con un especialista estadístico proporciona una gran libertad a los usuarios profesionales.

Pero es importante aclarar que la estadística se utiliza para validar o para matricular un modelo sugerido y preexistente, no para generarlo.

La data mining aventaja a la estadística en los siguientes supuesto indicados por Hand (1998):

- Las técnicas estadísticas se centran generalmente en técnicas confirmatorias, mientras que las técnicas de data mining son generalmente exploratorias. Así, cuando el problema al que pretendemos dar respuesta es refutar o confirmar una hipótesis, podremos utilizar ambas ciencias (diferentes conclusiones y más robusta la estadística).
- Sin embargo, cuando el objetivo es meramente exploratorio (para concretar un problema o definir cuáles son las variables más interesantes en un sistema de información) surge la necesidad de

delegar parte del conocimiento analítico de la empresa en técnicas de aprendizaje (inteligencia artificial), utilizando data mining.

- A mayor dimensionalidad del problema la data mining ofrece mejores soluciones.
- Cuantas más variables entran en el problema, más difícil resulta encontrar hipótesis de partida interesantes. O, aun cuando pudiera, el tiempo necesario no justificará la inversión.
- Las técnicas de data mining son menos restrictivas que las estadísticas. Una vez encontrado un punto de partida interesante y dispuestos a utilizar algún análisis estadístico en particular, puede suceder que los datos no satisfagan los requerimientos del análisis estadístico. Entonces, las variables deberán ser examinadas para determinar qué tratamiento permite adecuarlas al análisis, no siendo posible o conveniente en todos los casos. Aquí también destaca la data mining, puesto que es menos restrictivo que la estadística y permite ser utilizado con los mínimos supuesto posibles (permite ‘escuchar’ a los datos).
- Cuando los datos de la empresa son muy ‘dinámicos’ las técnicas de data mining inciden sobre la inversión y la actualización del conocimiento de nuestro negocio. Un almacén de datos poco ‘dinámico’ permite que una inversión en un análisis estadístico quede justificada -personal cualificado en estadística, metodología rígida y respuestas a preguntas muy concretas- dado que las conclusiones van a tener un ciclo de vida largo. Sin embargo, en un almacén ‘muy dinámico’ las técnicas de data mining permiten explorar cambios y determinar cuándo una regla de negocio ha cambiado. Permitiendo abordar diferentes cuestiones a corto / medio plazo.

Cabe destacar que ambas técnicas constituyen una sinergia y que no son excluyentes una de la otra. Así pues, data mining y estadística son técnicas complementarias que permiten obtener conocimiento inédito en nuestros almacenes de datos o dar respuestas a cuestiones concretas de negocio.

2.2.3.- Metodologías De Desarrollo

2.2.3.1. Metodología de desarrollo de Data Warehouse

Hoy en día existen muchas metodológicas de diseño y construcción de Data Warehouse (DW). Sin embargo, se imponen entre la mayoría dos metodologías, la de Kimball y la de Inmon.

- **Bill Inmon - Top-Down**

Para Inmon (2010), el diseño de un DWH comienza con la introducción de datos en el mismo. Además Inmon sustenta que el ambiente de origen de los datos y el ambiente de destino deben estar físicamente separados en distintas bases de datos.

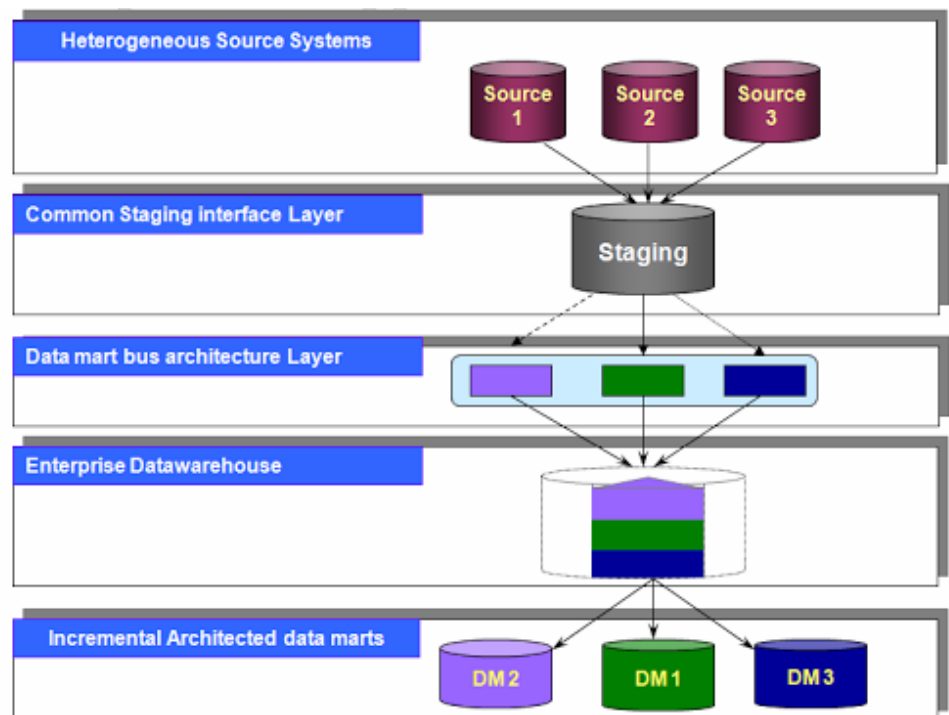
Su enfoque generalmente se asocia al nivel empresarial, es decir que involucra desde un comienzo a toda la organización. En su concepción un Data Mart es solo una parte de un DWH, y estos son dependientes del DWH central.

Bill Inmon ve la necesidad de transferir la información de los diferentes OLTP (sistemas transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis a la Fábrica de Información Corporativa (CIF)

Al tener este enfoque global, es más difícil de desarrollar en un proyecto sencillo (pues estamos intentando abordar el “todo”, a partir del cual luego iremos al “detalle”).

En la Figura N° 01 la metodología Bill Inmon - Top-Down

Figura N°01: Metodología Inmon



Fuente: Inmon, 2010

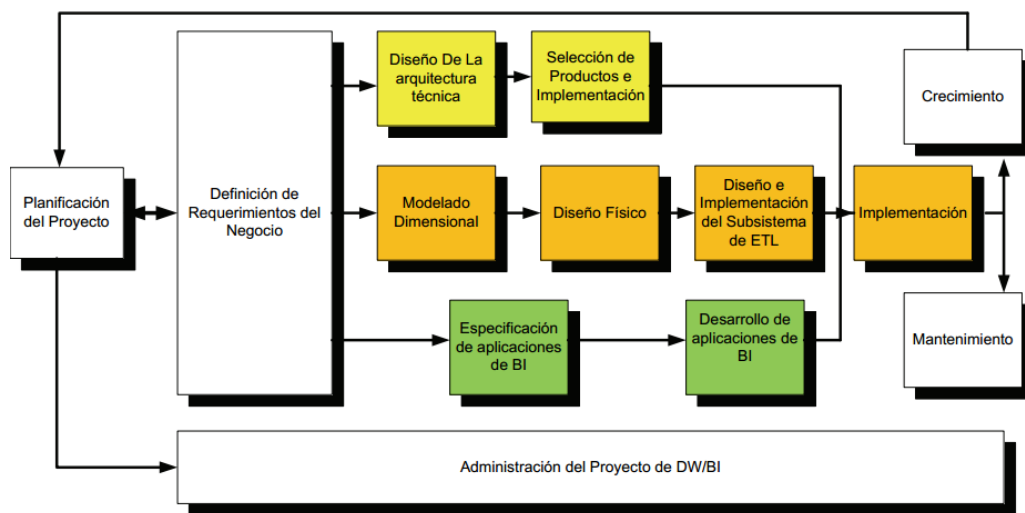
- **Ralph Kimbal - Bottom-Down.**

Ralph Kimball, es reconocido como uno de los padres del concepto de Data Warehouse. La metodología propuesta por

Ralph Kimball se ha convertido en el estándar en área de apoyo a las decisiones empresariales; se enfoca principalmente en el diseño de la base de datos que almacenara la información; este diseño está basado en la creación de tablas hechos, que son las que contienen información numérica o cuantitativa de cada uno de los indicadores que se va a analizar

La construcción de una solución de DW/BI ((Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball nos propone una metodología que nos ayuda a simplificar esa complejidad. Las fases de esta metodología se visualizan en la Figura N°03:

Figura N° 02.- Fases de Metodología de Kimball



Fuente: Kimball et al, 1998

Rivadera, describe la Metodología de Ralph Kimball de la siguiente manera:

Planificación del proyecto

-Identifica todas las tareas asociadas con el proyecto, identifica las áreas involucradas y su alcance

Definición de los Requerimientos del Negocio

-Define los requerimientos expresados por los diferentes usuarios a través de entrevistas, por tanto se debe entender claramente los procesos del negocio para poder trasladar los mismo hacia un diseño.

-Dentro de esta fase se incluye:

- Requerimientos de uso de información
- Tipo de información que las personas necesitan

Tipo de análisis
Requerimiento de datos
Fuente de datos
Calidad de datos y limpieza de datos
Almacenamiento de datos
Carga de datos
Matiz Bus
Que relaciona los procesos de la organización con los objetos de la misma.

Modelado Dimensional

Se comienza con una matriz donde se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle (atributos) dentro de cada concepto del negocio (dimensión), como así también la granularidad de cada indicador (variable o métrica) y las diferentes jerarquías que dan forma al Modelo Dimensional del Negocio

Análisis de los datos de un proceso de negocio para identificar la granularidad de las tablas de hechos, dimensiones y atributos asociados, hechos numéricos.

Mejora el entendimiento y desempeño de consultas al DW

-Tablas de hechos

Contiene métricas derivadas de un proceso de negocio o un evento.

Ventas, contabilidad, logística, etc.

El Modelado Dimensional debe ser estructurado alrededor de un proceso del negocio

Se diseña vistas similares y consistentes de los datos para toda la organización.

La granularidad de la tabla de hechos, debe ser el más atómico posible, esto permite mayor flexibilidad y extensibilidad.

-Tablas de dimensiones

Contiene la descripción de atributos y características asociadas con medidas de eventos tangibles y específicos, tales como clientes, productos, representantes de ventas.

Los atributos de dimensión son usados por limitar, agrupar, o rotular una pregunta.

Una forma de identificar las tablas de dimensiones es que sus atributos son posibles candidatos para ser encabezado en los

informes, tablas pivot, cubos, o cualquier forma de visualización, unidimensional o multidimensional.

-El modelado dimensional puede ser:

-Esquema Estrella, compuesta por una tabla hechos y varias dimensiones.

-Esquema Copo de Nieve, un esquema más complejo que el esquema estrella porque las tablas que describe las dimensiones están normalizadas.

Diseño Físico

Prepara el entorno de base de datos y la seguridad apropiada, algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y datos específicos del ambiente de la base de datos.

Diseño y Desarrollo de Presentación de Datos

Las principales sub etapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga (ETL process).

Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico acordado.

Se definen como procesos de transformación a los procesos para convertir o recodificar los datos fuente a fin poder efectuar la carga efectiva del Modelo Físico.

Los procesos de carga de datos son los procesos requeridos para poblar el Data Warehouse.

ETL Diseño y Desarrollo

Es la fase más importante, las capacidades de sistema ETL son:

- Extracción
- Limpieza y conformidad
- Entrega y administración

Los datos en bruto son extraídos de los sistemas operacionales y transformados en información significativa para el negocio, se verifica la calidad de los datos de entrada, las condiciones de calidad de datos se controlan continuamente.

Diseño de la Arquitectura Técnica

Los ambientes de DW requieren la integración de numerosas tecnologías.

Se debe tener en cuenta tres factores: Los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de DW.

Selección de Productos e Instalación

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura cómo será la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc.

Basado en la arquitectura técnica diseñada.

Evaluación y selección de

- Plataforma de hardware
- DBMS (base de datos)
- Herramienta ETL
- Herramientas de consultas (query tools)
- Herramienta de reportes.

Instalación de productos/componentes/herramientas.

Especificación de Aplicaciones para Usuarios Finales

Aplicaciones que consultan, analizan y presentan información desde el modelo dimensional.

Las aplicaciones BI entregan valor al negocio desde la solución DW/BI.

La meta es entregar capacidades al negocio para soportar y mejorar la toma de decisiones.

- Diseño de Aplicaciones BI.
- Desarrollo de aplicaciones BI.

Desarrollo de Aplicaciones para Usuarios Finales

Se realiza las configuraciones del metadata y construcción de reportes específicos

Despliegue

Representa la analogía de la tecnología, los datos y las aplicaciones de usuarios finales accesibles desde el escritorio del usuario del negocio.

Hay varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de feedback.

Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al DW.

Mantenimiento y crecimiento

DW es un proceso de naturaleza espiral que acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder seguir la evolución de las metas por conseguir. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

Administración del Proyecto DW/BI

Se encuentra el monitoreo del proyecto y la comunicación entre los requerimientos del negocio. Determina las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

- **Comparación de la Metodología de Bill Inmon - Top-Down VS la Metodología Ralph Kimbal - Bottom-Down**

Basado en lo descrito anteriormente, se pueden señalar las siguientes diferencias entre las dos metodologías indicadas en la Tabla N°01

Tabla N° 01: Comparación Bill Inmon - Ralph Kimbal

	INMON	KIMBALL
Enfoque	Top-down	Bottom up
Complejidad del método	Bastante Complejo	Complejidad simple
Orientación de la data	Maneja áreas	Orientado a procesos
Herramientas	Tradicional	Modelación Dimensional
Accesibilidad del usuario final	Bajo	Alto
Objetivos	IT	Usuarios Finales
Objetivos	Ofrecer una buena solución basada en métodos probados	Ofrecer una solución que hace que sea fácil para los usuarios finales para consultar sus datos y tener una respuesta razonable

Fuente: Ramos, 2011

2.2.3.2. Metodología de desarrollo de minería

- **SEMMA**

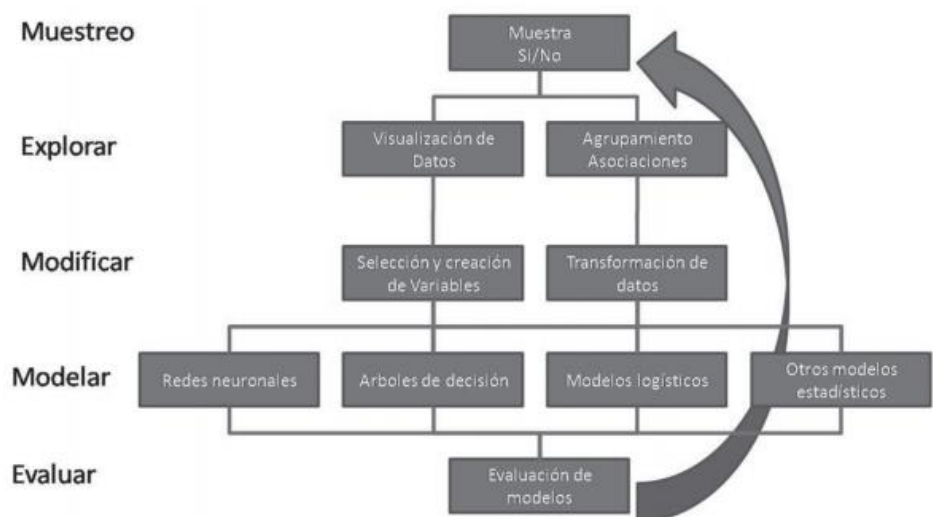
Carmargo and Silva, indican que, el acrónimo SEMMA surge de las iniciales de las palabras Sample (muestra), Explore (explorar),

Modify (modificar), Model (modelar) y Assess (evaluar). Es un proceso que se sigue para realizar minería de datos

Principalmente SEMMA es una organización lógica para el manejo de una herramienta funcional de SAS llamada Enterprise Manager para el manejo de tareas de minería de datos. SEMMA intenta hacer fácil de aplicar la exploración estadística y la visualización de técnicas, seleccionando y transformando las variables predictivas más relevantes, modelándolas para obtener resultados, y finalmente confirmar la precisión del modelo

SEMMA se enfoca en los aspectos de desarrollo del modelo de minería de datos: muestreo, explorar, modificar modelar y evaluar, detallados en la Figura N°03

Figura N°03: Fases de SEMMA



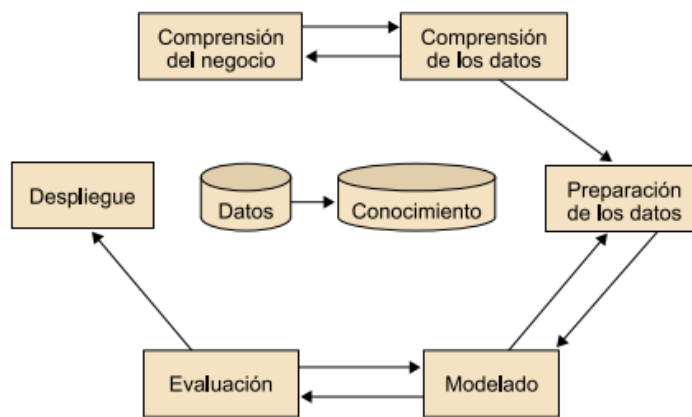
Fuente: Camargo and Silva

El proceso se inicia con la extracción de la población muestral sobre la cual se hará el análisis. Luego se realiza la exploración de los datos para simplificar el problema para optimizar la eficiencia del modelo. Se hacen uso de herramientas estadísticas. La tercera fase consiste en la manipulación de los datos, para definir y obtener el formato que serán introducidos en el modelo. Definido el modelo con el formato adecuado, se procede al análisis y modelado de los datos, para establecer una relación entre las variables explicativas y las variables objeto del estudio. Se emplean técnicas como análisis discriminante. Métodos de agrupamiento y análisis de regresión, entre otras. Finalmente, la última fase consiste en la valoración de los resultados mediante un análisis de bondad del modelo contrastado con métodos estadísticos.

- **CRISP DM (REFF)**

Girones, J., indica que, CRISP-DM (cross industry standard process for data mining) nació en el seno de dos empresas, DaimlerChrysler y SPSS, que en su día fueron pioneras en la aplicación de técnicas data mining en los procesos de negocio. CRISP-DM es una metodología basada en la práctica y experiencia real de analistas DM que han contribuido activamente al desarrollo de la misma. CRISP-DM se organiza en fases, procesos, documentos entregables y actividades. La Figura N°04 esquematiza el ciclo de fases que propone CRISP-DM

Figura N°04.- Fases de la metodología CRISP-DM



Fuente: Gironés, J.

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario. El modelo de CRISP-DM es flexible y se pueden personalizar fácilmente. Por ejemplo, si su organización intenta detectar actividades de blanqueo de dinero, es probable que necesite realizar una criba de grandes cantidades de datos sin un objetivo de modelado específico. En lugar de realizar el modelado, su trabajo se centrará en explorar y visualizar datos para descubrir patrones sospechosos en datos financieros. CRISP-DM permite crear un modelo de minería de datos que se adapte a sus necesidades concretas.

La primera fase corresponde al análisis del problema, que incluye la comprensión de los objetivos y requerimientos del proyecto desde la perspectiva empresarial.

La segunda fase análisis de los datos, comprende la recolección de los datos, identificando la calidad de los datos y estableciendo relaciones para percibir las primeras hipótesis.

Luego se procede a la preparación de los datos, para que estos puedan ser tratados por las técnicas de modelado. Para ello se seleccionan los datos a los que se les va a aplicar la técnica, pasan por un proceso de limpieza y cambios de formato, etc.

La fase de preparación de los datos se relaciona mucho con la fase de modelado, ya que en base a la técnica de modelado seleccionada, los datos necesitan ser procesados.

En la fase de evaluación, se evalúa el modelo, de acuerdo a los criterios de éxito del problema definidos en las primeras fases. Si el modelo generado es válido de acuerdo al cumplimiento de los criterios, se procede a la explotación del mismo.

- **Comparación entre la metodología CRISP – DM Y SEMMA**

Camargo y Silva realizan la comparación entre ambas metodologías vista en la Figura N°

Figura N°05: Comparación CRISP DM - SEMMA

CRISP - DM	SEMMA
Abierta	Cerrada (Abierta en los aspectos generales únicamente)
Funciona en cualquier esquema que aplique minería de datos. Permite que cualquier sistema informático pueda seguir estos pasos	Funciona específicamente en SAS
Implica retroalimentación, es cíclica	Implica retroalimentación, es cíclica
Fases: Entendimiento del negocio, Entendimiento de los datos, Preparación de los datos, Modelado, Evaluado, Despliegue	Fases: Muestreo, Explorar, Modificar, Modelar, Evaluar
Metodología	Secuencia Lógica
Permite aplicar cualquier modelo estadístico	Está obligado a los modelos estadísticos que tenga incorporados la herramienta Enterprise Miner
Enfocada a resultados empresariales	Enfocada a resultados del proceso
Sigue el esquema propuesto en KDD	Sigue el esquema propuesto en KDD
Libre distribución	Distribución en clientes SAS

Fuente: Camargo y Silva

2.2.4.- Herramientas de desarrollo

2.2.4.1. Plataforma SQL

Microsoft SQL Server ofrece un entorno integrado para crear modelos de minería de datos y trabajar con ellos. SQL Server Data Mining permite el acceso a la información necesaria para tomar decisiones inteligentes sobre problemas empresariales complejos.

a) **MicrosoftSQL Server Integration Services (SSIS).**- es una plataforma que permite generar soluciones de integración de datos de alto rendimiento, entre las que se incluyen paquetes de extracción, transformación y carga de datos (ETL) para el almacenamiento de datos.SSIS incluye herramientas gráficas y asistentes para generar y depurar paquetes; tareas para realizar funciones de flujo de datos tales como operaciones de FTP; ejecución de instrucciones SQL y envío de mensajes de correo electrónico; orígenes y destinos de datos para extraer y cargar datos; transformaciones para limpiar, agregar, combinar y copiar datos; un servicio de administración, el servicio Integration Services para administrar la ejecución y almacenamiento de paquetes; e interfaces de programación de aplicaciones (API) para programar el modelo de objetos de Integration Services.

b) **Microsoft SQL Server Analysis Services (SSAS).**- ofrece funciones de procesamiento analítico en línea (OLAP) y minería de datos para aplicaciones de Business Intelligence. Analysis Services admite OLAP y permite diseñar, crear y administrar estructuras multidimensionales que contienen datos agregados desde otros orígenes de datos, como bases de datos relacionales. En el caso de las aplicaciones de minería de datos, Analysis Services permite diseñar, crear y visualizar modelos de minería de datos que se construyen a partir de otros orígenes de datos mediante el uso de una gran variedad de algoritmos de minería de datos estándar del sector.

2.2.4.2.- Plataforma Pentaho

Pentaho es una completa herramienta Business Intelligence de código abierto que permite construir almacenes de datos y aplicaciones ricas y poderosas de BI a una fracción del costo de una solución propietaria. Pentaho ofrece una amplia gama de herramientas orientadas a la integración de información y al análisis inteligente de los datos de una organización. Cuenta con potentes capacidades para la gestión de procesos ETL (extracción, transformación y carga de datos), informes interactivos, análisis multidimensionales de información (OLAP). La web oficial de Pentaho, define las siguientes herramientas:

a) **Pentaho Data Integration (PDI).**- Transforma e integra datos entre sistemas de información existentes y los Data Marts que compondrán el sistema BI. Estas son algunas de sus características:

- Entorno gráfico de desarrollo.

- Uso de tecnologías estándar: Java, XML, JavaScript.
- Fácil de instalar y configurar.
- Multiplataforma.
- Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y Trabajos (colección de transformaciones).
- Herramientas para Big Data.

b) **Community Dashboard Editor (CDE)**.- permite el desarrollo y despliegue de dashboards avanzados. (CDE) nació para simplificar la creación, edición y los procesos de representación, es una herramienta muy poderosa y completa, que combina parte delantera con fuentes de datos y componentes personalizados de una manera perfecta.

III. MATERIALES Y MÉTODOS

3.1 Diseño de investigación

Tipo de investigación

Investigación Aplicada, Cuasi experimental, porque se va a tener control parcial de la variable dependiente, y por la existencia de grupos ya constituidos.

Esquema del diseño:

Diseño de un grupo con medición antes y después, porque se hará una medición previa y posterior de la variable dependiente.

O1 X O2

Donde:

X = Solución de Inteligencia de Negocios (Variable independiente).

O1 = Medición del proceso de toma de decisiones gerenciales (Variable Dependiente) antes de la utilización de la solución de Inteligencia de Negocios (Variable Independiente).

O2 = Medición del proceso de toma de decisiones gerenciales (Variable Dependiente) después de la utilización de la solución de Inteligencia de Negocios (Variable Independiente).

La evaluación de un antes y un después de la implementación de la solución de Inteligencia de Negocios, será para medir el efecto del sistema en el proceso de toma de decisiones gerenciales; pero como no se controlan variables extrañas o del medio (tales como condiciones económicas, políticas, etc.) que pueden afectar a la toma de decisiones, entonces es posible concluir que estas últimas podrían afectar los resultados del experimento.

Hipótesis

La implementación de una solución de Inteligencia de Negocios basado en el algoritmo de serie temporal ayudará a la mejora del proceso de toma de decisiones gerenciales en una empresa comercial.

Indicadores

Tabla N° 02.- Operacionalización de las variables

Objetivo Específico	Indicador	Definición conceptual	Unidad de Medida	Instrumento	Definición operacional
Reducir el tiempo en la elaboración de informes solicitados	Tiempo de elaboración de informes solicitados	Tiempo promedio de elaboración de informes solicitados. El tiempo de elaboración de informes solicitados se mide desde el momento en que el gerente solicita el informe al encargado hasta el momento en el que el informe es finalizado completamente	Horas	Observación y Timer	Σ Horas utilizadas para la elaboración de informes
Incrementar el Nivel de calidad de la información solicitada	Rango de valor del nivel de calidad de la información brindada en los informes	Nivel de calidad de información brindada en los informes , basado en los indicadores de gestión	Rango de valor	Cuestionarios	Rango de valor del nivel de calidad de la información solicitada
Incrementar el Nivel de satisfacción sobre la información solicitada	Rango de valor del nivel de satisfacción del gerente por los informes	Nivel de satisfacción del gerente basada en utilidad o ayuda que le brinde la información de los reportes a la hora de la toma de decisiones	Rango de valor	Cuestionarios	Rango de valor del nivel de satisfacción sobre la información solicitada

Población y muestra

Para el cálculo de la población en estudio y muestreo se ha considerado tener en cuenta la muestra censal ya que este tipo de muestra indica conocer las características poblacionales e incluso el tamaño del universo. En nuestro caso, la población estará constituida por las personas que forman parte del proceso de toma de decisiones en la empresa “El Ofertón” las cuales son 3; al contar con una población pequeña, la muestra será toda la población teniendo como margen de error 0%

Técnica de recolección de datos

Tabla N° 03: Técnica de recolección de datos

Técnica	Justificación	Instrumento	Aplicado a
Entrevista, Observación, Cuestionarios	Permitirá entender el dominio del negocio, personas, objetivos y problemas	Hojas de apuntes Cuestionarios Timer	-Gerente -Encargado de Sistemas -Encargado de finanzas

Técnicas de procesamiento de datos

El procesamiento para el análisis de los datos es de tipo estadístico, en el que se utiliza la herramienta SPSS de IBM, para calcular los niveles obtenidos y analizar las respuestas brindadas por los entrevistados.

Por lo que para el análisis de los datos se realizaron los siguientes pasos:

- 1.- Validación.- Para verificar que los cuestionarios se hayan elaborado adecuadamente
- 2.- Codificación.- Para agrupar o asignar respuestas abiertas individuales
- 3.- Introducción de datos.- Para transformar los datos en un formato electrónico como es la computadora.
- 4.-Representación de los resultados.- Para presentar estadísticos de los resultados.

Respecto a los indicadores cualitativos se ha hecho una evaluación he basado en dos indicadores incluidos en el Modelo de Éxito de DeLone y McLean (2003), los cuales son: Calidad de la Información y Satisfacción del Usuario

En este caso, el instrumento de medición es un cuestionario, defino por Zikmund (2003) como una técnica de investigación en la cual la información es obtenida

por medio de una encuesta y cuyos objetivos son identificar las características de un grupo en particular, medir actitudes y describir patrones de conducta.

Debido a que se cuenta con los recursos informáticos de hardware y software necesarios y se tiene la autorización de las personas involucradas en proceso de toma de decisiones para el llenado de dicho cuestionario tales como son el gerente, el encargado de sistemas y el encargado de finanzas, el cuestionario es el instrumento de medición seleccionado.

Entonces, a través de este cuestionario se pretende conseguir las percepciones de los usuarios acerca de los indicadores planteados para contestar las hipótesis formuladas.

Para la determinación del número de los ítems del cuestionario, no existe una regla al respecto, pero si el cuestionario es demasiado corto se pierde información y si es muy grande puede resultar tedioso y sería llenado de forma incompleta. En la práctica, Im y Grover (2004) recomiendan cuatro o cinco indicadores por cada indicador latente.

La medida usada en el instrumento es una escala ordinal de 5 puntos del estilo Likert (1 = Muy en desacuerdo, 2 = En desacuerdo, 3 = Ni de acuerdo ni en desacuerdo, 4 = De acuerdo y 5 = Muy de acuerdo), ya que de acuerdo a ha Dyba (2000) indica que si usa una escala muy pequeña no se obtiene el poder discriminatorio de quienes responden y una escala muy grande es para individuos con un nivel educativo muy alto por la percepción amplia de las situaciones.

Respecto a la fiabilidad de los ítems se utiliza el alfa de Cronbach, Dyba (2000) indica que un coeficiente superior a 0,7 se considera bueno, la cual asume pesos iguales en los ítems.

La obtención de los ítems se recolecta en base a estudios relativos al tema, vistos en la Tabla N° 04

Tabla N° 04: Matriz de Congruencias

Indicador	Item	Medida de Operacionalizacion (¿Qué se mide?)	Obtención de Items (Investigadores más importantes)
Nivel de calidad de la información solicitada.	1.- El sistema provee reportes de rápida comprensión 2.- El sistema provee reportes con información actualizada 3.- A través de los	<ul style="list-style-type: none"> ➤ Comprensibilidad ➤ Actual ➤ Completa ➤ Relevante ➤ Síntesis 	1.- DeLone y McLean (2003), 2.-Hightower (2002), Goodhue y Thompson (1995) 3.- DeLone y McLean (2003),

	<p>reportes obtengo información suficiente y necesaria para abarcar todos los factores críticos que influyen en la correcta ejecución de mis funciones</p> <p>4.- Los reportes son consistentes, es decir, la información que brinda es congruente con todos los objetivos de negocio</p> <p>5.- Los reportes son concisos, puntuales, selectivos y resumidos adecuadamente</p>		<p>Goodhue y Thompson (1995)</p> <p>4.- DeLone y McLean (2003), Goodhue (1995)</p> <p>5.- DeLone y McLean (2003)</p>
<p>Nivel de satisfacción de la gerencia sobre la información solicitada.</p>	<p>6.- Confío plenamente en los reportes que me brinda el sistema</p> <p>7.- Si me fuera a trabajar a una empresa de la competencia, recomendaría que implanten un sistema igual por la eficiencia de sus reportes</p> <p>8.- Los reportes me permiten adquirir un conocimiento valioso que de otra forma me resultaría muy difícil o imposible de alcanzarla</p>	<ul style="list-style-type: none"> ➤ Confianza ➤ Eficiencia ➤ Disfrute ➤ Satisfacción general 	<p>6.- Mahmood et al (2000)</p> <p>7.- DeLone y McLean (2003)</p> <p>8.- Chen. (2002)</p> <p>9.- DeLone y McLean (2003)</p> <p>10.- DeLone y McLean (2003)</p>

	<p>9.- Cuando estoy usando las funcionalidades de los reportes del sistema mi estado es de disfrute</p> <p>10.- En términos generales estoy satisfecho con el sistema de información</p>		
--	--	--	--

Fuente: Propia

3.2. Metodología

3.2.1 Descripción de la Metodología – Ralph Kimball

La metodología seleccionada para el diseño de la solución de inteligencia de negocio es la metodología de Ralph Kimball

Hoy en día existen muchas metodológicas de diseño y construcción de Data Warehouse (DW). Sin embargo, se imponen entre la mayoría dos metodologías, la de Kimball y la de Inmon.

Desde el punto de vista de la arquitectura, la mayor diferencia entre los dos autores es el sentido de la construcción del DW, esto es comenzado por los Data marts o ascendente (Bottom – up, Kimball) o comenzando con todo el DW desde el principio, o descendente (To Down, Imon).

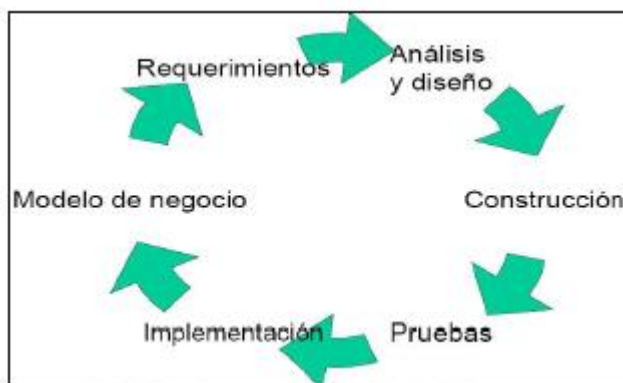
Para nuestro proyecto de investigación, la metodología que más se acopla a la empresa es la de Kimball, porque proporciona un enfoque de menor a mayor, muy versátil, se puede implementar pequeños datamarts en áreas específicas de las mismas (compras, ventas, etc.) con pocos recursos y de poco ir integrándolos en un gran almacén de datos.

Usar el enfoque Bottom – Up, permite implementaciones rápidas, de las cuales se obtengan resultados tempranos y éstos sirvan como retroalimentación para la implementación de los otros Datamarts.

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

Las fases de esta metodología se visualizan en la Figura N°06, y a continuación se explica cada una de ellas.

Figura N° 06- Fases de Metodología de Kimball



Fuente: The Data Warehouse Toolkit, 2002

1.- Modelo de Negocio.- En este punto se realiza el Análisis del Negocio. Es decir, se establece cuál es el objetivo de la realización de la Solución de Inteligencia de Negocios y se identifican las oportunidades de negocio, y todas las áreas asociadas que van a ser atendidas por el proyecto, incluye también la evaluación de factibilidad

2.- Requerimientos.- En este punto se realiza el Análisis de los requerimientos que deben ser plasmados en la solución de BI creada, de tal forma que satisfaga las necesidades de los usuarios. Las actividades relacionadas con este punto son:

- Identificación de usuarios
- Entrevistas
- Descubrir la información
- Listar requerimientos
- Verificación de los requerimientos

3.- Análisis y Diseño.- En este punto se realiza el Análisis y Diseño de la Solución:

Las actividades relacionadas con el Análisis Dimensional son:

- Establecer estándares para la creación del modelo dimensional
- Determinar la granularidad del datamart
- Escoger el esquema con el que se va a trabajar para el modelado del datamart
- Validar el modelo, verificando que se está cumpliendo con los requerimientos planteados.

El desarrollo de estas actividades se encuentra detallado en el “Documento de Análisis”

Las actividades relacionadas al Diseño de Extracción son:

- Identificar fuentes de datos: Sistemas transaccionales, base de datos, archivos, etc.
- Definir los atributos operacionales que se usarán
- Definir como se realizará la transformación de datos para el datamart
- Definir el mapeo entre los datos operacionales y los campos del datamart
- Definir como se obtendrán los campos derivados, agregados, totales, etc.
- Diseñar los procesos de ETL (Extracción, Transformación y Carga)

4.- Construcción.- En este punto se realiza la Construcción de la Solución. Las actividades relacionadas con este punto son:

- Construcción del modelo dimensional
- Construcción y ejecución de procesos ETL
- Explotación de los datos

5.- Pruebas.- En este punto se deben realizar las pruebas de los resultados obtenidos para validar si estos son correctos. Las pruebas a realizar deben validar los siguientes puntos

- Validar que los reportes cumplan con los requerimientos de los usuarios.
- Validar que los reportes muestren información correcta
- Validar que el diseño del datamart sea lo suficientemente flexible para realizar la explotación de los datos desde todas las perspectivas contempladas.

6.- Implantación.- En este punto se realizan las actividades destinadas a la implantación de la Solución de Inteligencia de Negocios en un ambiente de producción real, para que los usuarios accedan a trabajar con la información. Las actividades relacionadas con este punto son:

- Capacitación a los usuarios
- Preparar el ambiente de producción
- Establecer procedimientos de operación y mantenimiento
- Administración de los procesos de ETL y aplicaciones de explotación

3.2.2.- Descripción de la Metodología de CRISP – DM

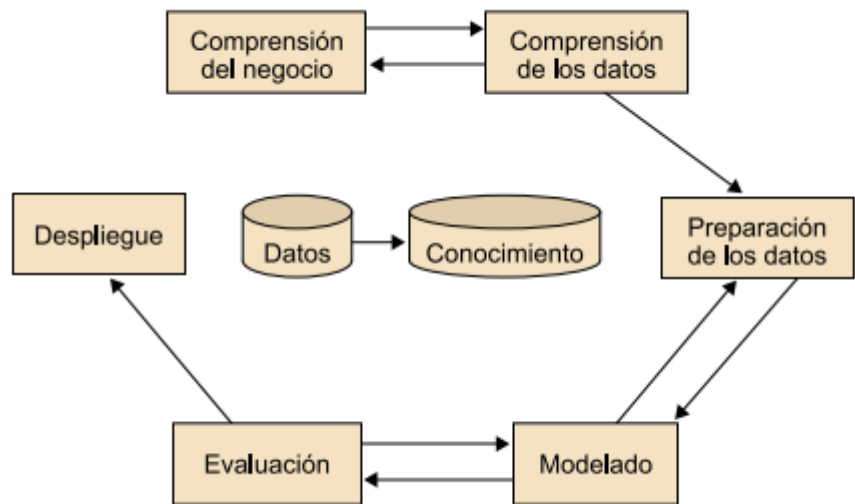
La metodología seleccionada para la aplicación de la minería de datos es CRISP – DM

La metodología CRISP- DM consta de cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos.

El proceso está organizado en seis fases, que recorren toda la vida del proyecto de datamining, desde la definición de los objetivos del negocio, hasta la vigilancia y mantenimiento del modelo que se propone. Cada fase está

estructurada en tareas generales, que se proyectan a tareas más específicas, con resultados concretos.

Figura N°07.- Fases de la metodología CRISP-DM



Fuente: Gironés, J.

La primera fase corresponde al análisis del problema, que incluye la comprensión de los objetivos y requerimientos del proyecto desde la perspectiva empresarial.

La segunda fase análisis de los datos, comprende la recolección de los datos, identificando la calidad de los datos y estableciendo relaciones para percibir las primeras hipótesis.

Luego se procede a la preparación de los datos, para que estos puedan ser tratados por las técnicas de modelado. Para ello se seleccionan los datos a los que se les va a aplicar la técnica, pasan por un proceso de limpieza y cambios de formato, etc.

La fase de preparación de los datos se relaciona mucho con la fase de modelado, ya que en base a la técnica de modelado seleccionada, los datos necesitan ser procesados.

En la fase de evaluación, se evalúa el modelo, de acuerdo a los criterios de éxito del problema definidos en las primeras fases. Si el modelo generado es válido de acuerdo al cumplimiento de los criterios, se procede a la explotación del mismo.

IV. RESULTADOS

4.1.-Planificación del proyecto

Las áreas incluidas en este proyecto son: el área de ventas y el área de compras, donde el objetivo global del proyecto es automatizar los procedimientos que realizan los usuarios para la creación de reportes de dichas áreas de forma precisa y oportuna y a través de estos tomar decisiones críticas en la organización La Tabla N°05 contiene la planificación del proyecto.

Tabla N°05: Cronograma de desarrollo

SEMANA		ENTREGABLE	% AVANCE
1	Jueves 5 Setiembre	Definición de los Requerimientos del Negocio	10%
		Análisis de las fuentes de datos	20%
2	Jueves 12 Setiembre	Modelado Dimensional y Diseño físico	25%
3	Jueves 19 Setiembre	Extracción, transformación y carga	50%
4	Jueves 26 Setiembre		
5	Jueves 3 Octubre		
6	Jueves 17 Octubre	Diseño de la Arquitectura Técnica y selección del producto de instalación	55%
7	Jueves 24 Octubre	Especificaciones de Aplicaciones para Usuarios Finales	60%
8	Jueves 31 Octubre	Diseño e Implementación de la Solucion Olap	70%
9	Jueves 7 Noviembre	Modelado de minería de datos	80%
10	Jueves 14 Noviembre	Evaluación del modelado	85%
		Despliegue	90%
TOTAL			90%

Fuente: Elaboración propia

4.2.-Definición de los Requerimientos del Negocio

Para realizar el proceso de recolección de requerimientos se realizaron entrevistas al personal encargado de finanzas, al encargado de Sistemas y a la Gerencia general, además de ello se recopiló documentos, bases de datos operacionales y reportes estadísticos de la organización.

Los requerimientos nos permiten identificar:

- ❖ Medidas
- ❖ Dimensiones

En base a toda la información brindada a través de las entrevistas y contrastada con la Base de Datos Operacional se definen los siguientes requerimientos en los que la aplicación de Inteligencia de Negocios se fundamenta:

- Cuál es el monto de ingreso generado por las ventas en determinados periodo de tiempo.
- Cuál es la cantidad de ingresos generadas por cada empleados a través de sus ventas en determinado periodo de tiempo.
- Cuál es la cantidad de unidades vendidas por tipo de producto en determinados periodos de tiempo.
- Cuáles son los clientes que nos generan más ingresos.
- Cuáles son las zonas que nos generan más montos ingresos por las ventas.
- Cuáles son los tipos de productos que se más se venden por campaña.
- Cuál es el monto de egresos generados por las compras en determinados periodo de tiempo.
- Cuál es la cantidad de unidades compradas por tipo de producto en determinados periodos de tiempo
- Cuáles son los producto que genera mayor rentabilidad al negocio
- Cuál es la cantidad de unidades compradas por proveedor
- Cuál será la cantidad de unidades vendidas por tipo de productos para los siguientes 3 meses

4.2.1. Dimensiones VS Hechos – Matriz Bus

La matriz bus mostrada en la Tabla N° 06 muestra las dimensionas que utiliza cada tabla de hechos.

Tabla N°06: Matriz Bus

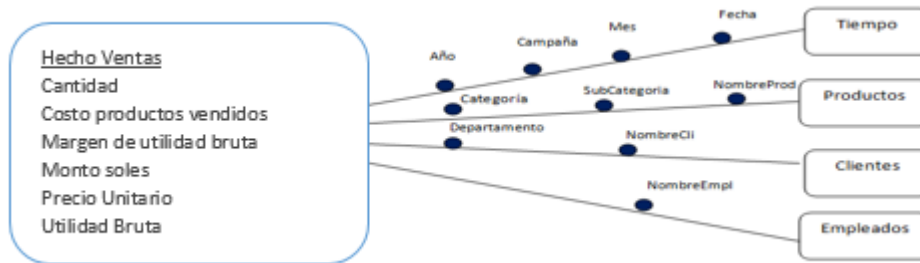
Hechos	Métricas	Dimensiones				
		Tiempo	Clientes	Producto	Proveedores	Empleados
Hecho Ventas	Cantidad Vendida	X	X	X		X
	Costo productos vendidos	X	X	X		X
	Margen de utilidad bruta	X	X	X		X
	Monto soles ventas	X	X	X		X
	Precio Unitario Ventas	X	X	X		X
	Utilidad Bruta	X	X	X		X
	Hecho Compras	Monto soles Compras	X		X	X
Cantidad Unidades Compras		X		X	X	
Precio Unitario Compra		X		X	X	

Fuente: Elaboración propia

4.2.2.- Start Mart

La figura N°08 muestra el hecho venta con sus respectivas dimensiones y los atributos de cada dimensión

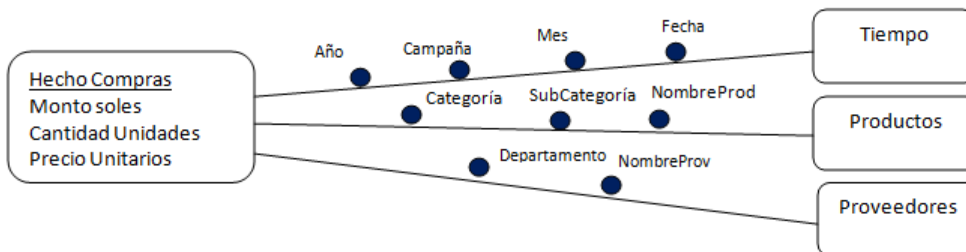
Figura N°08 : Start Mart Ventas



Fuente: Elaboración propia

La figura N°09 muestra el Hecho Compras con sus respectivas dimensiones y los atributos de cada dimensión

Figura N°09: Start Mart Compras



Fuente: Elaboración propia

4.3. Modelado Dimensional

Los modelos dimensionales han sido realizados siguiendo el Esquema Estrella. El Esquema Estrella plantea que se debe tener una tabla o tablas de Hechos o Fact Table, dicha tabla contiene los datos para el análisis y está rodeada de Tablas de Dimensiones que son tablas más pequeñas y con datos necesarios para la tabla hechos.

En la tabla N°07 se presentan las consideraciones de la creación de los modelos dimensionales respecto a los estándares de nombres usados:

Tabla N° 07: Consideraciones generales

Caso	Estándar
Los nombres de las Tablas de Hechos deben empezar con el prefijo	Hecho
Los nombres de las Tablas Dimensiones deben empezar con el prefijo	Dimensión
Los campos de las Tablas Dimensiones que son parte de la llave primaria deben comenzar con el prefijo	Key
Los campos que representan un código de la base de datos origen deben comenzar con el prefijo	Código OLTP
Los campos de las Tablas adicionales que representan un código deben comenzar con el prefijo	Id

Fuente: Elaboración propia

4.3.1.- Análisis de Procesos

A continuación se explica los procesos cubiertos por el datamart , y se muestra el nivel de granularidad de cada uno de ellos .

4.3.1.1- Procesos Ventas

Este tema es creado con el objetivo de analizar las ventas de la organización.

Los objetivos que busca cumplir este tema son:

- Analizar la Cantidad de Unidades de productos vendidos
- Analizar el Costo de los productos vendidos
- Analizar el Margen de utilidad de las ventas
- Analizar el Monto en Soles ingresado por las ventas.
- Analizar el Precio Unitario de los productos vendidos.
- Analizar la Utilidad Bruta de las ventas

Los resultados de este tema se miden por:

- Productos, y por su respectivas agrupaciones de Categorías y Sub Categorías.
- Periodos de tiempo, año y mes.
- Campañas, que son conjuntos de fechas en las que se ofrecen productos de temporada.
- Empleados.
- Clientes.

4.3.1.2.- Procesos Compras

Este tema es creado con el objetivo de analizar las compras de la organización.

Los objetivos que busca cumplir este tema son:

- Analizar el Monto en Soles egresados por las compras.
- Analizar la Cantidad de Unidades de productos comprados.
- Analizar el precio Unitario de los productos comprados.

Los resultados de este tema se miden por:

- Productos, y por su respectivas agrupaciones de Categorías y Sub Categorías.
- Periodos de tiempo, año y mes.
- Campañas, que son conjuntos de fechas en las que se ofrecen productos de temporada.
- Proveedores.

4.3.2.- Dimensiones

En el Tabla N°08 se presentan las dimensiones que son utilizadas en los temas que se presentaron en el punto anterior.

Tabla N° 08: Dimensiones

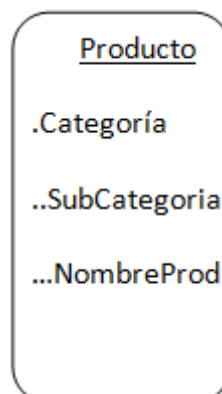
N°	Nombre
1	Dimensión Producto
2	Dimensión Tiempo
3	Dimensión Proveedor
4	Dimensión Cliente
5	Dimensión Empleado

Fuente: Elaboración propia

4.3.2.1.- Dimensión Producto

Esta dimensión contiene los datos de los productos para los cuales se registran las compras y las ventas. Los productos están clasificados por sub categorías y estas a su vez están agrupadas por Categorías. Esta dimensión presenta en su estructura una jerarquía, la cual es detallada en la Tabla N° 09:

Tabla N°09 : Dimensión Producto



Fuente: Elaboración propia

En la Tabla N° 10, se muestra la lista de atributos que componen la Dimensión Producto

Tabla N°10: Descripción producto

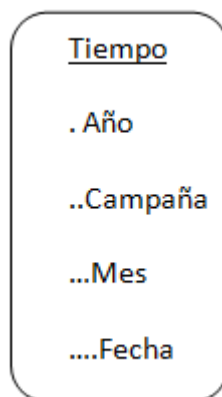
Atributo	Contenido	
	Descripción	Formato
key Producto	Identificador de Producto	Entero correlativo
CodigoOLTP	Código de producto de BD origen	Cadena
nombreProducto	Nombre de producto	Cadena
idCategoria	Código de categoría	Entero
nombreCategoria	Nombre de categoría	Cadena
nombreSubCategoria	Nombre de subcategoría	Cadena
precioUnitario	Precio unitario de producto	Decimal

Fuente: Elaboración propia

4.3.2.2.- Dimensión Tiempo

Esta dimensión contiene los periodos de tiempo por el cual se obtendrá información. Esta dimensión presenta en su estructura una jerarquía, la cual es detallada en el Tabla N 11:

Tabla N° 11: DimensiónTiempo



Fuente: Elaboración propia

En la tabla N°12 se muestra la lista de atributos que componen la Dimensión Productos

Tabla N° 12: Dimensión Producto

Atributo	Contenido	
	Descripción	Formato
Key Tiempo	Identificador de Tiempo	Entero correlativo
Año	Numero de año	Entero
Mes	Numero de mes	Entero
Fecha	Fecha	Fecha
Campaña	Nombre de campaña	Cadena

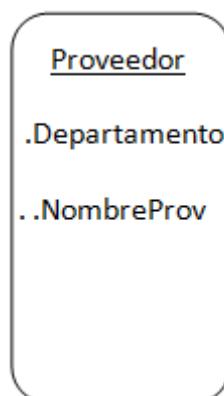
Fuente: Elaboración propia

4.3.2.3.- Dimensión Proveedor

Esta dimensión contiene todos los proveedores para los cuales se registra las compras.

Esta dimensión presenta en su estructura una jerarquía, la cual es detallada en la Tabla N° 13:

Tabla N° 13: Dimensión Proveedor



Fuente: Elaboración propia

En la Tabla N° 14, se muestra la lista de atributos que componen la Dimensión Proveedor

Tabla N° 14: Descripción Proveedor

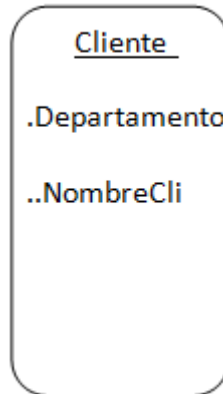
Atributo	Contenido	
	Descripción	Formato
KeyProveedor	Identificador de Proveedor	Entero correlativo
CodigoOLTP	Código de proveedor de BD origen	Entero
NombreProveedor	Numero de proveedor	Cadena
Departamento	Nombre de departamento	Cadena

4.3.2.4.- Dimensión Cliente

Esta dimensión contiene los datos de los clientes de la organización para los cuales se registran las ventas.

Esta dimensión presenta en su estructura una jerarquía, la cual es detallada en la Tabla N°15

Tabla N° 15: Dimensión Cliente



Fuente: Elaboración propia

En la Tabla N° 16 se muestra la lista de atributos que componen la Dimensión Cliente

Tabla N° 16: Descripción Cliente

Atributo	Contenido	
	Descripción	Formato
KeyCliente	Identificador de Proveedor	Entero correlativo
CodigoOLTP	Código de proveedor de BD origen	Entero
NombreCliente	Numero de cliente	Cadena
Departamento	Nombre de departamento	Cadena

Fuente: Elaboración propia

4.3.2.5.- Dimensión Empleado

Esta dimensión contiene los datos de los empleados de la organización quienes registran las ventas

En la Tabla N° 17, se muestra la lista de atributos que componen la Dimensión Proveedor

Tabla N° 17: Dimensión Empleado

Atributo	Contenido	
	Descripción	Formato
KeyEmpleado	Identificador de Empleado	Entero correlativo
CodigoOLTP	Código de Empleado de BD origen	Entero
NombreEmpleado	Nombre de Empleado	Cadena

Fuente: Elaboración propia

4.3.3.- Tabla de Hechos

En la tabla N° 18 se presenta la lista de la tabla de hechos (fact table) que conforman el modelo dimensional estrella

Tabla N° 18: Lista de Hechos

N°	Tema	Nombre
1	Ventas	HechoVentas
2	Compras	HechoCompras

Fuente: Elaboración propia

A continuación se muestra la descripción de cada una de las tablas de hechos:

4.3.3.1.- HechoVentas

Esta tabla de hechos permite mostrar las ventas de la organización por cliente, producto, empleado y por periodos de tiempo

4.3.3.2.- HechoCompras

Esta tabla de hechos permite mostrar las compras de la organización por proveedor, producto y por periodos de tiempo.

4.4. Diseño y Desarrollo de Presentación de Datos

En este punto se presentan los procesos de extracción y transformación de las tablas dimensiones y las tablas de hechos del modelo dimensional creado.

➤ Dimensión Producto

Las figuras N° 19, 20, 21, 22, 23, 24, 25, 26 describen los procesos diseño que permiten la extracción y carga de la Dimensión Producto.

❖ Descripción de Tablas Fuentes

Tabla N° 19: Producto Descripción

Nombre de Tabla	Descripción
Producto	Registro de productos comprados y vendidos por la organización.

Tabla N° 20: Categoría Descripción

Nombre de Tabla	Descripción
Categoría	Registro de categorías de los productos

Tabla N° 21: SubCategoría Descripción

Nombre de Tabla	Descripción
SubCategoría	Registro de SubCategorías de productos, cada subCategoría pertenece a una Categoría.

❖ Estandarización y Limpieza de Datos

Tabla N° 22: Producto Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
keyProducto	PK	Int	Numero autogenerado	No nulo	No tiene
CodigoOLTP		Nvarchar(15)	Texto(000000)	No nulo	No tiene
nombreProducto		Nvarchar(85)	Texto	No nulo	No tiene
idCategoría		Int	Número entero	No nulo	No tiene
nombreCategoría		Nvarchar (100)	Texto	No nulo	No tiene
nombreSubCategoría		Nvarchar(100)	Texto	No nulo	No tiene
precioUnitario		Float	Numero decimal	No nulo	No tiene

❖ Fuentes de Datos

Tabla N°23: Producto Fuente de Datos

Tabla: Producto				
Nombre	Llave	Tipo	Formato	Consideración
idProducto	PK	Nvarchar(15)	Numero autogenerado	No nulo

nomProducto		Varchar(85)	Texto	No nulo
precioCompra		float	Numero decimal	No nulo

Tabla N°24: Categoría Fuente de Datos

Tabla: Categoría				
Nombre	Llave	Tipo	Formato	Consideración
idCategoría	PK	Int	Numero autogenerado	No nulo
nombreCategoría		Varchar(85)	Texto	No nulo

Tabla N°25: SubCategoría Fuente de Datos

Tabla: SubCategoría				
Nombre	Llave	Tipo	Formato	Consideración
idSubCategoría	PK	Int	Numero autogenerado	No nulo
nombreCategoría		Varchar(85)	Texto	No nulo
IdCategoría		Int	Entero	No nulo

❖ Tabla Destino

Tabla N°26: DimensionProducto Destino

Tabla: DimensionProducto		
Campo	Tipo	Mapeo
keyProducto	Int	Identificador autogenerado
Código OLTP	Nvarchar(15)	Producto.idProducto
nombreProducto	Varchar(85)	Producto.nomProducto
idCategoría	Int	Categoría.idCategoría
nombreCategoría	Varchar(85)	Categoría.nombreCategoría
nombreSubCategoría	Varchar(85)	SubCategoría.nombreSubCategoría

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

Extraer los datos de las tablas Producto, Categoría, y SubCategoría

2.- Cargar Datos en el Destino de los Datos

Se recorre cada registro extraído, y se asigna la categoría y sub categoría a cada uno de los registros, resultado de la ejecución del componente script, y luego se

evalúa la existencia de dicho registro en la Tabla Destino, y de tratarse de uno nuevo, éste es cargado a la tabla DimensionProducto

➤ **Dimensión Tiempo**

Las figuras N°, 27, 28, 29, 30, 31, 32 describen los procesos diseño que permiten la extracción y carga de la Dimensión Tiempo

❖ Descripción de Tablas Fuentes

Tabla N° 27: Tiempo Venta Descripción

Nombre de Tabla	Descripción
Venta	Registro de las ventas realizadas por la organización.

Tabla N° 28: Tiempo Compra Descripción

Nombre de Tabla	Descripción
Compra	Registro de las compras realizadas por la organización.

❖ Estandarización y Limpieza de Datos

Tabla N° 29: Tiempo Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
keyTiempo	PK	Int	Número	No nulo	No tiene
Año		Int	Texto(000000)	No nulo	No tiene
Mes		Int	Número	No nulo	No tiene
Fecha		Smalldatetime	Fecha	No nulo	No tiene
Campaña		Varchar(100)	Texto	No nulo	No tiene

❖ Fuentes de Datos

Tabla N° 30: Tiempo Venta Fuente de Datos

Tabla: Venta				
Nombre	Llave	Tipo	Formato	Consideración
Fecha		dateTime	Fecha y hora	No nulo

Tabla N° 31: Tiempo Compra Fuente de Datos

Tabla: Compra				
Nombre	Llave	Tipo	Formato	Consideración
Fecha		dateTime	Fecha y hora	No nulo

❖ Tabla Destino

Tabla N° 32: DimensionTiempo Tabla Destino

Tabla: DimensionTiempo		
Campo	Tipo	Mapeo
keyTiempo	Int	Identificador autogenerado
Año	Int	Venta.fecha / compra.fecha
Mes	smalldatetime	Venta.fecha / compra.fecha
Fecha	Int	Venta.fecha / compra.fecha
Campaña	Varchar (100)	Venta.fecha / compra.fecha

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

Extraer los datos de las tablas Venta y Compra

2.- Cargar Datos en el Destino de los Datos

Se recorre cada registro extraído, y se asigna la campaña a cada uno de los registros, resultado de la ejecución del componente script, y luego se evalúa la existencia de dicho registro en la Tabla Destino, y de tratarse de uno nuevo, éste es cargado a la tabla DimensionTiempo

➤ Dimensión Proveedor

Las figuras N° 33, 34, 35, 36, 37 describe los procesos diseño que permiten la extracción y carga de la Dimensión Proveedor

❖ Descripción de Tablas Fuentes

Tabla N° 33 Proveedor Descripción

Nombre de Tabla	Descripción
CliPro	Registro de clientes y proveedores de la organización
Depa	Registro de Departamentos del Perú implicados en procesos de compra y venta de la organización

❖ Estandarización y Limpieza de Datos

Tabla N°34: Proveedor Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
keyProveedor	PK	Int	Numero autogenerado	No nulo	No tiene

CodigoOLTP		Nvarchar(15)	Texto(000000)	No nulo	No tiene
nombreProveedor		varchar(30)	Texto	No nulo	No tiene
Departamento		Varchar(85)	Texto	No nulo	No tiene

❖ Fuentes de Datos

Tabla N°35: Proveedor CliPro Fuentes de Datos

Tabla: CliPro				
Nombre	Llave	Tipo	Formato	Consideración
idCliPro	PK	Nvarchar(15)	Numero autogenerado	No nulo
nomCliPro		Varchar (85)	Texto	No nulo

Tabla N°36: Proveedor Depa Fuentes de Datos

Tabla: Depa				
Nombre	Llave	Tipo	Formato	Consideración
Nomdepa		Varchar (30)	Texto	No nulo

❖ Tabla Destino

Tabla N°37: DimensionProveedor Tabla Destino

Tabla: DimensionProveedor		
Campo	Tipo	Mapeo
keyCliente	Int	Identificador autogenerado
Código OLTP	Nvarchar(15)	Cliente.idCliPro
nombreCliente	Varchar(85)	Depa.nomdepa
Departamento	Varchar(30)	Cliente.nomCliPro

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

Extraer los datos de las tablas CliPro y Depa, seleccionado solo aquellos registros de CliPro que contengan en su columna Tipo el valor de “P” o “A” .

2.- Cargar Datos en el Destino de los Datos

Se recorre cada registro extraído, luego se evalúa la existencia de dicho registro en la Tabla Destino, y de tratarse de uno nuevo, éste es cargado a la tabla DimensionCliente.

➤ **Dimensión Cliente**

Las figuras N° 38, 39, 40, 41 describe los procesos diseño que permiten la extracción y carga de la Dimensión Cliente

❖ Descripción de Tablas Fuentes

Tabla N°38 Cliente Descripción

Nombre de Tabla	Descripción
CliPro	Registro de clientes y proveedores de la organización
Depa	Registro de Departamentos del Perú implicados en procesos de compra y venta de la organización

❖ Estandarización y Limpieza de Datos

Tabla N° 39: Cliente Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
keyCliente	PK	Int	Numero autogenerado	No nulo	No tiene
CodigoOLTP		Nvarchar(15)	Texto(000000)	No nulo	No tiene
Departamento		varchar(30)	Texto	No nulo	No tiene
nombreCliente		Varchar(85)	Texto	No nulo	No tiene

❖ Fuentes de Datos

Tabla N° 40: Clientes CliPro Fuente de Datos

Tabla: CliPro				
Nombre	Llave	Tipo	Formato	Consideración
idCliPro	PK	Nvarchar(15)	Numero autogenerado	No nulo
nomCliPro		Varchar (85)	Texto	No nulo

Tabla: Depa				
Nombre	Llave	Tipo	Formato	Consideración
Nomdepa		Varchar (30)	Texto	No nulo

❖ Tabla Destino

Tabla N° 41: DimensionClientes Tabla Destino

Tabla: DimensionCliente		
Campo	Tipo	Mapeo
keyCliente	Int	Identificador autogenerado
Código OLTP	Nvarchar(15)	Cliente.idCliPro
Departamento	Varchar(30)	Depa.nomdepa
nombreCliente	Varchar(85)	Cliente.nomCliPro

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

Extraer los datos de las tablas CliPro y Depa, seleccionado solo aquellos registros de CliPro que contengan en su columna Tipo el valor de “C” o “A” .

2.- Cargar Datos en el Destino de los Datos

Se recorre cada registro extraído, luego se evalúa la existencia de dicho registro en la Tabla Destino, y de tratarse de uno nuevo, éste es cargado a la tabla DimensionCliente.

➤ **Dimensión Empleado**

❖ Descripción de Tablas Fuentes

Las figuras N° 42, 43, 44, 45 describe los procesos diseño que permiten la extracción y carga de la Dimensión Empleado

Tabla N° 42: Empleado Descripción

Nombre de Tabla	Descripción
Trabajadores	Registro de los trabajadores de la organización

❖ Estandarización y Limpieza de Datos

Tabla N° 43: Empleado Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
keyEmpleado	PK	Int	Numero autogenerado	No nulo	No tiene
CodigoOLTP		varchar(15)	Texto(000000)	No nulo	No tiene
nomberEmpleado		varchar(85)	Texto	No nulo	No tiene

❖ Fuentes de Datos

Tabla N° 44: Empleado Fuente de Datos

Tabla: Trabajador				
Nombre	Llave	Tipo	Formato	Consideración
idTrabajador	PK	varchar(15)	Numero autogenerado	No nulo
nomTrabajador		Varchar (85)	Texto	No nulo

❖ Tabla Destino

Tabla N° 45: DimensionEmpleado Tabla Destino

Tabla: DimensionEmpleado		
Campo	Tipo	Mapeo
keyEmpleado	Int	Identificador autogenerado
Código OLTP	varchar(15)	Trabajador.idTrabajador
nombreEmpleado	Varchar(85)	Trabajador. nomTrabajador

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

Extraer los datos de la tabla Trabajador

2.- Cargar Datos en el Destino de los Datos

Se recorre cada registro extraído, luego se evalúa la existencia de dicho registro en la Tabla Destino, y de tratarse de uno nuevo, éste es cargado a la tabla DimensionEmpleado.

➤ **Hecho Ventas**

Las figuras N° 46, 47, 48, 49, 50 describe los procesos diseño que permiten la extracción y carga del Hecho Ventas.

❖ Descripción de Tablas Fuentes

Tabla N° 46: Venta Descripción

Nombre de Tabla	Descripción
Venta	Registro de ventas de la organización

Nombre de Tabla	Descripción
dVenta	Registro del detalle de ventas de la organización

❖ Estandarización y Limpieza de Datos

Tabla N° 47: Venta Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
KeyCliente	PK	Int	Número (0000)	No nulo	No tiene
keyTiempo	PK	Int	Número (0000)	No nulo	No tiene
keyProducto	PK	Int	Número (0000)	No nulo	No tiene
KeyEmpleado	PK	Int	Número (00000)	No nulo	No tiene
Numero Venta	PK	Char(100)	Texto	No nulo	No tiene
Monto soles		Float	Decimal	No nulo	No tiene
cantidadUnidades		Float	Decimal	No nulo	No tiene
precioUnitario		Float	Decimal	No nulo	No tiene

❖ Fuentes de Datos

Tabla N° 48: Venta Fuente de Datos

Tabla: venta				
Nombre	Llave	Tipo	Formato	Consideración
nVenta	PK	Char(15)	Numero autogenerado	No nulo
Fecha		dateTime	Fecha y hora	No nulo
IdCliPro		Nvarchar(15)	Texto	No nulo
idVendedor		Varchar(15)	Número	No nulo

Tabla N° 49: Venta Dventa Fuente de Datos

Tabla: Dventa				
Nombre	Llave	Tipo	Formato	Consideración
idProducto	PK	nVarchar(15)	Numero autogenerado	No nulo
nItem	PK	Int	Texto	No nulo
Cantidad		Float	Decimal	No nulo
Precio		Float	Decimal	No nulo
pIgv		Float	Decimal	No nulo

❖ Tabla Destino

Tabla N° 50: HechoVenta Tabla Destino

Tabla: HechoVentas		
Campo	Tipo	Mapeo
KeyCliente	Int	Se obtiene el campo venta.idTrabajador y se hace una búsqueda de este campo en la dimensión dimesionCliente y se obtiene el keyCliente
keyTiempo	Int	Se obtiene el campo venta.idTiempo y se hace una búsqueda de este campo en la dimensión dimesionTiempo y se obtiene el keyTiempo
keyProducto	Int	Se obtiene el campo dventa.idProducto y se hace una búsqueda de este campo en la dimensión dimesionProducto y se obtiene el keyProducto
KeyEmpleado	Int	Se obtiene el campo venta.idTrabajador y se hace una búsqueda de este campo en la dimensión dimesionEmpleado y se obtiene el keyEmpleado
Numero Venta	Char (100)	dVenta.nventa concatenado con dventa.nItem
Monto soles	Float	Se realiza la siguiente operación: dventa.cantidad * dventa.precio) +(dventa.cantidad*dventa.precio)* dventa.pIgv
cantidadUnidades	Foat	dventa.cantidad
PrecioUnitario	Float	dVenta.Precio

Extraer Datos de la Fuente de los Datos

Extraer los datos de la tabla venta y dVenta

2.- Cargar Datos en el Destino de los Datos

Recorre cada registro y guarda los datos en la tabla hechoVentas

➤ **Hecho Compras**

Las figuras N° 51, 52, 53, 54, 55 describe los procesos diseño que permiten la extracción y carga de la Dimensión Producto

❖ Descripción de Tablas Fuentes

Tabla N° 51: Compras Tablas Fuente

Nombre de Tabla	Descripción
Compra	Registro de compras de la organización

Nombre de Tabla	Descripción
dCompra	Registro del detalle de ventas de la organización

❖ Estandarización y Limpieza de Datos

Tabla N° 52: Compras Limpieza de Datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
KeyTiempo	PK	Int	Número (0000)	No nulo	No tiene
keyProducto	PK	Int	Número (0000)	No nulo	No tiene
keyProveedor	PK	Int	Número (0000)	No nulo	No tiene
NumeroCompra	PK	Varchar(100)	Texto	No nulo	No tiene
montoSoles		Float	Decimal	No nulo	No tiene
cantidadUnidades		Float	Decimal	No nulo	No tiene
precioUnitario		Float	Decimal	No nulo	No tiene

❖ Fuentes de Datos

Tabla N° 53: Compras Fuentes de Datos

Tabla: compra				
Nombre	Llave	Tipo	Formato	Consideración
nCompra	PK	varchar(15)	Numero autogenerado	No nulo
idCliPro		Nvarchar(15)	Texto (0000)	No nulo
Fecha		datetime	Fecha	No nulo

Tabla N° 54: Compras Dcompras Fuentes de Datos

Tabla: dCompra				
Nombre	Llave	Tipo	Formato	Consideración
idProducto	PK	nVarchar(15)	Numero autogenerado	No nulo
Cantidad		Float	Decimal	No nulo
Precio		Float	Decimal	No nulo
pIgv		Float	Decimal	No nulo

❖ Tabla Destino

Tabla N° 55: HechoCompras Tabla Destino

Tabla: HechoCompras		
Campo	Tipo	Mapeo
KeyTiempo	Int	Se obtiene el campo compra.fecha y se hace una búsqueda de este campo en la dimensión dimesionTiempo y se obtiene el keyTiempo
keyProducto	Int	Se obtiene el campo dCompra.idProducto y se hace una búsqueda de este campo en la dimensión dimesionProducto y se obtiene el keyProducto
keyProveedor	Int	Se obtiene el campo dCompra.idCliPro y se hace una búsqueda de este campo en la dimensión dimesionProveedor y se obtiene el keyProveedor
numeroCompra	Int	venta.nCompra
montoSoles	Char (100)	Se realiza la siguiente operación: $dCompra.cantidad * dCompra.precio$ $+(dCompra.cantidad*dCompra.precio)*$ $dCompra.pIgv$
cantidadUnidades	Float	dCompra.cantidad
precioUnitario	Foat	dCompra.precio

❖ Proceso

1.- Extraer Datos de la Fuente de los Datos

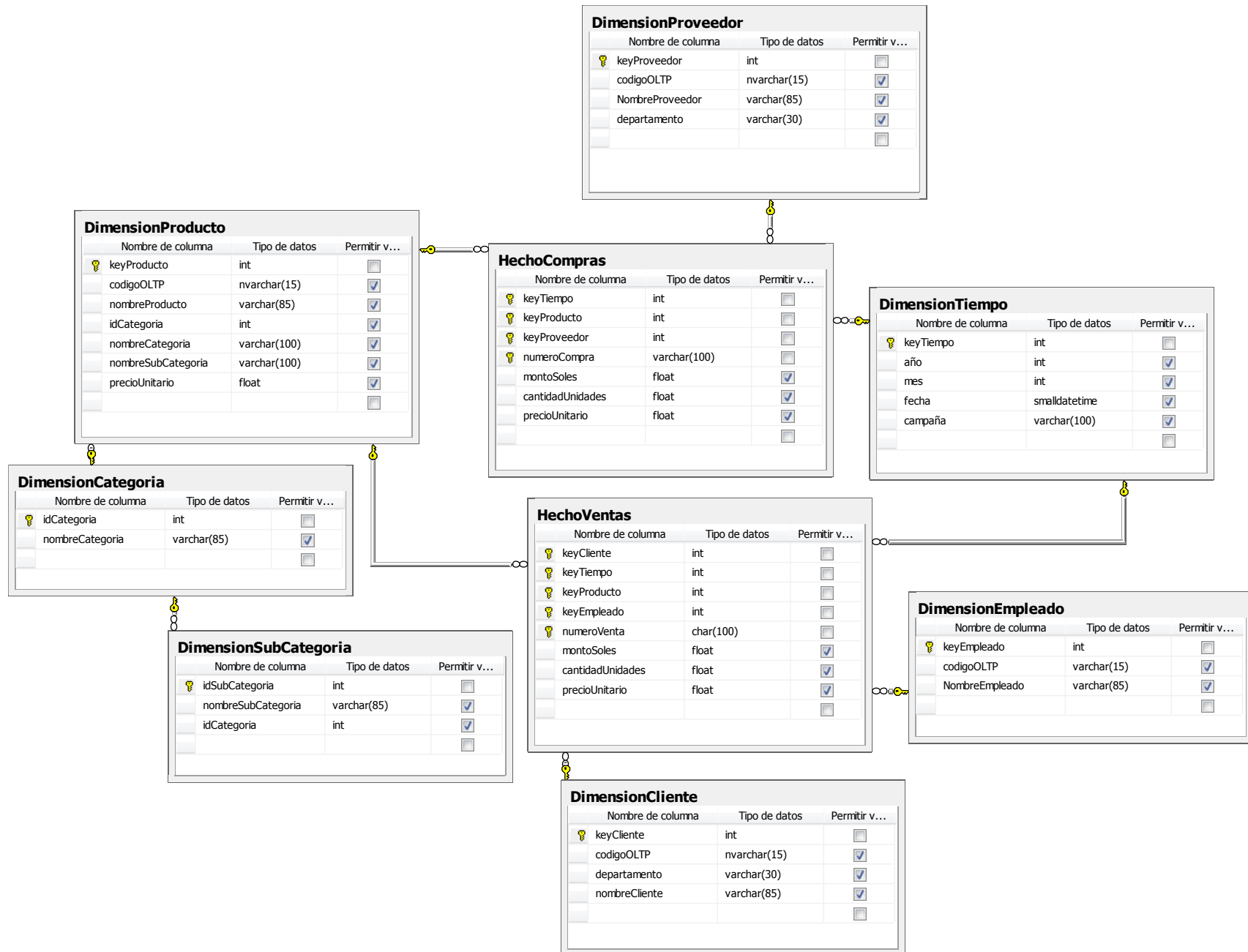
Extraer los datos de la tabla compra y dCompra

2.- Cargar Datos en el Destino de los Datos

Recorre cada registro y guarda los datos en la tabla hechoCompras

En la Figura N° 10 se muestran el modelo dimensional realizado.

Figura N° 10 : Modelo Dimensional



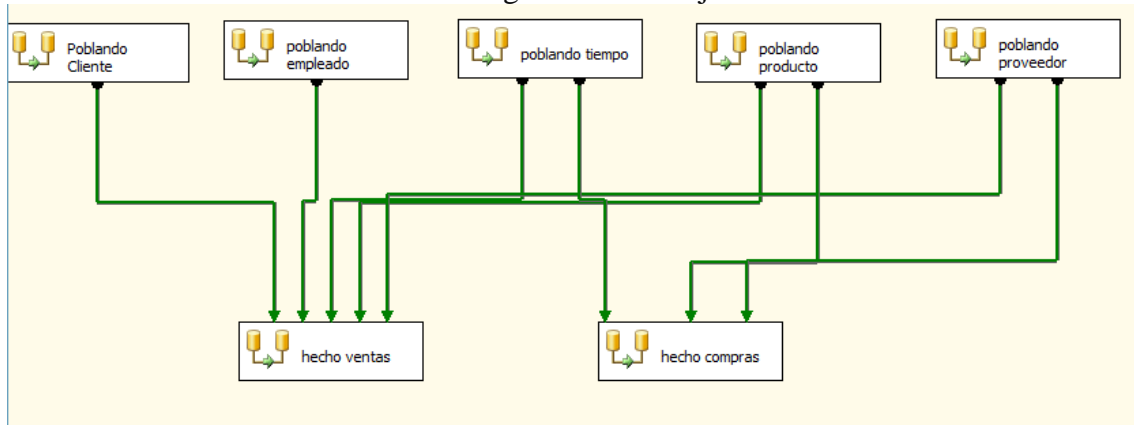
4.5. ETL Diseño y Desarrollo

La base de datos sirve como una estructura intermedia donde se van almacenar los datos que son extraídos de las fuentes de datos y luego serán procesadas por la herramienta OLAP

Se realizó de la siguiente forma:

- La herramienta Microsoft SQL Server Management Studio como una herramienta de administración de base de datos.
- A través de la herramienta se crea la base de datos DMSantaRita5
- Se realiza el proceso de carga de los datos desde la base de datos fuente transaccional hacia la base de datos recientemente creada, en la Figura N° 10 se visualizan el flujo de control de dicho proceso

Figura N° 11: Flujo de control



Para cada flujo de control se realiza un flujo de datos, observados en las figuras N° 12, 13, 14 :

Figura N° 12: Flujo de datos Cliente

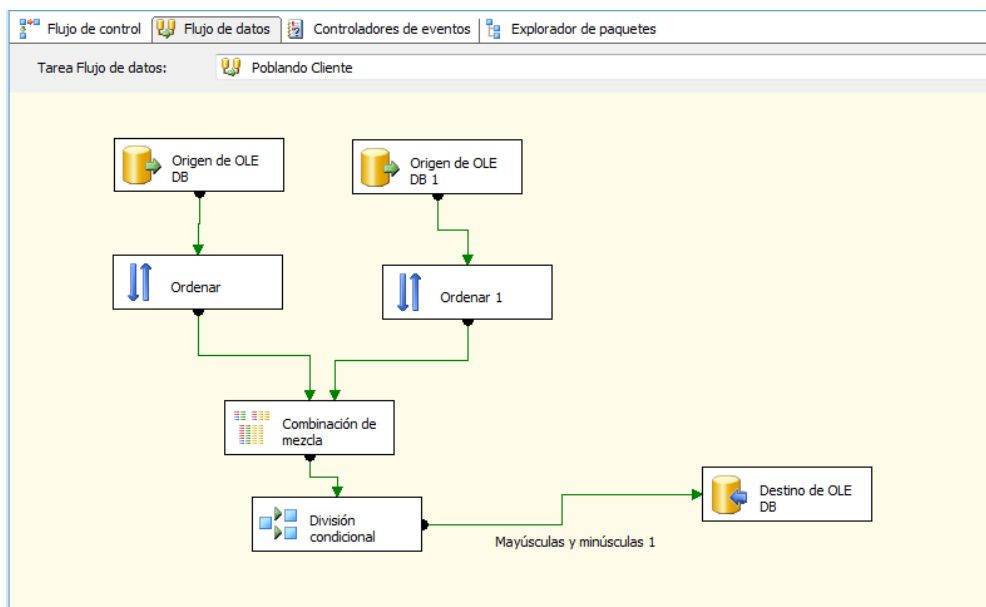


Figura N° 13: Flujo de Datos Empleado

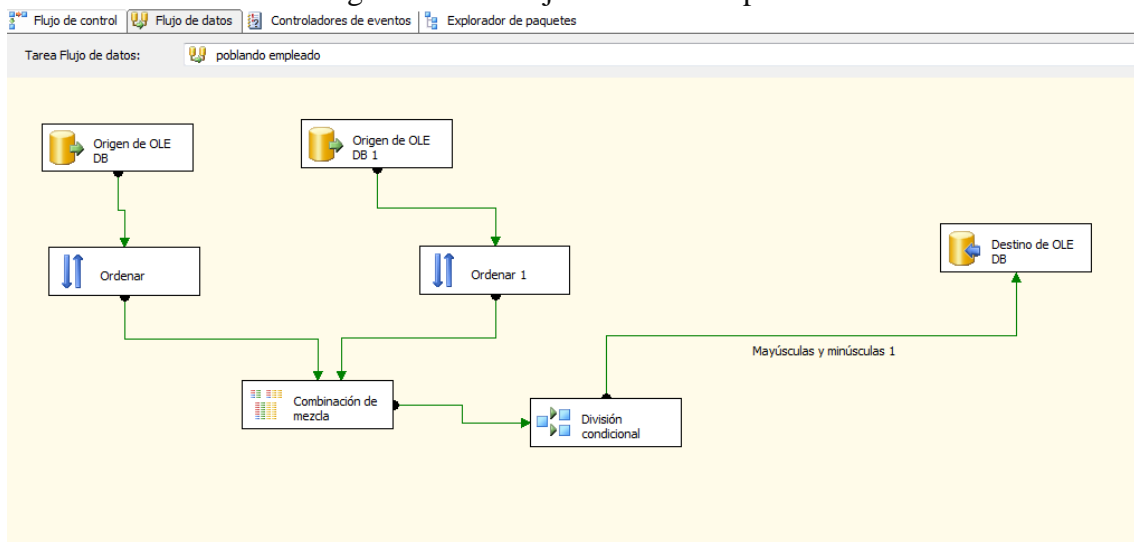


Figura N° 14 : Flujo de Datos HechoVentas

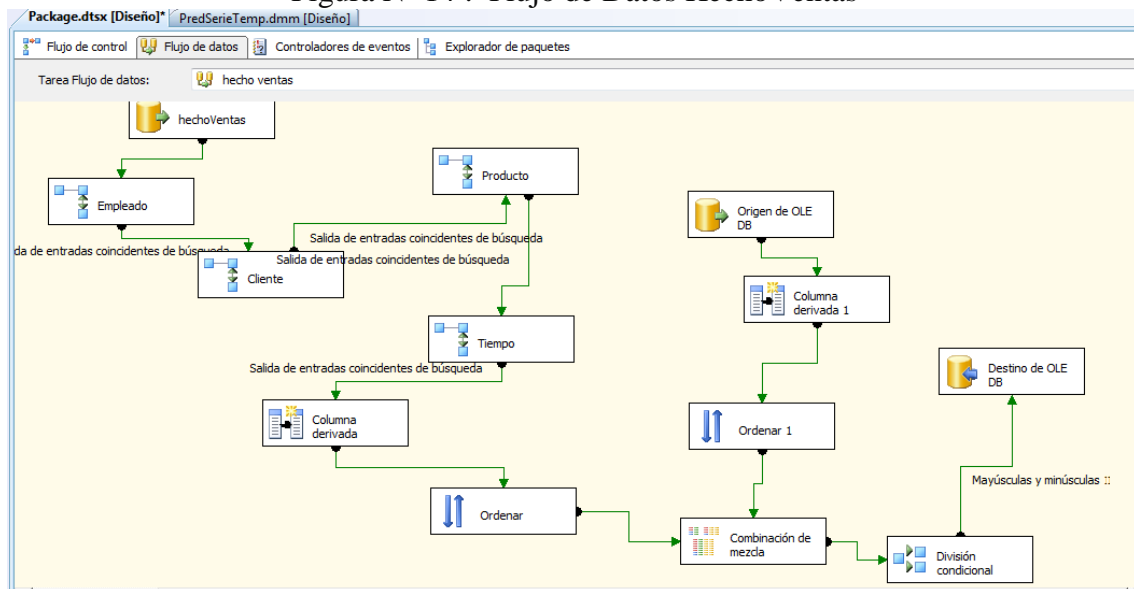
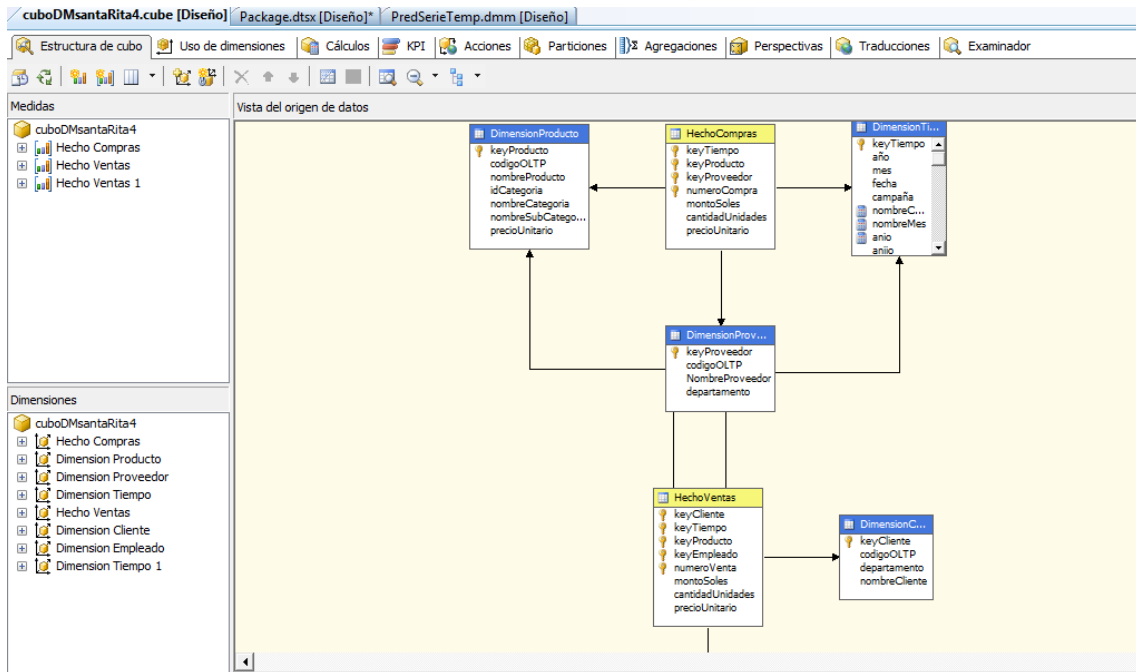


Figura N° 15: Cubo OLAP

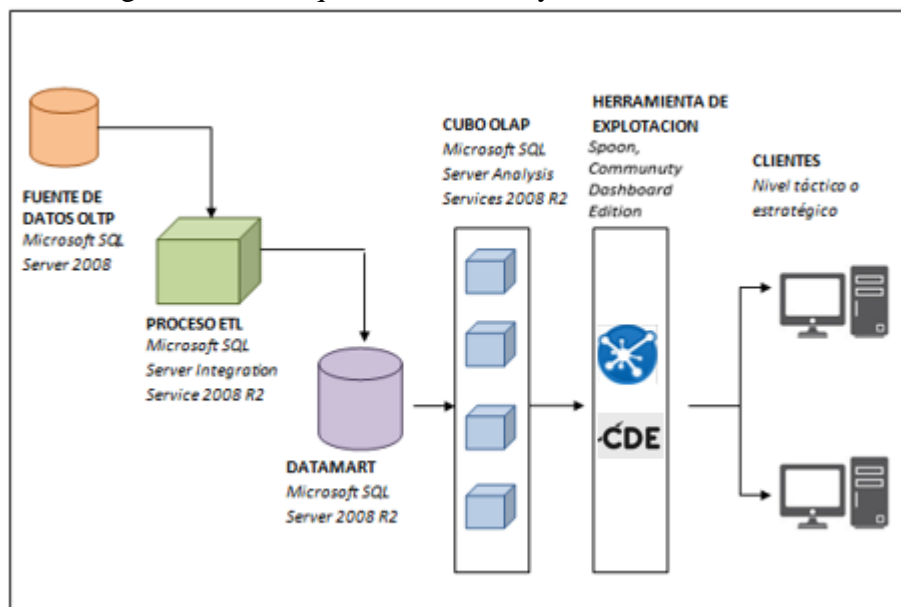


En la Figura N° 15 se visualiza la estructura de cubo creado en para su futura explotación

4.6. Diseño de la Arquitectura Técnica

En la Figura N° 16 se presenta la arquitectura del proyecto. Esta constituye el diseño de una aplicación a alto nivel, es decir, se realiza el estudio de la estructura de la aplicación desde el punto de vista de componentes que interactúan entre sí.

Figura N° 16: Arquitectura del Proyecto



A continuación se detalla cada uno de los componentes:

➤ Fuente de Datos OLTP

La fuente de datos que es usada en el proceso ETL es la base de datos del sistema transaccional de la organización, cuyos datos son soportados por una base de datos de SQL Server 2008

➤ Proceso ETL

El proceso ETL (Extracción, Transformación y Carga) del datamart ha sido desarrollado usando la herramienta Microsoft SQL Server Integration Services 2008 R2. Mediante esta herramienta se creó un paquete ETL para el modelo dimensional creado, para que de esta forma accedan a la fuente de datos mencionada anteriormente

➤ Datamart

A través de la herramienta administradora de base de datos Microsoft SQL Server 2008 R2 se crea una base de datos que contiene las tablas que conforman la estructura del modelo dimensional diseñado. En esta base de datos se depositaran toda la data de la base de datos OLTP cargada a mediante el proceso ETL

➤ Cubo OLAP

El procesamiento del datamart se realiza mediante la herramienta Microsoft SQL Server Analysis Services, donde se definen las jerarquías y nuevas columnas de las dimensiones para la posterior explotación de los datos que contienen.

➤ Proceso Explotación

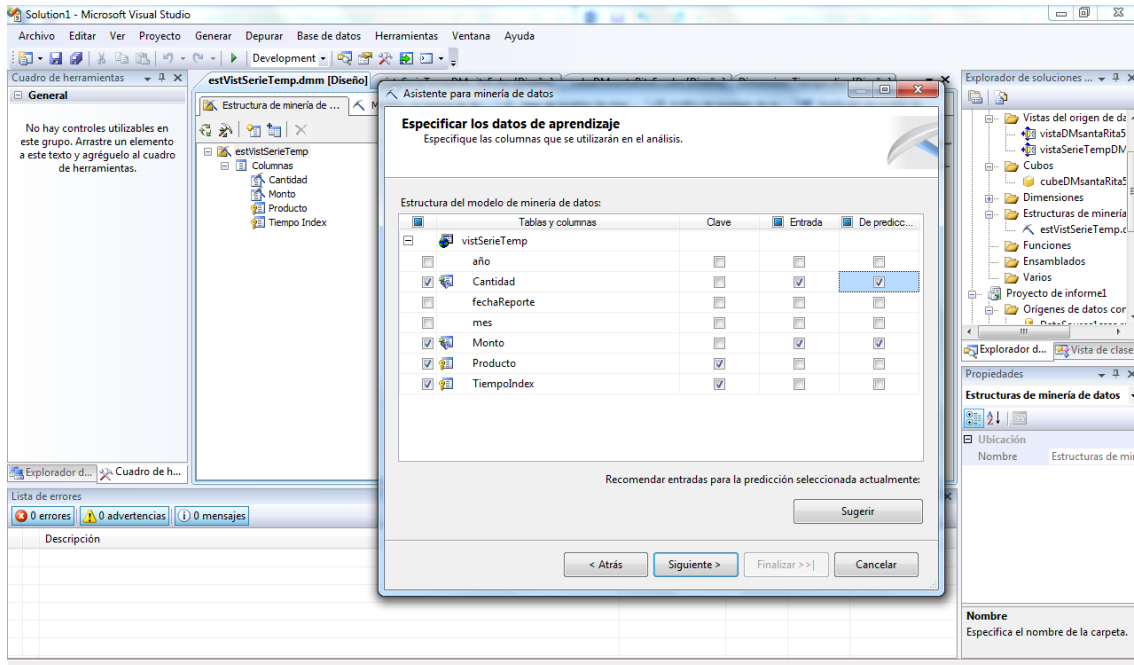
Para la explotación de cubo, primero se utiliza la herramienta Pentaho Data Integration para la conexión al Analysis Services y luego se crean los dashboard y kpi's los mediante la herramienta Community Dashboard Edition.

4.7. Modelado de la minería

Al utilizar una metodología híbrida, en que la integro la metodología de Ralph Kimbal - Bottom-Down (para el DataWarehouse) y la metodología CRIPS (para la minera), en esta parte de la metodología se desarrolla la fase de Modelado de la Minería.

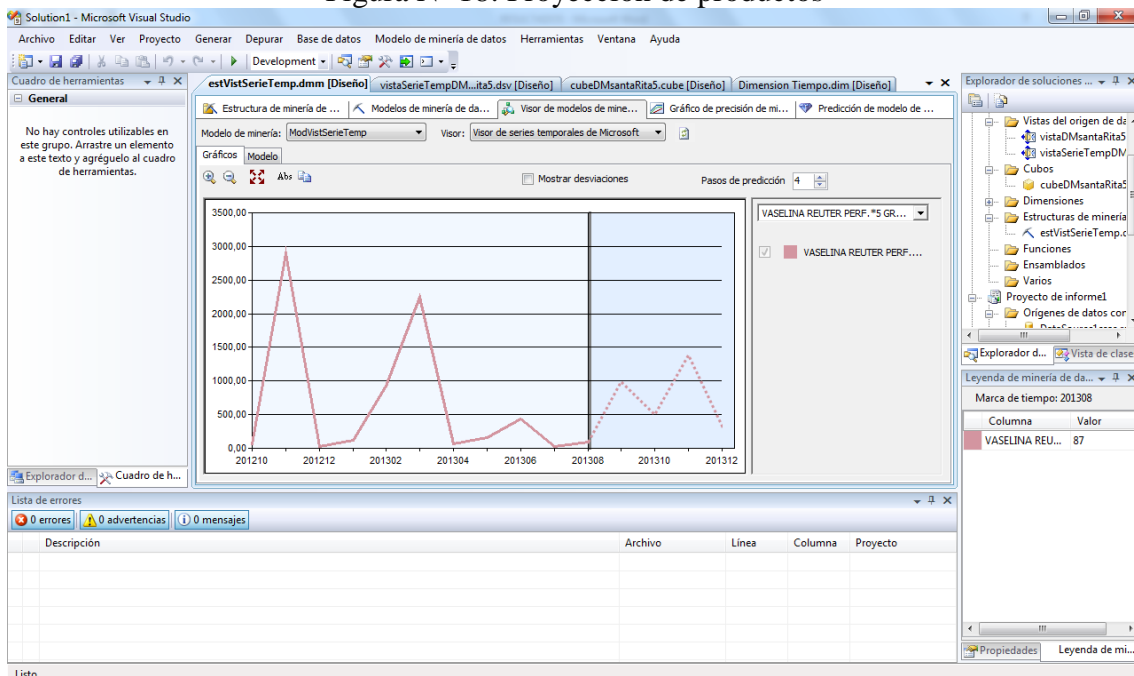
Para poder modelar la minería en Microsoft Analysis Services se crea una vista con los datos preparados, y a partir de ello seleccionar nuestros datos de aprendizaje, tal como se aprecia en la Figura N° 17.

Figura N° 17: Selección de datos de aprendizaje



En figura N°18 se puede observar, tanto el historial de ventas de los productos, como la proyección de los mismo

Figura N° 18: Proyección de productos



4.8. Evaluación de la minería

La mejor forma de comprobar la exactitud de predicción es la de dividir la serie de tiempo en dos conjuntos de datos: uno para la construcción por ejemplo, la formación del modelo y el otro para validarlo. El conjunto de datos de validación será la parte más reciente del conjunto de datos original e idealmente debería abarcar un período de tiempo igual a la

futura línea de tiempo de predicción. Para validar el modelo, los valores pronosticados se comparan con los valores reales. Teniendo en cuenta que después de producirse la validación, el modelo debe ser reconstruido utilizando toda la serie temporal para que las previsiones de futuro se pueden beneficiar de los valores reales más recientes.

El método de MAPE (Porcentaje promedio absoluto de error), mide la exactitud de los valores estimados de la serie de tiempo, las desviaciones medias de las previsiones se calculan como un porcentaje.

Microsoft Analysis Services presenta el algoritmo ARTXP, el algoritmo ARIMA y un algoritmo MIXED (ARTXP y ARIMA)

Para análisis del proyecto se realiza la proyección de las ventas para los 3 últimos meses del valor histórico y se compara con los valores reales que tenemos

En la Tabla N° 56 se presenta el MAPE para nuestras proyecciones utilizando el algoritmo MIXED (ARIMA Y ARTXP), que indica que existe un error de un 20% en el pronóstico.

Tabla N° 56: MAPE con el Algoritmo MIXED (ARTXP y ARIMA)

PRODUCTOS	AÑOMES PROYECTADO	MIX		ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
		REAL	PRONOSTIC			
BORRADOR BESKO * 20 NEGRO - -	201306	1	1	0	0	0%
	201307	1	1	0	0	0%
	201308	1	1	0	0	0%
COLA ARTESCO BARRA STICKY * 8 GR - Unida	201306	79	75,6132803	3,38671968	3,38671968	4%
	201307	126	39,0618834	86,9381166	86,9381166	69%
	201308	66	68,9136575	-2,91365748	2,91365748	4%
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201306	420	423,306317	-3,30631658	3,30631658	1%
	201307	339	281,422791	57,5772085	57,5772085	17%
	201308	540	356,209505	183,790495	183,790495	34%
ALFILER ARTESCO 28MM - - Unida	201306	108	104,2	3,8	3,8	4%
	201307	202	104,2	97,8	97,8	48%
	201308	138	104,2	33,8	33,8	24%
BLOCK AEREO - Unida	201306	12	10,5	1,5	1,5	13%
	201307	12	10,5	1,5	1,5	13%
	201308	9	10,5	-1,5	1,5	17%
CORRECTOR ARTESCO - CJ * 48 DOC - Unida	201306	682	810,428571	-128,428571	128,428571	19%
	201307	917	810,428571	106,571429	106,571429	12%
	201308	639	810,428571	-171,428571	171,428571	27%
BORRADOR LAYCONSA NEGRO LA-30 - -	201306	8	7	1	1	13%
	201307	6	7	-1	1	17%
	201308	6	7	-1	1	17%
CARTULINA ESCOLAR BLANCA - Unida	201306	7950	6936,09856	1013,90144	1013,90144	13%
	201307	6875	7839,95837	-964,958366	964,958366	14%
	201308	3875	3978,01441	-103,014412	103,014412	3%
CD COLGATE MAX. PROTEC. 75 GR - CJ*6 DOC - Unida	201306	18	10,5	7,5	7,5	42%
	201307	12	10,5	1,5	1,5	13%
	201308	12	10,5	1,5	1,5	13%
CLIPS ARTESCO 33 MM METALICOS - CJ*400 DSP*100 - Unida	201306	38	56,6695418	-18,6695418	18,6695418	49%
	201307	78	64,4854582	13,5145418	13,5145418	17%
	201308	216	60,4	155,6	155,6	72%
					MAPE	20%

En la tabla N° 57 se presenta el MAPE para nuestras proyecciones utilizando el algoritmo ARTXP, que indica que existe un error de un 37% en el pronóstico.

Tabla N° 57: MAPE con el Algoritmo ARTXP

PRODUCTOS	AÑOMES PROYECTADO	ARTXP		ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
		R	P			
BORRADOR BESKO * 20 NEGRO - -	201306	1	1	0	0	0%
	201307	1	1	0	0	0%
	201308	1	1	0	0	0%
COLA ARTESCO BARRA STICKY * 8 GR - Unida	201306	79	55,5555556	23,4444444	23,4444444	30%
	201307	126	55,5555556	70,4444444	70,4444444	56%
	201308	66	55,5555556	10,4444444	10,4444444	16%
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201306	420	355,777778	64,2222222	64,2222222	15%
	201307	339	355,777778	-16,7777778	16,7777778	5%
	201308	540	355,777778	184,2222222	184,2222222	34%
ALFILER ARTESCO 28MM - - Unida	201306	108	104,2	3,8	3,8	4%
	201307	202	104,2	97,8	97,8	48%
	201308	138	104,2	33,8	33,8	24%
BLOCK AEREO - Unida	201306	12	10,5	1,5	1,5	13%
	201307	12	10,5	1,5	1,5	13%
	201308	9	10,5	-1,5	1,5	17%
CORRECTOR ARTESCO - CJ * 48 DOC - Unida	201306	682	810,428571	-128,428571	128,428571	19%
	201307	917	810,428571	106,571429	106,571429	12%
	201308	639	810,428571	-171,428571	171,428571	27%
BORRADOR LAYCONSA NEGRO LA-30 - -	201306	8	7	1	1	13%
	201307	6	7	-1	1	17%
	201308	6	7	-1	1	17%
CARTULINA ESCOLAR BLANCA - Unida	201306	7950	7288,88889	661,111111	661,111111	8%
	201307	6875	7288,88889	-413,888889	413,888889	6%
	201308	3875	7288,88889	-3413,88889	3413,88889	88%
CD COLGATE MAX. PROTEC. 75 GR - CJ*6 DOC - Unida	201306	18	10,5	7,5	7,5	42%
	201307	12	10,5	1,5	1,5	13%
	201308	12	10,5	1,5	1,5	13%
CLIPS ARTESCO 33 MM METALICOS - CJ*400 DSP*100 - Unida	201306	38	196,4	-158,4	158,4	417%
	201307	78	196,4	-118,4	118,4	152%
	201308	216	196,4	19,6	19,6	9%
					MAPE	37%

En la tabla N° 58 se presenta el MAPE para nuestras proyecciones utilizando el algoritmo ARIMA, que indica que existe un error de un 39% en el pronóstico.

Tabla N° 58: MAPE con el Algoritmo ARIMA

PRODUCTOS	AÑOMES PROYECTADO	ARTXP		ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
		R	P			
BORRADOR BESKO * 20 NEGRO - -	201306	1	1	0	0	0%
	201307	1	1	0	0	0%
	201308	1	1	0	0	0%
COLA ARTESCO BARRA STICKY * 8 GR - Unida	201306	79	55,5555556	23,4444444	23,4444444	30%
	201307	126	55,5555556	70,4444444	70,4444444	56%
	201308	66	55,5555556	10,4444444	10,4444444	16%
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201306	420	490,834855	-70,8348554	70,8348554	17%
	201307	339	219,989745	119,010255	119,010255	35%
	201308	540	356,508733	183,491267	183,491267	34%
ALFILER ARTESCO 28MM - - Unida	201306	108	104,2	3,8	3,8	4%
	201307	202	104,2	97,8	97,8	48%
	201308	138	104,2	33,8	33,8	24%
BLOCK AEREO - Unida	201306	12	10,5	1,5	1,5	13%
	201307	12	10,5	1,5	1,5	13%
	201308	9	10,5	-1,5	1,5	17%
CORRECTOR ARTESCO - CJ * 48 DOC - Unida	201306	682	810,428571	-128,428571	128,428571	19%
	201307	917	810,428571	106,571429	106,571429	12%
	201308	639	810,428571	-171,428571	171,428571	27%
BORRADOR LAYCONSA NEGRO LA-30 - -	201306	8	7	1	1	13%
	201307	6	7	-1	1	17%
	201308	6	7	-1	1	17%
CARTULINA ESCOLAR BLANCA - Unida	201306	7950	6583,30823	1366,69177	1366,69177	17%
	201307	6875	8295,25906	-1420,25906	1420,25906	21%
	201308	3875	6762,54914	-2887,54914	2887,54914	75%
CD COLGATE MAX. PROTEC. 75 GR - CJ*6 DOC - Unida	201306	18	10,5	7,5	7,5	42%
	201307	12	10,5	1,5	1,5	13%
	201308	12	10,5	1,5	1,5	13%
CLIPS ARTESCO 33 MM METALICOS - CJ*400 DSP*100 - Unida	201306	38	191,617243	-153,617243	153,617243	404%
	201307	78	201,182757	-123,182757	123,182757	158%
	201308	216	196,4	19,6	19,6	9%
					MAPE	39%

El algoritmo seleccionado para las proyecciones es el algoritmo MIXED, ya que basado nuestros resultados, el algoritmo MIXED presenta un porcentaje de error (MAPE) menor con respecto a los demás.

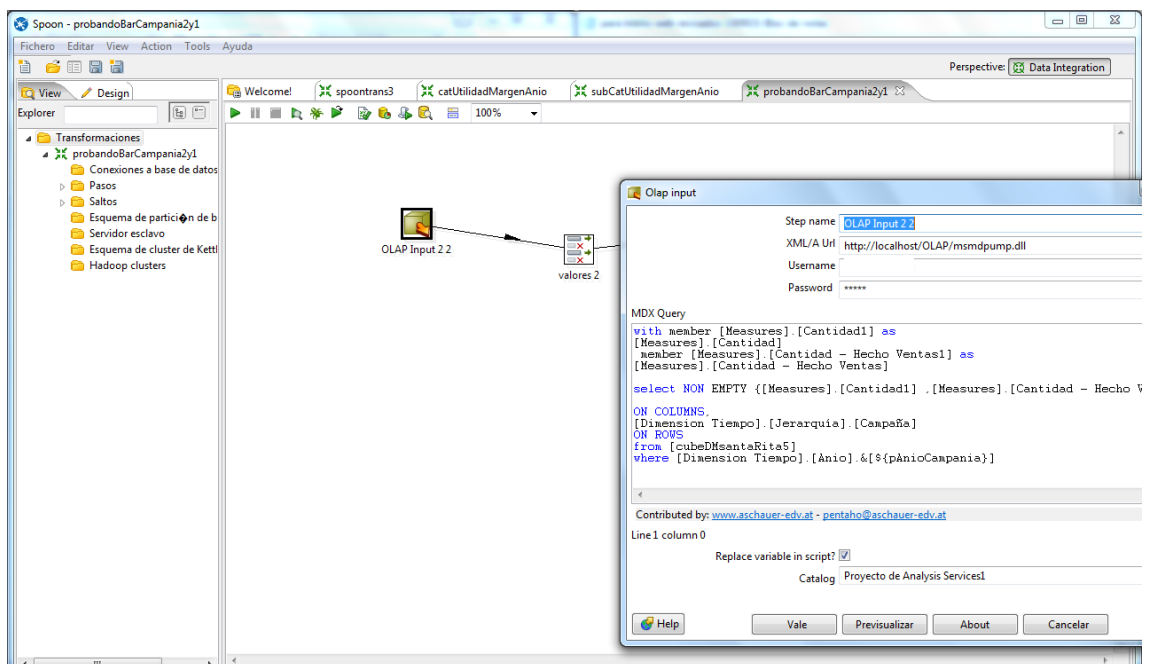
4.9. Despliegue

Teniendo el cubo y la minería implementada en el proyecto de Analysis Services, se procede a la creación de los dashboard y kpi del negocio.

Se utiliza la herramienta Pentaho Data Integration para la conexión con el Analysis Services, y se hace mediante la configuración acceso HTTP a Analysis Services en Internet Information Services.

En esta herramienta se realizan las consultas MDX al cubo, tal como se ve en la Figura N° 19

Figura N° 19: Consulta MDX es Pentaho Data Integration



Y se utilizan las consultas MDX para creación de los dashboard y KPI solicitados por la organización que a continuación se explican:

En el primer Dashboard, que se visualiza en la Figura N°75, contempla los 3 principales indicadores clave de desempeño:

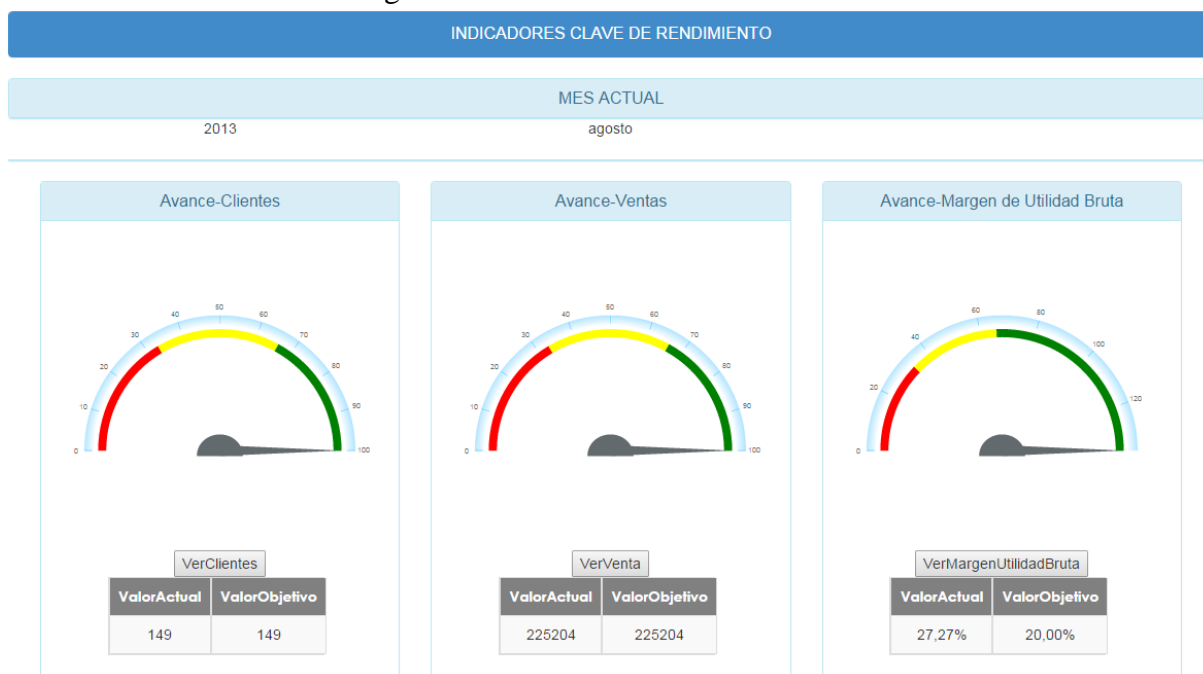
Avance- Clientes, indica el avance de crecimiento de clientes respecto al mismo mes del año anterior, mostrando el valor actual vs el valor objetivo

Avance-Ventas: Indica el avance del monto de ingreso por la ventas respecto al mismo mes del año anterior, mostrando el valor actual vs el valor objetivo

Avance-Margen de Utilidad Bruta: Indica el avance del porcentaje de utilidad bruta respecto al mismo mes del año anterior, mostrando el valor actual vs el valor objetivo

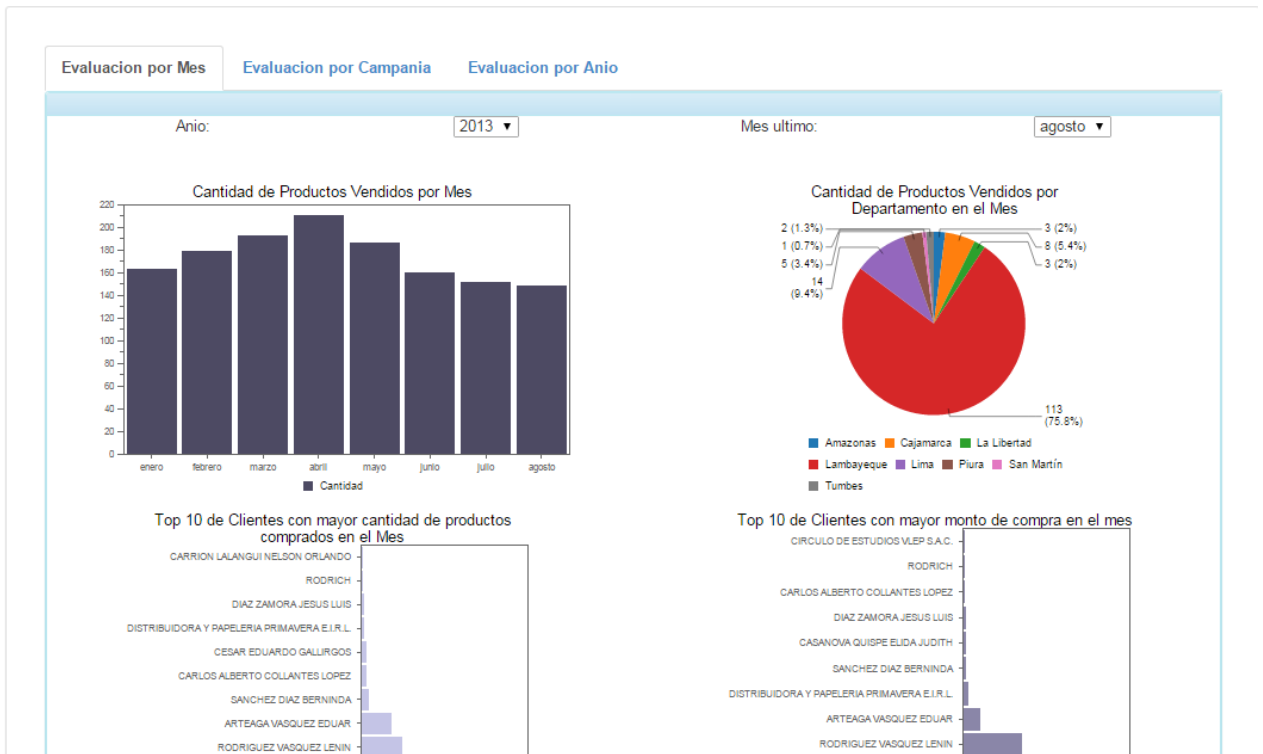
Para ver la el detalle de dichos valores, o la explicación de dichos resultados, se da clic en los botones verClientes, verVenta o verMargenUtilidadBruta según se desee.

Figura N°20: Dashboard 1



Al dar clic en VerClientes, se carga el segundo dashboard, observado en la figura N° 21. Este Dashboard se divide en 3 partes, evaluación por mes, evaluación por campaña y evaluación por año.

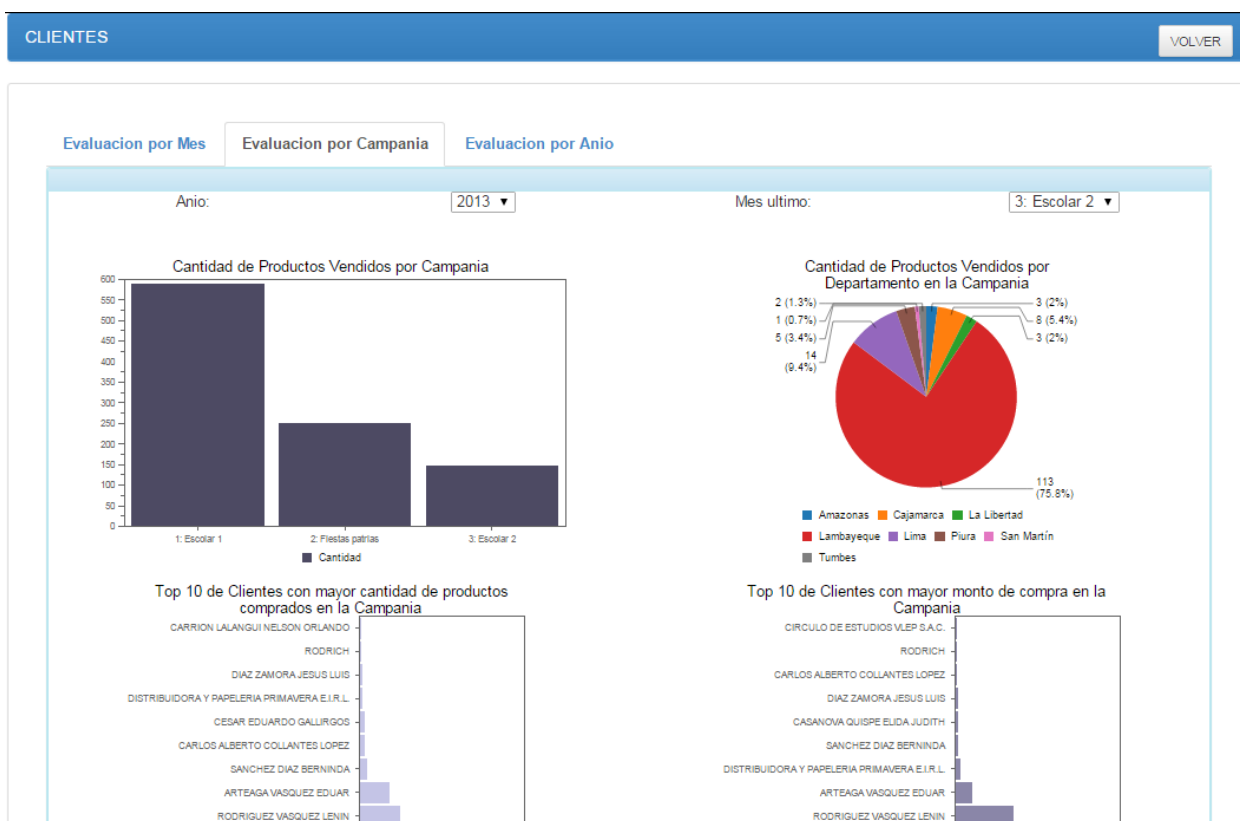
Figura N° 21: Dashboard 2 – Evaluacion por Mes



-Evaluación por Mes.-

- Se muestran un combo para seleccionar el año que contiene los meses que se desea evaluar.
- Del año seleccionado se muestra un gráfico de barras indicando la cantidad de clientes que ha tenido la organización en los respectivos meses
- Al dar clic en un determinado mes se carga:
 - Un gráfico de torta, que indica la cantidad de clientes por departamentos que ha tenido la organización en ese mes seleccionado.
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que compraron más cantidades en ese mes seleccionado
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que generaron un mayor ingreso monetario a la organización en ese mes seleccionado

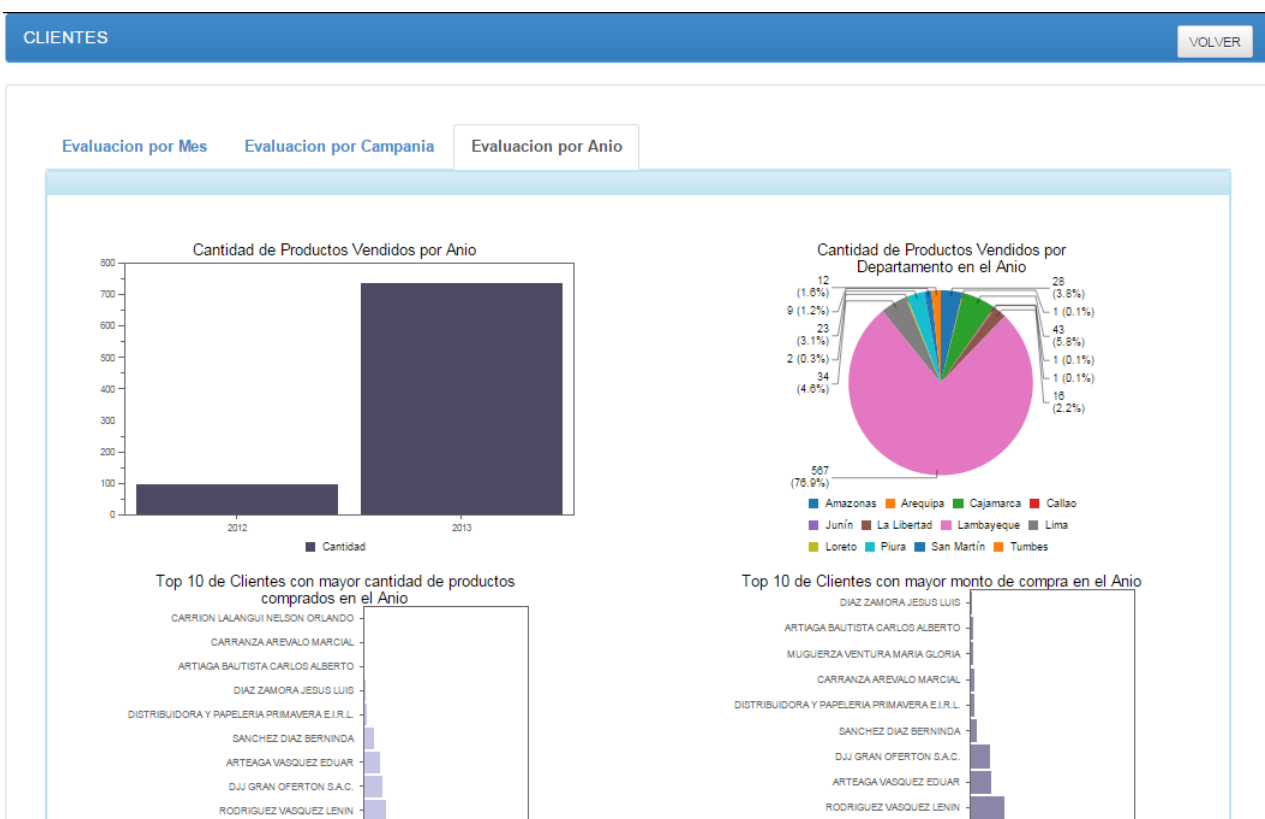
Figura N°22: Dashboard 2 – Evaluación por Campaña



-Evaluación por Campaña, que muestra en la figura N° 22 .-

- Se muestran un combo para seleccionar el año que contiene las campañas que se desea evaluar.
- De la campaña seleccionada se muestra un gráfico de barras indicando la cantidad de clientes que ha tenido la organización en las respectivas campañas.
- Al dar clic en una determinada campaña se carga:
 - Un gráfico de torta, que indica la cantidad de clientes por departamentos que ha tenido la organización en esa campaña seleccionada.
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que compraron más cantidades en esa campaña seleccionada
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que generaron un mayor ingreso monetario a la organización en esa campaña seleccionada

Figura N°23: Dashboard 2 – Evaluación por Año



-Evaluación por Año, que se observa en la figura N° 23

- Se muestra un gráfico de barras indicando la cantidad de clientes que ha tenido la organización en los respectivos años.
- Al dar clic en un determinado año se carga:
 - Un gráfico de torta, que indica la cantidad de clientes por departamentos que ha tenido la organización en ese año seleccionado.
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que compraron más cantidades en ese año seleccionado.
 - Un gráfico de barras horizontales, que muestra el top 10 de los clientes que generaron un mayor ingreso monetario a la organización en ese año seleccionado

Al dar clic en VerVenta, se carga el tercer dashboard, observado en la Figura N°24, este dashboard, al igual que el segundo, se divide en 3 partes, evaluación por mes, evaluación por campaña y evaluación por año.

Figura N°24: Dashboard 3 – Evaluación por Mes



-Evaluación por Mes, vista en la figura N° 24.-

- Se muestran un combo para seleccionar el año que contiene los meses que se desea evaluar.
- Del año seleccionado se muestran dos gráficos de barras:
 - Uno indicando los Montos en Soles de las Ventas VS los Montos en Soles en las Compras que ha tenido la organización en los respectivos meses.
 - El otro indicando la Cantidad de Productos Vendidos VS la Cantidad de Productos Comprados en la organización en los respectivos meses.
- Al dar clic en un determinado mes, de una determinada medida (montos en ventas, montos en compras, cantidades vendidas, cantidades compradas) se carga:
 - Un gráfico de anillo que indica las categorías de productos distribuidas en función de la medida seleccionada del mes seleccionado.

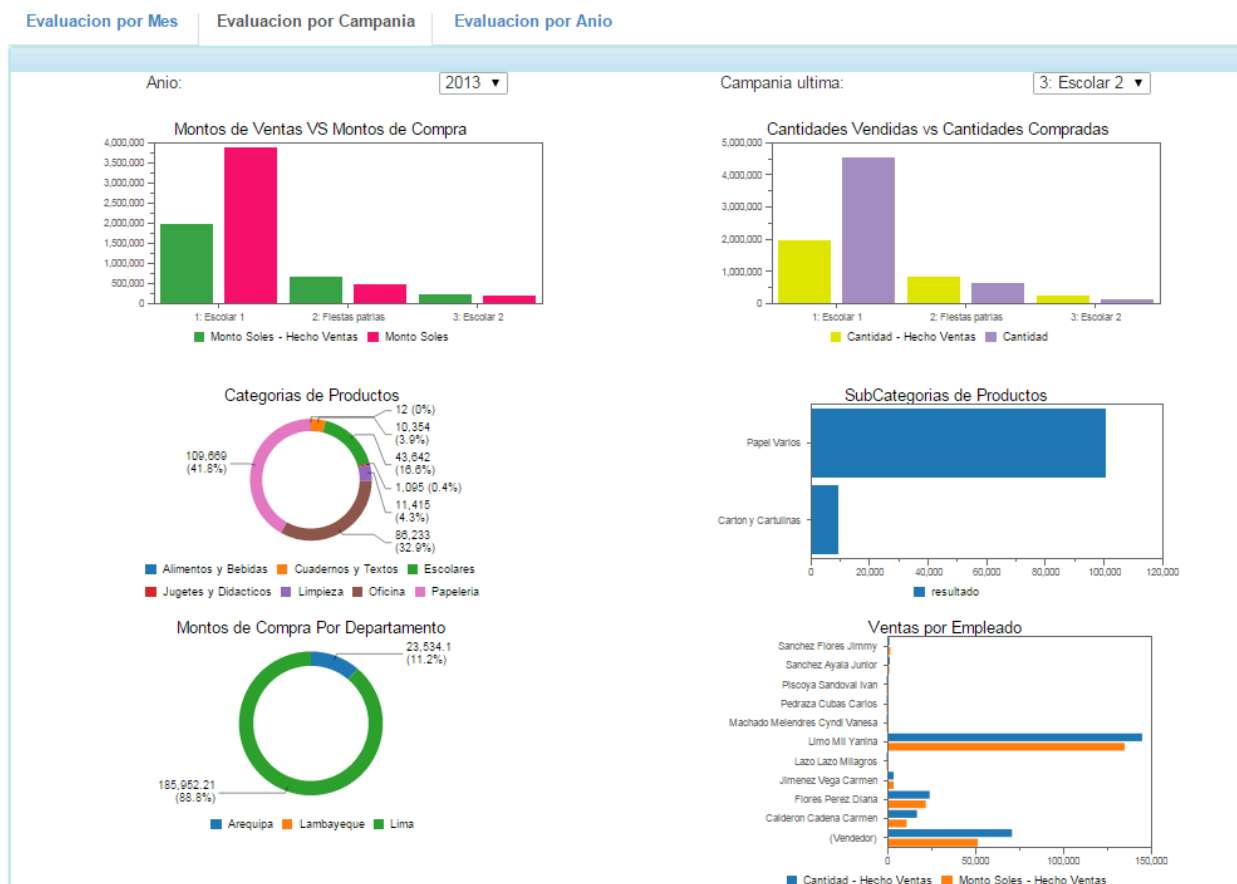
- Un gráfico de barras horizontal que indica, por defecto, las subcategorías de productos del top 1 de la categoría de productos de la medida seleccionada del mes seleccionado, este gráfico cambia de color de acuerdo al color de la medida seleccionada.
- Un gráfico de anillo que indica el monto de compra por departamento del mes seleccionado
- Un gráfico de barras horizontal que indica la cantidad de ventas y el monto de soles en ventas efectuadas por los empleados del mes seleccionado
- Al dar clic en una determinada categoría de producto se recalcula automáticamente el gráfico de la sub categoría de productos por la categoría seleccionada
- Al dar clic en una sub categoría de productos se muestra el top 5 de los productos en función de la medida preseleccionada del mes seleccionado.
- Si se da clic en VerProyeccionDeVentasPorProducto se muestra una tabla, que incluye la opción de búsqueda, con la proyección de las ventas por productos para los siguientes 3 meses como se observa en la figura N°23:

Figura N°25: Dashboard 3 – Proyección de Ventas

PROYECCION DE LAS VENTAS DE PRODUCTOS EN CANTIDADES			VOLVER
Show	10	entries	Search: <input type="text" value="agua"/>
Producto	t\$TIME	t.Cantidad	
ACONCAGUA DURAZNO * 822 GR - CJ * 2 DOC - Unida	201309	140.01525685948383	
ACONCAGUA DURAZNO * 822 GR - CJ * 2 DOC - Unida	201310	99.45182804565769	
ACONCAGUA DURAZNO * 822 GR - CJ * 2 DOC - Unida	201311	266.53830309633463	
AGUA FLORIDA*22 ML - CJ*288 UND - Unida	201309	708.3007394041077	
AGUA FLORIDA*22 ML - CJ*288 UND - Unida	201310	781.2703173233492	
AGUA FLORIDA*22 ML - CJ*288 UND - Unida	201311	895.2658422628568	
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201309	332.27871830410305	
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201310	380.5211132605792	
AGUA FLORIDA*270 ML - CJ*24 UND - Unida	201311	453.47401332940325	
AGUA FLORIDA*70 ML - CJ*72 UND - Unida	201309	1407.9125844093242	

Showing 1 to 10 of 18 entries (filtered from 3,756 total entries) Previous Next

Figura N°26: Dashboard 3 – Evaluación por Campaña



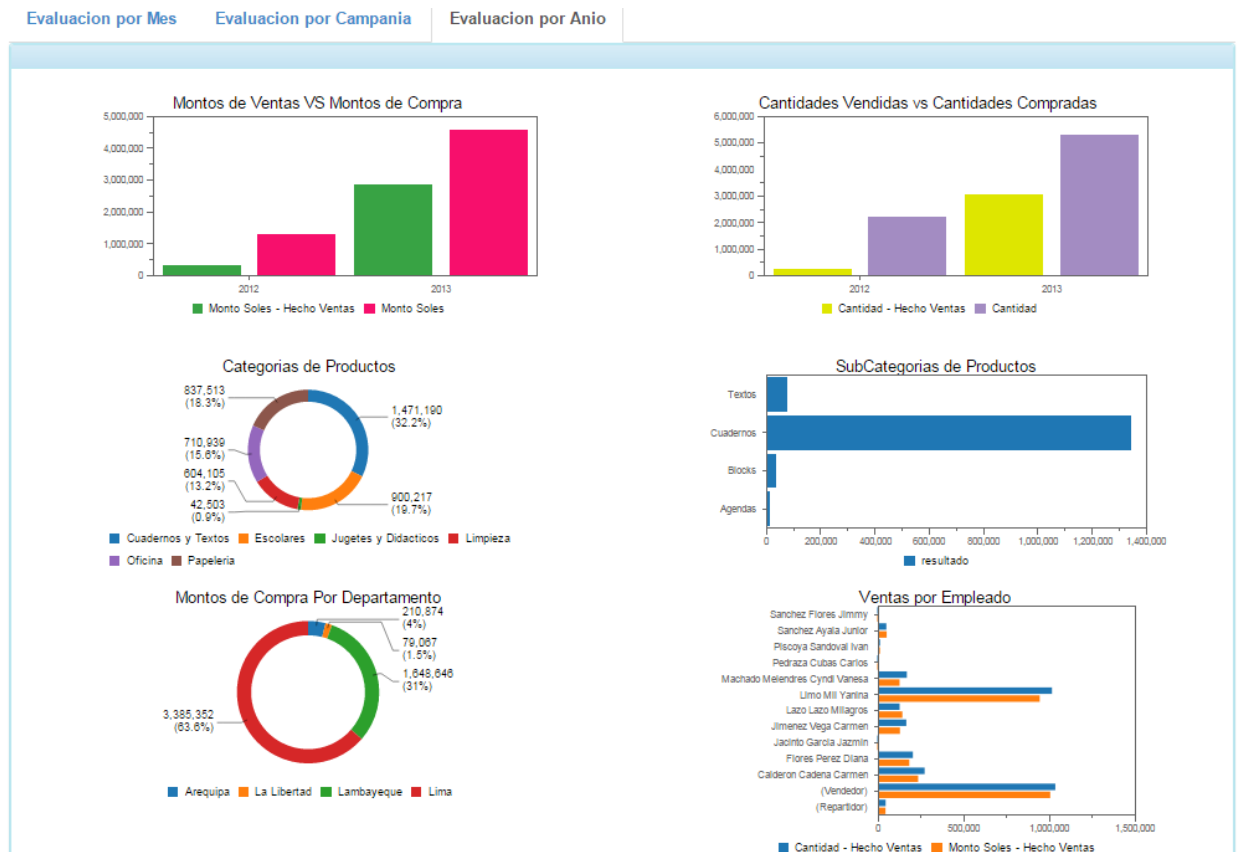
-Evaluación por Campaña, vista en la figura N° 26.

- Se muestran un combo para seleccionar el año que contiene las campañas que se desea evaluar.
- Del año seleccionado se muestran dos gráficos de barras:
 - Uno indicando los Montos en Soles de las Ventas VS los Montos en Soles en las Compras que ha tenido la organización en las respectivas campañas.
 - El otro indicando la Cantidad de Productos Vendidos VS la Cantidad de Productos Comprados en la organización en las respectivas.
- Al dar clic en un determinada campaña, de una determinada medida (montos en ventas, montos en compras, cantidades vendidas, cantidades compradas) se carga:
 - Un gráfico de anillo que indica las categorías de productos distribuidas en función de la medida seleccionada de la campaña seleccionada.
 - Un gráfico de barras horizontal que indica, por defecto, las subcategorías de productos del top 1 de la categoría de productos de la medida seleccionada de la campaña

seleccionada, este gráfico cambia de color de acuerdo al color de la medida seleccionada.

- Un gráfico de anillo que indica el monto de compra por departamento de la campaña seleccionada
- Un gráfico de barras horizontal que indica la cantidad de ventas y el monto de soles en ventas efectuadas por los empleados de la campaña seleccionada
- Al dar clic en una determinada categoría de producto se recalcula automáticamente el gráfico de la sub categoría de productos por la categoría seleccionada
- Al dar clic en una sub categoría de productos se muestra el top 5 de los productos en función de la medida preseleccionada del mes seleccionado.

Figura N°27: Dashboard 3 – Evaluación por Año



-Evaluación por Año, vista en la Figura N° 27.-

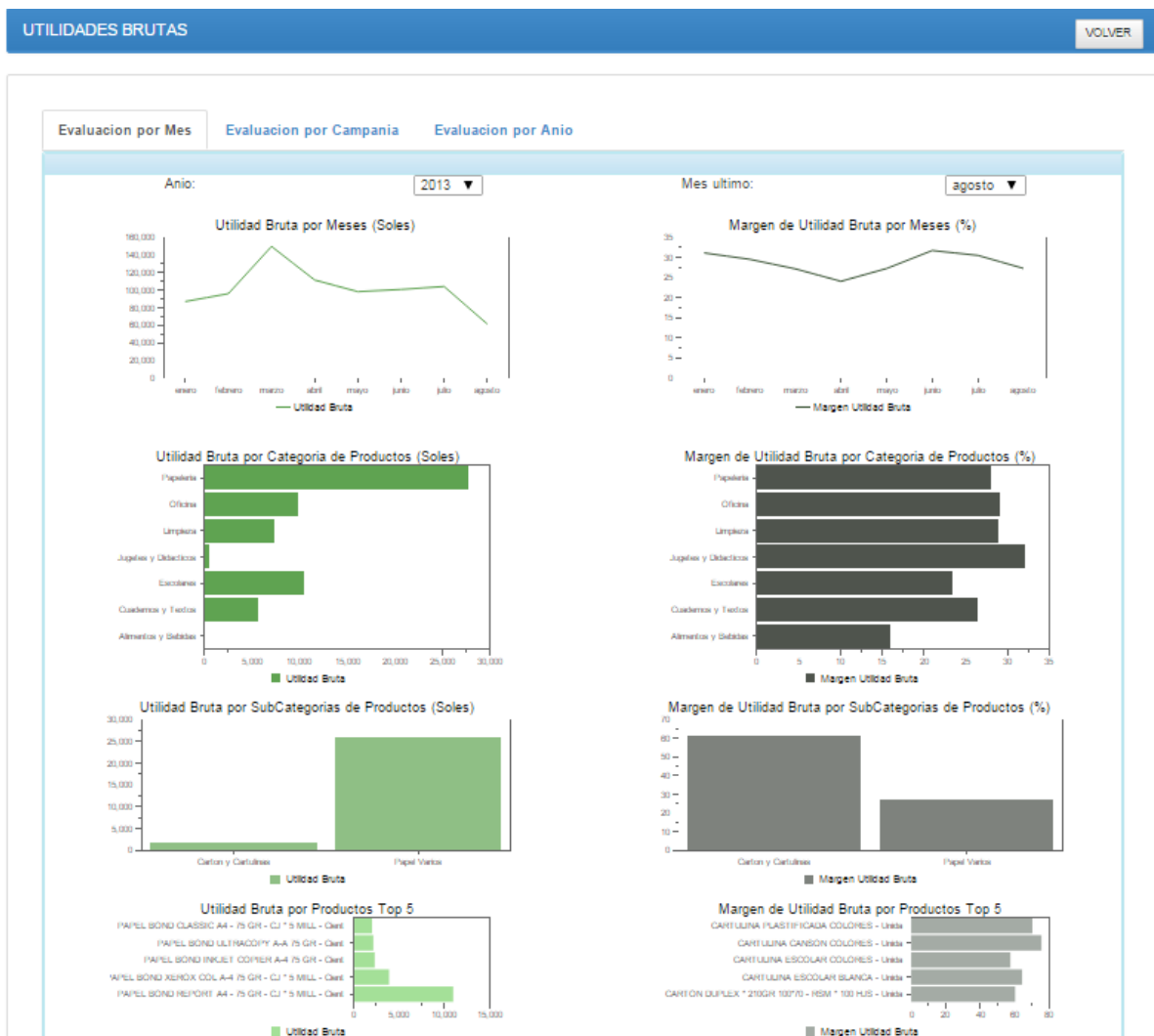
Se carga dos gráficos :

- Montos en Soles de las Ventas VS los Montos en Soles en las Compras que ha tenido la organización en las respectivos años.

- El otro indicando la Cantidad de Productos Vendidos VS la Cantidad de Productos Comprados en la organización en los respectivos años.
- Al dar clic en un determinada campaña, de una determinada medida (montos en ventas, montos en compras, cantidades vendidas, cantidades compradas) se carga:
 - Un gráfico de anillo que indica las categorías de productos distribuidas en función de la medida seleccionada del año seleccionado.
 - Un gráfico de barras horizontal que indica, por defecto, las subcategorías de productos del top 1 de la categoría de productos de la medida seleccionada del año seleccionado, este gráfico cambia de color de acuerdo al color de la medida seleccionada.
 - Un gráfico de anillo que indica el monto de compra por departamento del año seleccionado
 - Un gráfico de barras horizontal que indica la cantidad de ventas y el monto de soles en ventas efectuadas por los empleados del año seleccionado.
- Al dar clic en una determinada categoría de producto se recalcula automáticamente el gráfico de la sub categoría de productos por la categoría seleccionada
- Al dar clic en una sub categoría de productos se muestra el top 5 de los productos en función de la medida preseleccionada del año seleccionado.
- los productos en función de la medida preseleccionada del año seleccionado.

Al dar clic en VerMargenUtilidadBruta, se carga el tercer dashboard, observado en la figura N° 28. Este Dashboard se divide en 3 partes, evaluación por mes, evaluación por campaña y evaluación por año.

Figura N°28: Dashboard 4 – Evaluación por Mes



-Evaluación por Mes, vista en la figura N° 28.-

- Se muestran un combo para seleccionar el año que contiene los meses que se desea evaluar.
- Del año seleccionado se muestran dos gráficos de líneas:
 - Uno indicando la Utilidad Bruta en soles que ha tenido la organización en los respectivos meses.
 - El otro indicando el Margen de Utilidad Bruta en los respectivos meses.
- Al dar clic en un determinado mes de la medida Utilidad Bruta se carga:
 - Un gráfico de barras horizontales que muestra la Utilidad Bruta por Categoría de productos del mes seleccionado
 - Un gráfico de barras verticales que muestra la Utilidad Bruta por SubCategoría de Productos del mes seleccionado

- El top 5 de productos que mayor Utilidad Bruta del mes seleccionado
- Al dar clic en un determinado mes de la medida Margen de Utilidad Bruta se carga:
 - Un gráfico de barras horizontales que muestra el Margen de Utilidad Bruta por Categoría de productos del mes seleccionado
 - Un gráfico de barras verticales que muestra el Margen de Utilidad Bruta por SubCategoría de Productos del mes seleccionado
 - El top 5 de productos que mayor Margen de Utilidad Bruta del mes seleccionado

Figura N°29: Dashboard 4 – Evaluación por Campaña



-Evaluación por Campaña, vista en la figura N° 29.-

- Se muestran un combo para seleccionar el año que contiene las campañas que se desea evaluar.

- Del año seleccionado se muestran dos gráficos de líneas:
 - Uno indicando la Utilidad Bruta en soles que ha tenido la organización en las respectivas campañas.
 - El otro indicando el Margen de Utilidad Bruta en las respectivas campañas.
- Al dar clic en un determinada campaña de la medida Utilidad Bruta se carga:
 - Un gráfico de barras horizontales que muestra la Utilidad Bruta por Categoría de productos de la campaña seleccionada
 - Un gráfico de barras verticales que muestra la Utilidad Bruta por SubCategoría de Productos de la campaña seleccionada
 - El top 5 de productos que mayor Utilidad Bruta de la campaña seleccionada
- Al dar clic en una determinada campaña de la medida Margen de Utilidad Bruta se carga:
 - Un gráfico de barras horizontales que muestra el Margen de Utilidad Bruta por Categoría de productos de la campaña seleccionada
 - Un gráfico de barras verticales que muestra el Margen de Utilidad Bruta por SubCategoría de Productos de la campaña seleccionada
 - El top 5 de productos que mayor Margen de Utilidad Bruta de la campaña seleccionada

Figura N°30: Dashboard 4 – Evaluación por Año



-Evaluación por Año, vista en la figura N° 30.-

- Se muestran dos gráficos de barras verticales:
 - Uno indicando la Utilidad Bruta en soles que ha tenido la organización en los respectivos años.
 - El otro indicando el Margen de Utilidad Bruta en los respectivos años.
- Al dar clic en un determinado año de la medida Utilidad Bruta se carga:
 - Un gráfico de barras horizontales que muestra la Utilidad Bruta por Categoría de productos del año seleccionado.
 - Un gráfico de barras verticales que muestra la Utilidad Bruta por SubCategoría de Productos del año seleccionado.
 - El top 5 de productos que mayor Utilidad Bruta del año seleccionado.
- Al dar clic en un determinado año de la medida Margen de Utilidad Bruta se carga:

- Un gráfico de barras horizontales que muestra el Margen de Utilidad Bruta por Categoría de productos del año seleccionado
- Un gráfico de barras verticales que muestra el Margen de Utilidad Bruta por SubCategoría de Productos del año seleccionado
- El top 5 de productos que mayor Margen de Utilidad Bruta del año seleccionado.

4.10. Mantenimiento y crecimiento

En la Tabla N° 59 se indica aspectos a tener en cuenta para el correcto funcionamiento, mantenimiento y crecimiento del sistema

Tabla N° 59: Mantenimiento y crecimiento

DISPOSICIÓN	ELEMENTOS
Consistencia e Integridad de datos	0.5% de error
Precisión de la Información	100% precisa
Seguridad	Alta gerencia y usuarios autorizados
Mantenimiento de metadatos	Mensual
Desempeño del Sistema	En tiempo real
Mantenimiento y Actualización de datos	Finalizada cada jornada diaria
Conectividad y disponibilidad	24 horas del día

V.- DISCUSIÓN

EJECUCIÓN DEL DISEÑO DE CONTRASTACIONES PRE TEST Y POST TEST

Para contrastar la hipótesis se puso a prueba la implementación de una Solución de Inteligencia de Negocios basado en el algoritmo de serie temporal para la mejora del proceso de toma de decisiones gerencial en la empresa comercial el Oferton SAC., para lo cual se realizaron encuestas, aplicadas al gerente, al encargado de sistemas y al encargado de finanzas, estas encuestas se realizaron antes de la implementación del sistema y luego después de la implementación del sistema. Adicionalmente se hizo uso de un Timer para verificar el tiempo de elaboración de reportes antes y después del sistema propuesto.

INDICADORES CUANTITATIVOS

Indicador N°1 Tiempo de elaboración de informes solicitados

Tabla N° 60 Tiempo de elaboración de informes solicitados

Indicador	O ₁	O ₂	DI
Tiempo de la elaboración de informes solicitados	18 horas	0.6 horas	-15.4 Horas

Donde:

O₁: Es el tiempo promedio que demanda el encargado de sistemas para la elaboración de los informes solicitados por la gerencia, procedimiento que implica obtener los datos del sistema transaccional para luego pasarlos a la plataforma de Excel para elaborar los informes.

O₂: Es el tiempo promedio que demanda el encargado de sistemas para la elaboración de los informes solicitados por la gerencia, procedimiento que implica interactuar con la herramienta de inteligencia de negocios desarrollada.

DI: O₁ – O₂: Es la diferencia de tiempos que se obtiene en la elaboración de los informes solicitados por la gerencia tras la implementación de la solución de inteligencia de negocios desarrollada.

Como se observa en la Tabla N° 57, la implementación de la Solución de Inteligencia de Negocios basado en el Algoritmo de Serie Temporal ha permitido reducir en gran medida el tiempo de elaboración de informes solicitados por la alta gerencia, donde el tiempo reducido es de 15,4 horas. Estos reportes se realizan 2 veces al mes.

INDICADORES CUALITATIVOS

Tabla N° 61 Escala de medición de Likert

1	Muy en desacuerdo
2	En desacuerdo
3	Ni de acuerdo ni en desacuerdo
4	De acuerdo
5	Muy de acuerdo

Fuente: Castañeda, et al (2010)

Indicador N°2: Nivel de calidad de la información solicitada.

Tabla N° 62: Nivel de calidad de la información solicitada

		Media	N	Desviación típ.	Error típ. de la media
Par 1	MediaPreTest	1,8667	3	,57735	,33333
	MediaPostTest	4,6000	3	,52915	,30551

Hipótesis que respecta al segundo indicador

H0

Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal no incrementará el Nivel de calidad de la información

H1

Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal incrementará el Nivel de calidad de la información

Estadístico:

Tabla N°62: Prueba de hipótesis del segundo indicador

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	MediaPreTest - MediaPostTest	-2,73333	1,00664	,58119	-5,23398	-,23269	-4,703	2	,042

El Nivel de calidad de la información anterior era significativamente menor antes de la implementación de la Solución (M= 1,8667 SE = 0,33) que después de la misma (M= 4,60 SE = 0,30 , t(2) = -4,703 , p<.05)

M= Media

SE = Estándar Error

t= prueba T

() = Grados de Libertad

P = nivel de significación

No se acepta H0, y entonces podemos decir que Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal incrementa el Nivel de calidad de la información

Indicador N°4: Nivel de satisfacción sobre la información solicitada

Tabla N° 64: Nivel de satisfacción sobre la información solicitada

		Media	N	Desviación típ.	Error típ. de la media
Par 1	MediaPreTest	2,2000	3	,40000	,23094
	MediaPostTest	4,5333	3	,46188	,26667

Hipótesis que respecta al tercer indicador

H0

Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal no incrementará el Nivel satisfacción sobre la información solicitada

H1

Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal incrementará el Nivel de satisfacción sobre la información solicitada

Estadístico:

Tabla N°64: Prueba de hipótesis del tercer indicador

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	MediaPreTest - MediaPostTest	-2,33333	,23094	,13333	-2,90702	-1,75965	-17,500	2	,003

El Nivel de satisfacción sobre la información solicitada anterior era significativamente menor antes de la implementación de la Solución (M= 2,200 SE = 0,23) que después de la misma (M= 4,53 SE = 0,26 , t(2) = -17,500, p<.05)

M= Media

SE = Estándar Error

t= prueba T

() = Grados de Libertad

P = nivel de significación

No se acepta H_0 , y entonces podemos decir que Implementación de una solución de Inteligencia de Negocio basada en el Algoritmo de Serie Temporal incrementa el Nivel satisfacción sobre la información solicitada.

VI. CONCLUSIONES

- 1.- La implementación de la Solución de Inteligencia de Negocios basado en el algoritmo de Serie Temporal ha mejorado significativamente el proceso de toma de decisiones en la empresa “El Ofertón SAC”.
- 2.- Se redujo el tiempo promedio en la elaboración de informes solicitados por la alta gerencia en un 96,25% , es decir de 16 horas a 0.6 horas.
- 3.- Se incrementó el nivel de calidad de la información solicitada de 1,8667 a 4,600 basado en una escala de 5 puntos del estilo Likert.
- 4.- Se incrementó el nivel de satisfacción sobre la información solicitada de 2,2000 a 4,5333 basado en una escala de 5 puntos del estilo Likert.
- 5.- La metodología utilizada para desarrollar la Solución de Inteligencia de Negocios basado en el algoritmo de Serie Temporal es una metodología híbrida, entre la metodología Data Warehouse de Kimball (Bottom – up) y la metodología CRISP DM para el proceso de minería.
- 6.- Se realizaron Dashboards e Indicadores de Gestión para el proceso de compras y ventas de la organización, bajo diversas perspectivas para lograr un mejor análisis de la información y bajo una interfaz intuitiva.
- 7.- Se generó una proyección de las ventas para los próximos 3 meses utilizando el algoritmo de Serie Temporal Mixed por ser el más óptimo.

VII. REFERENCIAS BIBLIOGRÁFICAS

Aranceta, Javier and Gil, Ángel. (2010). *Alimentos funcionales y salud en la etapa infantil y juvenil*. Madrid: Ed. Médica Panamericana.

Carmargo, H and Silva, Mario. Revista Tecnológica. (Sin fecha). *Journal of Technology. Dos Caminos en la búsqueda de patrones por medio de Minería de Datos: SEMMA Y CRISP* . Vol 9 N°01.

Castañeda, et al (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS*. Brasil: Edipucrs.

Chen, Edward. (2002). “*The importance of Building a Foundation for Use Involvement in Information System Project*”.

Curto, Josep. (2012). *Introducción a Business Intelligence*. Barcelona: Editorial Uoc

De la Serna, Manuel. (2010). *Análisis, Diseño e Implementación de un Data Warehouse que de soporte a la toma de decisiones para la gestión de comercialización de la empresa comercial Cynthia* (Tesis de Pregrado). Universidad Católica Santo Toribio de Mogrovejo , Chiclayo.

DeLone, William H, McLean, Ephraim (2003). “*The DeLone and McLean Model of Information System Success: A Ten-year Update*”. Washington: The Institute of Management Sciences.

Galán, Carlos. (2011). *Desarrollo de una Solución de Business Intelligence para mejorar el proceso de toma de decisiones estratégicas en la gestión comercial de la empresa Truck and Motors del Perú S.A.C* (Tesis de Pregrado). Universidad Católica Santo Toribio de Mogrovejo , Chiclayo.

Girones, Jordi. (Sin fecha). *Metodologías y estándares*. Cataluña: Universidad Oberta de Catalunya.

Goodhue (2003). “*The Benffits of Data Warehousing al Wrippol*”. Reino Unido

Hand, David. (1998). *Data Mining: Statics and More?*. Reino Unido: The American Statistician.

Hernández, Jose.2004. *Introducción a la Minería de Datos*. Madrid: Editorial Pearson Prentice Hall.

Hightower, Ross. (2002). “*A validation of the End-User Computing Satisfaction Instrument in Taiwan*”. Information & Management.

ICA2 Innovación y Tecnología. (2011). *Caso de Éxito: Business Intelligence (desarrollo ad-hoc) – empresa del sector petrolero*. Madrid

Im, Kun, Varum (2004). *“The Use Structural Equation Modeling in Is Research; Review and Recommendations”*.

Inmon, W., Strauss, D., & Neushloss, G. (2010). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. USA: Editorial Morgan Kaufmann.

Kimball et al., (1998). *The Data Warehouse Lifecycle Toolkit*. New York: Wiley.

Kroenke, David. (2003). *Procesamiento de Base de Datos: Fundamentos, Diseño e Implementación*. México: Pearson.

Microsoft SQL Server. [https://technet.microsoft.com/es-es/library/bb677216\(v=sql.105\).aspx](https://technet.microsoft.com/es-es/library/bb677216(v=sql.105).aspx). (consultado: 15/01/2014).

Muiño, Alejandro. (2012). *Caso de éxito: Bodegas Chandon S.A*. Buenos Aires.

Núñez, Gace. (2010). *Análisis, Diseño e Implementación de una Solución de Inteligencia de Negocios para el área de finanzas de la Municipalidad Metropolitana de Lima* (Tesis de Pregrado). Pontificia Universidad Católica del Perú, Lima.

Pentaho. <http://www.webdetails.pt/ctools/cde> . Consultado (14/02/2014)

Pérez, Cesar. (2007). *Minería de Datos. Técnicas y Herramientas*. Madrid: Thomson.

Ramos, Salvador. 2011. *Microsoft Business Intelligence: vea el cubo medio lleno*. España: SolidQ Press

Mahmood et al (2000). *Variables Affecting Information Technology End-User Satisfaction”*.

Méndez, Luis. (2006). *Más allá del Business Intelligence*. Barcelona: GESTIÓN 2000.

Rivadera, Gustavo. Integración de datos. <http://www.ucasal.net/templates/unid-academicas/ingenieria/apps/5-p56-rivadera-formateado.pdf>. (Consultado: 20/01/2014)

Salazar, Leny. (2013). *Desarrollo de una Solución de Inteligencia de Negocios: aplicando la técnica de data profiling para mejorar la calidad de los datos en la empresa "Bioagro" S.R.L* (Tesis de Pregrado). Universidad Católica Santo Toribio de Mogrovejo , Chiclayo.

Siccha, Hober Willy (2012). *Minería de datos aplicados a las ventas con tarjeta de crédito clásica realizados en las tiendas Saga Falabella en la ciudad de Lima* (Tesis de Pregrado). Universidad Tecnológica del Perú. Lima.

SPPerú. (2011). *Caso de éxito: ARCOR*. Lima

The Data Warehousing Institute, Tdwi. <https://tdwi.org/portals/agile-bi.aspx>. (consultado: 12/01/2014).

Trujillo, Juan and Mazon, Jose. 2011. *Diseño y Explotación de Almacenes de Datos*. España: Editorial Club Universitario.

Tvrdivková, Milena. (2009) . *Support of Decision Making by Business Intelligence Tools, Departamente of Applied Informatics*. Republica Checa: VSB Technichal University Ostrava.

Vega, Gustavo. (2010). *Inteligencia de Negocios. Aplicación en la administración del presupuesto en una empresa del sector público* (Tesis de Pregrado). Instituto Politécnico Nacional, Mexico D.F.

Vieira, Luis and Ortiz, Luis. 2009. *Introducción a la Minería de Datos*. Brasil: E-papers

Vit, Michael. (2002). *Business Intelligence Técnicas de análisis para la toma de decisiones estratégicas*. España : Editorial UOC

Zikmund, William. (2003). "*Business Research Methods*". Ohio: Thomson South-Western Editorial

VIII.- ANEXOS

ANEXO N°01: Cuestionario para los usuario implicados en el proceso de toma decisiones de la empresa “El Ofertón S.A.C”

CUESTIONARIO PARA USUARIOS IMPLICADOS EN EL PROCESO DE TOMA DE DECISIONES DE LA EMPRESA “DISTRIBUCIONES Y REPRESENTACIONES EL OFERTÓN SAC”

Objetivo: Determinar el Nivel de Calidad de la Información y el Nivel de Calidad de Satisfacción de los Usuarios del Sistema de Información para el proceso de toma de decisiones en la empresa “Distribuciones y Representaciones El Ofertón SAC”.

Responda los cuestionamientos de acuerdo a su experiencia y percepción, marcando con un aspa o rellenando el recuadro que más se acerque a su respuesta, considerando la escala del 1 al 5, donde:

1: Totalmente en desacuerdo, 2: En desacuerdo, 3: Ni de acuerdo ni en desacuerdo, 4: De acuerdo, 5: Totalmente de acuerdo

Muchas gracias por su cooperación

CALIDAD DE LA INFORMACIÓN:

Con respecto al Sistema de Información de cara al proceso de toma de decisiones:

- 1.- El sistema provee reportes de rápida comprensión

1	2	3	4	5
---	---	---	---	---
- 2.- El sistema provee reportes con información actualizada

1	2	3	4	5
---	---	---	---	---
- 3.- A través de los reportes obtengo información suficiente y necesaria para abarcar todos los factores críticos que influyen en la correcta ejecución de mis funciones

1	2	3	4	5
---	---	---	---	---
- 4.- Los reportes son consistentes, es decir, la información que brinda es congruente con todos los objetivos de negocio

1	2	3	4	5
---	---	---	---	---
- 5.- Los reportes son concisos, puntuales, selectivos y resumidos adecuadamente

1	2	3	4	5
---	---	---	---	---

SATISFACCIÓN DEL USUARIO:

Con respecto al Sistema de Información de cara al proceso de toma de decisiones:

- 6.- Confío plenamente en los reportes que me brinda el sistema

1	2	3	4	5
---	---	---	---	---
- 7.- Si me fuera a trabajar a una empresa de la competencia, recomendaría que implanten un sistema igual por la eficiencia de sus reportes

1	2	3	4	5
---	---	---	---	---
- 8.- Los reportes me permiten adquirir un conocimiento valioso que de otra forma me resultaría muy difícil o imposible de alcanzarla

1	2	3	4	5
---	---	---	---	---
- 9.- Cuando estoy usando las funcionalidades de los reportes del sistema mi estado es de disfrute

1	2	3	4	5
---	---	---	---	---
- 10.- En términos generales estoy satisfecho con el sistema de información

1	2	3	4	5
---	---	---	---	---

**ANEXO N° 02:
Procesamiento de los resultados de los cuestionarios**

PRE-TEST del Indicador Nivel de calidad de la información solicitada y del indicador Nivel de Satisfacción de la Gerencia sobre la Información Solicitada

*spss12.sav [Conjunto_de_datos1] - IBM SPSS Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

	Comprensibilidad	Actual	Completa	Relevante	Sintesis	Confianza	Eficiencia	Trascendencia	Disfrute	Satisfaccion General
1	2	3	2	1	1	3	2	2	2	2
2	2	3	2	2	2	3	3	1	2	1
3	3	4	3	2	2	4	3	2	3	3

POST-TEST del Indicador Nivel de calidad de la información solicitada y del indicador Nivel de Satisfacción de la Gerencia sobre la Información Solicitada

*spss2CalidadInfoPost.sav [Conjunto_de_datos1] - IBM SPSS Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

	Comprensibilidad	Actual	Completa	Relevante	Sintesis	Confianza	Eficiencia	Trascendencia	Disfrute	Satisfaccion General
1	5	5	4	5	5	4	5	5	4	5
2	4	5	3	4	4	5	5	4	5	4
3	5	4	5	5	5	4	4	4	5	5

Para la fiabilidad de los ítems se utiliza el alfa de Cronbach, Dyba (2000) indica que un coeficiente superior a 0,7 se considera bueno, por lo que se analiza el alfa de Cronbach de cada indicador, como se aprecia en la Figura N° 03 y Figura N°04 todos son superiores a 0,7:

Alfa de Cronbach para el Indicador Nivel de calidad de la información solicitada

Escala: TODAS LAS VARIABLES

Resumen del procesamiento de los casos

		N	%
Casos	Válidos	3	100,0
	Excluidos ^a	0	,0
	Total	3	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
,921	,921	5

Alfa de Cronbach para el Indicador Nivel de satisfacción de la gerencia sobre la información solicitada

Escala: TODAS LAS VARIABLES

Resumen del procesamiento de los casos

		N	%
Casos	Válidos	3	100,0
	Excluidos ^a	0	,0
	Total	3	100,0

a. Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
,833	,839	5