

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



**SISTEMA DE BUSINESS INTELLIGENCE UTILIZANDO
ALGORITMO DE SERIE TEMPORAL PARA APOYAR LA
TOMA DE DECISIONES EN EL PROCESO DE VENTAS
DEL GRUPO EMPRESARIAL LEONCITO**

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

CHRISTIAN JESÚS MINO EGUSQUIZA

Chiclayo 05 de Septiembre de 2018

**SISTEMA DE BUSINESS INTELLIGENCE UTILIZANDO
ALGORITMO DE SERIE TEMPORAL PARA APOYAR LA
TOMA DE DECISIONES EN EL PROCESO DE VENTAS DEL
GRUPO EMPRESARIAL LEONCITO**

POR:

CHRISTIAN JESÚS MINO EGUSQUIZA

**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

**Mgtr. Marlon Eugenio Vilchez Rivas
PRESIDENTE**

**Ing. Héctor Miguel Zelada Valdivieso
SECRETARIO**

**Ing. Segundo José Castillo Zumarán
ASESOR**

DEDICATORIA

Dedico la presente tesis a Dios, a mis padres y a todos los que me rodean porque siento que me ayudaron a mejorar cada día como persona para llegar a ser quien soy.

AGRADECIMIENTOS

Agradezco a mis padres por el amor, la paciencia, el constante apoyo y por los consejos que me han dado a lo largo de mi vida.

También agradezco a todas a aquellas personas que han puesto su granito de arena y han hecho posible que culmine la presente tesis.

ÍNDICE

I.	INTRODUCCIÓN.....	11
II.	MARCO TEÓRICO	14
	2.1. Antecedentes.....	14
	2.2. Bases Teórico Científicas	17
	2.2.1. Toma de decisiones:	17
	2.2.2. Inteligencia de negocios:	17
	2.2.3. Data Warehouse:.....	20
	2.2.4. Minería de datos:	30
	2.2.5. Herramientas para el desarrollo del sistema de BI:	32
III.	MATERIALES Y MÉTODOS.....	37
	3.1. Diseño de investigación	37
	3.2. Metodología.....	41
IV.	RESULTADOS	42
	4.1. Planificación del proyecto.....	42
	4.1.1. Alcance:.....	42
	4.1.2. Propósito:.....	42
	4.2. Definición de requerimientos del negocio	42
	4.2.1. Requerimientos del proceso:	42
	4.2.2. Matriz bus:.....	43
	4.3. Modelo dimensional.....	44
	4.3.1. Elegir el proceso de negocio:.....	44
	4.3.2. Establecer el nivel de granularidad:.....	44
	4.3.3. Dimensiones y sus atributos:	45
	4.3.4. Identificar tablas hecho y sus medidas:	46
	4.4. Diseño físico	46
	4.5. Diseño y desarrollo de presentación de datos	48
	4.6. Especificaciones de aplicaciones para usuarios finales	52
	4.7. Desarrollo de aplicaciones para usuarios finales	52
	4.7.1. Creación de cubo OLAP:.....	53
	4.7.2. Creación de minería de datos:	57
	4.7.3. Construcción de reportes específicos:	65
	4.8. Despliegue.....	73
	4.9. Mantenimiento y crecimiento	73
V.	DISCUSIÓN.....	78
VI.	CONCLUSIONES.....	82
VII.	REFERENCIAS BIBLIOGRÁFICAS	83
VIII.	ANEXOS	85

ÍNDICE DE TABLAS

Tabla 1. Comparación entre la metodología de Kimball e Inmon.....	25
Tabla 2. Comparación entre OLTP y OLAP	29
Tabla 3. Contrastación de hipótesis – diseño cuasi experimental	37
Tabla 4. Indicadores	39
Tabla 5. Métodos, técnicas e instrumentos de recolección de datos	40
Tabla 6. Ejemplo de Matriz Bus	22
Tabla 7. Matriz bus	43
Tabla 8. Dimensiones y sus atributos	45
Tabla 9. Hecho con sus atributos y medidas	46
Tabla 10. Dimensión producto	46
Tabla 11. Dimensión cliente.....	46
Tabla 12. Dimensión local.....	46
Tabla 13. Dimensión tipo de venta.....	47
Tabla 14. Dimensión ubigeo.....	47
Tabla 15. Dimensión tiempo	47
Tabla 16. Dimensión vendedor.....	47
Tabla 17. Hecho venta	47
Tabla 18. Indicador 1.....	78
Tabla 19. Indicador 2 – Indicador 4	79
Tabla 20. Indicador 3.....	80
Tabla 21. Indicador 5.....	81
Tabla 22. Medición del tiempo de espera para acceder a la información por cada reporte.	88

ÍNDICE DE FIGURAS

Figura 1. Sistema de BI	19
Figura 2. Fases de la metodología Kimball	21
Figura 3. Metodología Inmon	24
Figura 4. Modelo dimensional de alto nivel	44
Figura 5. Conexión al origen y destino.....	48
Figura 6. Poblamiento – dimensión producto.....	49
Figura 7. Poblamiento – dimensión cliente	49
Figura 8. Poblamiento – dimensión local	50
Figura 9. Poblamiento – dimensión ubigeo	50
Figura 10. Consulta para poblar dimensión tiempo.....	51
Figura 11. Poblamiento – dimensión tiempo.....	51
Figura 12. Poblamiento – dimensión vendedor	51
Figura 13. Poblamiento – hecho venta	52
Figura 14. Conexión al data mart dm_gl	53
Figura 15. vista_dm_gl_ventas.....	53
Figura 16. Creación de cubo cubo_ventas.....	54
Figura 17. KPI_monto_ventas_año	55
Figura 18. KPI_monto_ventas_semestre	55
Figura 19. KPI_monto_ventas_trimestre.....	56
Figura 20. KPI_monto_ventas_mes	56
Figura 21. Consulta para predicción.....	57
Figura 22. Creación de tabla temporal.....	57
Figura 23. Creación de cursor 1.....	58
Figura 24. Creación de cursor 2.....	58
Figura 25. Creación de vista	58
Figura 26. Creación de vista para origen de datos.....	59
Figura 27. Selección de técnica de minería de datos	59
Figura 28. Selección de vista de origen de datos.....	60
Figura 29. Selección de escenario	60
Figura 30. Especificación de datos de aprendizaje.....	60
Figura 31. Especificación de contenido y tipo de datos	61
Figura 32. Finalización de creación de estructura de minería de datos	61
Figura 33. Estructura y modelo de minería de datos creado	61
Figura 34. Método ARIMA	62
Figura 35. Fiabilidad de método ARIMA	62
Figura 36. Método ARTXP	63
Figura 37. Fiabilidad de método ARTXP.....	63
Figura 38. Método MIXED	64
Figura 39. Fiabilidad de método MIXED.....	64
Figura 40. Interfaz para iniciar sesión	65
Figura 41. Semaforización por año.....	65
Figura 42. Semaforización por semestre	66
Figura 43. Semaforización por trimestre	66
Figura 44. Semaforización por mes	67
Figura 45. Reporte de producto por línea	67
Figura 46. Reporte de producto por marca	68
Figura 47. Reporte top de clientes por año.....	68
Figura 48. Reporte de vendedores por año y marca	69

Figura 49. Reporte de vendedores filtrados por marca, año y mes	69
Figura 50. Reporte de vendedor en año	70
Figura 51. Reporte de ventas por distrito	70
Figura 52. Reporte de ventas por local	71
Figura 53. Reporte de ventas por local y año – top productos	71
Figura 54. Reporte de predicción de ventas por distrito	72
Figura 55. Reporte de predicción de ventas por distrito	72
Figura 56. Agente SQL Server	73
Figura 57. Servicios – iniciar el Agente SQL Server	73
Figura 58. Creación de un nuevo TRABAJO – EjecutarETL	74
Figura 59. Creación de un nuevo PASO – PasoEjecutarETL	74
Figura 60. Creación de una nueva programación – ProgramacionEjecutarETL	75
Figura 61. Creación de script en portapapeles	76
Figura 62. Creación de un nuevo trabajo - EjecutarOLAP	76
Figura 63. Creación de un nuevo paso – PasoEjecutarOLAP	77
Figura 64. Creación de una nueva programación – ProgramacionEjecutarOLAP	77
Figura 65. Cuadrante mágico de Gartner en el 2014	89
Figura 66. Cuadrante mágico de Gartner en el 2015	89
Figura 67. Cuadrante mágico de Gartner en el 2016	90

RESUMEN

Las empresas de hoy se ven en la necesidad de analizar su información para la correcta y oportuna toma de decisiones, la cual ayuda a encaminarla hacia los objetivos planteados. En el caso del grupo empresarial Leoncito con respecto a su toma de decisiones en su proceso de ventas se pudo identificar que para obtener información solicitada existía una demora de hasta 2 días, se desconocía el comportamiento de sus ventas en el tiempo, no contaban con reportes que permitan monitorear metas establecidas, no analizaban sus ventas desde diferentes perspectivas de negocio. Y además no se explotaban los datos históricos de la empresa.

La presente tesis se desarrolló para apoyar la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito mediante la implementación de un sistema de inteligencia de negocios utilizando algoritmo de serie temporal. Se utilizó la metodología Ralph Kimball y se usaron las herramientas Microsoft. El resultado que se obtuvo fue un sistema que logró disminuir el tiempo promedio que tomaba la obtención de información sobre ventas, generó reportes los cuales brindaron información analítica desde diferentes perspectivas, además se hicieron proyecciones para sus ventas; para las cuales se empleó el algoritmo de serie temporal.

PALABRAS CLAVE: Inteligencia de negocios, análisis de información, toma de decisiones, serie temporal.

ABSTRACT

Companies today are in need of analyzing their information for the correct and timely decision making, which helps to direct it towards the objectives set. In the case of the Leoncito business group with respect to its decision-making process in its sales process, it could be identified that there was a delay of up to 2 days to obtain the requested information, the behavior of its sales was not known at the time, there were no reports that allow monitoring established goals, did not analyze their sales from different business perspectives. And also the historical data of the company was not exploited.

This thesis was developed to support decision making in the sales process of the Leoncito business group through the implementation of a business intelligence system using time series algorithm. The Ralph Kimball methodology was used and the Microsoft tools were used. The result obtained was a system that managed to reduce the average time it took to obtain sales information, generated reports that provided analytical information from different perspectives, and made projections for their sales; for which the time series algorithm was used.

KEYWORDS: Business intelligence, information analysis, decision making, time series.

I. INTRODUCCIÓN

Hoy en día las tecnologías de información y comunicación (TICs) se han convertido en un factor competitivo para las empresas y estas están tomando conciencia que el análisis de la información es crucial para la dirección de cualquier organización. Este análisis ayuda a conocer lo que pasa dentro y fuera de la empresa de una manera rápida y fiable. Es entonces que a raíz de esto surgió la necesidad de implementar sistemas de inteligencia de negocios (Business Intelligence, BI) los cuales ayudan a obtener y analizar información relevante para la correcta y oportuna toma de decisiones (Muinelo 2016).

En el Perú según un estudio realizado en el periodo 2011-2012 sobre la inversión de las empresas peruanas (se realizó el estudio entre 150 organizaciones de los sectores público y privado) respecto a sus prioridades de gestión de las TICs mostraron a los sistemas de BI en el segundo lugar dentro de un ranking en el cual mantiene su posición desde el 2010, lo cual mostró un importante interés respecto a estos sistemas por parte de las empresas peruanas (Santana 2012).

En la realidad de un grupo empresarial de la región Lambayeque llamado Leoncito dedicados a la venta de artículos para el hogar como: cocinas, refrigeradoras, lavadoras, muebles, etc. necesitaban un sistema que les permita tener un acceso rápido a reportes con información sobre sus ventas desde diferentes perspectivas. Si bien contaba con un sistema de procesamiento de transacciones en línea (OLTP), este solo podía responder a las acciones del día a día. Entonces se vieron en la necesidad de contar con sistemas de procesamiento analítico en línea (OLAP) que apoyen en su toma de decisiones. A través de dos entrevistas realizadas se pudo identificar que:

- Para obtener información solicitada existía una demora que podía ser de hasta 2 días, porque siempre se tenía que recurrir a los encargados del área de tecnologías de información para solicitar un reporte. Los cuales se encontraban realizando sus labores diarios. Esto retrasaba la toma de decisiones, pues no contaban con sistemas diseñados para usuarios menos técnicos (miembros del grupo empresarial). Además el sistema que tenían solo brindaba reportes que según el gerente administrativo únicamente ayudaba en tomar decisiones a corto plazo (Ver anexo n°1 – pregunta 2 - pregunta 3, Ver anexo n°2 – pregunta 1 - pregunta 3).
- Existía un único reporte sobre ventas en el transcurso del tiempo, sin embargo este reporte no les permitía analizar el comportamiento que estas presentaban. Motivo por el cual no se podía analizar las ventas en periodos paralelos y por consiguiente se desconocía la existencia de patrones estacionales en las ventas (Ver anexo n°1 – pregunta 7, Ver anexo n°2 – pregunta 4).
- No contaban con reportes que permitan monitorear constantemente su principal meta de incrementar las ventas en 30% por local (sucursal) respecto al mismo periodo anterior ya sea por mes, trimestre, semestre o año; tampoco podían monitorear el top de vendedores por mes actual el cual era importante porque el vendedor que vendía más productos de marca leoncito recibía un premio (esto para incentivar a sus vendedores); no podían monitorear lo antes mencionado porque solo utilizaban hojas de cálculo o registros a papel para realizar el seguimiento de la meta y el top de vendedores anteriormente mencionado (Ver anexo n°1 – pregunta 6).

- Existían cero reportes sobre productos top por líneas o marcas de productos, de igual manera sobre sus clientes top ya sea por clientes naturales o jurídicos. Información que era requerida por los miembros del grupo empresarial. En el caso de los productos para poder analizar sus ventas en las diferentes líneas y marcas; así conocer aquellas que tienen mayor demanda. En el caso de los clientes para poder entablar relaciones más cercanas con ellos, pues eran clientes que debían ser fidelizados según el gerente administrativo del grupo empresarial. Sin embargo no se contaban con reportes que muestren dicha información (Ver anexo n°1 – pregunta 4, Ver anexo n°2 – pregunta 4).
- Existían cero reportes sobre las ventas de los diferentes locales en los diferentes años y meses. Por lo tanto desconocían el comportamiento de sus ventas independientes para cada uno de estos, además de no tener el top de los productos más vendidos por local, lo cual no permitía tomar las estrategias de negocio más adecuadas para cada uno de los locales (Ver anexo n°1 –pregunta 5).
- No se explotaban los datos históricos que se encontraban almacenados a través de los años de vida de la empresa, datos que al no ser procesados no apoyaban en el análisis de información de sus ventas, por lo tanto no se podía realizar comparaciones sobre sus ventas en diferentes periodos de tiempo y tampoco realizar predicciones sobre sus ventas (Ver anexo n°1 – pregunta 8).

Analizando la situación en la cual se encontraba el grupo empresarial se planteó la siguiente interrogante: ¿Cómo apoyar la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito?, para resolver dicha situación se propuso la siguiente hipótesis: Con la implementación de un sistema de BI utilizando algoritmo de serie temporal se apoyará la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito.

El objetivo general de la tesis fue apoyar la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito mediante la implementación de un sistema de BI utilizando algoritmo de serie temporal. Y como objetivos específicos se plantearon los siguientes:

- Reducir el tiempo promedio de obtención de información solicitada (respecto a ventas) por los miembros del grupo empresarial.
- Incrementar el número de reportes que permita análisis comparativo de las ventas en el tiempo.
- Incrementar el número de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas.
- Incrementar el número de reportes para el análisis de las ventas a través de las diferentes dimensiones.
- Incrementar el número de reportes sobre predicciones de ventas futuras.

La presente tesis se justificó desde el punto de vista social porque benefició a los miembros del grupo empresarial, debido a que les permitió analizar información relevante sobre sus ventas y gracias a esto pudieron tomar mejores decisiones estratégicas basadas en información existente.

Desde el punto de vista tecnológico la investigación se justificó porque el sistema de BI que se desarrolló buscó demostrar que la información histórica se puede utilizar como base para llegar al conocimiento y así tomar adecuadas decisiones estratégicas, la cual permitió a los miembros del grupo empresarial tener un enfoque más amplio respecto a sus ventas y así optar por mejores estrategias. Para el sistema se utilizaron las herramientas de Microsoft la cual posee funcionalidades para procesos de extracción, transformación y carga (ETL), creación de cubos multidimensionales, elaboración de reportes personalizados y uso de algoritmos de minería de datos para el pronóstico de ventas el cual para este caso se utilizó el algoritmo de serie temporal que hizo predicciones basadas en el conjunto de datos de la base de datos transaccional que posee la empresa; logrando realizar predicciones sobre sus ventas. Los reportes para los miembros del grupo empresarial se mostraron en una plataforma web a través de la cual se pudo acceder desde cualquier navegador web con acceso a internet.

Se justificó económicamente debido a que no hubo gastos monetarios para la empresa. Esto porque se contaba con licencia de SQL Server 2008 R2, la cual trae consigo la suite de BI, entonces el sistema que se propuso se desarrolló en la herramienta que ya se tenían disponible. Además el sistema de BI fue de gran apoyo para los miembros del grupo empresarial porque les brindó información útil la cual les ayudó en el análisis de sus ventas, traducándose en el planteamiento de estrategias adecuadas y oportunas que posteriormente hicieron más eficiente su toma de decisiones.

Para alcanzar la justificación científica el sistema propuesto se evaluó a través de un pre test y post test el cual se justificó en la mejora de la toma de decisiones con la implementación del sistema de BI con respecto a la toma de decisiones en el proceso de ventas antes de la implementación del sistema de BI. Además la presente tesis servirá como antecedente a futuras investigaciones relacionadas con la problemática.

II. MARCO TEÓRICO

2.1. Antecedentes

En Ecuador Boada y Tituaña (2012) desarrollaron una investigación titulada “Desarrollo de una aplicación de business intelligence para la empresa EMPAQPLAST” en la cual descubrió que la empresa ecuatoriana EMPAQPLAST tenía dificultades para analizar su información en las áreas de compra, venta y almacén por lo cual se perdía tiempo en la toma de decisiones debido a la falta de información inmediata. Para darle solución a este problema se desarrolló un sistema de BI para así poder tener un acceso inmediato a la información de la empresa. Para dicho sistema se construyó un data warehouse el cual contenía tres data marts (una para cada área: compra, venta y almacén) en la suite de Pentaho BI y se utilizó la metodología de Kimball para la construcción del data warehouse. El sistema de BI que se desarrolló logró facilitar el acceso y el análisis de la información en tiempo real desde cualquier sitio ya que se mostró en un entorno web. La relación de esta investigación con la presente tesis es que del mismo modo se buscó facilitar el acceso a la información, sin embargo este antecedente de investigación empleó la suite de Pentaho BI, mientras que la presente tesis usó la suite de Microsoft porque se acopló a la realidad del negocio donde se desarrolló. Como valor agregado se aplicó minería de datos para las ventas del grupo empresarial Leoncito.

En Mexico Aguilar (2012) desarrolló una investigación titulada “Modelado de datos multidimensional en el ámbito pre-hospitalario en cruz roja mexicana delegación La Paz Baja California” en la cual descubrió que la cruz roja mexicana carecía de conocimiento sobre los servicios médicos de urgencia y en cuanto a la atención de pacientes del lugar, por lo que se desarrolló un sistema de BI que logró satisfacer las necesidades en las diferentes áreas de la organización brindando información sobre los servicios médicos de urgencias tales como estimación de sucesos relacionados a la población, clasificación de las enfermedades más presentes en la población y clasificación de pacientes por procedencia. Para la construcción de dicho sistema se utilizó SQL Server 2008 para los data marts y para la construcción de tableros de control se usó C# y ASP.NET 4.0. La relación de esta investigación con la presente tesis es que el antecedente mencionado desarrolló un sistema de BI en el sector salud mientras que la presente tesis desarrolló un sistema de BI a una empresa dedicada a la venta de artículos para el hogar. Este antecedente realizó estimaciones de sucesos relacionados a la población para generar campañas de concientización lo cual es parecido al pronóstico que hizo la presente tesis para las ventas del grupo empresarial Leoncito utilizando de minería de datos.

En Mexico Sánchez (2015) desarrolló una investigación titulada “Automatización del Sistema de Medición de Desempeño para la toma de decisiones estratégicas de negocio: caso de estudio” en la cual descubrió que la asociación civil llamada Amor en Ser A. C. contaba con una serie de conflictos en la administración de su información tales como: Excesiva redundancia de la información, además de una deficiente integración de la información de la organización; lo cual dificultaba las tareas de análisis de la información para la toma de decisiones. Por lo cual se desarrolló un sistema de BI para solucionar dicha problemática. El sistema que se desarrolló agilizó la administración de la información, automatizó el análisis de la información de la organización, optimizó procesos de almacenamiento, análisis y presentación de información relevante para la toma de decisiones. Este sistema se desarrolló en base al ciclo de vida del desarrollo de sistemas que propone el autor James A. Senn y el ciclo de vida de desarrollo del autor Kendall, mientras que la presente tesis utilizó la metodología de Kimball. El software que utilizó este antecedente fueron herramientas de libre distribución mientras que la

presente tesis usó la suite de Microsoft porque se acoplaba a la realidad del negocio donde se implementó. Este antecedente implementó BI en una empresa prestadora de servicios, mientras que la presente tesis lo implementó en una empresa dedicada a la venta de artículos para el hogar. El valor agregado de la presente tesis fue aplicar minería de datos para la predicción de ventas.

En Lima Coronel (2012) desarrolló una investigación titulada “Diseño de un datamart para seguros masivos” en la cual descubrió que la empresa de seguros llamada Hermes Asesores quería reducir costos, tiempo y optimizar sus procesos de riesgos masivos para poder lograr una ventaja competitiva por lo cual desarrolló un sistema de BI para seguros masivos el cual integró y automatizó los procesos de información masiva de pagos de seguros y así le permitió reducir los costos y el tiempo de procesamiento de información. Se utilizó SQL Server 2008, la metodología de Kimball y RUP de desarrollo de software. La relación de este antecedente con la presente tesis es que este implementó un sistema de BI para servicios de seguros mientras que la presente tesis lo implementó en una empresa dedicada a la venta de artículos para el hogar. El antecedente utilizó la metodología Kimball y SQL Server 2008. De la misma manera en la presente tesis se usó la metodología de Kimball y SQL Server 2008 R2 porque se acopla a las necesidades del negocio. Además se utilizó minería de datos para la predicción de ventas.

En Cajamarca Guillen (2012) desarrolló una investigación titulada “Desarrollo de un data mart para mejorar la toma de decisiones en el área de tesorería de la municipalidad provincial de Cajamarca” en la cual descubrió que la municipalidad provincial de Cajamarca no sabía cómo administrar adecuadamente la gran cantidad de información que se manejaba diariamente debido a que en aquel entonces no se contaba con un sistema que soporte dicho manejo de información por lo que desarrolló un sistema de BI que de soporte a las necesidades de información hacia los usuarios finales y así poder visualizar información relevante, la cual era vital para tomar decisiones adecuadas. Este antecedente utilizó MySQL y la suite de Pentaho BI para la construcción del data mart. La presente tesis utilizó las herramientas de Microsoft porque se acoplaban a la realidad del negocio donde se aplicó el sistema de BI. Como valor agregado la presente tesis utilizó algoritmo de minería de datos para realizar predicciones respecto a las ventas.

En Lima Córdova (2013) desarrolló una investigación titulada “Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de importaciones en una empresa comercializadora/importadora” en la cual descubrió que una empresa importadora comercializadora tenía inconvenientes en la obtención de información para la toma de decisiones relacionadas al área de importaciones, además no eran cubiertas de forma adecuada las áreas de compras, ventas y almacén; entonces con dicha problemática la empresa para resolver dichos inconvenientes desarrolló un sistema de BI con las herramientas de Microsoft y utilizó la metodología de Kimball. El sistema de BI creado permitió un manejo intuitivo y sencillo a los usuarios finales para generar sus propios reportes y así poder realizar el análisis correspondiente de las diferentes áreas para la adecuada toma de decisiones. La relación que tiene este antecedente con la presente tesis fue que utilizó de la misma manera las herramientas de Microsoft y también la metodología de Kimball debido a que se ajustaba a las necesidades del negocio. Aparte la presente tesis usó algoritmo de minería de datos para predicción de ventas.

En Chiclayo Millones (2012) desarrolló una investigación titulada “Implementación de un sistema de comercio electrónico basado en CRM y Balanced ScoreCard como herramienta para la toma de decisiones en la empresa Inversiones Vialsa de la localidad de Chiclayo” en la cual descubrió que la empresa Inversiones Vialsa tenía problemas y demoras en la obtención de reportes porque estos se elaboraban manualmente usando una calculadora la cual producía errores y demandaba tiempo extra en la verificación de documentos haciendo que la información no llegue en el tiempo oportuno al gerente para tomar decisiones en el momento adecuado. Además necesitaban un sistema que brinde reportes respecto a sus compras y ventas para mejorar su toma de decisiones y también se vieron en la necesidad de acercarse a sus clientes y fidelizarlos. Por esto fue que se desarrolló un sistema de comercio electrónico basado en CRM y BSC usando como lenguaje de programación Visual.Net y MySQL para reducir el tiempo y mejorar la calidad de los reportes y así apoyar en la toma de decisiones, también se logró obtener la información en tiempo real y con precisión. Este antecedente elaboró un BSC a partir de la información que se recopilaba del negocio, y tiene relación con la presente tesis porque el sistema de BI que se desarrolló pretende apoyar la toma de decisiones. Como valor agregado se utilizó minería de datos para la predicción de las ventas.

En Chiclayo Salazar (2013) desarrolló una investigación titulada “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.RL.” en la cual descubrió que la empresa tenía problemas de inadecuado almacenamiento de la información de sus transacciones comerciales por lo que se desarrolló una solución de BI para así mejorar el proceso de toma de decisiones en los procesos de venta. Se utilizó la técnica de Data Profiling para la limpieza de los datos del sistema transaccional para lo cual se usó la herramienta de Microsoft. Después de la limpieza de los datos estos se almacenaron en un data warehouse y se obtuvieron indicadores de crecimiento de un 5% en las ventas; esto debido a las mejores decisiones tomadas de acuerdo a la información que les brindó la solución de BI desarrollada. Este antecedente desarrolló una solución de BI para el sector comercial empleando la suite de BI de Microsoft así como en la presente tesis se desarrolló un sistema de BI a una empresa dedicada a la venta de artículos para el hogar utilizando la suite de BI de Microsoft porque se acoplaba a la realidad del negocio. Como valor agregado se usó algoritmo de minería de datos para predicción de ventas.

En Chiclayo Bravo (2016) desarrolló una investigación titulada “Solución de inteligencia de negocios utilizando algoritmos de serie temporal para apoyar la toma de decisiones en el proceso de ventas de una clínica de la ciudad de Chiclayo.” en la cual descubrió que en una clínica de Chiclayo habían problemas en la toma de decisiones, porque existía demora en la obtención de información sobre las ventas de sus servicios, satisfacción por parte de la gerencia de 37.5% sobre obtención de la información y además carecían de indicadores, proyecciones y análisis de su realidad desde diferentes perspectivas del negocio. Se construyó una solución de BI, se utilizó algoritmo de serie temporal para minería de datos y la metodología Kimball para el desarrollo de la solución con la cual se disminuyó el tiempo promedio para la obtención de información sobre ventas y se generó información que ayude a la toma de decisiones con la cual se incrementó el nivel de satisfacción de la gerencia. Este antecedente implementó BI para el sector salud, mientras que la presente tesis se desarrolló para una empresa dedicada a la venta de artículos para el hogar. Se utilizó algoritmo de serie temporal y la metodología de Kimball debido a que se acoplaban a la problemática encontrada.

2.2. Bases Teórico Científicas

2.2.1. Toma de decisiones:

2.2.1.1. Definición:

Según Miñano (2017) es el proceso a través del cual vamos a elegir entre alternativas o caminos a seguir sobre lo que viene ocurriendo en la gestión de la empresa; sin embargo, para llegar al punto de la decisión propiamente dicha, debemos pasar por una etapa en la cual evaluamos, nos informamos, comparamos o quizá incluso consultamos con nuestra intuición, antes de decir: “VA o NO VA” y de la forma X, Y, Z.

2.2.1.2. Tipos de toma de decisiones:

Según Roman (2004) en las empresas se toman diferentes tipos de decisiones dependiendo de la información, recursos disponibles y en función de los objetivos. Las cuales se pueden clasificar en decisiones estratégicas, tácticas y operativas.

- **Decisiones estratégicas:**

Implican asignación de recursos importantes con impacto a largo plazo sobre el conjunto de la empresa. Permite que la alta dirección de las empresas pueda analizar y monitorear objetivos estratégicos de la organización.

- **Decisiones tácticas:**

Tienen impacto a mediano plazo. Su objetivo es movilizar recursos de la organización para desarrollar las decisiones estratégicas que dependen de cada área de la empresa.

- **Decisiones operativas:**

Se trata de decisiones a corto plazo, o sea las decisiones rutinarias sobre tareas específicas. En caso de desviaciones o errores se pueden corregir rápidamente.

2.2.2. Inteligencia de negocios:

2.2.2.1. Definición:

Según Lluís (2008) es apoyar de forma sostenible a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones.

Para Sinnexus (2016) es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Una solución BI permite:

- Observar: ¿Qué está ocurriendo?
- Comprender: ¿Por qué ocurre?
- Predecir: ¿Qué ocurriría?
- Colaborar: ¿Qué debería hacer el equipo?
- Decidir: ¿Qué camino se debe seguir?

En mi opinión BI es la combinación de tecnología, herramientas y procesos que transforman los datos en información para finalmente transformarse en conocimiento y este ser dirigido a una estrategia empresarial. Para así tener la capacidad de tomar oportunas y correctas decisiones. Por lo tanto obtener mejores resultados en el futuro.

2.2.2.2. Beneficios:

Lluís (2008) describe tres tipos de beneficios que se pueden obtener a través del uso de BI, los cuales son:

- **Beneficios tangibles:**
Por ejemplo, reducción de costos, aumentar las ventas, reducción de tiempos para las distintas actividades del negocio, etc.
- **Beneficios intangibles:**
Por ejemplo, el hecho de que se tenga disponible la información para la toma de decisiones hará que los usuarios la utilicen para tomar mejores decisiones y mejorar la posición competitiva.
- **Beneficios estratégicos:**
Por ejemplo, la formulación de estrategias especificará a qué clientes, mercados o a qué productos dirigirse.

2.2.2.3. Características:

Según Rodero (2010) para que un sistema sea considerado de BI debe garantizar las siguientes características:

- **Accesibilidad:**
Debe permitir el acceso rápido de los usuarios a la información, siendo independiente de la procedencia de los mismos.
- **Orientada al usuario:**
Es independiente de los conocimientos técnicos que puedan poseer los distintos usuarios.
- **Apoyo a la toma de decisiones:**
No puede ser una simple presentación de información, sino que debe poseer herramientas de análisis que permitan seleccionar y manipular aquellos datos de análisis que sean necesarios para el usuario.

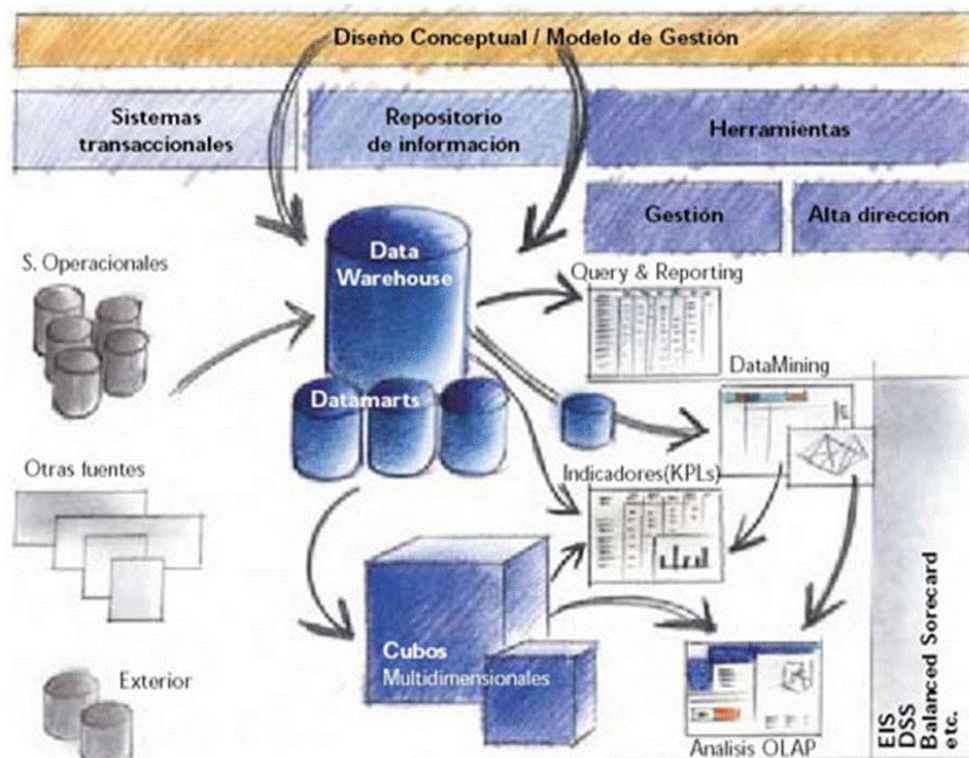


Figura 1. Sistema de BI

Fuente: (Ibermaticca 2006)

2.2.3. Data Warehouse:

2.2.3.1. Definición:

Para Rodríguez (2015) es como una base de datos corporativa donde se integra y depura información de una o varias fuentes distintas, que luego serán procesadas y analizadas desde distintas perspectivas y a grandes velocidades de respuesta.

Sinnexus (2016) define al data warehouse como una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde diferentes perspectivas y con grandes velocidades de respuesta.

2.2.3.2. Características:

Sinnexus (2016) dice que Bill Inmon, caracterizó al data warehouse por ser:

- **Integrado:**

Los datos almacenados en el data warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminados.

- **Temático:**

Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del data warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

- **Histórico:**

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

- **No volátil:**

El almacén de información de un data warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del data warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

2.2.3.3. Metodologías de desarrollo:

Investigando en diferentes fuentes se encontraron dos metodologías principales para el desarrollo de un data warehouse las cuales son la metodología de Ralph Kimball y metodología de Bill Inmon.

- **Metodologia de Ralph Kimball “Bottom up”:**

La metodología del ciclo de vida Kimball fue concebido a mediados de 1980 por los miembros del Grupo de Kimball y otros colegas en Metaphor Computer Systems.

La metodología de Ralph Kimball indica que el data warehouse es un conglomerado de todos los data marts dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo, al modelo dimensional (no normalizado) que incluyen las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar.

Este enfoque también se referencia como Bottom-up, pues al final el data warehouse corporativo no es más que la unión de los diferentes data marts, que están estructurados de una forma común, a través de la estructura bus. Esta característica permite que sea flexible y sencillo de implementar, pues se puede construir un data mart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones, ya definidas u otras nuevas.

La figura 2 muestra las fases de la metodología de Ralph Kimball.

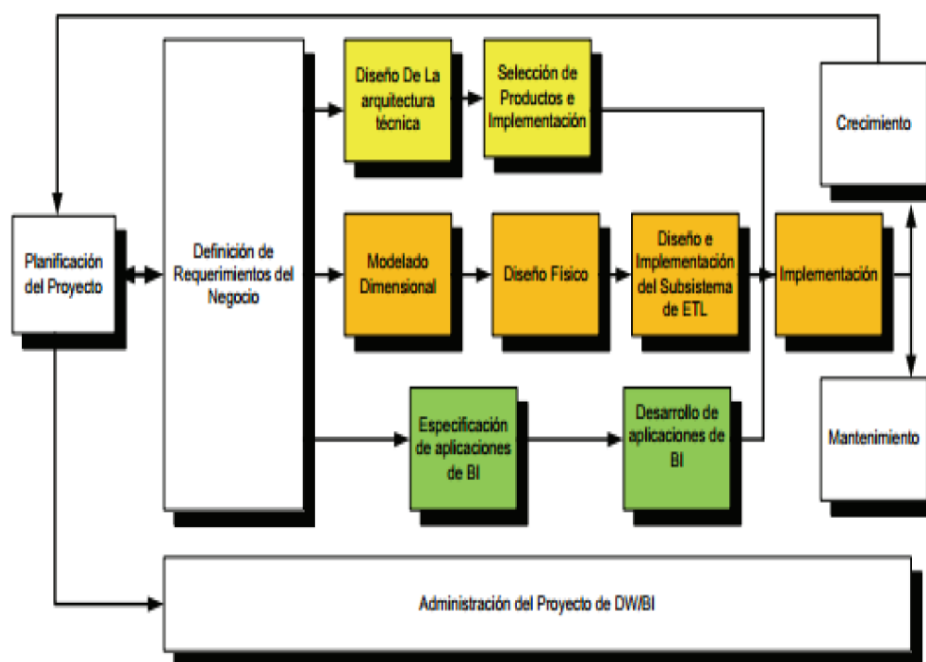


Figura 2. Fases de la metodología Kimball

Fuente: (Kimball 2002)

Fases de la metodología de Kimball:

A. Planificación del proyecto:

En esta fase se determina el propósito del proyecto de data warehouse, sus objetivos específicos y el alcance del mismo, evaluación de factibilidad y una aproximación inicial a las necesidades de información.

B. Definición de requerimientos del negocio:

Es la interpretación correcta de los diferentes niveles de requerimientos expresados por los diferentes niveles de usuarios. Se debe aprender tanto como se pueda sobre el negocio, los competidores, la industria y los clientes del mismo. A partir de la definición de requerimientos del negocio se puede construir una herramienta de la metodología denominada matriz de procesos/dimensiones (Bus Matrix en inglés). Una dimensión es una forma, vista o criterio por medio de cual se pueden resumir, cruzar o cortar datos numéricos a analizar, datos que se denominan medidas. Esta matriz tiene en sus filas los procesos de negocio identificados, y en las columnas, las dimensiones identificadas. Un ejemplo de esta matriz se puede observar en la tabla 6. Cada x en la intersección de las filas y columnas significa que en el proceso de negocio de la fila seleccionada se identifican las dimensiones propuestas.

Proceso de negocio	Dimensiones					
	Tiempo	Producto	Empleados	Clientes (Revendedores)	Geografía de ventas	Importes
Proyección de ventas	X	X	X	X	X	X
Compras	X	X	X	X	X	X
Control de llamadas	X	X	X	X	X	X
...						

Tabla 1. Ejemplo de Matriz Bus

Fuente: (Rivadera 2011)

C. Modelado dimensional:

La creación de un modelo dimensional es un proceso dinámico e iterativo. Comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados de la matriz bus. El proceso iterativo consiste en cuatro pasos: Elegir el proceso de negocio, establecer el nivel de granularidad, elegir las dimensiones e identificar medidas y las tablas de hechos.

D. Diseño físico:

Algunos de los elementos principales de este proceso son la definición de estándares de nombres y seteos específicos del entorno de la base de datos. Aquí es donde se prepara el entorno de base de datos y la seguridad apropiada de éste.

E. Diseño y desarrollo de presentación de datos:

En esta parte se encuentra el proceso ETL. Se define como proceso de extracción a la obtención de los datos que permitirán efectuar la carga del modelo físico acordado, proceso de transformación el convertir o recodificar los datos fuente, a fin poder efectuar la carga efectiva del modelo físico y como procesos de carga de datos son los requeridos para poblar el data warehouse.

F. Diseño de la arquitectura técnica:

Depende de tres factores importantes: Los requerimientos del negocio, el ambiente tecnológico actual y el plan estratégico planeado por la entidad donde se desarrollará el data warehouse.

G. Selección de productos e instalación:

De acuerdo al diseño de la arquitectura técnica se procederá a evaluar y seleccionar componentes específicos de la arquitectura tales como la plataforma de hardware, el motor de base de datos, la herramienta de ETL, la herramienta de reportes, etc. Una vez evaluados y seleccionados los componentes determinados se procede con la instalación y prueba de los mismos.

H. Especificación de aplicaciones para usuarios finales:

No todos los usuarios necesitan el mismo nivel de análisis, y es por esto que se determinan roles o perfiles de usuario de acuerdo al nivel de análisis que cada uno de ellos requiere.

I. Desarrollo de aplicaciones para usuarios finales:

El desarrollo de aplicaciones para usuarios finales involucrará la construcción de reportes específicos. Una vez que se ha cumplido con todos los pasos de la especificación y se tiene la posibilidad de trabajar con algunos datos de prueba, comienza el desarrollo de la aplicación.

J. Despliegue:

Consiste en el acceso que el usuario final tiene desde su lugar de trabajo a la aplicación construida. Para conseguirlo será necesario la realización de capacitaciones, soporte técnico, comunicación, retroalimentación, etc. Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al sistema.

K. Mantenimiento y crecimiento:

Dado que el ciclo de vida dimensional del negocio (Business Dimensional Lifecycle, BDL) de Kimball es un proceso de etapas con comienzo y fin pero de naturaleza espiral que acompaña la evolución de la organización durante toda su historia, el data warehouse construido será capaz de evolucionar y crecer, pero debe tenerse en cuenta el establecimiento de las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma conseguir crecimiento sostenido.

L. Gerenciamientos del proyecto:

Asegura que las actividades del BDL se lleven en forma sincronizada. El gerenciamiento acompaña todo el ciclo de vida. Aquí es donde se monitorea el proyecto y la comunicación entre los requerimientos del negocio. Asimismo, determina las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

- **Metogia de Bill Inmon “Top down”:**

Bill Inmon asocia frecuentemente los data warehouse a nivel empresarial, que involucran desde un inicio todo el ámbito corporativo, sin centrarse en un incremento específico hasta después de haber terminado completamente el diseño del data warehouse. En su filosofía, un data mart es sólo una de las capas del data warehouse y los data mart son dependientes del data warehouse corporativo y por lo tanto se construyen después del data warehouse (Tufiño 2011).

Inmon vio la necesidad de transferir data de diversos sistemas OLTP a un repositorio central, la cual pueda estar disponible para ser analizada. Los data marts son tratados como subconjuntos del data warehouse, cada uno es construido para un departamento individual y es optimizado para las necesidades de análisis de cada área para el cual es creado. Los datos son extraídos de las fuentes de datos operacionales y es cargada al extraer, validar y consolidar para asegurar el nivel de precisión, toda esta data es cargada en tercera forma normal. Los datos normalizados son usados para llenar repositorios adicionales de presentación de data (Coronel 2012).

La figura 3 muestra las fases de la metodología de Bill Inmon.

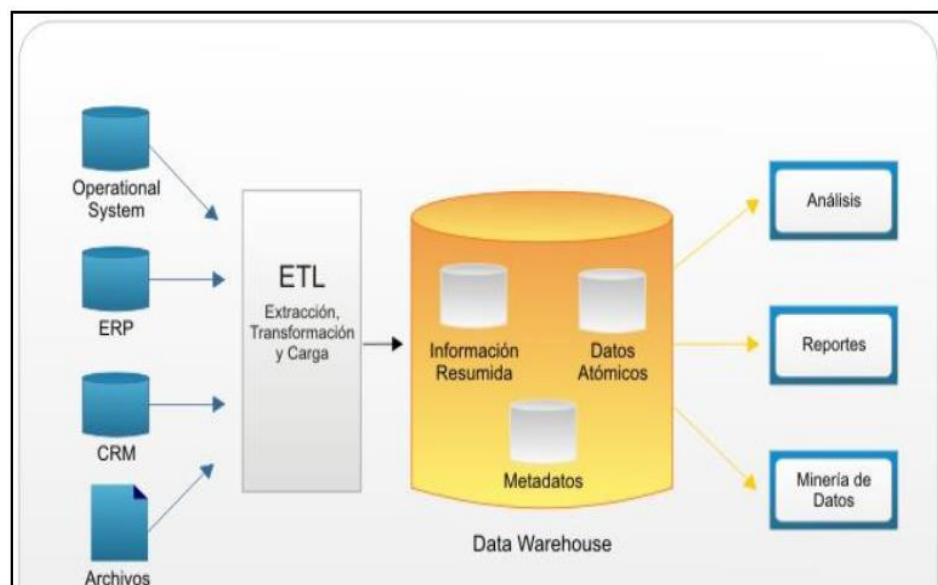


Figura 3. Metodología Inmon

Fuente: (Inmon 2005)

A continuación la tabla 1 muestra una comparación entre la metodología de Ralph Kimball, y la metodología de Bill Inmon.

	Kimball	Inmon
Filosofía de desarrollo	Se basa en la priorización de algunos procesos específicos del negocio. Desarrollo directo del Data Mart en los procesos seleccionados del negocio. Uso exclusivo de modelos dimensionales desnormalizados.	Se basa en el modelo de datos de toda la empresa. Desarrollo de un Data Warehouse empresarial basado en esquemas de base de datos normalizadas y los Data Mart se obtienen del Data Warehouse.
Enfoque	Tiene un enfoque por áreas del negocio. Trata de responder necesidades específicas según el tema. (Bottom-up)	Tiene un enfoque global de toda la empresa, no por áreas de negocio. (Top-Down)
En qué tipo de empresa se puede aplicar	En pequeñas y medianas empresas.	No ayuda mucho a las pequeñas y medianas empresas.
Tiempo de Implementación	Debido a que primero se implementan los data marts, el tiempo de implementación es rápido, por etapas.	Más tiempo para implementar debido a que son varios Data Mart por cada área.
La presentación de información para ser almacenada	La estructura de los datos requiere presentación dimensional.	Es necesario normalizar la información antes de almacenarla.
Costo	Costo es menor debido a que es aplicado a un proceso de negocio y el volumen de datos es menor.	Costo es mayor porque es aplicado a toda la organización y el volumen de datos es mayor.

Tabla 2. Comparación entre la metodología de Kimball e Inmon

Fuente: Adaptado de Espinoza y Palomino (2016) y Bravo (2016)

Al establecer una comparación entre las dos metodologías más importantes, en el caso de la presente tesis, se decidió utilizar la metodología de Ralph Kimball debido a que ésta permite el almacenamiento de la información en forma de dimensiones y sin normalizar la data (aspecto que sí considera Inmon) lo cual hace las consultas mucho más rápidas. Además proporciona un enfoque de menor a mayor donde se pueden implementar pequeños data marts en áreas específicas (en este caso: ventas) con pocos recursos y de a poco ir integrándolos en un data warehouse.

2.2.3.4. Proceso ETL:

Según Eckerson y White (2003) este proceso combina información proveniente de distintas fuentes de datos y la almacenan en el data Warehouse, permitiendo a los usuarios disponer de un almacén de datos único, con la información integrada, limpia, estructurada, consistente y preparada para el análisis. El Data Warehouse ofrece una sola versión de la verdad.

Curto y Conesa (2010) dicen que el proceso ETL es como el conjunto de aplicaciones, productos, técnicas y tecnologías que permiten una visión única consistente de los datos del negocio. En el contexto de BI, las herramientas ETL han sido la opción usual para alimentar el data warehouse.

Para SAS (2016) es un tipo de integración de datos que se refiere a los tres pasos (de extracción, transformación, carga) para mezclar datos de múltiples fuentes. A menudo se utiliza para construir un almacén de datos. Durante este proceso, se toma datos (extraído) a partir de un sistema de origen, convertida (transformada) en un formato que puede ser analizada, y se almacena (cargado) en un almacén de datos.

Según Lluís (2008) el proceso ETL se divide en 5 subprocesos:

- **Extracción:**
Recupera los datos físicamente de las distintas fuentes de información.
- **Limpieza:**
Recupera los datos y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores vacíos, es decir se transforman los datos siempre que sea posible para reducir los errores de carga.
- **Transformación:**
Recupera los datos limpios, de alta calidad, los estructura y sumarla en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios, consistentes, resumidos y útiles.
- **Integración:**
Valida que los datos que se cargan en el data warehouse sean consistentes con las definiciones y formatos del data warehouse. Los integra en los distintos modelos de las distintas áreas de negocio que se han definido en el mismo. Estos procesos pueden ser complejos.
- **Actualización:**
Es el que permite añadir los nuevos datos al data warehouse.

2.2.3.5. Data mart:

Para Ibermaticca (2006) es una base de datos especializada, orientada a satisfacer las necesidades específicas de un grupo particular de usuarios (por un departamento o grupo de usuarios en una compañía, para un conjunto definido de tareas).

Rodriguez y Mendoza (2011) definen al data mart como un conjunto de datos que son estructurados de una forma que facilite su posterior análisis, los data mart contienen la información referente a un área, un tema o una función en particular del negocio, con datos relevantes que provienen de las diferentes aplicaciones operacionales. Los data marts pueden ser de diversas bases de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar.

Según Vizuite y Yela (2006) en su investigación de tesis explica dos tipos de data mart:

- **Dependientes:**

Son los que se construyen a partir de un data warehouse central, es decir reciben sus datos de un repositorio empresarial central.

- **Independientes:**

Son aquellos que no dependen de un data warehouse central, ya que pueden recibir los datos directamente del ambiente operacional, ya sea mediante procesos internos de las fuentes de datos o de almacenes de datos operacionales.

De los dos tipos de data mart, en la presente tesis se creó un data mart de tipo independiente para las ventas del grupo empresarial Leoncito, esto debido a que se alimento al data mart que se creo directamente desde la base de datos trasaccional.

2.2.3.6. Cubo OLAP:

Para Microsoft (2016) es una estructura de datos que supera las limitaciones de las bases de datos relacionales y proporciona un análisis rápido de datos, los cubos pueden mostrar y sumar grandes cantidades de datos, a la vez que proporcionan a los usuarios acceso mediante búsqueda a los puntos de datos. De este modo, los datos se pueden resumir o reorganizar según sea necesario.

Rosado (2010) en su investigación clasifica al cubo olap en base a las siguientes categorías de acuerdo a su modo de almacenamiento.

- **ROLAP:**
Implementación que almacena los datos en un motor relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas.
- **MOLAP:**
Esta implementación almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado.
- **HOLAP (Hybrid OLAP):**
Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

De las diferentes categorías de acuerdo a su modo de almacenamiento OLAP se decidió optar por el modo de almacenamiento MOLAP, esto debido a que se construyó una base de datos multidimensional la cual se alimentó desde las diferentes fuentes de datos del grupo empresarial Leoncito.

A continuación la tabla 2 muestra una comparación entre OLTP y OLAP.

	OLTP	OLAP
Objetivo	Brindar soporte a las operaciones diarias del negocio (Tareas de lectura/escritura).	Brindar soporte en el análisis del negocio, identificando tendencias, comparando períodos, etc. mediante el almacenamiento de datos históricos.
Alineación de datos	Están alineados por aplicación. Diferentes sistemas tienen distintos tipos de datos, los cuales son estructurados por aplicación. Se focaliza en el cumplimiento de requerimientos de una aplicación en especial o una tarea específica.	Están alineados por dimensión. Todos los tipos de datos integrados en un solo sistema. Los datos son organizados definiendo dimensiones del negocio (áreas temáticas o sujetos). Se focaliza en el cumplimiento de requerimientos del análisis del negocio.
Integración de datos	Los datos se encuentran estructurados independientemente uno de otros, pudiendo tener diferentes estructuras de claves y nombres.	Los datos deben estar integrados, consolidados y orientados a un área de análisis.
Historicidad de información	Información cambiante en el tiempo.	Información almacenada como fotografías en el repositorio único de datos.

Acceso y manipulación de datos	Realizan una manipulación de datos registro por registro usando sentencias de INSERT, UPDATE y DELETE. Además, necesitan de rutinas de validación y transacciones a nivel registro como el COMMIT y el ROLLBACK	Tienen carga y acceso masivo de datos, no se usan sentencias de INSERT, UPDATE o DELETE. La validación de datos se realiza antes o después de la carga. Principalmente, se realizan sentencias de SELECT sobre varios registros y tablas, teniendo grandes volúmenes de datos involucrados en un único proceso o análisis. La estructura de la base de datos es desnormalizada de manera que reduce el tiempo de consulta y proceso de datos.
Granularidad de datos	Se encuentran los datos a nivel detallado o nivel transaccional. Una transacción incluye a nivel atómico cada uno de los componentes de su estructura (fecha, hora, código de cliente, código de movimiento, importe, etc.)	La granularidad de los datos viene dada por el uso de los mismos. Si bien un data warehouse puede tener información a nivel transaccional, el objetivo de esta granularidad mínima está asociado con el deseo de realizar ciertos tipos de análisis que requieren que la información esté a ese nivel de detalle, pero no significa que se vea la información a nivel transaccional.
Tiempo de respuesta	Largos tiempos de respuesta, ya que las consultas complejas de datos suelen implicar la unión de tablas de gran tamaño lo que se traduce en una incómoda espera.	Cortos tiempos de respuesta debido a que la solución de BI está orientada a las consultas.
Perfil de Usuario	El perfil de usuario que interactúa con dichos sistemas se encuadra dentro de los empleados operacionales de una organización (comunidad operativa).	El perfil de usuario sobre este tipo de sistemas corresponde a la comunidad gerencial, la cual está a cargo de la toma de decisiones.

Tabla 3. Comparación entre OLTP y OLAP

Fuente: Adaptado de Rojas (2014) y Bravo (2016)

Analizando el cuadro comparativo se concluye que optar por sistemas OLAP es adquirir una ventaja competitiva frente a los sistemas OLTP. Para superar las limitaciones que tiene los sistemas OLTP se tiene a los sistemas OLAP que se apoyan de un conjunto de herramientas que facilitan la extracción, la limpieza, y el almacenamiento de los datos generados en una organización, con la velocidad adecuada para mostrar la información requerida y apoyar en la toma de decisiones de los directivos.

2.2.4. Minería de datos:

2.2.4.1. Definición:

Según Lara (2012) es como una disciplina de la informática que estudia el análisis de grandes cantidades de datos con el objetivo de obtener conocimiento a partir de ellos.

Sinnexus (2016) lo define como el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Para Microsoft (2016) es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

2.2.4.2. Algoritmo de minería de datos:

Para Microsoft (2016) es conjunto de cálculos y reglas heurísticas que permite crear un modelo de minería de datos a partir de los datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias.

El algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo de minería de datos. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas.

El modelo de minería de datos es un conjunto de datos, estadísticas y patrones que se pueden aplicar a los nuevos datos para generar predicciones y deducir relaciones. Puede tomar diversas formas, incluyendo:

- Un conjunto de clústeres que describe cómo se relacionan los casos de un conjunto de datos.
- Un árbol de decisión que predice un resultado y que describe cómo afectan a este los distintos criterios.
- Un modelo matemático que predice las ventas.
- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción, y las probabilidades de que dichos productos se adquieran juntos.

2.2.4.3. Tareas de la minería de datos:

Según Hernández, Ramírez y Ferri (2004) dentro de la minería de datos se debe de distinguir tipos de tareas las cuales pueden considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Estas tareas pueden ser predictivas o descriptivas.

- **Tareas predictivas:**

- **Clasificación:**

En esta cada registro de la base de datos pertenece a una clase, la cual se indica mediante el valor de un atributo llamada la clase del registro. El objetivo es maximizar la razón de precisión de clasificación de los nuevos registros. Predicen una o más variables discretas.

- **Regresión:**

Consiste en aprender una función real que asigna a cada registro un valor real. El objetivo es minimizar el error entre el valor predicho y el valor real. Predicen una o más variables numéricas continuas, como pérdidas o ganancias, basándose en otros atributos del conjunto de datos.

- **Tareas descriptivas:**

- **Agrupamiento:**

Consiste en obtener grupos “naturales” a partir de los datos. Se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Se suele llamar segmentación porque segmenta o parte los datos en grupos que pueden ser o no disjuntos. Dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares.

- **Correlación:**

Se usa para examinar el grupo de similitud de los valores de dos variables numéricas. Puede ser muy útil para establecer reglas de ítems correlacionados.

- **Asociación:**

Similar a las correlaciones, su objetivo es identificar relaciones no explícitas entre atributos categóricos. Se utiliza frecuentemente en el análisis de la cesta de la compra, para identificar productos que son frecuentemente comprados juntos, información que puede usarse para ajustar los inventarios, para la organización física del almacén o campañas publicitarias.

➤ **Asociación secuencial:**

Es un caso especial de reglas de asociación, se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo. Resumen las secuencias frecuentes o episodios en los datos, como una serie de clics en un sitio web o una serie de eventos de registro que preceden al mantenimiento del equipo.

En la presente tesis se realizó una tarea predictiva de regresión, esto porque se buscó predecir el monto de las ventas y las unidades vendidas por producto del grupo empresarial Leoncito, las cuales son variables numéricas continuas.

2.2.5. Herramientas para el desarrollo del sistema de BI:

Para el desarrollo del sistema de BI se usó la suite de BI de Microsoft SQL Server 2008 R2 por las siguientes razones:

- El grupo empresarial Leoncito ya contaba con Microsoft SQL Server 2008 R2 en la cual viene incluida la suite de BI, por lo tanto no habría costos en licencia.
- Se tuvo como requerimiento no funcional por parte del grupo empresarial Leoncito desarrollar el sistema a través de la suite de BI de SQL Server 2008 R2, debido a que ya contaban con una arquitectura empresarial definida.
- Microsoft a lo largo de los años se ha mantenido como una de las plataformas líderes en el cuadrante mágico de Gartner. Mostrando como estas se enfocan al negocio respondiendo a las necesidades de las organizaciones para una mayor accesibilidad y visión analítica de información relevante. (Ver anexo 4).

Debido a que se usaron las herramientas de Microsoft para desarrollar el sistema de BI se usó: Integration Services, Analysis Services y Reporting Services.

2.2.5.1. Integration Services:

Es una plataforma para la construcción de soluciones de integración de datos y flujo de trabajo de alto rendimiento. Permite la creación de paquetes SSIS que se componen de las tareas que se pueden mover los datos desde el origen al destino y modificarla si es necesario. Es básicamente una herramienta cuyo objetivo principal es hacer la extracción, transformación y carga de datos (Arshad 2016).

Permite extraer y transformar datos de muchos orígenes distintos, como archivos de datos XML, archivos planos y orígenes de datos relacionales, y posteriormente cargarlos en uno o varios destinos. 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 mensajería de correo electrónico, orígenes y destinos de datos para extraer y cargar datos, transformaciones para limpiar, agregar, combinar y copiar datos (Microsoft 2016).

2.2.5.2. Analysis Services:

Permite diseñar, crear y administrar estructuras multidimensionales que contienen datos de detalle y agregados procedentes de varios orígenes de datos, como bases de datos relacionales, en un único modelo lógico unificado compatible con los cálculos integrados. Almacenan cubos OLAP, los cuales se pueden considerar como la última pieza del rompecabezas para una solución de almacenamiento de datos, que se genera mediante bases de datos OLAP para permitir el análisis casi instantáneo de datos. Proporciona análisis rápido, intuitivo y descendente de grandes cantidades de datos generados en este modelo de datos unificado, que se puede entregar a los usuarios finales, además de admitir el análisis de datos históricos y en tiempo real. También contiene las características y herramientas necesarias para crear complejas soluciones de minería de datos (Microsoft 2016).

Para la minería de datos en el caso de la presente tesis se propuso predecir el monto de ventas y las unidades vendidas por producto para los siguientes 6 meses.

- **Algoritmo de serie temporal:**

El algoritmo de serie temporal de Microsoft proporciona varios algoritmos optimizados para la previsión en el tiempo de valores continuos. Un modelo de serie temporal puede predecir tendencias basadas únicamente en el conjunto de datos original utilizado para crear el modelo. En este algoritmo se puede usar 2 algoritmos:

Algoritmo ARTxp (AutoRegressive Tree model): Utilizado para la predicción a corto plazo.

Algoritmo ARIMA (AutoRegressive Integrated Moving Average): Utilizado para la predicción a largo plazo.

De forma predeterminada, el algoritmo de serie temporal de Microsoft utiliza una mezcla de los dos algoritmos al analizar patrones y realizar predicciones. El algoritmo entrena dos modelos independientes sobre los mismos datos: uno de los modelos usa el algoritmo ARTXP y el otro modelo usa el algoritmo ARIMA. A continuación, el algoritmo combina los resultados de los dos modelos para obtener la mejor predicción sobre un número variable de intervalos de tiempo. Dado que ARTXP obtiene mejores resultados en las predicciones a corto plazo, se le da mayor importancia al principio de una serie de predicciones. Sin embargo, a medida que los intervalos de tiempo que se están prediciendo se adentran en el futuro, se va dando más importancia a ARIMA. Se puede especificar el algoritmo a usar:

- Utilizar solo ARTXP para la predicción a corto plazo.
- Utilizar solo ARIMA para la predicción a largo plazo.
- Utilizar la mezcla predeterminada de los dos algoritmos (MIXED).

Requisitos para un modelo de serie temporal:

➤ **Una única columna Key Time:**

Cada modelo debe contener una columna numérica o de fecha que se utilizará como serie de casos y que define los intervalos de tiempo que utilizará el modelo. El tipo de datos para la columna de clave temporal puede ser un tipo de datos datetime o numérico. Sin embargo, la columna debe contener valores continuos y deben ser únicos para cada serie. La serie de casos para un modelo de serie temporal no pueden estar almacenada en dos columnas. Por ejemplo una columna Año y una columna Mes.

➤ **Una columna predecible:**

Cada modelo debe contener por lo menos una columna predecible alrededor de la que el algoritmo generará el modelo de serie temporal. El tipo de datos de la columna predecible debe contener valores continuos. Por ejemplo, es posible predecir la manera en que los atributos numéricos tales como ingreso, ventas o temperatura, varían con el tiempo. Sin embargo, no es posible utilizar como columna predecible una columna que contenga valores discretos tales como el estado de las compras o el nivel de educación.

➤ **Una columna de clave de serie opcional:**

Cada modelo puede tener una columna de clave adicional que contenga valores únicos que identifiquen a una serie. La columna de clave de serie opcional debe contener valores únicos. Por ejemplo, un solo modelo puede contener ventas de muchos modelos de producto, siempre y cuando haya un solo registro para cada nombre del producto para cada intervalo de tiempo.

2.2.5.3. Reporting Services:

Microsoft (2016) dice que es una plataforma de creación de informes basada en servidor que ofrece una completa funcionalidad de creación de informes para una gran variedad de orígenes de datos. Contiene un completo conjunto de herramientas para crear, administrar y entregar informes interactivos, tabulares, gráficos, así como interfaces de programación de aplicaciones con las que los desarrolladores podrán integrar o extender el procesamiento de los datos y los informes en aplicaciones personalizadas. Sus informes presentan las siguientes características:

- **Orígenes de datos relacionales, multidimensionales, XML y personalizados:**
Se puede generar informes utilizando datos de cualquier tipo de origen de datos que tenga un proveedor de datos administrado por Microsoft .NET Framework, un proveedor OLE DB o un origen de datos ODBC. También crear informes que utilizan datos relacionales y multidimensionales de SQL Server y Analysis Services, Oracle, Hyperion y otras bases de datos.
- **Diseños de informes tabulares, matriciales, de gráficos y de formato libre:**
Se puede crear informes tabulares para datos basados en columnas, informes matriciales para datos resumidos, informes de gráficos para datos gráficos, e informes de formato libre para todos los demás tipos de datos. El uso de la región de datos *tablix* permite crear informes que combinan diseños de tabla y matriz. Los informes pueden incrustar otros informes y gráficos, junto con listas, gráficos y controles para las aplicaciones basadas en web dinámicas.
- **Informes ad hoc:**
Se puede crear y guardar informes sobre la marcha, y guárdelos localmente o publíquelos directamente en un servidor de informes o en un sitio de SharePoint, donde se pueden administrar como los demás informes. La creación de informes ad hoc se admite en modelos y plantillas de informe mediante el uso de una aplicación de ClickOnce denominada Generador de informes 1.0. Los informes ad hoc basados en modelos proporcionan una funcionalidad click-through infinita para que los usuarios puedan explorar los datos disponibles.
- **Informes detallados e interactividad:**
Se puede agregar características interactivas proporcionando vínculos a subinformes e informes detallados. Utilice parámetros para filtrar los datos de las vistas personalizadas.
- **Formatos de presentación:**
Se puede elegir un formato de presentación al abrir el informe o después de abrirlo. Puede elegir entre formatos orientados a web o a página, o formatos de aplicaciones de escritorio. Entre estos formatos se incluyen los siguientes: HTML, MHTML, PDF, XML, CSV, TIFF, Word y Excel.

- **Controles y elementos de informe personalizados:**
Se puede incrustar controles o elementos de informe personalizados creados por usted o adquiridos de otro fabricante. Un control personalizado requiere una extensión de procesamiento de informes personalizada. Puede agregar expresiones con scripts en Microsoft Visual Basic.
- **Navegación:**
Se puede agregar marcadores y mapas de documentos para proporcionar opciones de navegación en informes de gran tamaño.
- **Agregaciones:**
Se puede agregar y resumir datos mediante controles y expresiones.
- **Elementos gráficos:**
Se puede incrustar o hacer referencia a imágenes y a otros recursos, incluso a contenido externo.

III. MATERIALES Y MÉTODOS

3.1. *Diseño de investigación*

3.1.1. **Tipo de investigación:**

El tipo de estudio de investigación de la presente tesis es experimental ya que se pretende establecer, el efecto (variable dependiente) de una causa (variable independiente) que se manipula al generar una situación para tratar de explicar cómo afecta a quienes participan en ella. En este caso a quienes analizaremos son a los miembros del grupo empresarial Leoncito.

3.1.2. **Hipótesis:**

La hipótesis planteada para esta investigación fue: Con la implementación de un Sistema de BI utilizando algoritmo de serie temporal se apoyará la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito.

3.1.3. **Diseño de contrastación:**

El diseño de contrastación de hipótesis en este caso fue cuasi-experimental, el cual es un diseño de grupo con medición antes y después, para conocer la mejoría que alcanzaron los miembros del grupo empresarial Leoncito con respecto a la toma de decisiones en su proceso de ventas. La Tabla 3 muestra la contrastación de hipótesis en el diseño cuasi experimental de un grupo con medición antes y después.

Variable dependiente	Sistema de BI	Variable dependiente
01: Toma de decisiones en el proceso de ventas del grupo empresarial Leoncito antes de la implementación de un sistema de BI utilizando algoritmo de serie temporal. A través de entrevistas a los miembros del grupo empresarial Leoncito se pudo recaudar información de la realidad de su toma de decisiones.	X: Sistema de Business Intelligence	02: Toma de decisiones en el proceso de ventas del grupo empresarial Leoncito después de la implementación de un sistema de BI utilizando algoritmo de serie temporal. A través de entrevistas a los miembros del grupo empresarial Leoncito después de la implementación de un sistema de BI se pudo recaudar información de la realidad de su toma de decisiones.

Tabla 4. Contrastación de hipótesis – diseño cuasi experimental

3.1.4. **Variables:**

- **Variable independiente:** Sistema de BI utilizando algoritmo de serie temporal.
- **Variable dependiente:** Toma de decisiones en el proceso de ventas del grupo empresarial Leoncito.

3.1.5. **Indicadores:**

La Tabla 4 muestran los objetivos específicos, indicadores, descripciones, unidades de medida, instrumentos y definiciones operacionales.

OBJETIVO ESPECÍFICO	INDICADOR	DESCRIPCION	UNIDAD MEDIDA	INSTRUMENTO	DEFINICION OPERACIONAL
Reducir el tiempo promedio de obtención de información solicitada (respecto a ventas) por los miembros del grupo empresarial.	Tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas.	Identificación del tiempo promedio desde que los miembros del grupo empresarial solicitan información (reporte sobre ventas) hasta que la reciben.	Minutos	Cronómetro Hoja de medición de tiempos (Ver anexo N° 3)	Suma de tiempos por reporte / Total de mediciones
Incrementar el número de reportes que permita análisis comparativo de las ventas en el tiempo.	Cantidad de reportes sobre ventas que permita análisis comparativo en el tiempo.	Identificación de la cantidad de reportes para análisis comparativo y reportes sobre el progreso de ventas en el tiempo.	Unidad	Cuestionario entrevista (Ver anexo N° 1) Cuestionario entrevista (Ver anexo N° 5)	N° reportes con sistema – N° reportes sin sistema
Incrementar el número de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas.	Cantidad de reportes con indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas.	Identificación de la cantidad de reportes con indicadores que permiten monitorear el crecimiento y cumplimiento de las metas mensuales, trimestrales, semestrales y anuales establecidas respecto a las ventas.	Unidad	Cuestionario entrevista (Ver anexo N° 1) Cuestionario entrevista (Ver anexo N° 5)	N° reportes con sistema – N° reportes sin sistema

Incrementar el número de reportes para el análisis de las ventas a través de las diferentes dimensiones.	Cantidad de reportes para el análisis de las ventas a través de las diferentes dimensiones.	Identificación de la cantidad de reportes que permiten analizar las ventas desde las dimensiones: producto, clientes, local, tipo de venta, ubigeo, tiempo, vendedor.	Unidad	Cuestionario entrevista (Ver anexo N° 1) Cuestionario entrevista (Ver anexo N° 5)	N° reportes con sistema – N° reportes sin sistema
Incrementar el número de reportes sobre predicciones de ventas futuras.	Cantidad de reportes sobre predicción de ventas futuras.	Identificación de la cantidad de reportes sobre la predicción respecto a la cantidad y monto de ventas en los siguientes seis meses.	Unidad	Cuestionario entrevista (Ver anexo N° 1) Cuestionario entrevista (Ver anexo N° 5)	N° reportes con sistema – N° reportes sin sistema

Tabla 5. Indicadores

3.1.6. Unidades de investigación:

- **Miembros del grupo empresarial Leoncito:** Encargados de tomar las decisiones estratégicas las cuales decidirán las estrategias que seguirá la empresa.
- **Analista programador de sistemas del grupo empresarial Leoncito:** Es quien elabora los reportes solicitados por los miembros del grupo empresarial Leoncito.

3.1.7. Población y muestra:

3.1.7.1. Población:

Como población se tomará los miembros del grupo empresarial Leoncito, quienes deciden qué acciones tomar frente a las diferentes situaciones que ocurran en la empresa. También se considera al analista-programador de sistemas quien es el que recibe las peticiones de reportes.

3.1.7.2. Muestra de estudio:

- **Miembros del grupo empresarial Leoncito:** Es total son tres personas; pero se trabajo directamente con el gerente administrativo de la empresa. La muestra, por ser pequeña, será igual a la población.
- **Analista-programador de sistemas del grupo empresarial Leoncito:** Es una persona. La muestra, por ser pequeña, será igual a la población.

3.1.8. Métodos y técnicas de recolección de datos:

La Tabla 5 muestra los métodos, técnicas e instrumentos de recolección de datos que se emplearon en la presente investigación.

MÉTODO	TÉCNICAS E INSTRUMENTOS	ELEMENTOS DE LA POBLACIÓN
Entrevista	Guía de entrevista (Ver anexo n° 1)	Gerente administrativo del grupo empresarial Leoncito
Entrevista	Guía de entrevista (Ver anexo n° 2)	Analista-programador de sistemas

Tabla 6. Métodos, técnicas e instrumentos de recolección de datos

Los datos se obtuvieron a través de la elaboración de dos entrevistas, mediante el entendimiento mutuo entre entrevistador y entrevistados para el adecuado entendimiento de la información brindada por parte de los entrevistados. Mediante los indicadores planteados anteriormente se podrá evaluar si se están logrando los objetivos establecidos para posteriormente poder formular las conclusiones respectivas.

3.2. Metodología

Para el desarrollo de la presente tesis se ha considerado la metodología de Kimball, la cual se evidencia en cada una de sus etapas:

- **Planificación del proyecto:**
En esta etapa se realizó entrevistas con el gerente administrativo para plantear un acuerdo respecto a lo que se desarrollara en el sistema.
- **Definición de requerimientos del negocio:**
Se revisó en conjunto con el gerente administrativo sus requerimientos funcionales y no funcionales que debía de tener el sistema de BI
- **Modelado dimensional:**
Se eligió el proceso de ventas al cual se le identificaron dimensiones, atributos, medias y tabla hecho para posteriormente realizar su modelado dimensional.
- **Diseño físico:**
Se identificaron para cada dimensión y la tabla hecho los atributos que se utilizaran para el ETL.
- **Diseño y desarrollo de presentación de datos:**
Se realizó el ETL para poblar cada una de las dimensiones que se identificaron y también se pobló el hecho VENTA.
- **Diseño de la arquitectura técnica:**
La empresa ya contaba con una arquitectura empresarial definida por lo que se utilizaron dichas herramientas en las cuales se implementaron cada uno de sus requerimientos.
- **Selección de productos e instalación:**
La empresa donde se desarrolló el sistema ya contaba con la plataforma de hardware, el motor de base de datos, la herramienta de ETL, la herramienta de reportes, etc. por lo cual no hubo la necesidad de seleccionar nuevos productos.
- **Especificación de aplicaciones para usuarios finales:**
En esta parte se determinó que quienes podían acceder a la información brindada por los reportes que se desarrollaron fueran únicamente los miembros del grupo empresarial Leoncito.
- **Desarrollo de aplicaciones para usuarios finales:**
Se desarrolló el sistema de BI con reportes solicitados, los cuales apoyan en la toma de decisiones a los miembros del grupo empresarial Leoncito.
- **Despliegue:**
Se brindó la documentación correspondiente sobre el desarrollo del sistema de BI desarrollado y además se mostró el funcionamiento del mismo.
- **Mantenimiento y crecimiento:**
En esta etapa se conversó con el gerente sobre los nuevos requerimientos que pudieran surgir. Para lo cual los encargados del área de TI deberían de estar atentos a estas necesidades para poder añadirlas al sistema desarrollado.
- **Gerenciamientos del proyecto:**
Para el desarrollo del sistema de BI se estuvo en constante comunicación con el grupo empresarial Leoncito para cumplir con sus requerimientos y también puedan monitorear los avances del sistema.

IV. RESULTADOS

4.1. Planificación del proyecto

4.1.1. Alcance:

Apoyar la toma de decisiones en el proceso de ventas del grupo empresarial Leoncito.

4.1.2. Propósito:

Desarrollar para los miembros del grupo empresarial un sistema a través del cual puedan analizar sus ventas desde diferentes perspectivas. Brindando acceso a información útil a través de un navegador web con acceso a internet, con el fin de convertir dicha información en conocimiento que servirá para tomar decisiones adecuadas y oportunas sobre sus ventas.

4.2. Definición de requerimientos del negocio

4.2.1. Requerimientos del proceso:

Los miembros del grupo empresarial tenían la necesidad de analizar sus ventas. Contaban con un sistema OLTP, sin embargo existía la necesidad de contar con un sistema de BI que apoye en la toma de decisiones estratégicas. A continuación se mencionan los requerimientos funcionales y no funcionales.

Requerimientos funcionales:

- Reporte de semaforización que incluya:
 - Un indicador de metas mensuales, un indicador de metas trimestrales, un indicador de metas semestrales y otro de metas anuales evaluados por el monto de venta por local.

- Reporte de productos que incluya:
 - Top de líneas y marcas evaluadas por las unidades y monto vendido. Además de mostrar para ambos sus unidades y montos vendidos filtrado por año.

- Reporte de clientes que incluya:
 - Top clientes naturales y clientes jurídicos mostrando las unidades y montos de compra. Ambos top filtrados por año.

- Reporte de vendedores que incluya:
 - Top vendedores por unidades vendidas de la marca seleccionada en un determinado año. Aparte poder evaluar del top vendedores las unidades y monto vendido en los diferentes meses y poder visualizar sus ventas en los diferentes distritos en los cuales tenga ventas registradas.
 - Top vendedores por unidades vendidas de la marca seleccionada en un determinado mes de un año. Aparte poder evaluar del top vendedores las unidades y monto vendido en los diferentes meses y poder visualizar sus ventas en los diferentes distritos los cuales tenga ventas registradas.

- Reporte de ventas que incluya:
 - Monto por tipo de venta en los diferentes distritos. Además del monto de venta en todos los meses. Evaluados por año.
 - Monto por tipo de venta en los diferentes locales evaluados por año. Y al seleccionar un determinado local mostrar el monto vendido en los diferentes meses, además mostrar los productos con mayor cantidad de ventas.
 - Predicción de monto de venta en los diferentes distritos para los siguientes seis meses.
 - Predicción de unidades de venta por producto para los siguientes seis meses.
- Crear un sistema web con inicio de sesión para autorizar el acceso únicamente a los miembros del grupo empresarial y estos puedan visualizar los reportes.

Requerimientos no funcionales:

- Disponibilidad las 24 horas del día.
- Uso único por los miembros del grupo empresarial.
- Interfaz amigable en el aspecto visual para el usuario.
- Desarrollo del sistema a través de la suite de BI de SQL Server 2008 R2.
- Desarrollo de la página web en asp.net.
- Actualización constante de las tablas dimensiones, la tabla hecho.
- Actualización constante del cubo.

4.2.2. Matriz bus:

La matriz bus se utilizó para poder identificar las diferentes dimensiones que tienen relación con el proceso del negocio seleccionado. La Tabla 7 muestra la matriz bus para el proceso del negocio seleccionado. En este caso: VENTAS.

PROCESO DEL NEGOCIO	DIMENSIONES						
	PRODUCTO	CLIENTE	LOCAL	TIEMPO	UBI GEO	TIPO DE VENTA	VENDEDOR
VENTAS	X	X	X	X	X	X	X

Tabla 7. Matriz bus

4.3. Modelo dimensional

4.3.1. Elegir el proceso de negocio:

El proceso de estudio elegido fue ventas. A continuación en la figura 4 se muestra modelo dimensional de alto nivel:



Figura 4. Modelo dimensional de alto nivel

4.3.2. Establecer el nivel de granularidad:

Para cada dimensión se identificó el nivel de detalle más bajo que se desea analizar. A continuación las dimensiones de análisis del proceso ventas y su granularidad:

- Dimensión producto: Nombre del modelo
- Dimensión cliente: Nombre del cliente
- Dimensión local: Nombre del local
- Dimensión tipo de venta: Tipo de venta
- Dimensión ubigeo: Distrito
- Dimensión tiempo: Fecha
- Dimensión vendedor: Nombre del vendedor

4.3.3. Dimensiones y sus atributos:

Cada dimensión tiene atributos relevantes para el proceso ventas del grupo empresarial. A continuación la tabla 8 muestra las dimensiones con sus atributos y jerarquías.

DIMENSIONES	ATRIBUTOS	JERARQUIAS
PRODUCTO	<ul style="list-style-type: none"> • key_producto • codigo_oltp • nombre_modelo • nombre_marca • nombre_tipo • nombre_linea 	* nombre_linea ** nombre_tipo *** nombre_modelo *nombre_marca **nombre_modelo
CLIENTE	<ul style="list-style-type: none"> • key_cliente • codigo_oltp • nombre_cliente • tipo_cliente 	*tipo_cliente **nombre_cliente
LOCAL	<ul style="list-style-type: none"> • key_local • codigo_oltp • nombre_local • nombre_empresa 	*nombre_empresa ** nombre_local
TIPO DE VENTA	<ul style="list-style-type: none"> • key_tipoventa • tipo_venta 	No tiene
UBIGEO	<ul style="list-style-type: none"> • key_ubigeo • codigo_oltp • distrito • provincia • departamento 	*departamento **provincia ***distrito
TIEMPO	<ul style="list-style-type: none"> • key_tiempo • fecha • mes • trimestre • semestre • año 	*año **semestre ***trimestre ****mes *****fecha
VENDEDOR	<ul style="list-style-type: none"> • key_vendedor • codigo_oltp • nombre_vendedor 	No tiene

Tabla 8. Dimensiones y sus atributos

4.3.4. Identificar tablas hecho y sus medidas:

La tabla hecho surge del proceso del negocio escogido, en este caso VENTA, la cual agrupa las medidas que se desean analizar. A continuación la tabla 9 muestra el hecho VENTA con sus respectivos atributos y medidas.

HECHO	ATRIBUTOS	MEDIDAS
VENTA	<ul style="list-style-type: none"> • Key_hechoventa • Key_producto • Key_cliente • Key_local • Key_tipoventa • Key_ubigeo • Key_tiempo • Key_vendedor •Codigo_oltp 	<ul style="list-style-type: none"> • unidades_vendidas • monto

Tabla 9. Hecho con sus atributos y medidas

4.4. Diseño físico

A continuación le modelo físico de la base de datos multidimensional:

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_producto	key_producto	int		No	Si	PK
	codigo_oltp	int		No	No	
	nombre_modelo	varchar	100	No	No	
	nombre_marca	varchar	60	No	No	
	nombre_tipo	varchar	80	No	No	
	nombre_linea	varchar	40	No	No	
Nota: Para la dimensión producto el atributo nombre_modelo es el producto.						

Tabla 10. Dimensión producto

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_cliente	key_cliente	int		No	Si	PK
	codigo_oltp	int		No	No	
	nombre_cliente	varchar	121	No	No	
	tipo_cliente	varchar	20	No	No	
Nota: Para poblar el atributo tipo_cliente se evaluó en la tabla transaccional los códigos de los clientes y si se encontraban en la tabla jurídica significa que es cliente jurídico y se les registraba CLIENTE JURIDICO, en caso de que se encontraban en la tabla natural significa que se clientes naturales y se les registraba CLIENTE NATURAL.						

Tabla 11. Dimensión cliente

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_local	key_local	int		No	Si	PK
	codigo_oltp	int		No	No	
	nombre_local	varchar	60	No	No	
	nombre_empresa	varchar	80	No	No	

Tabla 12. Dimensión local

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_tipo_venta	key_tipoventa	int		No	Si	PK
	tipo_venta	varchar	20	No	No	
Nota: Por ser 2 tipos de venta, se registraron directamente en la base de datos.						

Tabla 13. Dimensión tipo de venta

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_ubigeo	key_ubigeo	int		No	Si	PK
	codigo_oltp	int		No	No	
	distrito	varchar	60	No	No	
	provincia	varchar	60	No	No	
	departamento	varchar	30	No	No	

Tabla 14. Dimensión ubigeo

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_tiempo	key_tiempo	int		No	Si	PK
	fecha	date		No	No	
	mes	int		No	No	
	trimestre	int		No	No	
	semestre	int		No	No	
	año	int		No	No	
Nota: Para poblar la dim_tiempo se hizo uso de una consulta para generar el tiempo. Desde la fecha más antigua, hasta la fecha final de las fechas de la copia de seguridad.						

Tabla 15. Dimensión tiempo

Dimensión	Atributos	Tipo De Dato	Longitud	Nulo	Auto Incremental	Restricción
dim_vendedor	key_vendedor	int		No	Si	PK
	codigo_oltp	int		No	No	
	nombre_vendedor	varchar	100	No	No	

Tabla 16. Dimensión vendedor

Hecho	Atributos	Tipo de Dato	Longitud	Nulo	Auto Incremental	Restricción
hecho_venta	key_hechoventa	int		No	Si	PK
	key_producto	int		No	No	FK
	key_cliente	int		No	No	FK
	Key_local	int		No	No	FK
	key_tipoventa	int		No	No	FK
	key_ubigeo	int		No	No	FK
	key_tiempo	int		No	No	FK
	key_vendedor	Int		No	No	FK
	codigo_oltp	Int		No	No	
	unidades_vendidas	int		No	No	
	monto	numeric	(12,2)	No	No	

Tabla 17. Hecho venta

4.5. Diseño y desarrollo de presentación de datos

En esta parte se hizo el proceso ETL en donde se realiza una población paralela de las dimensiones y al último se hace la población del hecho VENTA. A continuación los procesos realizados para la población del Data Mart.

- Poblamiento de la dimensión PRODUCTO.
- Poblamiento de la dimensión CLIENTE.
- Poblamiento de la dimensión LOCAL.
- Poblamiento de la dimensión TIPO DE VENTA.
- Poblamiento de la dimensión UBIGEO.
- Poblamiento de la dimensión TIEMPO.
- Poblamiento de la dimensión VENDEDOR.
- Poblamiento del hecho VENTA.

Primeramente para comenzar con el proceso ETL se realizan las conexiones al origen u orígenes (dependiendo desde dónde se extraigan los datos que se necesitan para el poblamiento del data mart) y al destino, en este caso se conecta al origen de datos llamado **LocalHost.trabajo_GrupoLeoncito** y al destino llamado **LocalHost.dm_gl**. A continuación en la parte inferior de la figura 5 se muestran las conexiones al origen y al destino.

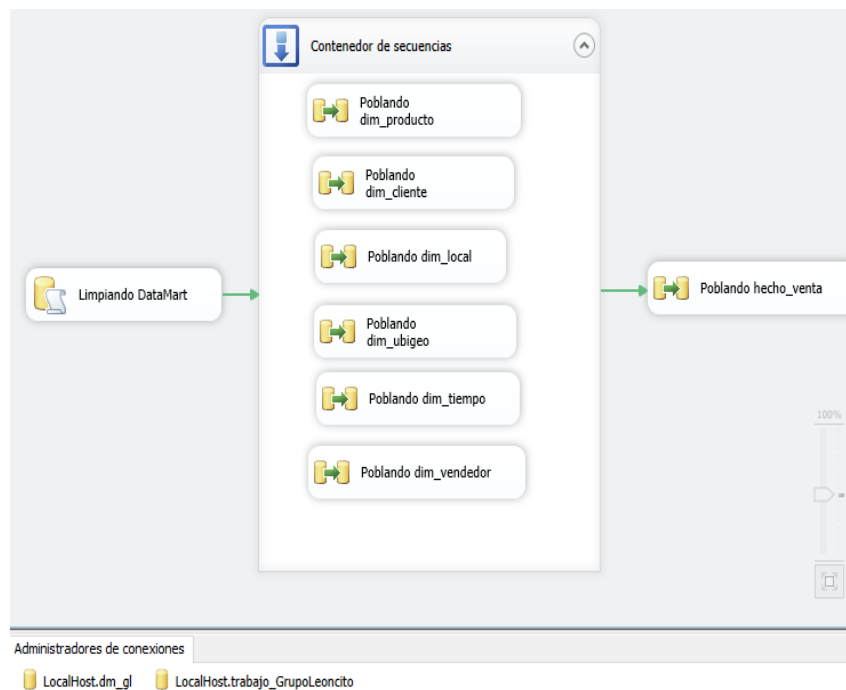


Figura 5. Conexión al origen y destino

Posteriormente a la conexión se hizo limpieza del data mart en la cual se eliminan las tablas hecho_venta, dim_producto, dim_cliente, dim_local, dim_vendedor. A continuación se muestra el proceso realizado para poblar las diferentes dimensiones y el hecho venta:

✓ **Poblamiento de la dimensión PRODUCTO:**

Para realizar el proceso ETL de los productos se utilizó un Origen de OLE DB para poder extraer los datos de la base de datos transaccional. Los datos necesarios se extrajeron a través de un comando SQL para poder extraer por ejemplo su línea de producto, tipo de producto, marca y modelo (Para este caso el modelo es el producto). Se utilizó Columna derivada en la cual se usó la función trim para limpiar los espacios que puedan existir, también se realizó conversión de datos para que los datos que se extrajeron sean iguales al tipo de dato que espera la tabla dim_producto y finalmente se pobló tabla dim_producto. La figura 6 muestra el proceso de ETL para la dimensión producto.



Figura 6. Poblamiento – dimensión producto

✓ **Poblamiento de la dimensión CLIENTE:**

Para realizar el proceso ETL de los clientes se utilizó un Origen de OLE DB para poder extraer los datos de la base de datos transaccional. Los datos necesarios se extrajeron a través de un comando SQL para poder extraer por ejemplo su nombre y también se hizo una validación en la consulta para saber si es un CLIENTE NATURAL o CLIENTE JURIDICO. Se utilizó Columna derivada en la cual se usó la función trim para limpiar los espacios que puedan existir, se usó conversión de datos para que los datos que se extrajeron sean iguales al tipo de dato que espera la tabla dim_cliente y finalmente se pobló tabla dim_cliente. La figura 7 muestra el proceso de ETL para la dimensión cliente.

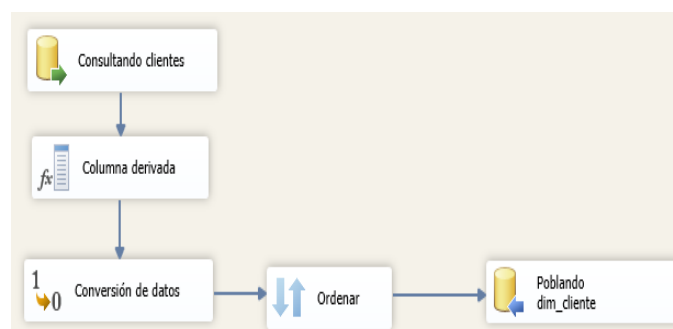


Figura 7. Poblamiento – dimensión cliente

Poblamiento de la dimensión LOCAL:

Para realizar el proceso ETL de los locales se utilizó un Origen de OLE DB para poder extraer los datos de la base de datos transaccional. Los datos necesarios se extrajeron a través de un comando SQL para poder extraer por ejemplo su nombre y la empresa a la cual pertenece el local (El grupo empresarial cuenta con varias empresas y cada empresa con sus respectivos locales = sucursales). Se utilizó conversión de datos para que los datos que se extrajeron sean los mismos al tipo de dato que espera la tabla dim_local y finalmente se pobló tabla dim_local. La figura 8 muestra el proceso de ETL para la dimensión local.

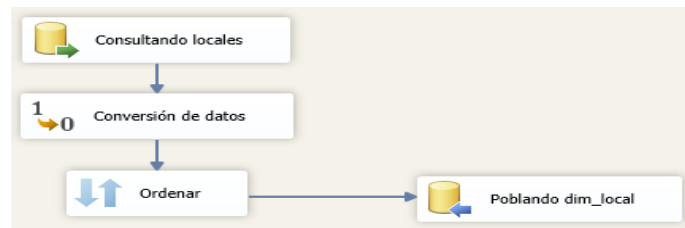


Figura 8. Poblamiento – dimensión local

✓ Poblamiento de la dimensión TIPO DE VENTA:

Para poblar la dimensión tipo de venta se realizó a través de dos consultas. Dos insert para poder agregarlos directamente en la tabla **dim_tipo_venta**, Los insert fueron:

- Insert into dim_tipo_venta (tipo_venta) values ('VENTAS AL CONTADO');
- insert into dim_tipo_venta (tipo_venta) values ('VENTAS AL CREDITO');

✓ Poblamiento de la dimensión UBIGEO:

Para realizar el proceso ETL del ubigeo se utilizó un Origen de OLE DB para poder extraer los datos de la base de datos transaccional. Los datos necesarios se extrajeron a través de un comando SQL para poder extraer por ejemplo el departamento, provincia y distrito. Se utilizó Columna derivada en la cual se usó la función trim para limpiar los espacios en blanco, se usó conversión de datos para que los datos se extrajeron sean los mismos al tipo de dato que espera la tabla dim_ubigeo. Además se hizo una validación para que solo puedan pasar aquellos ubigeos que sean nuevos en la tabla dim_ubigeo. La figura 9 muestra el proceso de ETL para la dimensión ubigeo.



Figura 9. Poblamiento – dimensión ubigeo

✓ **Poblamiento de la dimensión TIEMPO:**

Para poblar la dimensión tiempo se creó una consulta en la cual se declaró una fecha de inicio (@f1) y una fecha fin (@f2), las cual se puede visualizar en la figura 10. Se pobló desde comienzos del 2011 hasta finales de Mayo del 2016. Pero también se tomó en cuenta una validación para cuando existan nuevas fechas solo puedan pasar aquellas fechas que son nuevas y estas se guarden en la tabla dim_tiempo. La figura 11 muestra el proceso de ETL para para la dimensión tiempo.

```
alter table dim_tiempo add unique (fecha)
declare @f1 smalldatetime='01/01/2011'
declare @f2 smalldatetime='31/05/2016'
while @f1<=@f2 begin insert into dim_tiempo values (@f1,MONTH(@f1), datepart(quarter,@f1),
(case when MONTH(@f1) <7 then 1 else 2 end), YEAR(@f1)) set @f1=@f1+1 end
```

Figura 10. Consulta para poblar dimensión tiempo

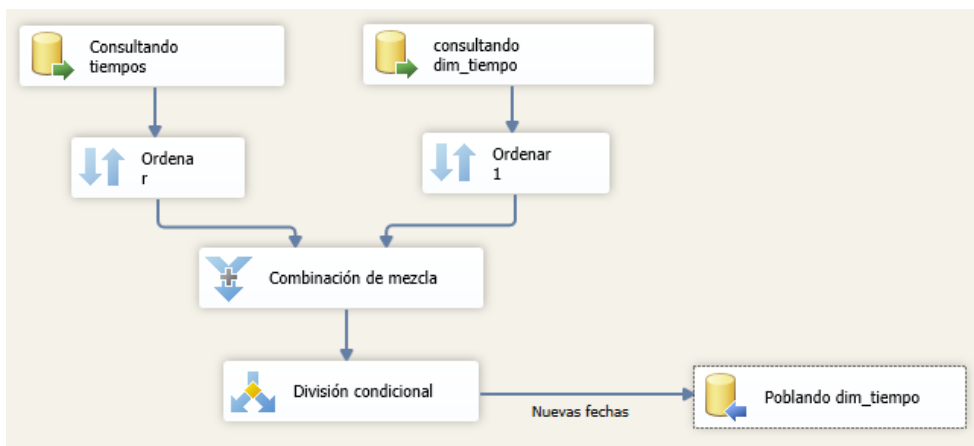


Figura 11. Poblamiento – dimensión tiempo

✓ **Poblamiento de la dimensión VENDEDOR:**

Para realizar el proceso ETL de los vendedores se utilizó un Origen de OLE DB para poder extraer los datos de la base de datos transaccional. Los datos necesarios se extrajeron a través de un comando SQL para poder extraer en este caso únicamente su nombre. Se utilizó Columna derivada en la cual se usó la función trim para limpiar los espacios que puedan existir, se usó conversión de datos para que el dato que se extrajo sea igual al tipo de dato que espera la tabla dim_vendedor y finalmente se pobló la tabla dim_vendedor. La figura 12 muestra el proceso de ETL para la dimensión vendedor.



Figura 12. Poblamiento – dimensión vendedor

✓ Poblamiento del Hecho VENTA

Finalmente cuando se ha terminado de poblar todas las tablas dimensiones se comienza a poblar la tabla hecho_venta. Para realizar el proceso ETL de la tabla hecho_venta se utilizó un Origen de OLE DB para poder extraer los datos que se necesitan de la base de datos transaccional. Se usó conversión de datos para que los datos que se extrajeron sean iguales a los datos a través de los cuales se buscará el key de cada dimensión (se buscará a través de los códigos oltp), la cual se almacenará en la tabla hecho_venta como clave foránea y también para que los demás datos que se necesitan sean iguales al tipo de dato que espera la tabla hecho_venta y finalmente se pobló tabla hecho_venta. La figura 13 muestra el proceso de ETL para el hecho venta.

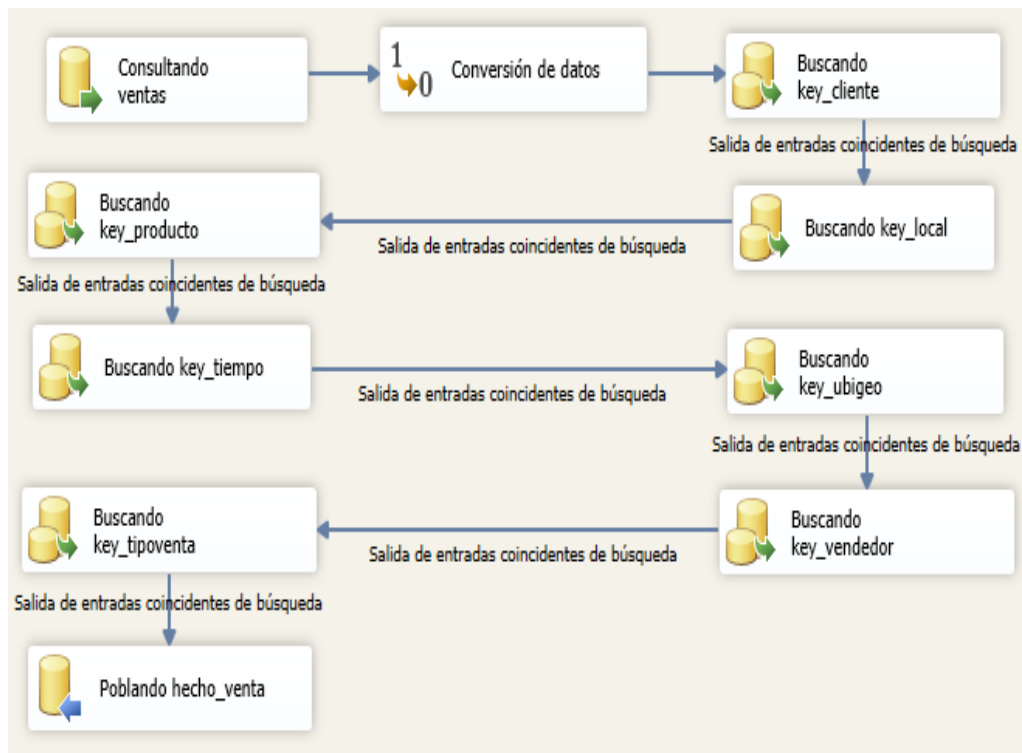


Figura 13. Poblamiento – hecho venta

4.6. Especificaciones de aplicaciones para usuarios finales

En esta parte se crearon los perfiles de usuarios que podrán acceder a la información brindada por los reportes que se desarrollaron, y quienes solo podrían acceder son los miembros del grupo empresarial. Este acceso se realizó a través de un logueo en el cual se tiene que ingresar un usuario y una clave. Los miembros del grupo empresarial ya poseen perfiles debido a que en la base de datos transaccional de la empresa existe una tabla usuario y una tabla perfiles de usuarios. Los miembros del grupo empresarial siempre y cuando ingresen correctamente su usuario y clave podrán acceder a un sistema web donde podrán visualizar los reportes.

4.7. Desarrollo de aplicaciones para usuarios finales

En esta parte se realiza la creación del cubo olap, para lo cual se necesita crear una conexión con el data mart, después se crea una vista al origen que viene a ser el data mart, luego se crean las dimensiones y posteriormente se crea el cubo. También se realiza minería de datos, para lo cual se hace uso de un algoritmo para aplicar la predicción. En este caso se van a predecir las ventas. Y por último se elaboran reportes.

4.7.1. Creación de cubo OLAP:

Para crear el cubo OLAP se utilizó SQL Server Analysis Services, esto después de haber terminado el poblamiento del data mart mediante el proceso ETL.

Primeramente para comenzar con este proceso se realiza la conexión al origen, en este caso se conecta al origen de datos llamado **LocalHost.dm_gl**. La figura 14 muestra la conexión con el data mart **LocalHost.dm_gl**.

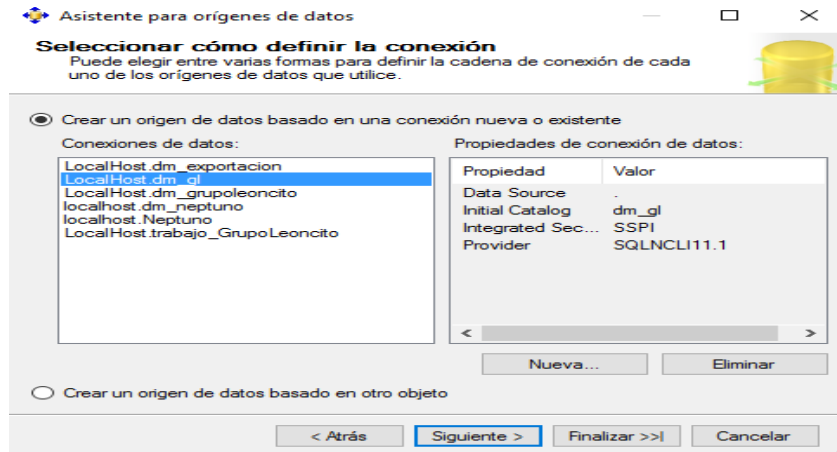


Figura 14. Conexión al data mart dm_gl

Después de crear la conexión con el data mart se crea una vista del origen de datos, en este caso llamada **origen_dm_gl_ventas**, La figura 15 muestra la vista creada, la cual en este caso se llamó **vista_dm_gl_ventas**.

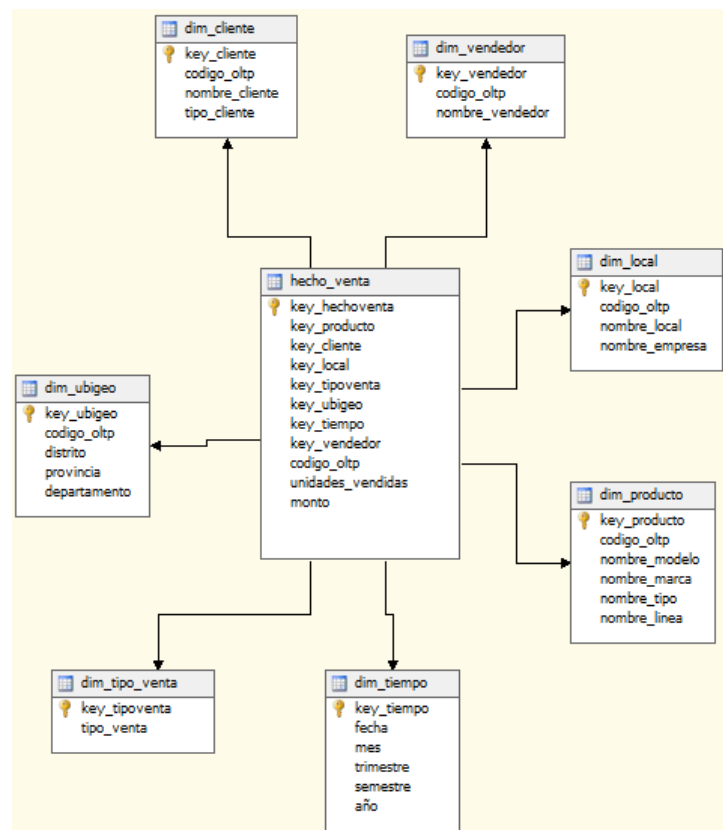


Figura 15. vista_dm_gl_ventas

Luego de crear la vista se crean las dimensiones y sus jerarquías. Para posteriormente crear el cubo, en este caso llamado **cubo_ventas**. La figura 16 muestra al cubo ya implementado.

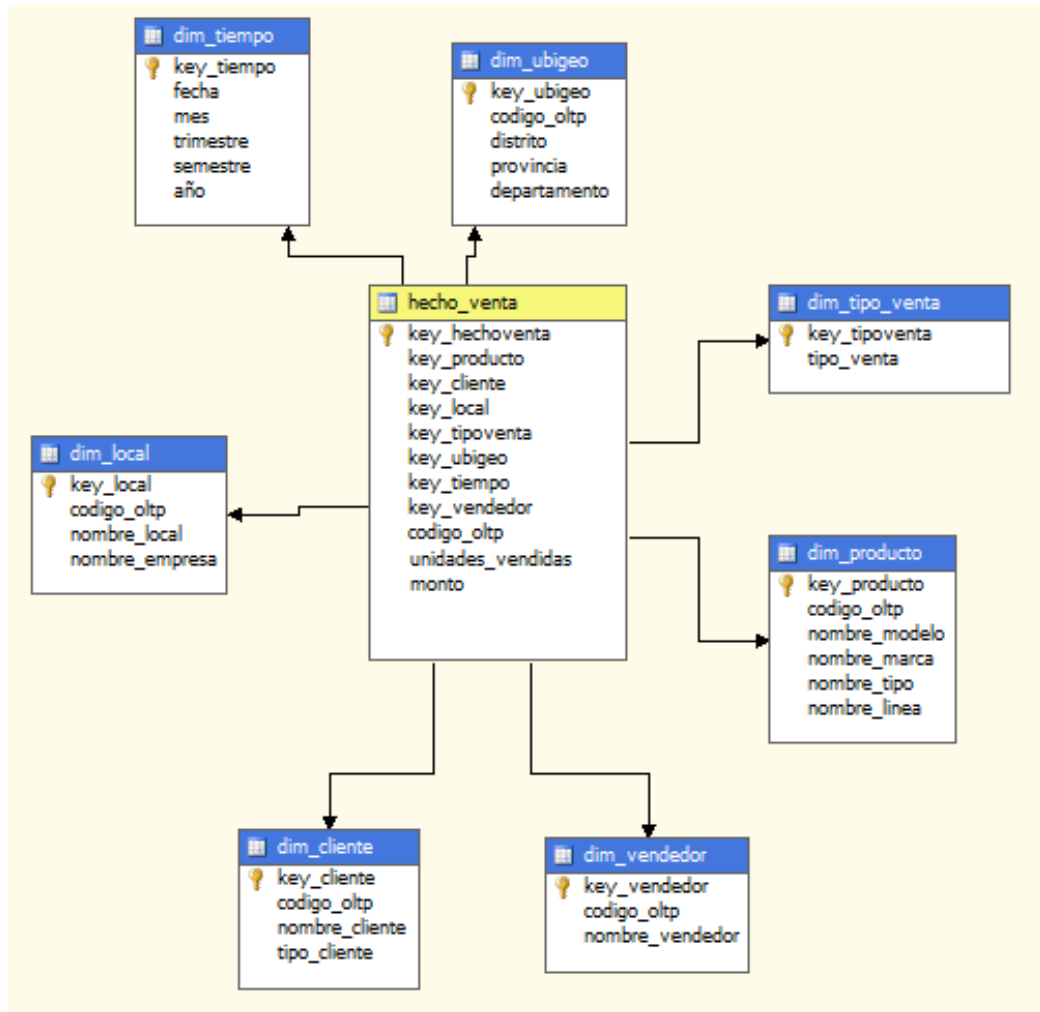


Figura 16. Creación de cubo cubo_ventas

Se crearon KPIs en el cubo para los reportes con indicadores. Los KPIs fueron los cuatro siguientes: **KPI_monto_ventas_año**, **KPI_monto_ventas_semestre**, **KPI_monto_ventas_trimestre** y **KPI_monto_ventas_mes**. Se conversó con el gerente administrativo para definir las expresiones objetivo de los KPI, las cuales son las metas a las cuales se desea llegar.

Para visualizar el cumplimiento de estas se usó semaforización (verde, ámbar y rojo) para monitorear el cumplimiento de las metas. A continuación la creación de los KPIs:

✓ **KPI_monto_ventas_año:**

Este KPI se refirió al monto de las ventas realizadas por año. En este caso el objetivo es llegar o superar en 30% el monto de las ventas vendidas respecto al año anterior. La semaforización se determinó de la siguiente manera: Si el monto de ventas del año actual es igual o mayor en un 95% con respecto al objetivo el color será verde, si el monto de ventas del año actual es igual o mayor en un 90% con

respecto al objetivo el color será ámbar y si el monto de ventas del año actual es menor en un 90% con respecto al objetivo el color será rojo. La figura 17 muestra la generación del KPI llamado **KPI_monto_ventas_año**.

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

```
CASE WHEN ISEMPTY((([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[año],1))) THEN 0 ELSE ([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[año],1)))*1.3 END
```

Estado
Indicador de estado:

Expresión de estado:

```
CASE WHEN KPIVALUE("KPI_monto_ventas_año") >= KPIGOAL("KPI_monto_ventas_año")*0.95 THEN 1 WHEN KPIVALUE("KPI_monto_ventas_año") < KPIGOAL("KPI_monto_ventas_año")*0.90 THEN -1 ELSE 0 END
```

Figura 17. KPI_monto_ventas_año

✓ **KPI_monto_ventas_semestre:**

Este KPI se refirió al monto de las ventas realizadas por semestre. En este caso el objetivo es llegar o superar en 30% el monto de las ventas vendidas respecto al semestre anterior en el mismo periodo. La semaforización se determinó de la siguiente manera: Si el monto de ventas del semestre actual es igual o mayor en un 95% con respecto al objetivo el color será verde, si el monto de ventas del semestre actual es igual o mayor en un 90% con respecto al objetivo el color será ámbar y si el monto de ventas del semestre actual es menor en un 90% con respecto al objetivo el color será rojo. La figura 18 muestra la generación del KPI llamado **KPI_monto_ventas_semestre**.

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

```
CASE WHEN ISEMPTY((([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[semestre],2))) THEN 0 ELSE ([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[semestre],2)))*1.3 END
```

Estado
Indicador de estado:

Expresión de estado:

```
CASE WHEN KPIVALUE("KPI_monto_ventas_semestre") >= KPIGOAL("KPI_monto_ventas_semestre")*0.95 THEN 1 WHEN KPIVALUE("KPI_monto_ventas_semestre") < KPIGOAL("KPI_monto_ventas_semestre")*0.90 THEN -1 ELSE 0 END
```

Figura 18. KPI_monto_ventas_semestre

✓ **KPI_monto_ventas_trimestre:**

Este KPI se refirió al monto de las ventas realizadas por trimestre. En este caso el objetivo es llegar o superar en 30% el monto de las ventas vendidas respecto al trimestre anterior en el mismo periodo. La semaforización se determinó de la siguiente manera: Si el monto de ventas del trimestre actual es igual o mayor en un 95% con respecto al objetivo el color será verde, si el monto de ventas del trimestre actual es igual o mayor en un 90% con respecto al objetivo el color será ámbar y si el monto de ventas del trimestre actual es menor en un 90% con respecto al objetivo el color será rojo. La figura 19 muestra la generación del KPI llamado **KPI_monto_ventas_trimestre**.

The screenshot shows the configuration interface for the KPI 'KPI_monto_ventas_trimestre'. It includes the following fields and code:

- Nombre:** KPI_monto_ventas_trimestre
- Grupo de medida asociado:** Hecho Venta
- Expresión de valor:** [Measures].[Monto]
- Expresión objetivo:**

```
CASE WHEN ISEMPY([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[trimestre],4)) THEN 0 ELSE ([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[trimestre],4))*1.3 END
```
- Estado:** Medidor
- Expresión de estado:**

```
CASE WHEN KPIVALUE("KPI_monto_ventas_trimestre") >= KPIGOAL("KPI_monto_ventas_trimestre")*0.95 THEN 1 WHEN KPIVALUE("KPI_monto_ventas_trimestre") < KPIGOAL("KPI_monto_ventas_trimestre")*0.90 THEN -1 ELSE 0 END
```

Figura 19. KPI_monto_ventas_trimestre

✓ **KPI_monto_ventas_mes:**

Este KPI se refirió al monto de las ventas realizadas por mes. En este caso el objetivo es llegar o superar en 30% el monto de las ventas vendidas respecto al mes anterior en el mismo periodo. La semaforización se determinó de la siguiente manera: Si el monto de ventas del mes actual es igual o mayor en un 95% con respecto al objetivo el color será verde, si el monto de ventas del mes actual es igual o mayor en un 90% con respecto al objetivo el color será ámbar y si el monto de ventas del mes actual es menor en un 90% con respecto al objetivo el color será rojo. La figura 20 muestra la generación del KPI llamado **KPI_monto_ventas_mes**.

The screenshot shows the configuration interface for the KPI 'KPI_monto_ventas_mes'. It includes the following fields and code:

- Nombre:** KPI_monto_ventas_mes
- Grupo de medida asociado:** Hecho Venta
- Expresión de valor:** [Measures].[Monto]
- Expresión objetivo:**

```
CASE WHEN ISEMPY([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[mes],12)) THEN 0 ELSE ([Measures].[Monto],PARALLELPERIOD([dim_tiempo].[Fechas por año].[mes],12))*1.3 END
```
- Estado:** Medidor
- Expresión de estado:**

```
CASE WHEN KPIVALUE("KPI_monto_ventas_mes") >= KPIGOAL("KPI_monto_ventas_mes")*0.95 THEN 1 WHEN KPIVALUE("KPI_monto_ventas_mes") < KPIGOAL("KPI_monto_ventas_mes")*0.90 THEN -1 ELSE 0 END
```

Figura 20. KPI_monto_ventas_mes

4.7.2. Creación de minería de datos:

Sabiendo que el proceso para realizar la minería de datos para la predicción del monto de las ventas por distrito y las unidades vendidas por producto es la misma se escogió a la predicción del monto de las ventas por distrito como ejemplo.

Para la creación de minería de datos primero se tiene que seleccionar la información que formará parte del modelo que el algoritmo tomará para la predicción de los datos que se desean predecir.

En este caso para el pronóstico del monto de las ventas por distrito se realizó una consulta a la tabla **dim_ubigeo**. A continuación la figura 21 muestra la consulta realizada.

```
--consulta para prediccion de monto por distrito
select t.año, t.mes, u.distrito, sum(hv.monto) as monto
from dim_ubigeo u
inner join hecho_venta hv on u.key_ubigeo=hv.key_ubigeo
inner join dim_tiempo t on t.key_tiempo=hv.key_tiempo
group by u.distrito,t.año,t.mes
order by 1,2,3]
```

	año	mes	distrito	monto
1	2011	1	CHICLAYO	65013.25
2	2011	1	FERREÑAFE	119792.00
3	2011	1	JAYANCA	16521.00
4	2011	1	LA VICTORIA	267775.05
5	2011	1	LAMBAYEQUE	110010.00
6	2011	1	MOCHUMI	39897.00
7	2011	1	MOTUPE	27883.00
8	2011	1	OLMOS	54853.00
9	2011	2	CHICLAYO	70097.04
10	2011	2	FERREÑAFE	113763.00
11	2011	2	JAYANCA	9838.00
12	2011	2	LA VICTORIA	217250.99
13	2011	2	LAMBAYEQUE	85748.00

Figura 21. Consulta para predicción

Como se observa en la figura 21 se consideró el año, mes, distrito y monto de venta por cada distrito.

Después de tener la consulta que se utilizó para la extracción de la información que fue parte del modelo de minería de datos se creó una tabla temporal con la misma estructura: distrito, año, mes y monto. La figura 22 muestra la creación de la tabla temporal, en este caso llamada tp_distritos.

```
create table tp_distritos
(
distrito varchar(60) not null,
año int not null,
mes int not null,
monto numeric (12,2) not null
)
select * from tp_distritos order by 2,3,1
```

	distrito	año	mes	monto
1	CHICLAYO	2011	1	65013.25
2	FERREÑAFE	2011	1	119792.00
3	JAYANCA	2011	1	16521.00
4	LA VICTORIA	2011	1	267775.05
5	LAMBAYEQUE	2011	1	110010.00
6	MOCHUMI	2011	1	39897.00
7	MOTUPE	2011	1	27883.00
8	OLMOS	2011	1	54853.00
9	CHICLAYO	2011	2	70097.04
10	FERREÑAFE	2011	2	113763.00
11	JAYANCA	2011	2	9838.00
12	LA VICTORIA	2011	2	217250.99
13	LAMBAYEQUE	2011	2	85748.00
14	MOCHUMI	2011	2	27336.50
15	MOTUPE	2011	2	23519.00
16	OLMOS	2011	2	27302.00
17	CHICLAYO	2011	3	49762.61

Figura 22. Creación de tabla temporal

Revisando no todos los distritos tenían monto para todos los meses, entonces no tenían registros en todos los meses de todos los años. Para esto se creó un cursor el cual llenó la tabla temporal **tp_distritos** con los distritos para todos los meses de todos los años de los cuales se tenían información y colocando al monto el valor “0” para todos ellos. La figura 23 muestra el primer cursor creado, en este caso llamado **cursor1**.

```
--creacion de cursor para registrar monto 0 a los distritos
declare @distrito varchar(60), @año int, @mes int
declare cursor1 cursor for
select distinct u.distrito
from dim_ubigeo u
inner join hecho_venta hv on u.key_ubigeo=hv.key_ubigeo
open cursor1;
fetch next from cursor1 into @distrito
set @mes=1
while @@FETCH_STATUS=0
begin
    set @año=2011
    while @año<=2016
        begin
            set @mes=1
            while @mes<=12
                begin
                    insert into tp_distritos values(@distrito,@año,@mes,0)
                    set @mes=@mes+1
                end
            set @año=@año+1
        end
    fetch next from cursor1 into @distrito
end
close cursor1;
deallocate cursor1;
```

Figura 23. Creación de cursor 1

Después se creó otro cursor para actualizar el monto que el primero cursor le registró a todos los distritos con el valor “0” por el monto real de cada uno de los distritos. De esta manera siempre que un distrito no tenga monto en un mes determinado vaya una fila pero con valor “0”. La figura 24 muestra el segundo cursor creado, en este caso llamado **cursor2**.

```
--creacion de cursor para registrar monto real a los distritos
declare @distrito_ varchar(60), @año_ int, @mes_ int, @monto numeric(12,2)
declare cursor2 cursor for
select t.año, t.mes, u.distrito, sum(hv.monto) as monto
from dim_ubigeo u
inner join hecho_venta hv on u.key_ubigeo=hv.key_ubigeo
inner join dim_tiempo t on t.key_tiempo=hv.key_tiempo
group by u.distrito, t.año, t.mes order by 1,2,3;
open cursor2;
fetch next from cursor2 into @año_,@mes_,@distrito_,@monto
while @@FETCH_STATUS=0
begin
    update tp_distritos set monto=@monto
    where distrito=@distrito_ and año=@año_ and mes=@mes_
    fetch next from cursor2 into @año_, @mes_, @distrito_, @monto
end
close cursor2
deallocate cursor2;
```

Figura 24. Creación de cursor 2

Luego, se creó una vista a la cual se realizó una consulta de la tabla temporal que se pobló. La figura 25 muestra la creación de la vista, en este caso llamada **v_tp_distritos**.

```
--creacion de vista
create view v_tp_distritos as
select cast('1/'+right('00'+cast(mes as varchar),2)+'/'+cast(año as char(4)) as date) as fecha, distrito, monto from tp_distritos
```

Figura 25. Creación de vista

Esto se hizo para juntar el año con el mes y concatenarlos con un día; esto para darle el formato date y el modelo culmine con un modelo de fecha, distrito y monto. Y este modelo sea utilizado para la minería de datos.

Después para crear la estructura de minería de datos se tuvo que crear una nueva vista en la cual se incluye a la vista llamada **v_tp_distritos** que se creó anteriormente. La figura 26 muestra la vista creada, en este caso llamado **vista_v_tp_distritos**.

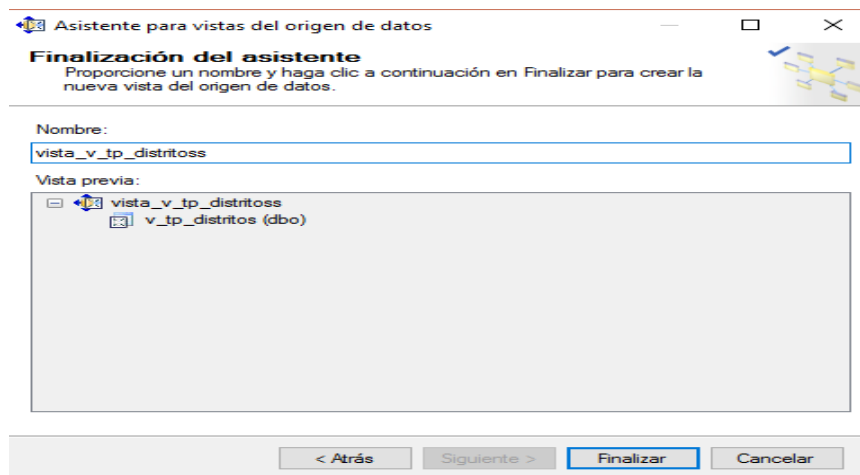


Figura 26. Creación de vista para origen de datos

Ya teniendo la vista se crea una nueva estructura de minería de datos. A continuación se muestra la creación de la estructura de minería de datos:

Primero se escoge la técnica que se desea crear. Se usó la técnica serie temporal de Microsoft. La figura 27 muestra la selección de la técnica.

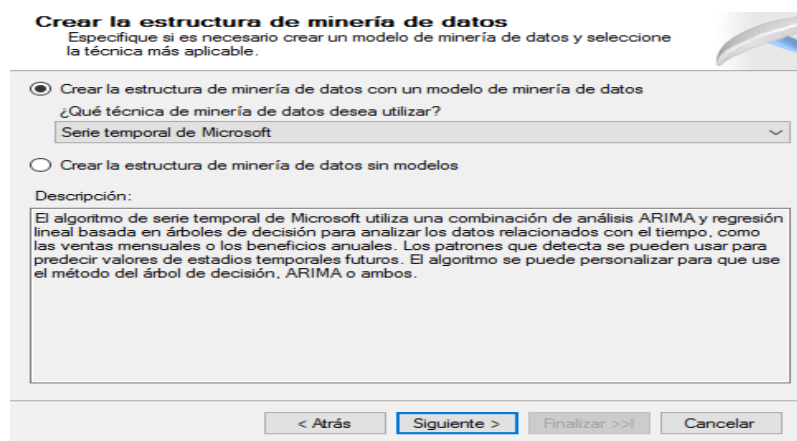


Figura 27. Selección de técnica de minería de datos

Después se selecciona la vista de origen, para lo cual se seleccionó la vista llamada **vista_v_tp_distritos** la cual contiene la tabla **v_tp_distritos** que tiene los datos para la estructura de minería de datos. La figura 28 muestra la selección de la vista.



Figura 28. Selección de vista de origen de datos

Luego se escoge el escenario que se utilizará para la estructura de minería de datos. La figura 29 muestra la selección del escenario.

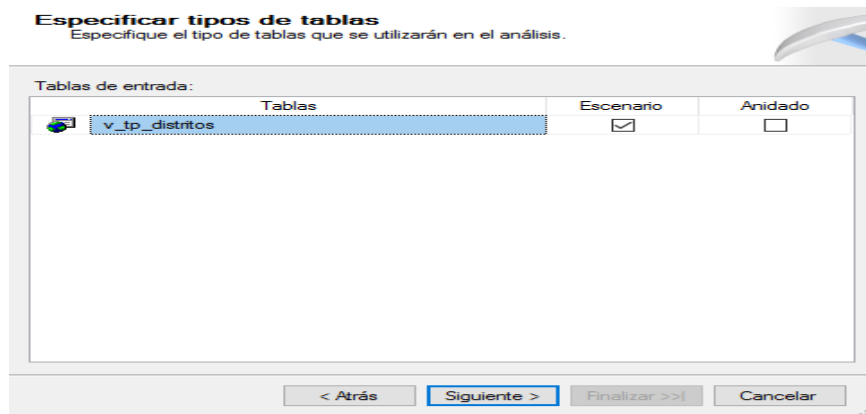


Figura 29. Selección de escenario

El siguiente paso fue escoger los datos de aprendizaje. La figura 30 muestra la selección de la clave, la entrada y el atributo que se predeciría.

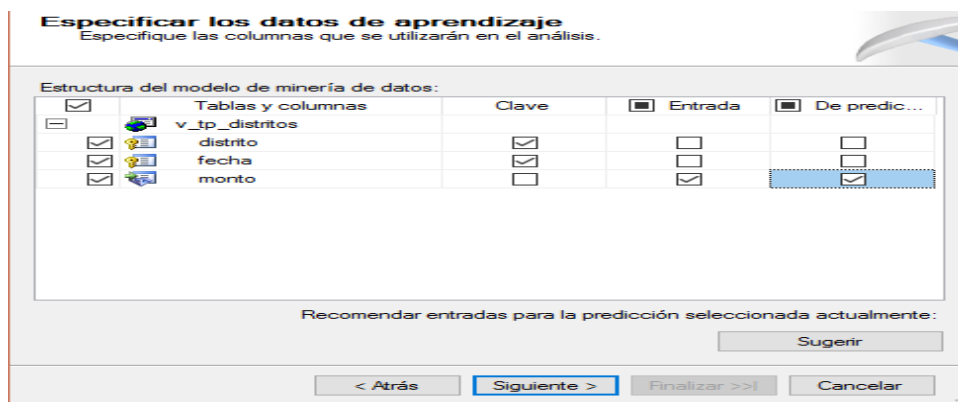


Figura 30. Especificación de datos de aprendizaje

Posteriormente se escoge el tipo de contenido y tipo de datos para los datos de aprendizaje previamente seleccionados. La figura 31 muestra la selección del tipo de contenido y tipo de dato para las columnas de la estructura de minería de datos.

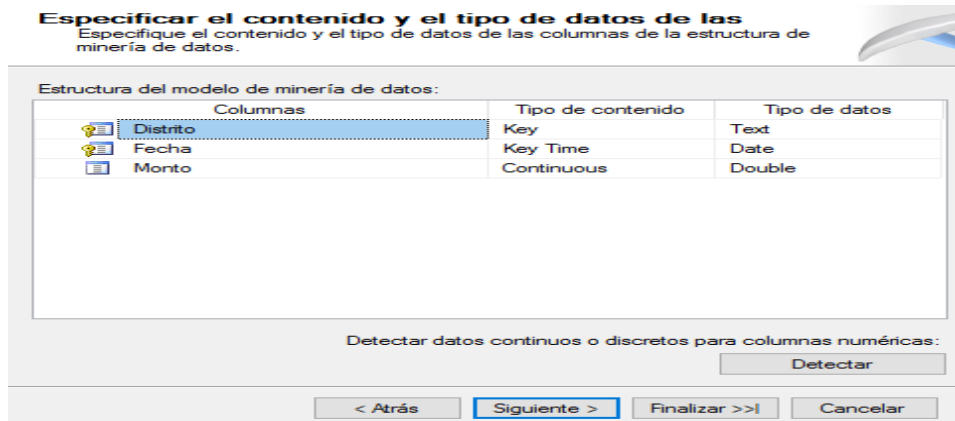


Figura 31. Especificación de contenido y tipo de datos

Finalmente se le da un nombre a la estructura y al modelo de minería de datos y se da clic en finalizar. La figura 32 muestra la finalización de la estructura de minería de datos.

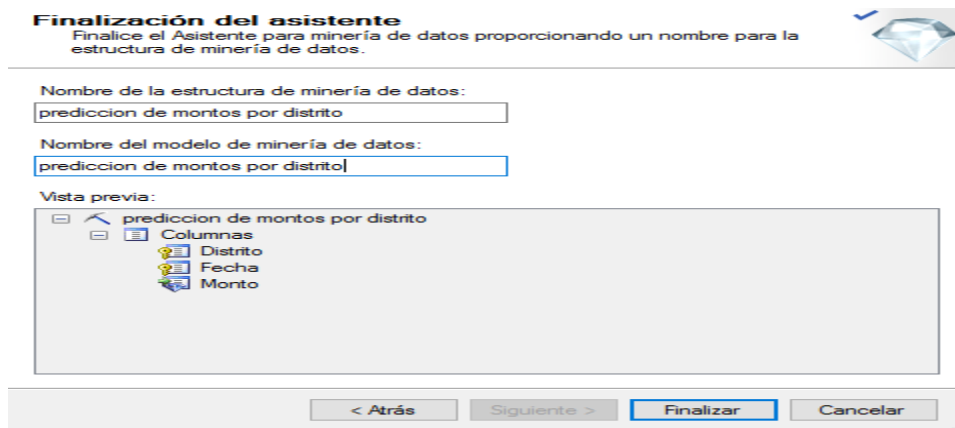


Figura 32. Finalización de creación de estructura de minería de datos

Una vez creada la estructura y modelo de minería de datos, estos se podrán visualizar. La figura 33 muestra la estructura y modelo de minería de datos.

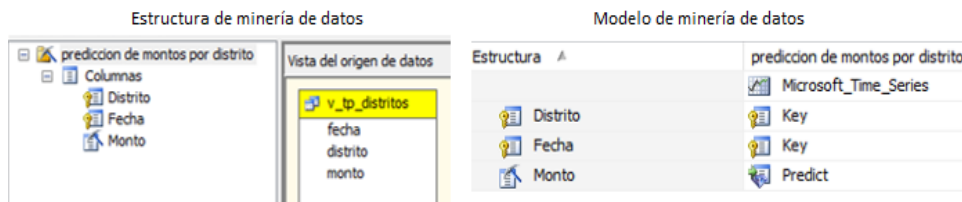


Figura 33. Estructura y modelo de minería de datos creado

Después teniendo el modelo de minería de datos se procede a establecer parámetros para el algoritmo. Primeramente se procesó un modelo con el método ARIMA y se evaluó su fiabilidad. La figura 34 muestra el modelo con el método ARIMA y la figura 35 muestra su fiabilidad.

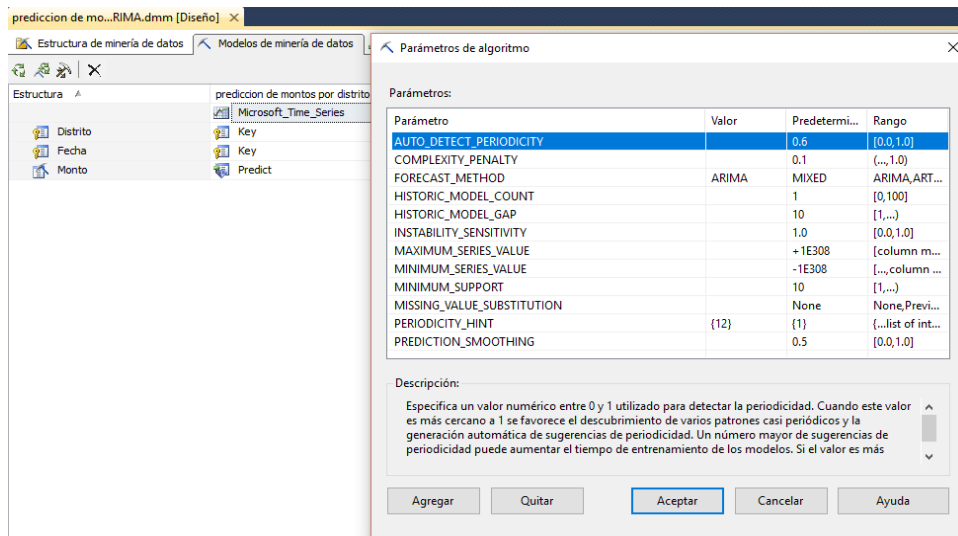


Figura 34. Método ARIMA

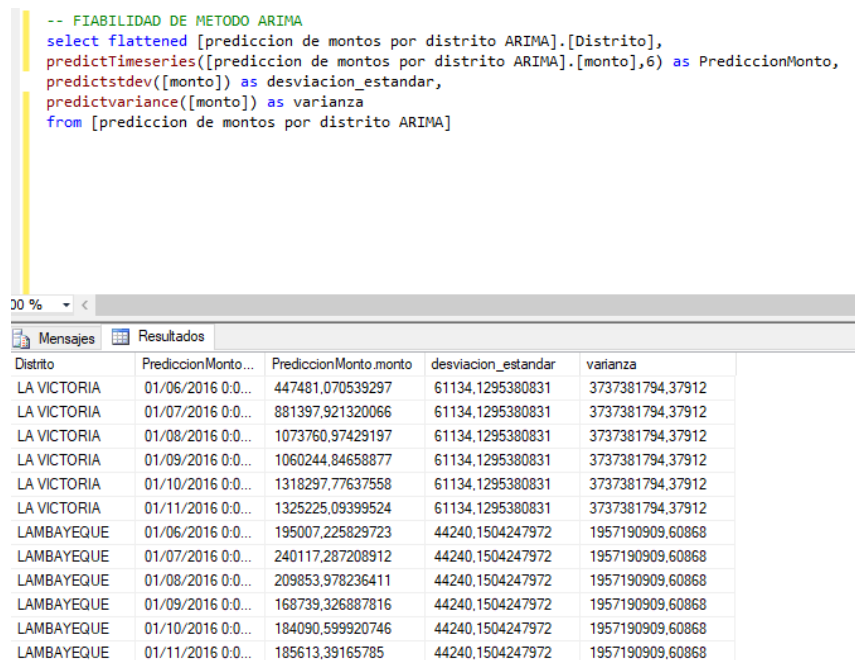


Figura 35. Fiabilidad de método ARIMA

Después se procesó un modelo con el método ARTXP y se evaluó su fiabilidad. La figura 36 muestra el modelo con el método ARTXP y la figura 37 muestra su fiabilidad.

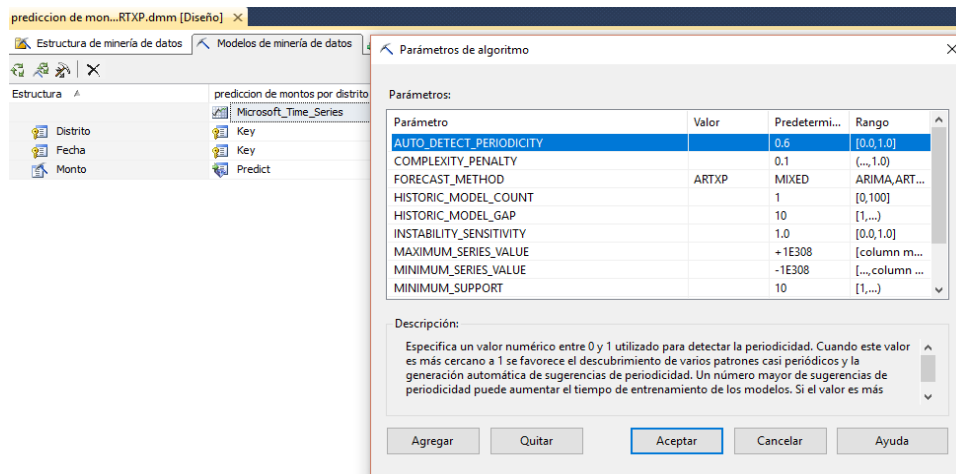


Figura 36. Método ARTXP

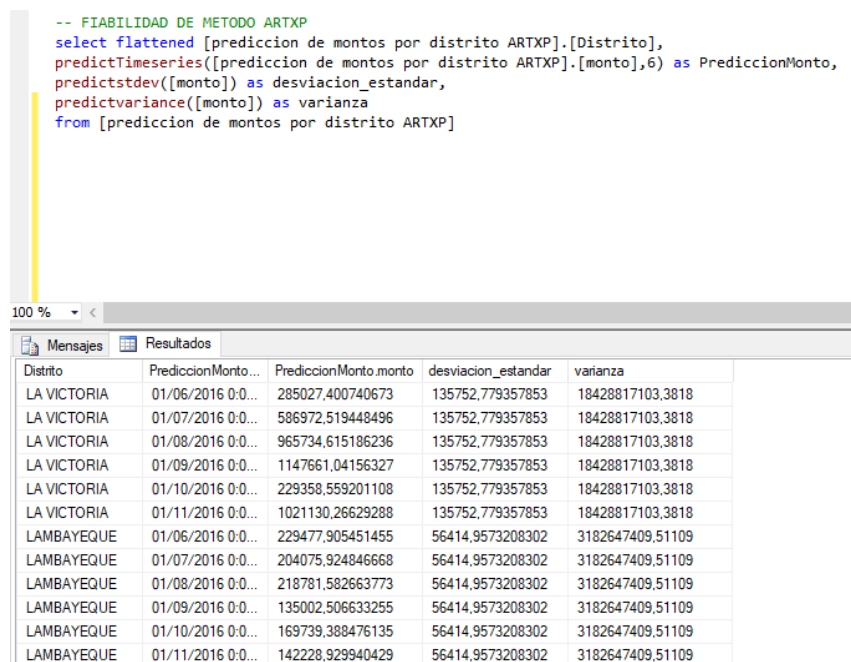


Figura 37. Fiabilidad de método ARTXP

Por último se procesó un modelo con el método MIXED y se evaluó su fiabilidad. La figura 38 muestra el modelo con el método MIXED y la figura 39 muestra su fiabilidad.

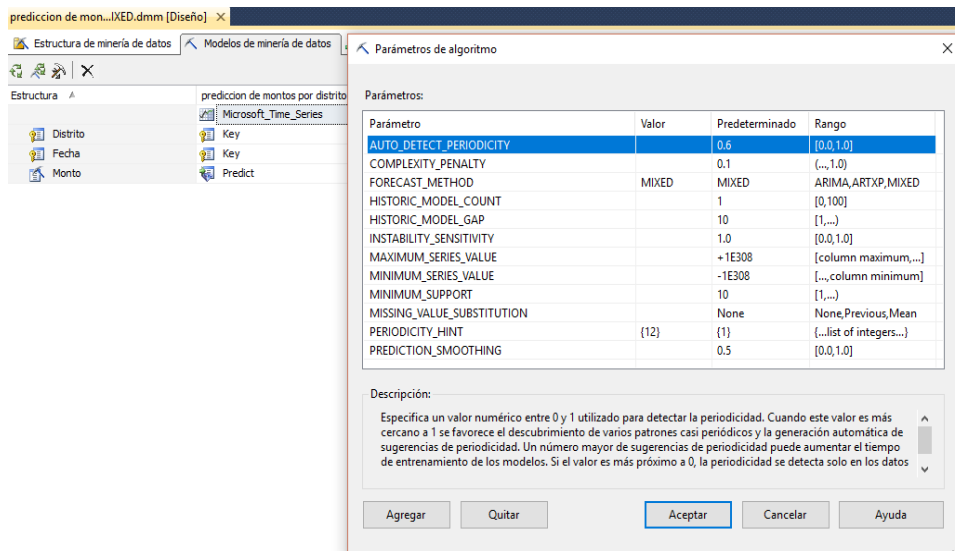


Figura 38. Método MIXED

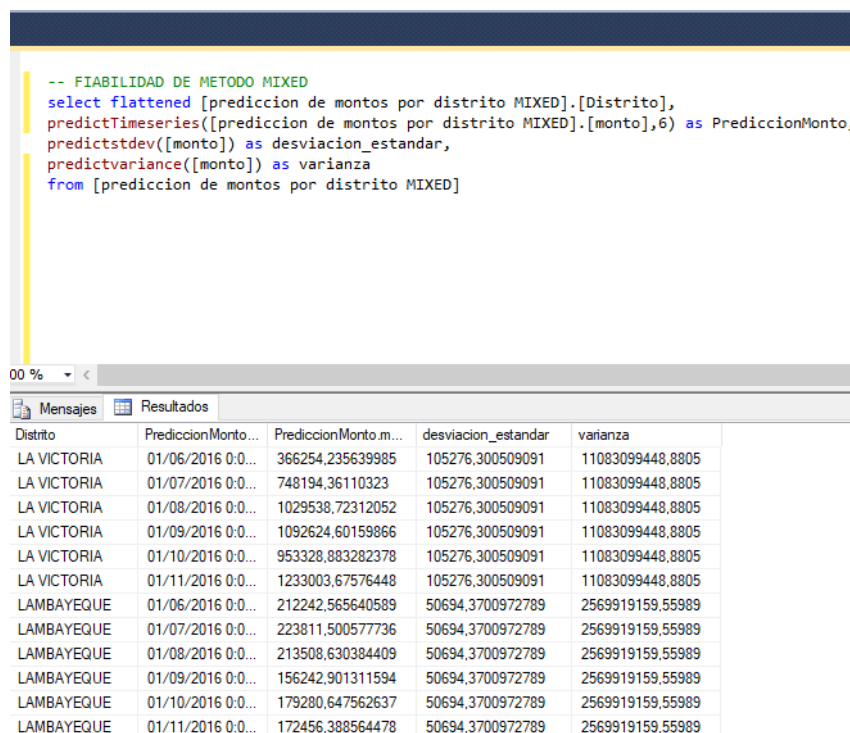


Figura 39. Fiabilidad de método MIXED

Dados los resultados por algoritmo. De acuerdo a su desviación estándar (mide cuánto se separan los datos con respecto a su media.) y varianza (desviación estándar elevada al cuadrado) se optó por tomar al método ARIMA. Por ser el método que tiene una menor desviación estándar y menor varianza, por lo tanto es el de mayor fiabilidad. Entonces los resultados del método ARIMA serán los que se muestre en el reporte de predicción de monto para los distritos donde se vende.

4.7.3. Construcción de reportes específicos:

Los reportes se realizaron de acuerdo a los requerimientos de los miembros del grupo empresarial, para la construcción de los reportes se utilizó SQL Server Reporting Services. Para que los usuarios finales accedan a los reportes construidos se creó un sistema web en asp.net con un logueo para identificación del usuario. La figura 40 muestra la interfaz de legueo para poder iniciar sesión.

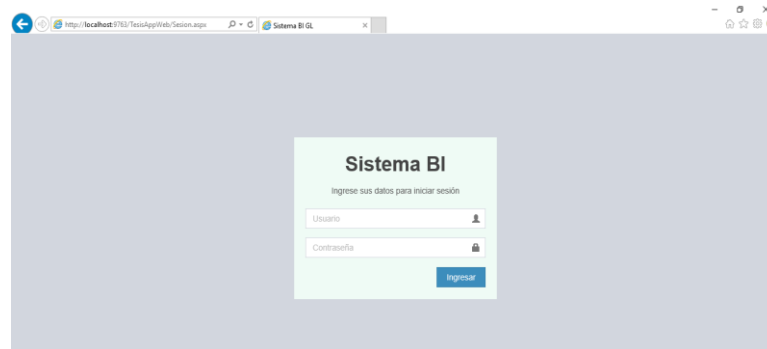


Figura 40. Interfaz para iniciar sesión

Una vez comprobado al usuario se podrá acceder a los reportes. A continuación se presentan los reportes desarrollados:

✓ Reportes de semaforización:

- Semaforización por año:

Para este reporte se evaluó el monto vendido por cada local, en el cual se tiene como meta aumentar en 30% el monto con respecto al monto del año anterior. Si el monto de ventas del año evaluado es igual o mayor en un 95% con respecto al objetivo el color del indicador será verde, si el monto de ventas del año actual es igual o mayor en un 90% con respecto al objetivo el color del indicador será ámbar y si el monto de ventas del año actual es menor en un 90% con respecto al objetivo el color del indicador será rojo. Además se muestra el crecimiento con respecto al año anterior y el cumplimiento de la meta planteada. La figura 41 muestra la interfaz de semaforización por año.

Nombre Local	Monto Real	Monto Objetivo	Crecimiento	Cumplimiento	Indicador
Local1	S/. 5.266.930,85	S/. 6.491.225,94	5,48 %	81,14%	↓
Local12	S/. 3.480.194,77	S/. 4.246.711,47	6,54 %	81,95%	↓
Local13	S/. 1.325.086,60	S/. 1.618.584,83	6,43 %	81,87%	↓
Local14	S/. 19.257,89	S/. 290.083,81	-91,37 %	6,64%	↓
Local15	S/. 2.226.791,93	S/. 2.023.077,68	43,09 %	110,07%	↑
Local16	S/. 1.416.652,76	S/. 1.367.531,49	34,67 %	103,59%	↑
Local17	S/. 754.410,71	S/. 775.655,21	26,44 %	97,26%	↑
Local18	S/. 4.678.910,48	S/. 5.124.648,01	18,69 %	91,30%	→
Local19	S/. 764.460,05	S/. 169.440,70	486,52 %	451,17%	↑
Local20	S/. 1.051.019,63	S/. 856.519,12	59,52 %	122,71%	↑
Local21	S/. 342.422,94	S/. 3.021.075,85	-85,27 %	11,33%	↓
Local26	S/. 49.020,04	S/. 30.295,20	110,35 %	161,81%	↑
Local3	S/. 1.050.234,19	S/. 1.605.593,60	-14,97 %	65,41%	↓
Local7	S/. 1.310.979,03	S/. 1.589.791,97	7,20 %	82,46%	↓

Figura 41. Semaforización por año

-SemafORIZACIÓN por semestre:

Para este reporte se evaluó el monto vendido por cada local, en el cual se tiene como meta aumentar en 30% el monto con respecto al monto del semestre paralelo anterior. Si el monto de ventas del semestre evaluado es igual o mayor en un 95% con respecto al objetivo el color del indicador será verde, si el monto de ventas del semestre evaluado es igual o mayor en un 90% con respecto al objetivo el color del indicador será ámbar y si el monto de ventas del semestre evaluado es menor en un 90% con respecto al objetivo el color del indicador será rojo. Además se muestra el crecimiento con respecto al semestre paralelo anterior y el cumplimiento de la meta planteada. La figura 42 muestra la interfaz de semafORIZACIÓN por semestre.

TABLA METAS POR SEMESTRE					
Nombre Local	Monto Real	Monto Objetivo	Crecimiento	Cumplimiento	Indicador
Local1	S/. 2.623.957,28	S/. 4.113.482,95	-17,07 %	63,79%	↓
Local12	S/. 1.723.865,64	S/. 2.376.779,67	-5,71 %	72,53%	↓
Local13	S/. 629.219,71	S/. 926.702,66	-11,73 %	67,90%	↓
Local14	S/. 16.776,00	S/. 46.921,42	-53,52 %	35,75%	↓
Local15	S/. 1.180.867,69	S/. 1.149.168,70	33,59 %	102,76%	↑
Local16	S/. 715.176,58	S/. 839.789,73	10,71 %	85,16%	↓
Local17	S/. 362.626,10	S/. 486.986,57	-3,20 %	74,46%	↓
Local18	S/. 2.827.523,12	S/. 2.557.200,39	43,74 %	110,57%	↑
Local20	S/. 644.472,74	S/. 286.443,25	192,49 %	224,99%	↑
Local21	S/. 1.080,00	S/. 1.473.336,58	-99,90 %	0,07%	↓
Local26	S/. 44.610,02	S/. 30.295,20	91,43 %	147,25%	↑
Local3	S/. 480.444,04	S/. 860.141,10	-27,39 %	55,86%	↓
Local7	S/. 737.873,00	S/. 795.596,10	20,57 %	92,74%	→

Figura 42. SemafORIZACIÓN por semestre

-SemafORIZACIÓN por trimestre:

Para este reporte se evaluó el monto vendido por cada local, en el cual se tiene como meta aumentar en 30% el monto con respecto al monto del trimestre paralelo anterior. Si el monto de ventas del trimestre evaluado es igual o mayor en un 95% con respecto al objetivo el color del indicador será verde, si el monto de ventas del trimestre evaluado es igual o mayor en un 90% con respecto al objetivo el color del indicador será ámbar y si el monto de ventas del trimestre evaluado es menor en un 90% con respecto al objetivo el color del indicador será rojo. Además se muestra el crecimiento con respecto al trimestre paralelo anterior y el cumplimiento de la meta planteada. La figura 43 muestra la interfaz de semafORIZACIÓN por trimestre.

TABLA METAS POR TRIMESTRE					
Nombre Local	Monto Real	Monto Objetivo	Crecimiento	Cumplimiento	Indicador
Local1	S/. 1.753.161,77	S/. 2.570.633,65	-11,34 %	68,20%	↓
Local12	S/. 993.885,05	S/. 1.280.340,75	0,91 %	77,63%	↓
Local13	S/. 359.085,80	S/. 493.564,76	-5,42 %	72,75%	↓
Local14	S/. 16.377,00	S/. 8.942,00	138,09 %	183,15%	↑
Local15	S/. 653.834,69	S/. 680.609,83	24,89 %	96,07%	↑
Local16	S/. 433.258,08	S/. 494.913,51	13,80 %	87,54%	↓
Local17	S/. 246.027,00	S/. 281.304,47	13,70 %	87,46%	↓
Local18	S/. 1.693.263,59	S/. 1.351.143,27	62,92 %	125,32%	↑
Local20	S/. 373.571,69	S/. 91.044,47	433,41 %	410,32%	↑
Local26	S/. 26.360,02	S/. 16.710,20	105,07 %	157,75%	↑
Local3	S/. 305.228,00	S/. 455.819,00	-12,95 %	66,96%	↓
Local7	S/. 448.162,60	S/. 451.012,90	29,18 %	99,37%	↑

Figura 43. SemafORIZACIÓN por trimestre

-SemafORIZACIÓN por mes:

Para este reporte se evaluó el monto vendido por cada local, en el cual se tiene como meta aumentar en 30% el monto con respecto al monto del mes paralelo anterior. Si el monto de ventas del mes evaluado es igual o mayor en un 95% con respecto al objetivo el color del indicador será verde, si el monto de ventas del mes evaluado es igual o mayor en un 90% con respecto al objetivo el color del indicador será ámbar y si el monto de ventas del mes evaluado es menor en un 90% con respecto al objetivo el color del indicador será rojo. Además se muestra el crecimiento con respecto al mes paralelo anterior y el cumplimiento de la meta planteada. La figura 44 muestra la interfaz de semaforización por mes.

Nombre Local	Monto Real	Monto Objetivo	Crecimiento	Cumplimiento	Indicador
Local1	S/. 849.317,97	S/. 1.218.588,72	-9,39 %	69,70%	↓
Local12	S/. 503.212,00	S/. 624.958,10	4,68 %	80,52%	↓
Local13	S/. 172.577,60	S/. 221.451,10	1,31 %	77,93%	↓
Local14	S/. 659,00	S/. 3.946,77	-78,29 %	16,70%	↓
Local15	S/. 320.895,11	S/. 296.151,70	40,86 %	108,35%	↑
Local16	S/. 233.723,98	S/. 234.222,30	29,72 %	99,79%	↑
Local17	S/. 122.089,00	S/. 139.924,20	13,43 %	87,25%	↓
Local18	S/. 919.876,78	S/. 537.001,52	122,69 %	171,30%	↑
Local20	S/. 142.300,54	S/. 29.359,37	530,09 %	484,69%	↑
Local26	S/. 7.140,00	S/. 8.840,00	5,00 %	80,77%	↓
Local3	S/. 130.274,50	S/. 191.457,50	-11,54 %	68,04%	↓
Local7	S/. 219.939,10	S/. 214.094,40	33,55 %	102,73%	↑

Figura 44. Semaforización por mes

✓ **Reportes de productos:**

-Reporte por líneas:

En este reporte se realizó un top de líneas por unidades vendidas y un top de líneas por monto vendido, ambos filtrados por un año seleccionado. Además se realizaron dos gráficos: Para poder visualizar las unidades vendidas y otro para el monto vendido. Esto en los diferentes meses filtrado por una línea y año seleccionado. La figura 45 muestra el reporte de productos por línea.

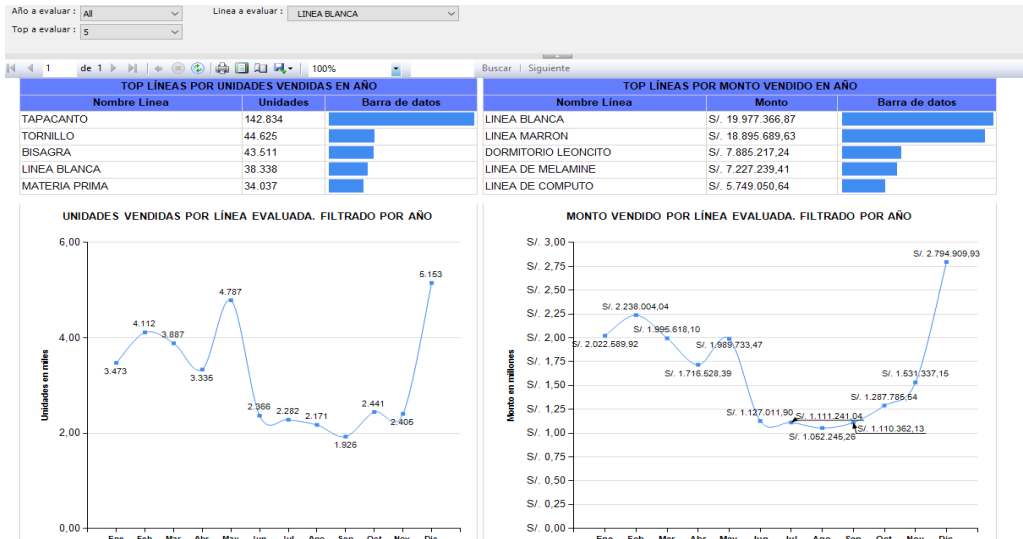


Figura 45. Reporte de producto por línea

- Reporte por marcas:

En este reporte se realizó un top de marcas por unidades vendidas y un top de marcas por monto vendido, ambos filtrados por un año seleccionado. Además se realizaron dos gráficos: Para poder visualizar las unidades vendidas y otro para el monto vendido. Esto en los diferentes meses filtrado por una marca y año seleccionado. La figura 46 muestra el reporte de productos por marca.

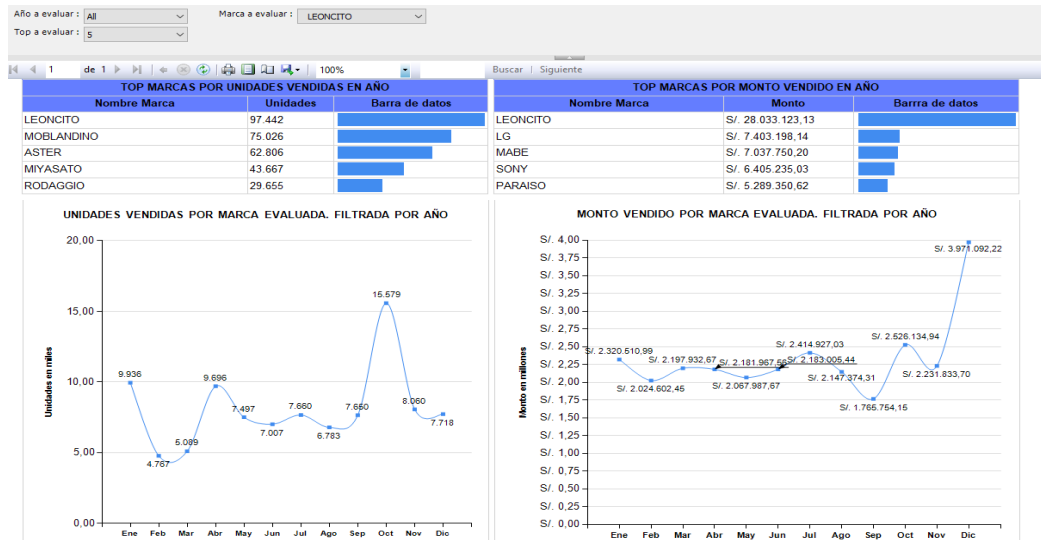


Figura 46. Reporte de producto por marca

✓ **Reportes de clientes:**

-Top clientes por año:

El presente reporte muestra el top clientes naturales y clientes jurídicos mostrando las unidades y montos de compra. Ambos top filtrados por año. La figura 47 muestra el reporte de top de clientes por año.

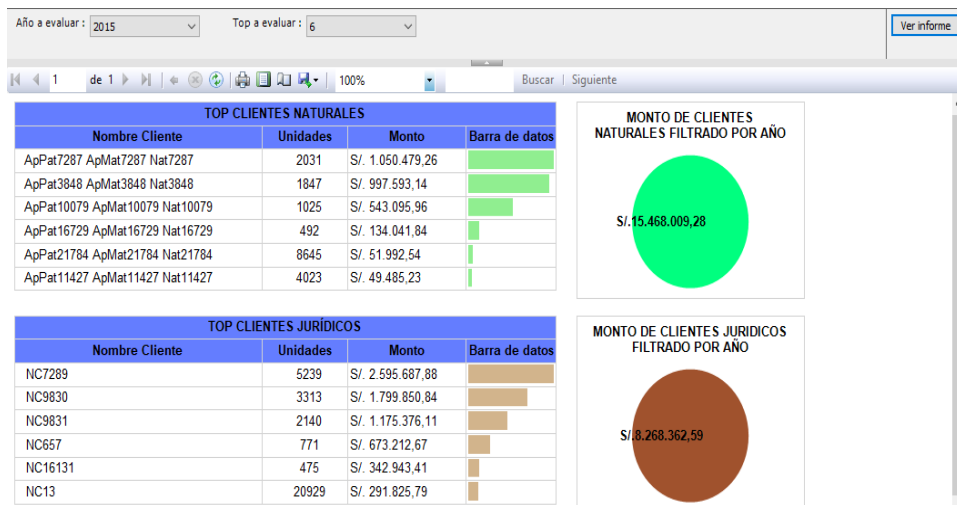


Figura 47. Reporte top de clientes por año

✓ **Reportes de vendedores:**

-Top vendedores por año:

El reporte muestra al top de vendedores por unidades vendidas filtrado por un año y marca seleccionada. La figura 48 muestra el reporte de vendedores por año y marca.

Nombre Vendedor	Unidades	Barra de datos	Monto	Barra de datos
ApPat12391 ApMat12391 Nat12391	5.895		S/ 8.565,33	
ApPat23393 ApMat23393 Nat23393	4.515		S/ 995.222,69	
ApPat15870 ApMat15870 Nat15870	3.983		S/ 2.931.905,62	
ApPat1541 ApMat1541 Nat1541	600		S/ 417.247,94	
ApPat25347 ApMat25347 Nat25347	537		S/ 424.397,99	
ApPat2657 ApMat2657 Nat2657	420		S/ 138.982,89	
ApPat5102 ApMat5102 Nat5102	338		S/ 311.859,80	
ApPat26281 ApMat26281 Nat26281	325		S/ 145.887,67	
ApPat745 ApMat745 Nat745	322		S/ 293.768,23	
ApPat658 ApMat658 Nat658	239		S/ 220.712,00	

Figura 48. Reporte de vendedores por año y marca

-Top vendedores por mes:

El reporte muestra al top de vendedores por unidades vendidas filtrado por mes, año y marca seleccionada. La figura 49 muestra el reporte de vendedores por mes y marca.

Nombre Vendedor	Unidades	Barra de datos	Monto	Barra de datos
ApPat15870 ApMat15870 Nat15870	559		S/ 511.189,15	
ApPat23393 ApMat23393 Nat23393	310		S/ 189.709,49	
ApPat5102 ApMat5102 Nat5102	46		S/ 45.306,10	
ApPat745 ApMat745 Nat745	39		S/ 39.077,00	
ApPat218 ApMat218 Nat218	38		S/ 48.098,00	
ApPat15914 ApMat15914 Nat15914	27		S/ 33.561,00	
ApPat658 ApMat658 Nat658	27		S/ 26.211,00	
ApPat1541 ApMat1541 Nat1541	25		S/ 30.036,00	
ApPat20006 ApMat20006 Nat20006	25		S/ 17.938,00	
ApPat9755 ApMat9755 Nat9755	25		S/ 27.596,00	

Figura 49. Reporte de vendedores filtrados por marca, año y mes

Para ambos reportes de vendedores al dar clic en el nombre de un vendedor (sea en el reporte top vendedores por año o en el reporte top vendedores por mes) se podrá visualizar otro reporte en el cual se podrá ver del vendedor las unidades y monto vendido de un determinado año en los diferentes meses del año seleccionado. También las unidades y monto vendido en los diferentes distritos en los cuales tiene ventas registradas. La figura 50 muestra el reporte de un vendedor en el año.

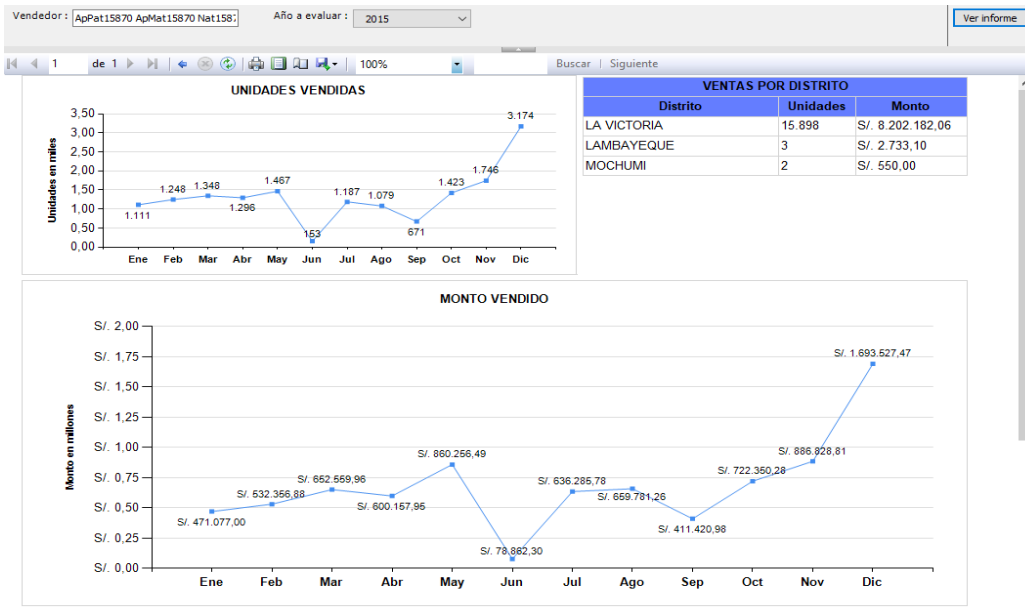


Figura 50. Reporte de vendedor en año

✓ Reportes de ventas:

-Ventas por distrito:

Este reporte brinda información sobre las ventas e los diferentes distritos ya sea diferenciado por el tipo de venta o las ventas en general y también muestra el progreso de las ventas en los diferentes meses. Esto evaluado por el año o años seleccionados. La figura 51 muestra el reporte de ventas por distrito.

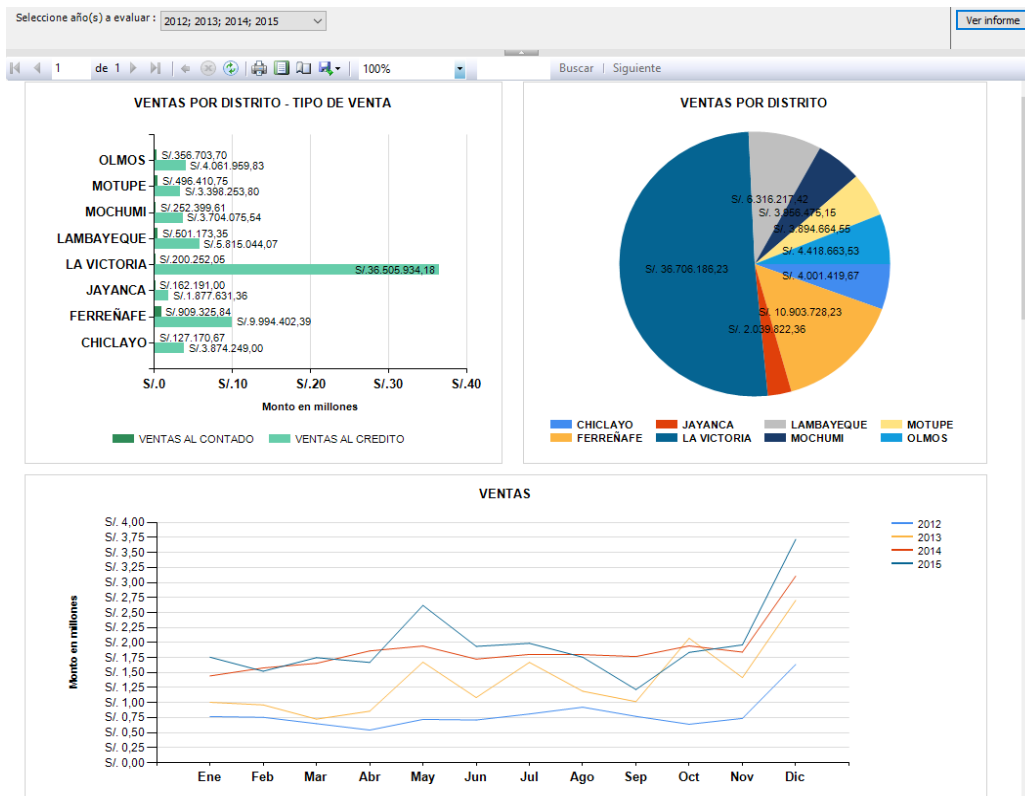


Figura 51. Reporte de ventas por distrito

-Ventas por local:

En este reporte se puede visualizar el monto de venta por tipo de venta en los diferentes locales en un año seleccionado. La figura 52 muestra el reporte de ventas por local.

Selección año a evaluar : 2015			
Ver informe			
1 de 1			
VENTAS			
NOMBRE LOCAL	VENTAS AL CONTADO	VENTAS AL CREDITO	TOTAL
Local1	S/. 4.559,00	S/. 5.262.371,85	S/. 5.266.930,85
Local12	S/. 246.185,77	S/. 3.234.009,00	S/. 3.480.194,77
Local13	S/. 80.376,11	S/. 1.244.710,49	S/. 1.325.086,60
Local14	S/. 0,00	S/. 19.257,89	S/. 19.257,89
Local15	S/. 175.656,62	S/. 2.051.135,31	S/. 2.226.791,93
Local16	S/. 158.082,20	S/. 1.258.570,56	S/. 1.416.652,76
Local17	S/. 57.635,00	S/. 696.775,71	S/. 754.410,71
Local18	S/. 892,53	S/. 4.678.017,95	S/. 4.678.910,48
Local19	S/. 16.301,70	S/. 748.158,35	S/. 764.460,05
Local20	S/. 0,00	S/. 1.051.019,63	S/. 1.051.019,63
Local21	S/. 0,00	S/. 342.422,94	S/. 342.422,94
Local26	S/. 0,00	S/. 49.020,04	S/. 49.020,04
Local3	S/. 18.893,25	S/. 1.031.340,94	S/. 1.050.234,19
Local7	S/. 111.796,50	S/. 1.199.182,53	S/. 1.310.979,03
TOTAL	S/. 870.378,68	S/. 22.865.993,19	S/. 23.736.371,87

Figura 52. Reporte de ventas por local

Además al dar clic en el nombre de un local automáticamente se dirige a otro reporte con las ventas en los diferentes meses del local al cual se le dio clic y los meses que se mostraran serán los del año previamente seleccionado, también se muestra el top de producto ordenados por las unidades vendidas en dicho local. La figura 53 muestra el reporte de ventas por local en un determinado año, además el top de productos del local.

Top a evaluar : 10			
Ver informe			
1 de 1			
VENTAS			
MESES	VENTAS AL CONTADO	VENTAS AL CREDITO	TOTAL
Enero	S/. 13.272,00	S/. 150.698,20	S/. 163.970,20
Febrero	S/. 12.293,00	S/. 116.447,20	S/. 128.740,20
Marzo	S/. 20.791,10	S/. 161.897,40	S/. 182.688,50
Abril	S/. 8.121,00	S/. 121.933,00	S/. 130.054,00
Mayo	S/. 17.701,05	S/. 216.886,10	S/. 234.587,15
Junio	S/. 4.815,10	S/. 201.069,09	S/. 205.884,19
Julio	S/. 24.082,05	S/. 153.755,20	S/. 177.837,25
Agosto	S/. 16.644,15	S/. 187.707,00	S/. 204.351,15
Septiembre	S/. 10.071,00	S/. 134.773,60	S/. 144.844,60
Octubre	S/. 10.984,07	S/. 155.922,00	S/. 166.906,07
Noviembre	S/. 12.099,00	S/. 153.934,51	S/. 166.033,51
Diciembre	S/. 24.783,10	S/. 296.112,01	S/. 320.895,11
TOTAL	S/. 175.656,62	S/. 2.051.135,31	S/. 2.226.791,93
TOP PRODUCTOS POR UNIDADES VENDIDAS			
PRODUCTO	Unidades	Monto	
DE RESORTES MEDALLON ERGO ONE SIDE 2 PLZS	103	S/. 62.210,00	
ZEBRA 18 RIBETADO DE 18 KG/M3 2 PLZ X 8 ACOLCHADO	50	S/. 19.902,00	
DE RESORTES MEDALLON ERGO ONE SIDE 2 PLZ COMPRIMIDO	46	S/. 27.411,01	
COLCHON ZEBRA 18 RIBETADO DE 18KG/M3 1.5 PLZ X8 ACOLCHADO	42	S/. 12.502,50	
DVD USB NUEVO LG MOD. DP132-NU	36	S/. 4.587,00	
RESORTES EDEN * 1.5 PLZ	33	S/. 9.955,05	
LEO SKU-DM3P22 ROPER CON ILLUM. TRES CUERPOS	31	S/. 44.034,40	
LEO SKU-DMCAI2P22 CAMA ITALIA 2 PLAZAS	31	S/. 31.887,40	
DE RESORTES MEDALLON ERGO ONE SIDE 1.5 PLZS	30	S/. 16.378,00	
LED 32" SAMSUG 32FH4005	26	S/. 24.142,00	

Figura 53. Reporte de ventas por local y año – top productos

-Predicción por monto:

Para el reporte de predicción se tomó en cuenta a los distritos ya que cada fin de mes se evalúa el monto que venden por cada uno de ellos, por lo cual conversando con el gerente administrativo se vio conveniente predecir el monto que se venderá en cada uno de estos en los siguientes seis meses. La figura 54 muestra el reporte de predicción de ventas por distrito.

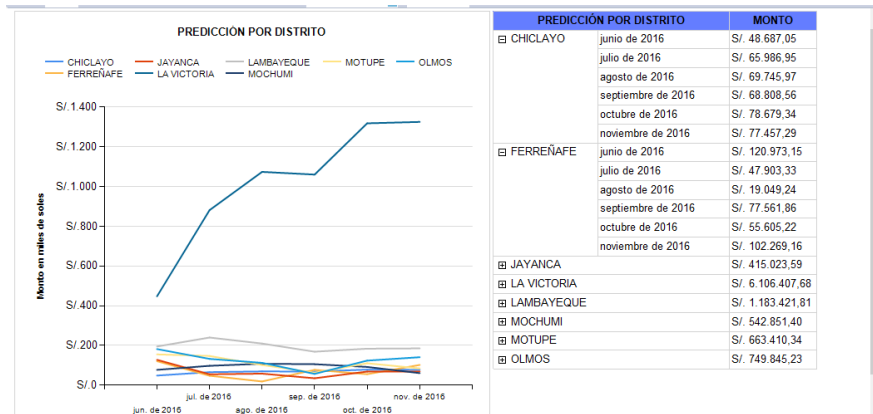


Figura 54. Reporte de predicción de ventas por distrito

-Predicción de unidades de venta por producto:

Para el reporte de predicción se tomó en cuenta a los productos para poder saber el stock que se debe tener para cada uno de ellos, por lo cual conversando con el gerente administrativo se vio conveniente predecir en cada uno de estos en los siguientes seis meses. La figura 55 muestra el reporte de predicción de unidades por producto.

PRODUCTO	UNIDADES
ACCESORIO	21
	28
	31
	32
	32
	32
ACEITE	1
	1
	1
	1
	3
ALAMBRE	73
	55
	60
	76
	74
	18
ALMOHADA	459
ANTENA	14

Figura 55. Reporte de predicción de ventas por distrito

4.8. Despliegue

Esta parte implica probar el sistema BI construido y que este funcione correctamente. Para esto se brindó la documentación correspondiente.

4.9. Mantenimiento y crecimiento

Con respecto al crecimiento es importante establecer prioridades para poder manejar nuevos requerimientos de los usuarios finales que puedan surgir en el futuro y para esto los encargados que laboren en la empresa deben de estar atentos a dichas necesidades que se presenten para que estos puedan cumplir con los nuevos requerimientos.

Con respecto al mantenimiento del data mart en este caso llamado **dm_gl**, para que este se encuentre actualizado se crearon tareas programadas. De esta manera Integration Services y Analysis Services se ejecutarán todos los días a la media noche para que se actualice la información que se muestre a los usuarios finales.

Primeramente el Agente SQL Server debe de estar iniciado. La figura 56 muestra al Agente SQL Server iniciado.

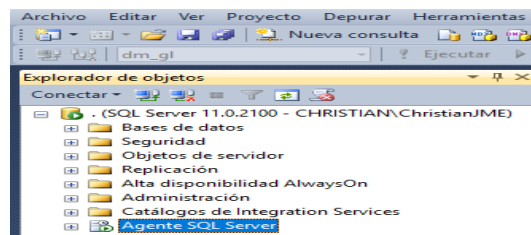


Figura 56. Agente SQL Server

En caso de que el servicio no se encuentre iniciado, se debe acceder a servicios, se busca al Agente SQL Server, dar clic secundario y se inicia el servicio. La figura 57 muestra como iniciar el servicio Agente SQL Server.

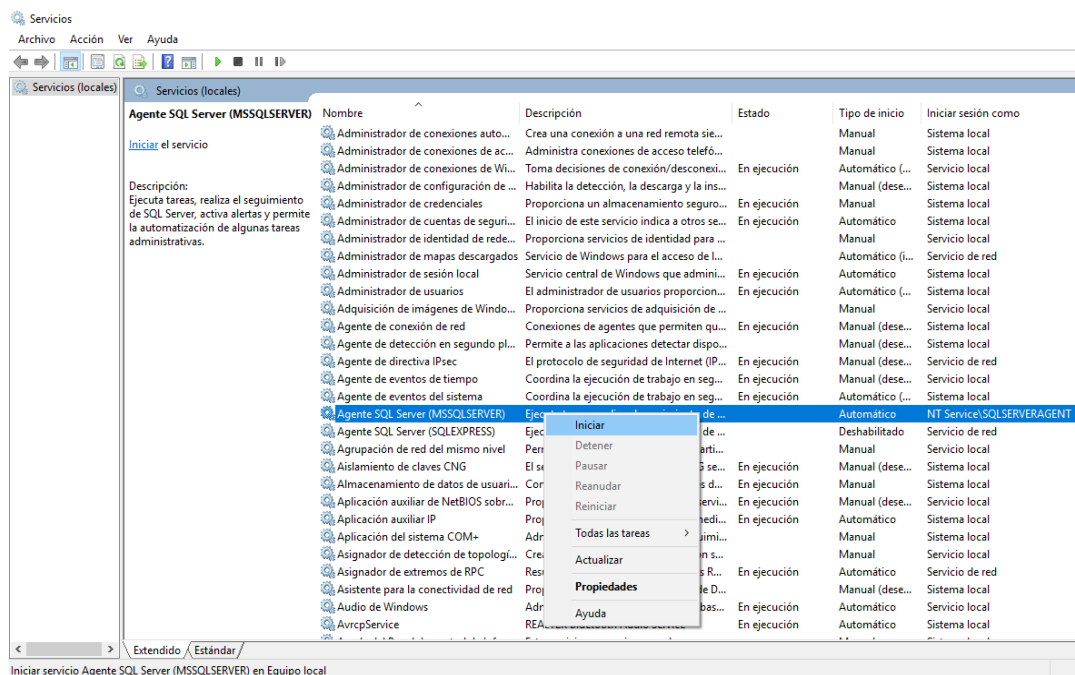


Figura 57. Servicios – iniciar el Agente SQL Server

Con el servicio Agente SQL Server iniciado se crea un nuevo trabajo al cual se le dio el nombre **EjecutarETL** para el procesamiento del paquete de Integration Services llamado en este caso **PoblamientoDataMart**. La figura 58 muestra la creación de un nuevo trabajo llamado **EjecutarETL**.

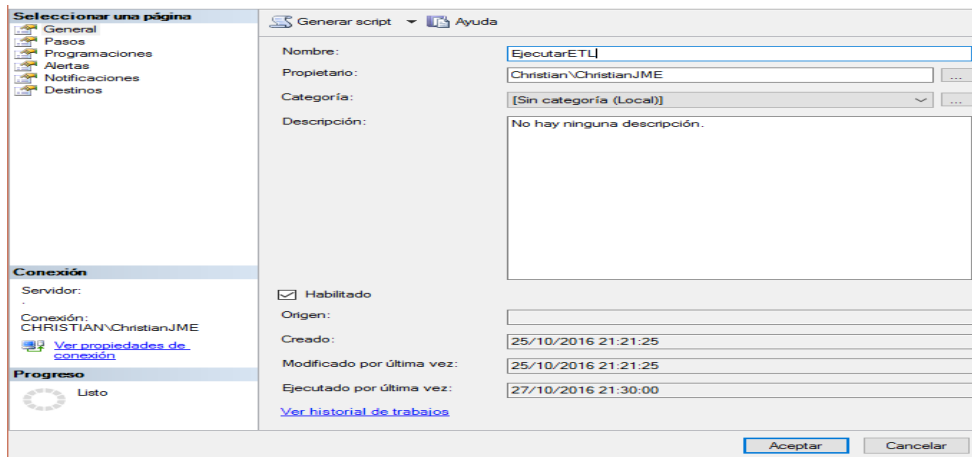


Figura 58. Creación de un nuevo TRABAJO – EjecutarETL

Después se accede a Pasos y se define uno nuevo. En este caso se le dio el nombre de **PasoEjecutarETL** y el tipo es Paquete SQL Server Integration Services (paquete donde se encuentra todo el proceso de ETL) el cual se ejecutará como Cuenta del servicio del Agente SQL Server. Después en el combo el origen del paquete elegimos Sistema de archivos y se coloca la dirección donde está ubicado el paquete, en este caso llamado **PoblamientoDataMart.dtsx** y se da clic en aceptar. La figura 59 muestra la creación del paso **PasoEjecutarETL**.

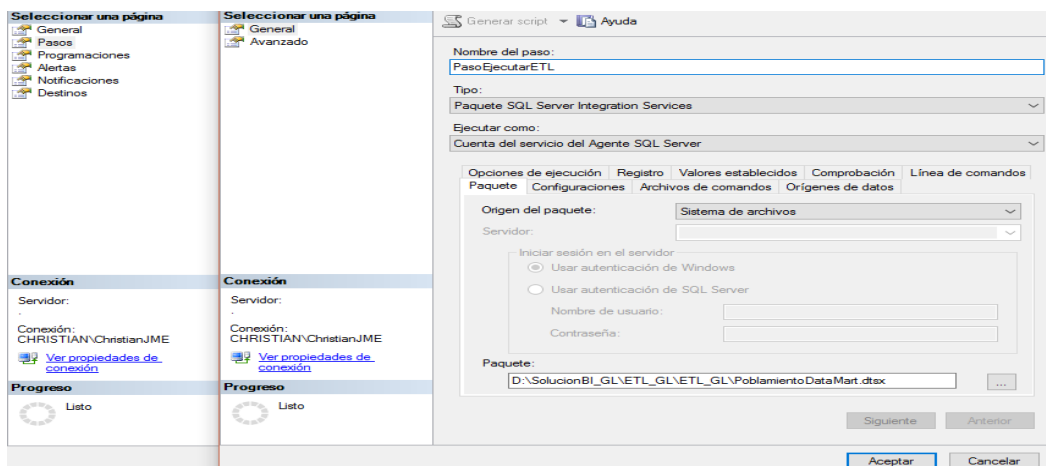


Figura 59. Creación de un nuevo PASO – PasoEjecutarETL

Luego se accede a Programaciones y se define una nueva. En este caso se le dio el nombre de **ProgramacionEjecutarETL**. En el combo Tipo de programación se elegimos Periódica, la Frecuencia que suceda Diaria que se repita una vez, la frecuencia diaria que suceda una vez a la media noche, la duración que sea sin fecha de finalización y se da clic en aceptar. La figura 60 muestra la creación de la programación **ProgramacionEjecutarETL**.

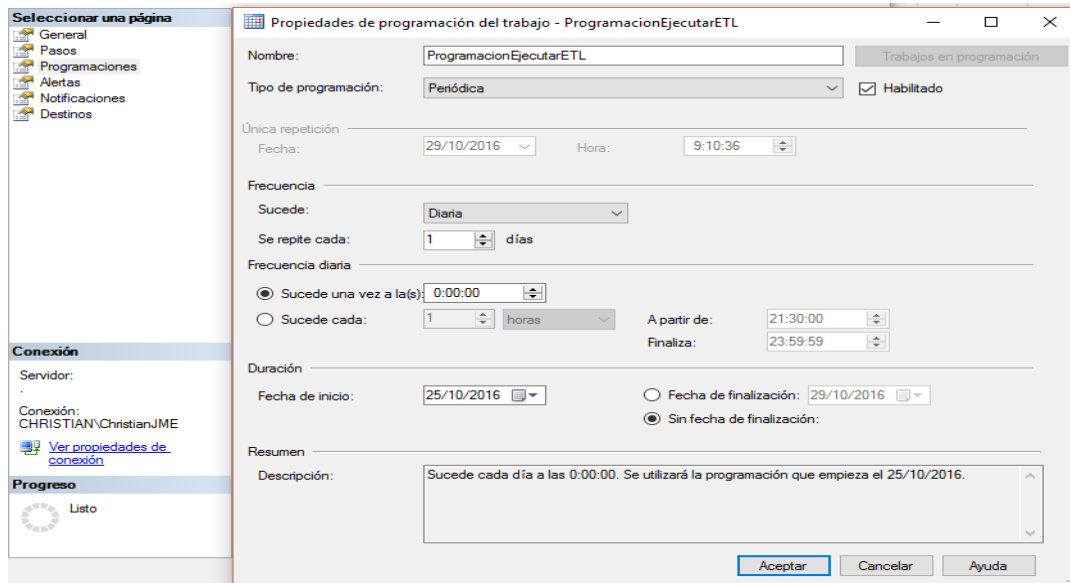


Figura 60. Creación de una nueva programación – ProgramacionEjecutarETL

Concluido con el trabajo programado el Data Mart en este caso llamado **dm_gl** se actualizará y a partir de éste tendrá que procesarse el cubo en este caso llamado **cubo_ventas** para actualizar la base de datos en Analysis Services, a través del cual se conectarán los reportes en Reporting Services.

Para crear el trabajo programado que se encargará de procesar el cubo cada noche después de ejecutar el ETL, se debe de ingresar a Analysis Services desde SQL Server Management Studio. Ya estando en Analysis Services damos clic secundario al cubo **cubo_ventas** y elegimos la pestaña Procesar. Estando allí vamos a la opción Generar script y escogemos la opción Generar script de acción en Portapapeles y de manera se obtiene script para enviarlo al trabajo programado en el Agente SQL Server. La figura 61 muestra la generación del script en el portapapeles.

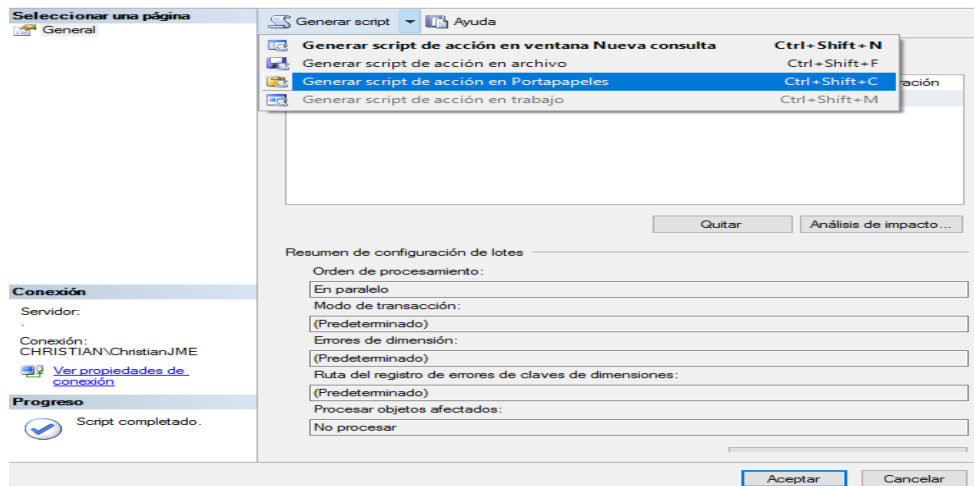


Figura 61. Creación de script en portapapeles

Después de tener el script en el portapapeles creamos un nuevo trabajo al cual se le dio el nombre **EjecutarOLAP** para el procesamiento del cubo llamado en este caso **cubo_ventas**. La figura 62 muestra la creación de un nuevo trabajo llamado **EjecutarOLAP**.

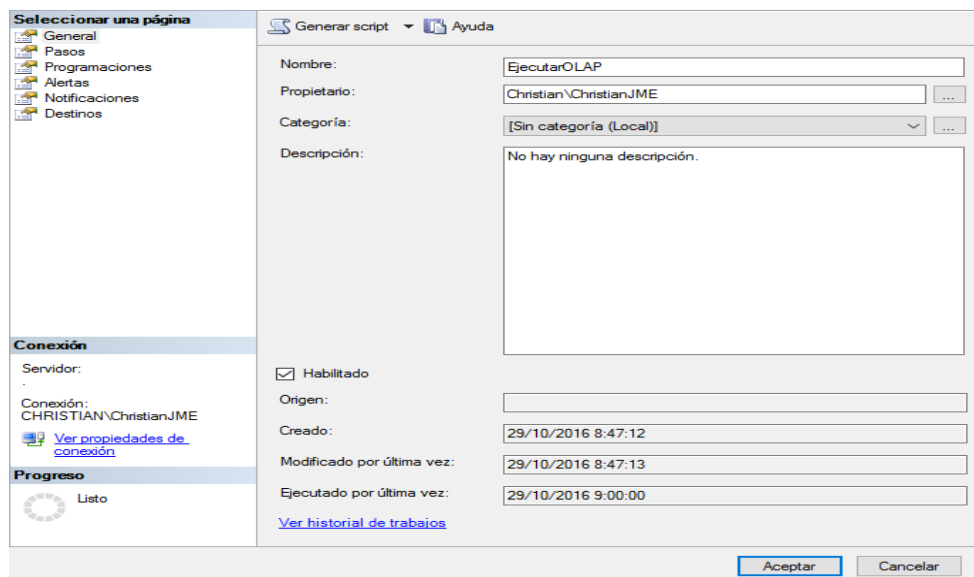


Figura 62. Creación de un nuevo trabajo - EjecutarOLAP

Después se accede a Pasos y se define uno nuevo. En este caso se le dio el nombre de **PasoEjecutarOLAP** y el tipo es Comando de SQL Server Analysis Services (debido a que se ejecutará una instrucción XMLA) el cual se ejecutará como Cuenta del servicio del Agente SQL Server, en el servidor se coloca localhost o simplemente un punto (.) que representa lo mismo, se pegan las líneas XMLA que se copiaron al portapapeles y se da clic en aceptar. La figura 63 muestra la creación del paso **PasoEjecutarOLAP**.

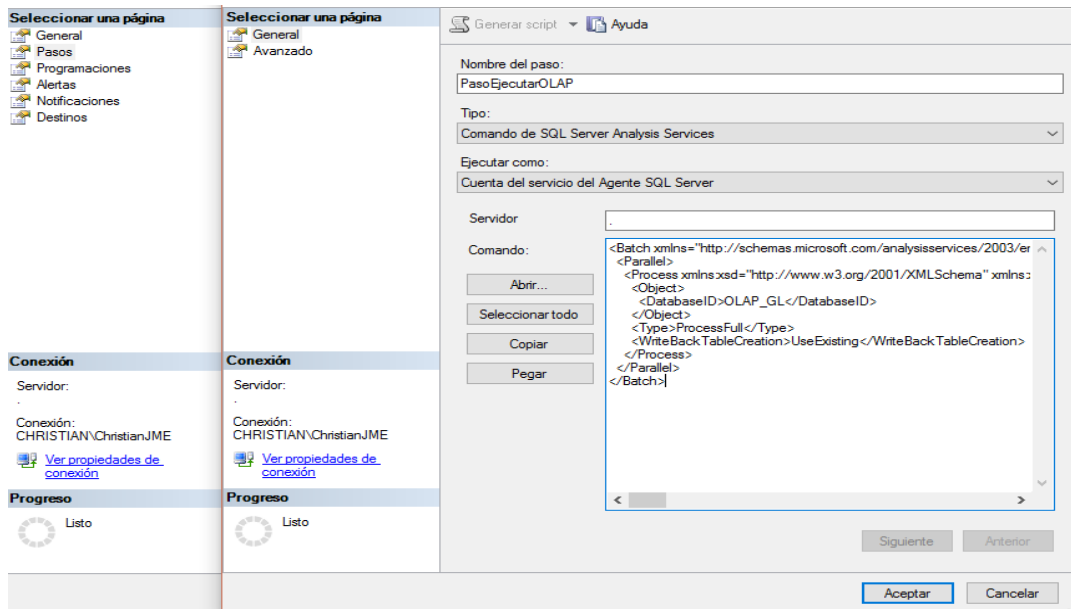


Figura 63. Creación de un nuevo paso – PasoEjecutarOLAP

Luego se accede a Programaciones y se define una nueva. En este caso se le dio el nombre de **ProgramacionEjecutarOLAP**. En el combo Tipo de programación se elegimos Periódica, la Frecuencia que suceda Diaria que se repita una vez, la frecuencia diaria que suceda una vez a la media noche, la duración que sea sin fecha de finalización y se da clic en aceptar. La figura 64 muestra la creación de la programación.

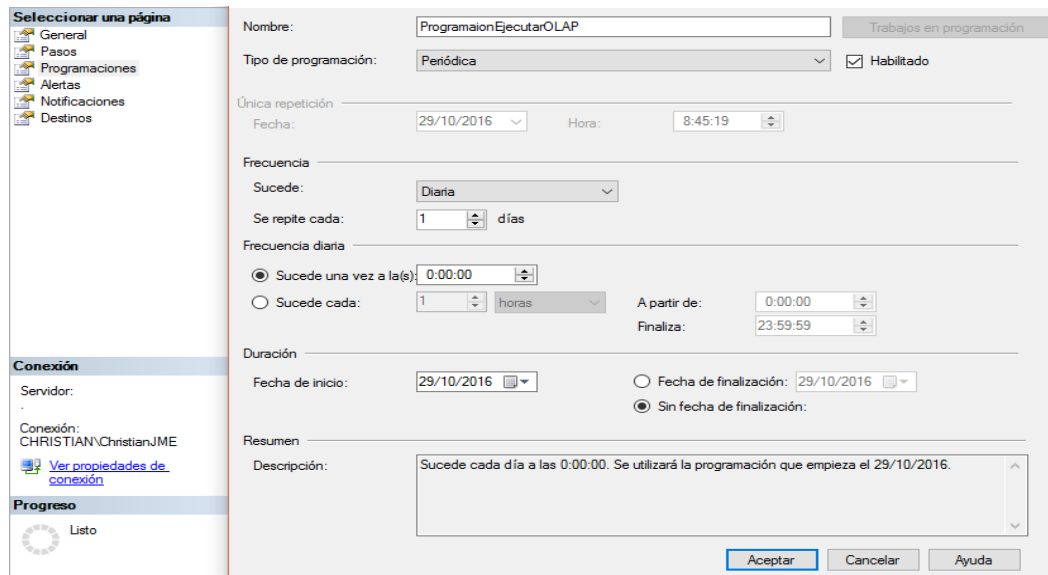


Figura 64. Creación de una nueva programación – ProgramacionEjecutarOLAP

V. DISCUSIÓN

5.1. Indicador 1:

Tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas.

Indicador	O1	O2	Diferencia
Tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas.	2 880'	0'031''	2 879'969''

Tabla 18. Indicador 1

Para calcular el **O2** se aplicó una formula. La aplicación de la fórmula se encuentra en el Anexo 3.

O2 = Tiempo promedio que toma la obtención de información sobre ventas

O2 = Suma de tiempos por reporte / Total de mediciones

O2 = **0'031''**

- **O1:** Es el tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas sin el uso del sistema de BI.
- **O2:** Es el tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas con el uso del sistema de BI.
- **Diferencia (O2 – O1):** Durante el análisis de la realidad problemática del grupo empresarial Leoncito se identificó que había una demora con respecto a la obtención de información sobre ventas, según dos entrevistas realizadas con el gerente administrativo y con el analista - programador de sistemas. La demora podía ser de hasta 2 días (2 879 minutos) porque el área de tecnologías de información a quien se solicitaba dicha información ya contaba con actividades diarias, aparte de estar pendiente de los inconvenientes que puedan ocurrir en el día a día, agregar nuevas funcionalidades y darle mantenimiento constante al sistema transaccional con el que se cuenta. Ahora, con el sistema de BI que se construyó, el tiempo promedio se redujo a **0'031''** debido a que se tiene acceso rápido a los reportes a través de un navegador web el cual es uso exclusivo para los miembros del grupo empresarial. Por lo tanto se cumplió con el objetivo propuesto: Reducir el tiempo promedio que toma la obtención de información sobre ventas. Lo cual se traduce en información en tiempo oportuno para la toma de decisiones para así convertirse en ventaja competitiva (Muineló 2016).

5.2. Indicador 2 y 4:

Por estar fuertemente relacionados se evaluaron conjuntamente.

- 5.2.1. Indicador 2: Cantidad de reportes sobre ventas que permita análisis comparativo en el tiempo.
- 5.2.2. Indicador 4: Cantidad de reportes para el análisis de las ventas a través de las diferentes dimensiones.

Indicador	O1	O2	Diferencia
Cantidad de reportes sobre ventas que permita análisis comparativo en el tiempo.	1	9	8
Cantidad de reportes para el análisis de las ventas a través de las diferentes dimensiones.			

Tabla 19. Indicador 2 – Indicador 4

- **O1:** Es la cantidad de reportes sobre ventas que permita análisis comparativo en el tiempo y la cantidad de reportes para el análisis de las ventas a través de las diferentes dimensiones sin el uso del sistema de BI.
- **O2:** Es la cantidad de reportes sobre ventas que permita análisis comparativo en el tiempo y la cantidad de reportes para el análisis de las ventas a través de las diferentes dimensiones con el uso del sistema de BI.

Diferencia (O2 – O1): Durante el análisis de la realidad problemática del grupo empresarial Leoncito se identificó que tenían 1 reporte que permitía análisis comparativo de sus ventas por distrito a través de una tabla la cual mostraba los montos de venta durante los años de vida de la empresa, sin embargo existía una demora aproximada de 13 minutos para cargar el reporte. Ahora, con el sistema de BI que se construyó, el grupo empresarial cuenta con 9 reportes. Los cuales cuentan con tablas y dashboards que permiten apreciar de una manera fácil y amigable la información. Permiten analizar sus ventas desde diferentes dimensiones ya sea por producto, cliente, local, etc. en los diferentes meses y años para así poder detectar alzas y bajas en las ventas, patrones estacionales, etc. Se tuvo un incremento del 100% del total de reportes. Por lo tanto se cumplió con el objetivo propuesto: Incrementar el número de reportes que permita análisis comparativo de las ventas en el tiempo e incrementar el número de reportes para el análisis de las ventas a través de las diferentes dimensiones. Esto gracias al procesamiento analítico en línea (OLAP) lo cual permite a los usuarios una fácil y amigable navegación por la información (Córdova Yupanqui 2013). Además se logró optimizar los tiempos de respuesta con respecto al reporte del sistema transaccional. Reduciendo la demora de 13' a 0'031'' debido a que esta información esta calculada previamente en el cubo OLAP.

5.3. Indicador 3:

Cantidad de reportes de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas.

Indicador	O1	O2	Diferencia
Cantidad de reportes de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas.	0	4	4

Tabla 20. Indicador 3

- **O1:** Es la cantidad de reportes de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas sin el uso del sistema de BI.
- **O2:** Es la cantidad de reportes de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas con el uso del sistema de BI.
- **Diferencia (O2 – O1):** Durante el análisis de la realidad problemática del grupo empresarial Leoncito se identificó que tenía 0 reportes de indicadores que permitan monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas. Ahora, con el sistema de BI que se construyó, el grupo empresarial tiene reportes con indicadores mensuales, trimestrales, semestrales y anuales. Los cuales muestran el monto vendido, su objetivo y porcentaje de crecimiento para poder monitorear el crecimiento de sus ventas. Y para monitorear el cumplimiento de las metas establecidas respecto a sus ventas se cuenta con semaforización y un cálculo porcentual del cumplimiento de la meta respecto a la meta establecida. Se tuvo un incremento del 100% del total de reportes. Por lo tanto se cumplió con el objetivo propuesto: Incrementar el número de indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas. Lo cual era vital para saber cuándo introducir los ajustes o cambios necesarios y proponer acciones para el logro de sus objetivos (Otto Valle y Otto Rivera 2015).

5.4. Indicador 5:

Cantidad de reportes sobre predicción de ventas futuras.

Indicador	O1	O2	Diferencia
Cantidad de reportes sobre predicción de ventas futuras.	0	2	2

Tabla 21. Indicador 5

- **O1:** Es la cantidad de reportes sobre predicción de ventas futuras sin el uso del sistema de BI.
- **O2:** Es la cantidad de reportes sobre predicción de ventas futuras con el uso del sistema de BI.
- **Diferencia (O2 – O1):** Durante el análisis de la realidad problemática del grupo empresarial Leoncito se identificó que tenía 0 reportes sobre predicción de sus ventas. Ahora, con el sistema de BI que se construyó, el grupo empresarial puede tomar decisiones en base a la tendencia de sus ventas, pues se generaron 2 reportes el cual uno es para pronosticar el monto de sus ventas en los diferentes distritos donde tiene ventas registradas para los siguientes seis meses y el otro reportes es para pronosticar la cantidad de productos de marca Leoncito que se van a vender en los próximos seis meses. Se tuvo un incremento del 100% del total de reportes. Por lo tanto se cumplió con el objetivo propuesto: Incrementar el número de reportes sobre predicciones de ventas futuras. Contar con predicción es fundamental porque a través de esta información se puede respaldar la toma de decisiones en algo más que la intuición de los empresarios (Torres Barron 2011).

VI. CONCLUSIONES

- 6.1. Se redujo el tiempo promedio para que los miembros del grupo empresarial reciban la información solicitada respecto a ventas con el uso del sistema de BI logrando que tengan acceso rápido y puedan analizar información sobre sus ventas cada vez que lo requieran y así posteriormente tomar decisiones correctas y en el tiempo oportuno.
- 6.2. Se logró incrementar la cantidad de reportes que permitan análisis comparativo en el tiempo, permitiendo que los miembros del grupo empresarial puedan analizar el comportamiento y progreso de sus ventas logrando de esta manera tener conocimiento de sus alzas y bajas en sus ventas, conocer la existencia de patrones estacionales y en que periodos de tiempo se dan.
- 6.3. Se logró incrementar la cantidad de reportes sobre indicadores para monitorear el crecimiento y cumplimiento de las metas establecidas respecto a las ventas, logrando que los miembros del grupo empresarial puedan conocer mediante la semaforización el estado en las que estas se encuentran y así poder optar por estrategias para que se logre su cumplimiento.
- 6.4. Se logró incrementar la cantidad de reportes que permitan analizar sus ventas a través de las diferentes dimensiones, logrando que los miembros del grupo empresarial puedan analizar sus ventas desde diferentes perspectivas y a través de ello tener un conocimiento más amplio sobre los factores que son más influyentes, aprovecharlos para aumentar su crecimiento y plantear estrategias con respecto a los factores que no apoyan significativamente a las ventas.
- 6.5. Se logró incrementar la cantidad de reportes sobre predicciones de ventas futuras, logrando que los miembros del grupo empresarial tengan al alcance una aproximación del monto de sus ventas en los diferentes distritos, y también una aproximación de la cantidad de productos que se venderán en los próximos seis meses. Para así poder realizar presupuestos, comprar insumos, estimar los planes de ventas, etc. con fundamento basado en datos históricos de la empresa.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Aguilar, Emmanuel. «Modelado de datos multidimensional en el ámbito pre-hospitalario en cruz roja mexicana delegación La Paz Baja California.» Tesis Pregrado, Mexico, 2012.
- Arshad, Ali. *SQLSERVER*. 2016. <http://www.sql-server-performance.com/2009/ssis-introduction-part1/> (último acceso: 7 de Junio de 2016).
- Boada Vargas Machuca, Byron Alejandro, y Alvaro Arturo Tituaña Burgos. «Desarrollo de una aplicación de business intelligence para la empresa EMPAQLAST.» Tesis Pregrado, Ecuador, 2012.
- Bravo Aguinaga, Juan Arturo. «Solución de inteligencia de negocios utilizando algoritmos de serie temporal para apoyar la toma de decisiones en el proceso de ventas de una clínica de la ciudad de Chiclayo.» Tesis Pregrado, Chiclayo, 2016.
- Córdova Yupanqui, José Eduardo. «Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de importaciones en una empresa comercializadora/importadora.» Tesis Pregrado, Lima, 2013.
- Coronel, Nohely. «Diseño de un datamart para seguros masivos.» Tesis Pregrado, Lima, 2012.
- Curto, Josep, y Jordi Conesa. *Introducción a los Business Intelligence*. Barcelona: UOC, 2010.
- Eckerson, Wayne, y Colin White. *Evaluating ETL and Data Integration Platforms. The Data Warehouse Institute (TDWI)*. 2003.
- Espinoza Vargas, Jenny Marina, y Carlos Alberto Palomino Ruiz. «Desarrollo de un datamart para optimizar la generación de información estratégica de apoyo a la toma de decisiones en la vicepresidencia de banca comercial de Interbank Perú.» Tesis Pregrado, Lima, 2016.
- Gartner. *Gartner*. s.f. <https://www.gartner.com>.
- Guillen Rodríguez, Fiorelly Shirley. «Desarrollo de un data mart para mejorar la toma de decisiones en el área de tesorería de la municipalidad provincial de Cajamarca.» Tesis Pregrado, Cajamarca, 2012.
- Hernández Orallo, José, José Ramírez Quintana, y César Ferri Ramírez. *Introducción a la Minería de datos*. Madrid: Pearson Educación, 2004.
- Ibermatica. *Ibermatica*. 2006. <http://www.ibermatica.com/ibermatica/whitepapers> (último acceso: 15 de Mayo de 2016).
- Inmon, William. *Building the Data Warehouseagin toolkit*. London: Wiley, 2005.
- Kimball, Ralph. *The Data Warehouse Toolkit*. New York: Wiley, 2002.
- Lara Torralbo, Juan Alfonso. *Minería de datos*. Madrid, 2012.
- Lluís Cano, Josep. *Business Intelligence: Competir con información*. Barcelona: Dataprix, 2008.
- Microsoft. *Microsoft*. 2016. <https://msdn.microsoft.com/> (último acceso: 12 de 03 de 2016).
- Millones, Sammy. «Implementación de un sistema de comercio electrónico basado en CRM y Balanced ScoreCard como herramienta para la toma de decisiones en la empresa Inversiones Vialsa de la localidad de Chiclayo.» Tesis Pregrado, Chiclayo, 2012.

- Muinelo, Abraham. *iwsconsultores*. 09 de 08 de 2016. <http://www.iwsconsultores.com/el-business-intelligence-como-ventaja-competitiva/>.
- Otto Valle, y Otto Rivera. «Monitoreo e indicadores.» 2015.
- Rivadera, Gustavo. «La metodología de Kimball para el diseño de almacenes de datos (Data warehouses).» 2011: 7.
- Rodero, Héctor. *Desarrollo de una aplicación de business intelligence para la toma de decisiones con QlikViwe*. Madrid, 2010.
- Rodríguez Cabanillas, K, y A Mendoza Peña. «Análisis, Diseño e Implementación de una Solución de Inteligencia de Negocios para el Área de Compras y Ventas de una Empresa Comercializadora de Electrodomésticos.» Tesis Pregrado, Lima, 2011.
- Rodríguez, Jorge. «¿Qué es y para qué sirve a las empresas el Data Warehouse?» Blog, 2015.
- Rojas Zaldívar, Alejandro. «Implementación de un data mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones en el departamento de finanzas de la contraloría general de la república.» Tesis Pregrado, Chiclayo, 2014.
- Rosado, Alveiro. «Inteligencia de negocios: Estado del arte.» *Scientia et Technica (Sc)*, 2010: 3.
- Salazar, Leny. «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.RL.» Tesis Pregrado, Chiclayo, 2013.
- Sánchez Pérez, Angel Roberto. «Automatización del Sistema de Medición de Desempeño para la toma de decisiones estratégicas de negocio: caso de estudio.» Tesis Pregrado, Mexico, 2015.
- Santana Ormeño, Martín. *¿En qué invierten las empresas peruanas cuando se trata de TI?* 14 de 02 de 2012. <http://www.esan.edu.pe/conexion/actualidad/2012/02/14/en-que-invierten-las-empresas-peruanas-cuando-se-trata-de-ti/>.
- SAS. *SAS the power to know*. 2016. https://www.sas.com/en_my/insights/data-management/what-is-etl.html (último acceso: 20 de Junio de 2016).
- Sinnexus. *Sinnexus Business Intelligence Informática estratégica*. 2016. http://www.sinnexus.com/business_intelligence/ (último acceso: 11 de Agosto de 2016).
- Torres Barron, María de Jesús. «PRONÓSTICOS, UNA HERRAMIENTA CLAVE PARA LA PLANEACIÓN DE LAS EMPRESAS.» 2011.
- Tufiño, Jorge. «Desarrollo del datamart para el sistema nacional de vigilancia tecnológica del software libre.» Tesis Pregrado, 2011.
- Vizuete Naranjo, Michael Wladimir, y Carlos Patricio Yela Shinin. «Análisis, diseño e implementación de un datamart para el área de sismología del departamento de geofísica de la escuela plitécnica nacional.» Tesis Pregrado, Quito, 2006.

VIII. ANEXOS



Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y
Computación

ANEXO N° 1

ENTREVISTA

Nombre: Ruben Barrantes Becerra

Cargo: Gerente Administrativo

Fecha: 14/05/2016

- 1. ¿Qué tan satisfecho se siente usted con la información que obtiene del sistema actual respecto a ventas?**
 - A. Totalmente Satisfecho
 - B. Muy Satisfecho
 - C. Satisfecho
 - D. Poco Satisfecho
 - E. Nada Satisfecho

- 2. ¿Los reportes que obtiene del sistema actual le ayudan a tomar decisiones estratégicas sobre las ventas?**

No, solo nos ayuda a tomar de decisiones a corto plazo. Los reporte que nos brinda el sistema actual son los típicos reportes como listado detallado de ventas entre fechas determinadas, listado de ventas anuladas, listado de ventas por tipo (ventas al contado, ventas al crédito), etc. Pero no contamos con reportes que nos ayuden en el análisis de las ventas para la posterior toma de decisiones.

- 3. ¿Cuándo solicita reportes para tomar decisiones estratégicas, cuánto tiempo aproximadamente tarda en recibir los reportes solicitados?**

Dependiendo de la cantidad de reportes y del nivel detallado de información que se solicite. Esto puede demorar hasta 2 días debido a que las personas encargadas del área de tecnologías de información muchas veces se encuentran atendiendo diferentes problemas o inconvenientes que puedan ocurrir en el día a día y es por esto que primeramente se atienden dichos inconvenientes para luego realizar el reporte solicitado.

- 4. ¿Qué información le ayudaría a tomar mejores decisiones respecto a ventas?**
Información como el monto de ventas por local, línea, marca, mis clientes top (estos servirán para fidelizar a los clientes con mayor monto de compra), reportes de ventas mensuales y anuales, mis productos top. Además analizar las ventas visualizando información por tipo de venta (contado – crédito), tipo de clientes; desearía ver todo este análisis desde los diferentes puntos de vista mencionados en el tiempo para poder analizar su evolución en los diferentes meses y años.

- 5. Actualmente ¿Qué elementos de las ventas son evaluados en la información que obtiene respecto a ventas?**
Se evalúa la información de ventas por local en fechas determinadas, el monto total de ventas que realizan los diferentes locales por mes; sin embargo esto no es suficiente pues se debería de contar al menos con los productos top de los locales para poder implementar la estrategia más conveniente para cada uno de estos. También conocer las ventas de los locales por tipo de venta (ventas al contado - crédito) y las ventas de cada local en los diferentes meses y años.

- 6. ¿De qué manera monitorean el cumplimiento de sus metas respecto a ventas?**
Se utilizan hojas de cálculo o registros a papel para realizar el seguimiento de la meta de incrementar las ventas en 30% por local respecto al mismo periodo anterior ya sea por mes, trimestre, semestre o año; pero falta contar con indicadores para monitorear el cumplimiento de la meta establecida en tiempo real. Además no se monitorea adecuadamente el top de vendedores por mes actual evaluados por el número de productos vendidos respecto a la marca leoncito lo cual es importante pues la empresa realiza una especie de concurso en el cual el vendedor con mayor número de ventas respecto a los productos marca leoncito reciben un premio.

- 7. ¿De qué manera realizan el seguimiento del progreso de ventas?**
Solo se realiza un seguimiento a las metas mensuales; sin embargo falta seguir su evolución en el transcurso del tiempo para poder saber el comportamiento que estas y saber si existen patrones estacionales.

- 8. ¿Se analiza la información histórica de las ventas? ¿Cómo cree que ayuda el análisis de la información histórica?**
No se analiza la información histórica que contamos en nuestra base de datos y es por esto que no se tiene conocimiento sobre cómo se han evolucionado las ventas en el tiempo y por lo tanto tampoco sabemos por qué han venido teniendo ese comportamiento. Además al no analizar la información histórica no podemos predecir cómo serán las ventas a futuro lo cual no nos permite plantear estrategias de expansión en el mercado.

ANEXO N° 2

ENTREVISTA

Nombre: Juan Aldana

Cargo: Analista - programador de sistemas

Fecha: 14/05/2016

- 1. ¿Qué reportes brinda el sistema actual respecto a ventas?**
El sistema actual brinda reportes como listado de ventas entre fechas determinadas, monto de ventas mensuales, listadas de ventas anuladas, etc.
- 2. ¿Con cuánta frecuencia (en días) le solicitan reportes para la toma de decisiones estratégicas sobre ventas? ¿Cuáles son los reportes que le solicitan con mayor frecuencia respecto a ventas?**
Usualmente todos los fines de semana y sobre todo a fin de mes se ve el avance de las ventas. Los reportes más comunes que me solicitan son los reportes respecto al monto de venta por local, línea y marca.
- 3. ¿Aproximadamente cuánto tiempo promedio (en minutos) demora en generar los reportes que le solicitan?**
Depende mucho del tiempo el cual me tome analizar el reporte que me solicitan. A veces me ha tomado hasta dos días para el análisis y luego para generar el reporte solicitado aproximadamente entre una a dos horas.
- 4. ¿Considera que los reportes que Ud. genera para la gerencia son suficientes para la buena toma de decisiones estratégicas respecto a ventas?**
No es suficiente, a mi parecer los reportes que me solicitan apoyan pero son reportes con ayuda muy limitada, esto podría mejorar si se agregara información que apoye en el análisis de las ventas como reportes sobre el progreso de las ventas en el tiempo, productos top por línea y marca, clientes top por tipo de cliente, etc. También creo que ayudaría a la gerencia reportes con gráficos que les permita un mejor análisis de la información.

ANEXO N° 3

Resultados de la medición del tiempo de espera para acceder a la información por cada reporte respecto a ventas.

	TIEMPO 1	TIEMPO 2	TIEMPO 3	PROMEDIO
Semaforización por año	0'02''	0'03''	0'02''	0'023''
Semaforización por semestre	0'02''	0'03''	0'03''	0'027''
Semaforización por trimestre	0'02''	0'03''	0'03''	0'027''
Semaforización por mes	0'02''	0'03''	0'03''	0'027''
Reporte por líneas	0'03''	0'03''	0'03''	0'030''
Reporte por marcas	0'03''	0'03''	0'03''	0'030''
Top clientes por año	0'08''	0'07''	0'09''	0'080''
Top vendedores por año	0'03''	0'02''	0'03''	0'027''
Top vendedores por mes	0'02''	0'03''	0'03''	0'027''
Reporte de vendedor en año	0'03''	0'03''	0'03''	0'030''
Ventas por distrito	0'02''	0'03''	0'02''	0'023''
Ventas por local	0'03''	0'02''	0'02''	0'023''
Reporte de ventas por local y año – top productos	0'03''	0'03''	0'03''	0'030''
Predicción por monto	0'03''	0'02''	0'03''	0'027''
PROMEDIO				0'031''

Tabla 222. Medición del tiempo de espera para acceder a la información por cada reporte.

Tiempo 1: Carga inicial de la página del reporte.

Tiempo 2: Carga del reporte por segunda vez incluyendo selección de parámetros en aquellos reportes que tienen parámetros, caso contrario carga del reporte por segunda vez.

Tiempo 3: Carga del reporte por tercera vez incluyendo selección de parámetros en aquellos reportes que tienen parámetros, caso contrario carga del reporte por tercera vez.

ANEXO N° 4

Existe una gran gama de plataformas de Business Intelligence. Gartner lanza cada año un cuadrante clasificándolas en cuatro grupos. Estos cuatro grupos son: En la parte superior izquierda a los retadores, parte superior derecha a los líderes, parte inferior izquierda a los jugadores de nicho y parte inferior derecha a los visionarios.



Figura 65. Cuadrante mágico de Gartner en el 2014



Figura 66. Cuadrante mágico de Gartner en el 2015



Figura 67. Cuadrante mágico de Gartner en el 2016

ANEXO N° 5

ENTREVISTA

Nombre: Ruben Barrantes Becerra

Cargo: Gerente Administrativo

Fecha: 27/05/2017

1. ¿Qué tan satisfecho se siente usted con la información que obtiene del sistema de BI respecto a ventas?

A. Totalmente Satisfecho

B. Muy Satisfecho

C. Satisfecho

D. Poco Satisfecho

E. Nada Satisfecho

2. ¿Los reportes que obtiene del sistema de BI le ayudan a tomar decisiones estratégicas sobre las ventas?

Por supuesto, cada uno de los reportes nos ayuda a tomar decisiones a mediano y largo plazo. Los reportes que nos brinda el sistema de BI nos ayudan en el análisis de las ventas desde diferentes perspectivas, y esto nos sirve para tomar mejores decisiones estratégicas para la empresa.

3. ¿Cuánto tiempo aproximadamente tarda en recibir la información solicitada al sistema de BI?

Entre 3 a 5 segundos tarda el sistema en mostrar los reportes. Carga más rápido en comparación con el reporte que se tiene del sistema transaccional. El cual podía tener una demora de hasta 13 minutos. Lo cual para mi gusto es demasiado tiempo de espera.

4. ¿De qué manera pueden monitorear el cumplimiento de sus metas respecto a ventas con el sistema de BI?

A través de los 4 reportes elaborados con semaforización se puede monitorear el cumplimiento de las metas establecidas. Además podemos monitorear los top de productos, top de clientes y el top de vendedores.

5. ¿Le ayuda el análisis de la información histórica?

Nos ayuda a conocer cómo han evolucionado las ventas en el tiempo y por lo tanto podemos saber el comportamiento que estas han tenido a lo largo de los años. Y gracias a las predicciones realizadas permite plantear estrategias respaldadas en información ya existente.