

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



**SISTEMA BI CON PREDICCIÓN DE VENTAS BASADO EN EL
ALGORITMO DE SERIES TEMPORALES PARA APOYAR LA
GESTIÓN EN LA EMPRESA MOTOFUERZA S.A.C.**

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

AUTOR

LUIS FERNANDO JUNIORS SANTA CRUZ MONTAÑO

ASESOR

SEGUNDO JOSÉ CASTILLO ZUMARAN

<https://orcid.org/0000-0001-5613-5519>

Chiclayo, 2021

**SISTEMA BI CON PREDICCIÓN DE VENTAS BASADO
EN EL ALGORITMO DE SERIES TEMPORALES PARA
APOYAR LA GESTIÓN EN LA EMPRESA
MOTOFUERZA S.A.C.**

**PRESENTADA POR:
LUIS FERNANDO JUNIORS SANTA CRUZ MONTAÑO**

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:

Juan Antonio Torres Benavides
PRESIDENTE

Marlon Eugenio Vílchez Rivas
SECRETARIO

Segundo José Castillo Zumaran
VOCAL

DEDICATORIA

Dedicar en primer lugar mi tesis a Dios por darme la salud y el conocimiento necesario para poder seguir siempre adelante, a mis padres por darme la fuerza y el apoyo para continuar hasta el final y a mi asesor ya que con su experiencia y seguimiento pude lograr el objetivo.

AGRADECIMIENTOS

A Dios por haberme dado el valor y el esfuerzo necesario para continuar, y por haberme permitido que a pesar de todos los tropiezos llegue al final de la carrera

A mi familia por haberme brindado su apoyo incondicional para iniciar y terminar mi carrera universitaria

A mi asesor de tesis Ing. Segundo José Castillo Zumaran porque sin su ayuda, exigencia y conocimiento no estaría hoy donde me encuentro

A la empresa por haberme brindado la confianza e información suficiente para poder llevar al cabo esta investigación.

ÍNDICE

I.	INTRODUCCIÓN	14
II.	MARCO TEÓRICO	18
2.1.	ANTECEDENTES	18
2.1.1.	ANTECEDENTES INTERNACIONALES	18
2.1.2.	ANTECEDENTES NACIONALES.....	19
2.1.3.	ANTECEDENTES LOCALES.....	21
2.2.	BASES TEÓRICO CIENTÍFICAS	23
2.2.1.	TOMA DE DECISIONES.....	23
2.2.2.	SISTEMAS DE INFORMACIÓN	23
2.2.3.	INVENTARIOS	24
2.2.3.1.	TIPOS.....	24
2.2.4.	SISTEMA DE CONTROL DE INVENTARIOS.....	25
2.2.5.	DEMANDA	25
2.2.5.1.	DEMANDA DEPENDIENTE	26
2.2.5.2.	DEMANDA INDEPENDIENTE.....	26
2.2.5.2.1	DEMANDA DETERMINISTICA	26
2.2.5.2.2	DEMANDA PROBABILISTICA.....	27
2.2.6.	MINERÍA DE DATOS	27
2.2.7.	DATA WHEREHOUSE.....	29
2.2.8.	DATA MART	31
2.2.9.	PROCESO ETL	32
2.2.10.	ALGORITMO DE SERIE TEMPORAL.....	33
2.2.11.	SQL SERVER	35

III. METODOLOGÍA	36
3.1. TIPO Y NIVEL DE INVESTIGACIÓN	36
3.1.1. TIPO DE INVESTIGACIÓN	36
3.1.2. NIVEL DE INVESTIGACIÓN	36
3.2. DISEÑO DE INVESTIGACIÓN	36
3.3. POBLACIÓN, MUESTRA Y MUESTREO	36
3.3.1. POBLACIÓN	36
3.3.2. MUESTRA	37
3.3.3. MUESTREO	37
3.4. CRITERIOS DE SELECCIÓN	37
3.5. OPERACIONALIZACIÓN DE VARIABLES	37
3.5.1. VARIABLES	37
3.5.1.1. Variable independiente	37
3.5.1.2. Variable dependiente	37
3.5.2. INDICADORES (OPERACIONALIZACIÓN DE VARIABLES)	38
3.6. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	39
3.7. PROCEDIMIENTOS	39
3.7.1. METODOLOGÍA DE DESARROLLO	39
3.7.2. ANÁLISIS DE RIESGOS	39
3.7.3. PRODUCTO ACREDITABLE	40
3.7.4. MANUAL DE USUARIO	40
3.8. PLAN DE PROCESAMIENTO Y ANÁLISIS DE DATOS.....	40
3.9. MATRIZ DE CONSISTENCIA	41
3.10. CONSIDERACIONES ÉTICAS.....	42
IV. RESULTADOS	43
4.1. EN BASE A LA METODOLOGÍA UTILIZADA	43
4.1.1. ITERACIÓN #1: PLANIFICACIÓN DEL PROYECTO (PLAN DE SISTEMAS)	43
4.1.2. ITERACIÓN #2: DEFINICIÓN DE LOS REQUERIMIENTOS DE MINERÍA DE DATOS Y BI.....	44
4.1.3. ITERACIÓN #3: COMPRESION DE DATOS	45
4.1.4. ITERACIÓN #4: PREPARACIÓN DE LOS DATOS DE MINERÍA DE DATOS.....	50
4.1.5. ITERACIÓN #5: MODELADO DIMENSIONAL	56

4.1.6.	ITERACIÓN #6: MODELADO DE MINERÍA DE DATOS	60
4.1.7.	ITERACIÓN #7: DISEÑO FÍSICO	61
4.1.8.	ITERACIÓN #8: ETL (EXTRAER, TRANSFORMAR Y CARGAR)	63
4.1.9.	ITERACIÓN #9: DISEÑO DE LA ARQUITECTURA TECNICA.....	75
4.1.10.	ITERACIÓN #10: IMPLEMENTACIÓN	77
4.1.11.	ITERACIÓN #11: EVALUACIÓN	89
4.2.	EN BASE A LOS OBJETIVOS DE LA INVESTIGACIÓN	91
4.2.1.	DESCRIPCIÓN DE OBJETIVO 1	91
4.2.2.	DESCRIPCIÓN DE OBJETIVO 2	92
4.2.3.	DESCRIPCIÓN DE OBJETIVO 3	93
4.2.4.	DESCRIPCIÓN DE OBJETIVO 4	94
4.3.	IMPACTOS ESPERADOS	94
4.3.1.	IMPACTOS ECONÓMICOS.....	94
4.3.2.	IMPACTOS SOCIALES	94
4.3.3.	IMPACTOS EN TECNOLOGÍA	94
4.3.4.	IMPACTOS AMBIENTALES.....	94
4.3.5.	IMPACTOS EN LA FORMACIÓN DE CADENAS PRODUCTIVAS	94
V.	DISCUSIÓN	95
VI.	CONCLUSIONES.....	98
VII.	RECOMENDACIONES.....	99
VIII.	LISTA DE REFERENCIAS	100
IX.	ANEXOS.....	101
	ANEXO N° 01. CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS	102
	ANEXO N° 02. ANÁLISIS DE RIESGOS	103
	ANEXO N° 03. INSTRUMENTOS DE RECOLECCIÓN DE DATOS	113
	ANEXO N° 04. INSTRUMENTOS DE RECOLECCIÓN DE DATOS	115

LISTA DE TABLAS

TABLA I. DISEÑO DE LA INVESTIGACIÓN	36
TABLA II. METODOS, TÉCNICAS E INSTRUMENTOS	37
TABLA III. INDICADORES	38
TABLA IV. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	39
TABLA V. MATRIZ DE CONSISTENCIA	41
TABLA VI. RESULTADO DE VALORES CON ERROR DE MEDICIÓN, VALORES PERDIDOS, Y VALORES NULOS.....	49
TABLA VII. VENTAS DE MOTOCICLETAS POR MESES EN EL AÑO 2014.....	50
TABLA VIII. VENTAS DE MOTOCICLETAS POR MESES EN EL AÑO 2015	51
TABLA IX. VENTAS DE MOTOCICLETAS POR MESES EN EL AÑO 2016.....	51
TABLA X. VENTAS DE MOTOCICLETAS POR MESES EN EL AÑO 2017	52
TABLA XI. VENTAS DE MOTOCICLETAS DESDE ENERO HASTA MAYO EN EL 2018.....	52
TABLA XII. MODELO DE ARTÍCULO Y CANTIDAD VENDIDA.....	53
TABLA XIII. MATRIZ BUS	56
TABLA XIV. DIMENSIONES	57
TABLA XV. DIMENSIÓN ARTICULO	58
TABLA XVI. DIMENSIÓN TIEMPO	58
TABLA XVII. DIMENSIÓN CLIENTE	59
TABLA XVIII. DIMENSIÓN VENDEDOR	59
TABLA XIX. DIMENSIÓN FORMA PAGO	59
TABLA XX. LISTA DE HECHOS	60
TABLA XXI. MODELADO FÍSICO DE LA TABLA DIMENSIÓN CLIENTE.....	61
TABLA XXII. MODELADO FÍSICO DE LA TABLA DIMENSIÓN ARTICULO.....	61
TABLA XXIII. MODELADO FÍSICO DE LA TABLA DIMENSIÓN FORMA DE PAGO	61
TABLA XXIV. MODELADO FÍSICO DE LA TABLA DIMENSIÓN TIEMPO.....	62
TABLA XXV. MODELADO FÍSICO DE LA TABLA DIMENSIÓN TIPO DE DOCUMENTO.....	62
TABLA XXVI. MODELADO FÍSICO DE LA TABLA DIMENSIÓN VENDEDOR..	62
TABLA XXVII. HECHO VENTA.....	63
TABLA XXVIII. INDICADOR 1 – INDICADOR 2.....	95
TABLA XXIX. INDICADOR 3	97
TABLA XXX. INDICADOR 4	97
TABLA XXXI. INTERESADOS INTERNOS	103

TABLA XXXII. INTERESADOS EXTERNOS	103
TABLA XXXIII. MATRIZ DE RIESGOS DE PLANIFICACIÓN DEL PROYECTO	105
TABLA XXXIV. MATRIZ SALVAGUARDA DE RIESGOS DE LA PLANIFICACIÓN DE PROYECTO	106
TABLA XXXV. MATRIZ DE RIESGOS DE LA ETAPA DEFINICIÓN DE LOS REQUERIMIENTOS DE MINERÍA DE DATOS Y BI	107
TABLA XXXVI. MATRIZ DE SALVAGUARDA DE RIESGOS DE LA ETAPA DEFINICIÓN DE REQUERIMIENTOS DE MINERIA DE DATOS Y BI	108
TABLA XXXVII. MATRIZ DE RIESGOS DE LA ETAPA DE MODELADO DIMENSIONAL	109
TABLA XXXVIII. MATRIZ DE SALVAGUARDA DE RIESGOS DE LA ETAPA MODELADO DIMENSIONAL.....	110
TABLA XXXIX. MATRIZ DE RIESGO PARA LA ETAPA DE ARQUITECTURA TÉCNICA.....	111
TABLA XL. MATRIZ DE SALVAGUARDA DE RIESGOS PARA LA ETAPA DE ARQUITECTURA TÉCNICA.....	112

LISTA DE FIGURAS

FIG. 1. TIPOS DE INVENTARIOS MAS CONOCIDOS SEGÚN SU NATURALEZA	24
FIG. 2. MODELO DE PREVISION EN EL TIEMPO DE LAS VENTAS DE UN PRODUCTO.....	34
FIG. 3. VENTAS POR ARTÍCULO DETALLE MOTOS	46
FIG. 4. PRIMERA LIMPIEZA DE LA TABLA VENTAS	46
FIG. 5. PRIMERA LIMPIEZA DE LA TABLA ARTÍCULO	47
FIG. 6. COMPRAS POR ARTÍCULO DETALLE	47
FIG. 7. MODELADO DIMENSIONAL	55
FIG. 8. START MART VENTAS	56
FIG. 9. MOLDEAMIENTO DE LOS DATOS PARA ADECUAR AL MODELO DE PREDICCIÓN.....	60
FIG. 10. CONEXIÓN AL ORIGEN Y DESTINO	64
FIG. 11. CONEXIÓN AL ORIGEN Y DESTINO	64
FIG. 12. FLUJO DE DATOS VENDEDOR	65
FIG. 13. FLUJO DE DATOS ARTÍCULO	66
FIG. 14. FLUJO DE DATOS TIPO DOCUMENTO	67
FIG. 15. FLUJO DE DATOS FORMA DE PAGO.....	68
FIG. 16. FLUJO DE DATOS CLIENTE	69
FIG. 17. CONSULTA PARA EL POBLAMIENTO DE LA DIMENSIÓN TIEMPO.....	69
FIG. 18.FLUJO DE DATOS HECHO VENTAS	70
FIG. 19. VISTA DEL ORIGEN DE DATOS VISTA_DM VENTA MOTOS	71
FIG. 20. CREACION DEL CUBO	72
FIG. 21. ELECCION DE DIMENSIONES Y UNIDADES DE MEDIDA PARA EL CUBO	73
FIG. 22. CREACION DEL CUBO CUBO_DM VENTAS MOTOS	74
FIG. 23. ARQUITECTURA DEL PROYECTO	75
FIG. 24. CONSULTA PARA EL PRONOSTICO.....	77
FIG. 25. CREACION DE LA TABLA TEMPORAL TP_ARTÍCULOS	78
FIG. 26. CREACION DEL CURSOR DENOMINADO ARTÍCULOS	79
FIG. 27. CREACION DEL SEGUNDO CURSOR DENOMINADO ARTÍCULOS2	79
FIG. 28. CREACION DE LA VISTA DENOMINADA V_TP_ARTÍCULOS.....	80
FIG. 29. CREACION DE LA VISTA DENOMINADA VISTAPREDICCIÓN POR MEDIO DEL ORIGEN DE DATOS DM_VENTAS MOTOS.DS	80
FIG. 30. SELECCIÓN DE LA TÉCNICA DE MINERÍA DE DATOS	81
FIG. 31. SELECCIÓN DE LA VISTAPREDICCIÓN	81

FIG. 32. SELECCIÓN DE LA TABLA MOLDEADA PARA EL ALGORITMO DE SERIES TEMPORALES	82
FIG. 33 SELECCIÓN DE DATOS DE APRENDIZAJE	82
FIG. 34. ESPECIFICACION DE TIPO DE DATO Y CONTENIDO	83
FIG. 35. FINALIZACION DE LA ESTRUCTURA DE MINERÍA DE DATOS	83
FIG. 36. ESTRUCTURA Y MODELADO DE MINERÍA DE DATOS	84
FIG. 37. PARAMETROS PARA EL ALGORITMO CON EL METODO MIXED	84
FIG. 38. VISOR DEL MODELO DE MINERÍA DE DATOS MIXED	85
FIG. 39. PARAMETROS DEL ALGORITMO CON EL METODO ARTXP	86
FIG. 40. VISOR DE MODELO DE MINERÍA DE DATOS ARTXP	86
FIG. 41. PARAMETROS DEL ALGORITMO CON EL METODO ARIMA	87
FIG. 42. VISOR DEL MODELO DE MINERÍA DE DATOS ARIMA	87
FIG. 43. CREACION DE MODELOS DE MINERÍA DE DATOS BASADO EN EL ALGORITMO DE SERIES TEMPORALES	88
FIG. 44. CONSULTA DMX PARA CREAR LOS DIFERENTES TIPOS DE PREDICCIONES DE UN ARTÍCULO ESPECÍFICO.....	88
FIG. 45. EVALUACIÓN DEL ALGORITMO ARIMA CON EL METODO MAPE	89
FIG. 46. EVALUACIÓN DEL ALGORITMO MIXED CON EL METODO MAPE	90
FIG. 47. EVALUACIÓN DEL ALGORITMO ARTXP CON EL METODO MAPE	90
FIG. 48. PREDICCIÓN DE UNIDADES VENDIDAS EN EL 2018 POR MESES.....	91
FIG. 49. REPORTE TOP CLIENTE POR AÑO.....	92
FIG. 50. REPORTE TOP UNIDADES MAS VENDIDAS POR AÑO	93

RESUMEN

En el presente trabajo de investigación tuvo como objetivo apoyar el proceso de gestión en la empresa “MotoFuerza SAC”. Se utilizó una metodología híbrida mezclando Ralph Kimball con Crisp DM una para el proceso de BI y consecutivamente para desarrollar minería de datos, también se usaron las herramientas Microsoft

Dentro de los problemas identificados se encontró que no tienen una coherente administración con respecto a sus ventas en lapsos de tiempo, así tener estimación a una predicción de ventas, al igual que carecen de análisis en base a información histórica para la toma de decisiones estratégicas por parte de la gerencia. Esto también nos lleva a poder evaluar modelos de predicción de ventas con técnicas de error probabilístico y porcentual

La implementación del sistema BI basado en el algoritmo de series temporales ha mejorado significativamente el proceso de toma de decisiones estratégicas, la fiabilidad y la madurez en la empresa “MotoFuerza SAC”.

Palabras clave: BI, Series temporales, predicción de ventas, toma decisiones.

ABSTRACT

In this research work, the objective was to support the management process in the company "MotoFuerza SAC". A hybrid methodology was developed by mixing Ralph Kimball with Crisp DM one for the BI process and consecutively to develop data mining, Microsoft tools can also be used. Among the problems identified are that they do not have a coherent administration with respect to their sales in time periods, thus having a prediction of sales, as well as the analysis falls based on historical information for strategic decision making by of management.

This also leads us to evaluate sales prediction models with probabilistic and percentage error techniques.

The implementation of the BI system based on the time series algorithm has significantly improved the strategic decision-making process, reliability and maturity in the company "MotoFuerza SAC".

Keywords: BI, Time series, sales prediction, decision making.

I. INTRODUCCIÓN

Hoy en día, las organizaciones de distintos rubros, buscan que su información sea eficiente y confiable, de tal manera, generan estrategias de software que no solo brinden apoyo físico a las existencias, sino también de acuerdo al tipo de inventario que se está utilizando, pues la gestión dentro de la empresa, es uno de los elementos logísticos con mayor complejidad dentro de áreas como producción y distribución de bienes. Así mismo, hay estrategias para poder tomar decisiones, que implican decisiones sobre cuando realizar alguna actividad, cuanto tiempo llevará organizar, balancear inventarios, así como el establecimiento de tiempos de entregas. Gutiérrez [3]

Es por ello, que la deficiencia de no contar con un sistema de inventario que se amolde a los requerimientos del negocio, podría causar pérdidas significativas en éste. Haciendo mención a un ejemplo de una empresa reconocida como Nike, se puede mencionar que esta organización estadounidense, pasó por pérdidas millonarias de dólares, en el año 2000, a causa de la deficiente planeación de la demanda en el software que empleaban.

El mundo de TI ve a Nike como la empresa que echó a perder su cadena de suministro, especialmente el de planeación de demanda que arrojó miles de más de productos que los solicitados en el mercado. Koch [6]

Ante esto, se puede mencionar que el autor Valle, especialista en tecnologías de información, considera que éstas cumplen un propósito, en el manejo y tratamiento de la información, entendiéndose como el conjunto de datos, señales o ilustraciones, que están registrados o transportados en base a soportes físicos, cuya tipología es diversa. Estas tecnologías incluyen estrategias, y metodologías que facilitan la obtención, transmisión, transformación y combinación de los datos, signos o conocimientos que necesita la organización.

Por otro lado, en el rubro de las motocicletas, la empresa Honda se ha posicionado como líder en todo el Perú, reconociéndose como la única marca que cuenta con una planta de fabricación de sus productos, obteniendo una producción de 53.000 unidades anuales, en cuestión de motos y trimotos, siendo una de las empresas que cuenta con certificaciones y estándares de calidad, emisiones que requieren en todo el mundo. El registro de trimotos lo lidera Honda con 9487 unidades vendidas en el 2014, entre motos lineales y trimotos, la marca Honda en el Perú vendió 35,444 entre los meses de Enero a Septiembre del 2014.

Hoy en día, en cuanto al registro de ventas de motos lineales en estos meses que van del año lo lidera la empresa Honda, con 25,957 unidades.

Cabe resaltar, que Perú es uno de los países con mayor crecimiento, se considera que el mercado continuara avanzando, pues existe una adecuada perspectiva en cuanto a la economía del país, así como la baja inserción, tanto del automóvil como de la motocicleta, a diferencia de otros países latinoamericanos. Mizogu [3]

Así mismo, existe una moto por cada 70 personas, por lo que hay un gran potencial de crecimiento en este mercado [El comercio, 2016].

De acuerdo al reporte de la Asociación Automotriz del Perú (AAP) el monto de ventas de motos y trimotos a nivel nacional estima los 187,013 entre los meses de enero y agosto del año 2014.

Acercándonos a nuestra región, observamos que ciudades como Chiclayo, Piura, Tarapoto y Pucallpa existen más comercialización de motos.

En nuestra región, durante el periodo 2014, las ventas de motos honda ha logrado la cifra de 2675 ventas. Por otro lado, de esta cantidad la empresa MotoFuerza S.A.C. obtuvo una venta de 1113 unidades, representándose en el 41.6 % del total de las ventas en la ciudad.

Por esta razón, observamos la problemática existente en la empresa MotoFuerza S.A.C. siendo una concesionaria dedicada a la venta de motocicletas cuyo proveedor principal es “HONDA”; su nombre comercial es MotoFuerza S.A.C., esta es un tipo de empresa cerrada, de condición activa, dedicada a la comercialización, venta, mantenimiento y reparación de motocicletas.

MotoFuerza S.A.C. cuenta con una cartera de clientes debido a su marca reconocida y a la garantía de sus productos, liderando el mercado a nivel regional, dado que en el año 2015 vendió 848 motos alcanzando una participación del 58% del mercado.

- No obstante, se ha observado que la empresa, no contaba con un sistema especializado de información, a fin de poder gestionar los procesos de inventario ocasionando problemas como no tener información a tiempo de inventario actualizado y relativo a sus ventas.
- Con respecto a la predicción de venta, Esta situación ocasionaba muchas veces deserción de clientes, pérdidas de dinero por ventas, tiempo adicional para consultar inventario por el gerente de ventas, estancamiento de productos, e incomodidad por parte de los clientes; generando no solo deficiencias en cuanto a software a corto plazo, sino también a largo plazo, de no haber realizado un diseño informático o software que permita mejorar la condición que presentó la organización.
- El gerente de ventas no se siente satisfecho con la información que el utiliza para las diferentes funciones que se realizan en el proceso de ventas lo cual es muy complicado tomar decisiones y estrategias concretas ya que al no tener un sistema que desarrolle análisis con la data histórica no se puede tener eficiencia en sus distintos modelos de proyección de ventas de este mismo.

- El gerente de ventas no tiene entendimiento sobre la evolución de las ventas al transcurrir del tiempo, lo que necesita es información sobre como proyectar estas mismas a futuro de tal manera que se establezcan estrategias de mejora en la empresa, por consiguiente carecen de reportes que ayuden a monitorear. No cuentan con las herramientas adecuadas para poder evaluar datos e información histórica lo cual es muy importante frente a ocasiones futuras y aumentar la certeza en la toma de decisiones

Finalmente, la investigación científica, es relevante a nivel científico, pues servirá como aporte científico a diferentes estudios vinculados con este tipo de actividad; así mismo se justifica a nivel financiero, ya que el sistema se diseñó para registrar los artículos que se ingresen en almacén, ingresar la demanda conocida, el costo de mantenimiento de inventario y costo por ordenar un pedido de sus productos, generando reportes de los costos por cada producto, garantizando el stock a tener en almacén para cada producto (motocicletas y motokar), donde se garantizó la visualización del estado de los productos, contribuyendo a mejorar la gestión de inventarios, así como la imagen de la empresa ante sus clientes, en cuanto a innovación y competencia local, por lo que se estaría hablando también de una justificación social y tecnológica.

Bajo los argumentos referidos, el presente estudio científico, pretende beneficiar a la empresa, representado por la reducción de costos por mantenimiento de inventario, la reducción de pérdida de clientes, ventas perdidas, el aprovechamiento máximo de recursos para gestionar sus ventas y finalmente, el reconocimiento de la mejora que reflejará la empresa al aplicar estos modelos.

II. MARCO TEÓRICO

2.1. Antecedentes

Se han considerado para esta investigación los siguientes antecedentes:

2.1.1. Antecedentes internacionales

Vargas [5] en su investigación denominada “Diseño e implementación de una aplicación web de control de inventario y rastreo de puntos de ventas en la empresa Pos Managed Services C.A.” Universidad Simón Bolívar. Sartenejas, Colombia. (Tesis de pre grado), tuvo como finalidad la creación de un modelo de inventario de rastreo de puntos de venta para su sistema PMis (Pos Managed information System), el cual permitió a la organización llevar un registro de los equipos que ingresan y salen diariamente de sus almacenes, registrando a su vez datos históricos de estos mismos. Finalmente, se diseñó e implementó la solución de un sistema web, utilizando las tecnologías establecidas por la organización. Se tomó en consideración esta tesis, pues se observa que se utilizó un sistema web a fin de apoyar el control de la gestión de inventarios para la organización PMS, guardando relación con la finalidad que como investigador propongo en mi estudio, ya que también pretendo apoyar el control de la gestión de inventarios.

Muñoz [6], en su investigación de título “Implementación de una aplicación web para el control de inventario y facilitación de material de trabajo para la empresa Maderas BSC Ltda.”. Universidad del Bio – Bio. Chillan, Chile. (Tesis de pre grado), tuvo como finalidad implementar y desarrollar una aplicación web para el área de la bodega de la organización. La metodología que empleo el estudio fue interactiva e incremental, por las condiciones en las que se realizó el proyecto, trabajando con el lenguaje web PHP, con el enfoque orientado a objetos, aplicando la arquitectura de tres capas MVC, y el modelo UML. Se concluyó que la implementación de la aplicación web, cumplió con los requerimientos planteados por el cliente, satisfaciendo las expectativas y necesidades de la empresa. Finalmente se tomó esta investigación, pues el objetivo fue controlar y ajustar los gastos en adquisición de existencias,

como la prevención de irregularidades en los ingresos y egresos de insumos.

Lucas [7], en su investigación denominada “Desarrollo e implementación de aplicación web para el control de inventario del local comercial Maquinas Hidalgo”. Universidad Politécnica Salesiana Sede Guayaquil. Guayaquil, Ecuador. (Tesis de pre grado), tuvo como objetivo desarrollar una aplicación web que permita ejecutar la gestión de las existencias de mercadería, a fin de optimizar la toma de decisiones. El enfoque que se empleo fue cuantitativo, de diseño experimental, de nivel aplicativo. Este estudio logró resolver los inconvenientes que padecían los usuarios, debido al manejo manual del inventario de mercadería; así mismo esta investigación sirve de respaldo al presente estudio que se ha realizado, pues se desarrolla un sistema web para el apoyo del control de inventarios, permitiendo optimizar el proceso de toma de decisiones dentro de la organización.

2.1.2. Antecedentes nacionales

Rodríguez [8], en su tesis denominada “Análisis, diseño e implementación de un sistema de formación para una tienda de ropa con enfoque al segmento juvenil”. Pontificia Universidad Católica del Perú. Lima, Perú. (Tesis de pre grado), tuvo como objetivo realizar el análisis, diseño e implementación de un sistema de información que permita la gestión de ventas e inventario de ventas para la tienda del sector juvenil. El enfoque que se utilizó fue cuantitativo, de diseño experimental, nivel aplicativo. Para el desarrollo de la investigación se emplearon las buenas prácticas para la gestión de proyectos por el PMI. Se concluyó que se logró el desarrollo de un sistema de información que permitió gestionar ventas y el almacén de ventas, de esta manera se alcanzó la organización, el control y administración de los productos de la organización, mejorando la interacción entre clientes. Esta investigación sirve de respaldo para el estudio presente, debido a que la finalidad también se basa en el apoyo de la gestión de inventarios de la empresa.

Rodríguez [9], en su tesis denominada “Implementación de un sistema informático web para el control de ventas e inventario en la empresa Calzados Winner E.I.R.L.” Universidad Católica Los Ángeles de Chimbote. Trujillo, Perú. (Tesis de pre grado). Tuvo como objetivo implementar un sistema informático web para el control de ventas e inventario en la empresa antes mencionada. El enfoque que se empleó fue cuantitativo, de diseño no experimental, de nivel descriptivo y documental. Respecto a la muestra, se contó con la participación de 20 trabajadores, con quienes se utilizó tanto la técnica de la observación directa, como la aplicación de un cuestionario a fin de medir las variables. Se concluyó que el 95.00% de los trabajadores encuestados si muestran necesidad hacia la implementación y manejo de un sistema informático web para los proceso de ventas e inventario; en cambio solo el 20.00% y 5.00% indicó que no perciben la necesidad de implementación y manejo de un sistema informático web. Finalmente, esta investigación guarda relación con el presente estudio, debido a la finalidad y propósito que se pretende abordar.

Izquierdo [10], en su investigación denominada “Sistema web para el control de inventario en la empresa MC Air Servis S.A.C.”. Universidad Cesar Vallejo. Lima, Perú. (Tesis de pre grado). Se tuvo como propósito determinar de qué manera influye un sistema web para el control de inventario en la empresa MC Air Servis S.A.C. En la investigación, se utilizó el enfoque cuantitativo, de diseño pre experimental y de nivel aplicativo. La muestra estuvo compuesta por 26 fichas de registro de los pedidos, donde se aplicó la técnica de fichaje y el instrumento de ficha de registro, los cuales obtuvieron validación de expertos. Se concluyó que la implementación del sistema web, facilito la disminución de la rotura de stock de productos terminados a un 37.50%, incrementándose la materia prima. Se tomó esta investigación se relaciona con el estudio presente, debido a que se pretende también facilitar el proceso de control de inventarios para disminuir costos.

2.1.3. Antecedentes locales

Vidaurre [11], en su tesis sobre “Aplicación de las redes neuronales artificiales para el pronóstico de la demanda del agua potable en la empresa Epsel S.A de la ciudad de Lambayeque”. Universidad Santo Toribio de Mogrovejo”. Chiclayo, Perú. (Tesis de pre grado), tuvo como objetivo demostrar que el empleo de técnicas computacionales basadas en inteligencia artificial, reducen los niveles de error en las predicciones de demanda de agua potable. El enfoque de la investigación fue cuantitativa, de diseño experimental, de nivel descriptivo, donde la población se trabajó a través de las mediciones de agua (Enero 1990 Diciembre 2010). La técnica que se utilizó fueron las técnicas neuronales. Así mismo, la investigación logró predecir la cantidad de demandada de agua potable a mediano plazo, mediante la aplicación de técnicas de inteligencia artificial. Se concluyó finalmente, que las redes neuronales poseen un valor como método eficaz. Se tomó en cuenta esta investigación, pues se relaciona con la investigación que se realizó en cuanto a la predicción de ventas de la organización (demanda de la compra de motos Honda).

Sánchez [12], en su tesis sobre “Implementación de un sistema informático para mejorar la gestión y control de inventarios de bienes muebles, aplicando tecnología RFID en Gobierno Regional Lambayeque”. Universidad Santo Toribio de Mogrovejo. Chiclayo, Perú. (Tesis de pre grado). Se tuvo como finalidad diseñar un sistema informático que permitió mejorar la gestión y control de bienes muebles utilizando tecnología RFID. En la investigación se utilizó el enfoque cuantitativo, de diseño cuasi experimental, de nivel aplicativo. La muestra estuvo compuesta por 375 bienes; para la ejecución de la investigación se utilizó como técnica la entrevista y observación, y como instrumento la guía de entrevista semi estructurada y la ficha de observación. La investigación logró mejorar la gestión y control de bienes muebles en la institución donde fue realizada, concluyéndose que se logró reducir un 95% del tiempo destinado a la realización de la toma de inventario, con lo cual pasó

a reducirse de 60 días a tan solo 3 días. Se tomó en cuenta este estudio, debido a su relación con los objetivos de la investigación presente.

Santos [13], en su tesis de título “Aplicación móvil y web responsiva utilizando el algoritmo de planificación basado en prioridades para apoyar en la atención de solicitudes del servicio delivery en la Lavandería Quin”. Universidad Santo Toribio de Mogrovejo. Chiclayo, Perú. (Tesis de pre grado), tuvo como finalidad apoyar la atención de solicitudes del servicio delivery, a través de la implementación de una aplicación móvil y web responsiva, empleando el algoritmo de planificación basado en prioridades. El enfoque de la investigación fue cuantitativo, de diseño cuasi experimental, de nivel aplicativo. La población con la que se trabajó fueron 6 trabajadores, quienes conforman toda la organización, tratándose de una empresa pequeña. En cuanto a los instrumentos que se emplearon, se utilizó la entrevista y la encuesta. Así mismo, con la investigación se logró disminuir el tiempo promedio que tomaba ordenar las solicitudes de servicios en la lavandería Quin. Se concluye que se logró reducir el tiempo que se invertía, en 9 minutos 55 segundos, pues la velocidad con la que el sistema procesa los datos es más rápida y eficaz que la de una persona. Esta investigación se relaciona con el presente estudio, pues se trata de buscar la reducción del tiempo a través de un sistema web, objetivo que también forma parte de este estudio.

2.2. Bases teórico científicas

2.2.1. Toma de decisiones:

La toma de decisiones, se contempla como el origen dentro del proceso administrativo, pues las actividades direccionadas hacia la producción de bienes o prestaciones de servicios deben tener un orden y planificación, así como la organización, dirección control dentro una empresa. El proceso de toma de decisiones, necesita de todo el equipo de personas dentro de la empresa, encargados de la tecnología e infraestructuras clasificadas en estratos en niveles jerárquicos que están a cargo de distintas actividades dentro de la empresa, siendo un proceso que se traduce en el comportamiento lógico de las empresas, de hacer cosas mediante las personas o bajo la opinión de los individuos que conforman la organización Álvarez [14]

Así mismo, se puede mencionar que una decisión, se traduce a un juicio, por ello, comúnmente no se puede contemplar a una decisión como correcta o incorrecta. Siendo positivos, se puede mencionar que la toma de decisiones al involucrar el compromiso de las demás personas, puede llegar a ser casi correcto, o lo que se conoce como una solución óptima, en matemáticas. Moody [15]

2.2.2. Sistemas de información:

Se trata de un conjunto ordenado de elementos (no necesariamente computacionales) Que facilitan la manipulación de toda la información que se necesita para poder implementar elementos específicos para la toma de decisiones, cabe resaltar que todo sistema de información, nace del requerimiento de información que experimenta una empresa a fin de efectuar un conjunto específico para la toma de decisiones (autor).

2.2.3. Inventarios

Se refieren a las materias primas, en el que las cantidades de estas (stock), son las que están disponibles para su venta, o para su consumo en la producción de bienes a fin de comercializarse posteriormente.

Así mismo, representan un componente para amortiguar las diversas fases comerciales con la finalidad de lograr un funcionamiento que no genere pérdida económica o deficiencias respecto al sistema de producción.

Los tipos de inventarios según su naturaleza pueden clasificarse de acuerdo a la fig.1.



Fig. 1. Tipos de inventarios más conocidos según su naturaleza

- **Materias primas:**

Estos son convertidos en productos terminados, mediante un proceso de manufactura.

- **Productos terminados:**

Aquellos que son transformados para estar disponibles a su venta.

- **Materiales en proceso**

Al encontrarse en una etapa intermedia, necesitan un trabajo adicional a fin de ser transformados en producto terminado y ser vendidos.

- **Componentes o suministros**

Se encuentran en una etapa intermedia y requieren algún trabajo adicional para convertirse en productos terminados listos para la venta. Taha [10].

2.2.4. Sistema de control de inventarios

Este sistema facilita el responder ¿Cuánto comprar? y ¿Cuándo comprar? Al calcular el EOQ, se responderá a la primera de estas dos interrogantes. Si utilizamos un sistema de control de inventarios debemos estar en capacidad de responder a ambas interrogantes.

A fin de determinar el sistema de control, se debe de conocer el tipo de demanda [16], indicaron que si se tiene demanda independiente, se obtienen dos sistemas de control de inventarios: De revisión continua o sistema Q y el sistema de revisión periódica o sistema P.

2.2.5. Demanda:

Demanda se refiere al monto de unidades que se han requerido y no a las despachadas. Así mismo, la demanda será igual al consumo, si es que hay suficiente inventarios, debido a que cada unidad solicitada es despachada. La demanda será superior al consumo, al requerirse materiales tras una ruptura de inventario. Existen diferentes tipos de demanda los cuales se pueden clasificar de la siguiente manera:

2.2.5.1. Demanda Dependiente:

Está compuesta por la materia prima, los componentes y sub-ensambles que se utilizan para fabricar artículos que se utilizan en la fabricación de productos finales. Esta demanda debe predecirse, en el momento en el que se ejecuta un proceso de planeación de producción. No obstante, en ocasiones al calcular a partir de la demanda de artículos finales y las decisiones de producción, no es necesario predecirla.

2.2.5.2. Demanda Independiente:

Existe desconocimiento sobre la dependencia de la demanda con otro producto. Las condiciones del mercado fuera del control de operaciones, influyen sobre la demanda independiente; incluye los productos terminados, partes de repuestos y ciertos artículos cuya demanda aumenta más directamente del ambiente incierto del mercado. Por lo general requiere de alguna clase de pronósticos. Esta demanda puede ser de los siguientes tipos:

2.2.5.2.1. Demanda Determinística:

Es la demanda de un artículo que se conoce con certeza. Esta a su vez, puede ser:

- Estática

Es aquella donde la tasa de consumo permanece constante durante el transcurso del tiempo.

- Dinámica

Es aquella en la cual la demanda se conoce con certeza, pero varía de un periodo al siguiente.

2.2.5.2.2. Demanda Probabilística

Es cuando la demanda de un artículo está sujeta a incertidumbre y variabilidad, y se describe en términos de una función de probabilidad. Esta puede ser:

- Estacionaria

En la cual la función de densidad de probabilidad de la demanda, se mantiene sin cambio con el tiempo.

- No Estacionaria

Donde la función de densidad de probabilidad varía con el tiempo.

2.2.6. Minería de datos:

Servente y García-Martínez [17], denominan al termino minería de datos, al conjunto de técnicas y herramientas que se aplican al proceso no trivial de extraer y presentar conocimiento que no está propiamente dicho, que no se conoce, y que es potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con la finalidad de predecir a manera computarizada, tendencias y comportamientos; lo que permite describir de forma automatizada modelos previamente desconocidos. Piatetski-Shapiro et al., 1991; Chen et al., 1996; Mannila, 1997 [18]

Según Evangelos y Han [19], refieren que la minería de datos, esta específicamente relacionada a la aplicación de métodos de aprendizaje automático con la finalidad de descubrir y enumerar patrones presentes en los datos, para lo cual, se desarrollaron un gran número de métodos de análisis de datos basados en la estadística. Michalski [20].

Al momento en el que se incrementaba la cantidad de información almacenada en las bases de datos, estos métodos ya no poseían eficiencia y escalabilidad, dando origen al concepto de minería de

datos. La diferencia entre el análisis de datos que conocemos comúnmente y la minería de datos, es que el primer término se basa en las hipótesis que ya están construidas y validadas contra los datos, por el contrario minería de datos, se basa en los patrones e hipótesis que son extraídos de manera automática, de los datos Hernández [21].

La Minería de Datos es un proceso completo de descubrimiento de conocimiento que involucra varios pasos Morales [22]:

1. Entendimiento del dominio de aplicación, el conocimiento relevante a utilizar y las metas del usuario.
2. Seleccionar un conjunto de datos en donde realizar el proceso de descubrimiento.
3. Limpieza y pre procesamiento de los datos, diseñando una estrategia adecuada para manejar ruido, valores incompletos, valores fuera de rango, valores inconsistentes, etc.
4. Selección de la tarea de descubrimiento a realizar, por ejemplo, clasificación, agrupamiento o clustering, reglas de asociación, etc.
5. Selección de los algoritmos a utilizar.
6. Transformación de los datos al formato requerido por el algoritmo específico de explotación de datos, hallando los atributos útiles, reduciendo las dimensiones de los datos, etc.
7. Llevar a cabo el proceso de minería de datos para encontrar patrones interesantes.
8. Evaluación de los patrones descubiertos y presentación de los mismos mediante técnicas de visualización. Quizás sea

necesario eliminar patrones redundantes o no interesantes, o se necesite repetir algún paso anterior con otros datos, con otros algoritmos, con otras metas o con otras estrategias.

9. Utilización del conocimiento descubierto, ya sea incorporándolo dentro de un sistema o simplemente para almacenarlo y reportarlo a las personas interesadas.

Dentro de los pasos mencionados, es fundamental la etapa del pre-procesamiento de datos y su metamorfosis al formato que necesita el algoritmo, pues según como se ejecuten las tareas, beneficiara la calidad final de los patrones que están descubiertos. Un patrón es sugestivo, si es fácil de entender para los individuos, contenidamente útil, innovador o da origen a la validación de alguna hipótesis, que los usuarios necesitan confirmar. Un patrón interesante es un conocimiento

Las principales técnicas de minería de datos, se clasifican por su tarea de descubrimiento, en:

- Agrupación o clustering.
- Clasificación.
- Asociación.

2.2.7. Data Warehouse:

Según [24], Lo que trata de exponer un Data Warehouse, es un repositorio de datos que generan un panorama global, habitual y completado por los datos de la empresa, lo cual es independiente al cómo se utilicen a futuro, por los clientes o usuarios, mediante las siguientes propiedades: firme, coherente, confiable y con información histórica. Cuando se toma el ámbito global de la organización, así como la información histórica, el volumen de los datos será más grande, (hablamos de centenas de terabytes). Las bases de datos

relacionales, son la base técnica que se utiliza frecuentemente a fin de acumular estructuras de datos y sus volúmenes.

Por otro lado, Ramos [25] señala que el Data Warehouse se caracteriza por ser:

- **Orientado a temas:** los datos se organizan a través de temas, a fin de facilitar el entendimiento por parte de los usuarios, de tal manera que los datos referentes a un mismo dispositivo de la vida real, queden unidos entre ellos. Por ejemplo, todos los datos de un comprador, se pueden consolidar en una misma tabla, dando origen en todos los datos de los productos en otra y así continuamente.
- **Integrado:** Los datos deben ser integrados en una estructura consistente, al que deberá eliminarse las inconsistencias que existen en los diversos sistemas de operación. La información se organiza en variedad de niveles de consulta para los usuarios. Las inconsistencias más frecuentes que encontramos son: en nomenclaturas, unidades de medida, formatos de fechas, múltiples tablas con información similar (existen variedad de aplicaciones con tablas de clientes).
- **Histórico (variante en el tiempo):** los datos pueden cambiar con el tiempo, por lo que deberán quedar registrados de forma que al consultarse, muestren los cambios que se han dado, y no se altere la realidad existente al momento de almacenarlos, de esta manera se está evitando el problema que existe en diversos sistemas de operación, cuyas funciones solo muestran un estado actual del negocio. Un Data Warehouse tiene que reunir diversos valores que toma una variable al paso del tiempo. Por ejemplo, si un usuario ha vivido en tres ciudades diferentes, debe almacenar el periodo que vivió en cada una de ellas y asociar los hechos (ventas, devoluciones, incidencias, etc.) que se produjeron en cada momento a la ciudad en la que vivía cuando se produjeron, y no asociar todos los hechos históricos a la ciudad en la que vive actualmente.

- **No volátil:** la información de un Data Warehouse, al momento de ser encajada, debe ser de sólo para la lectura, no deberá ser modificada ni se eliminada, siendo permanente, manteniéndose para futuras consultas.

2.2.8. Data Mart:

Curto [24], menciona al Data Mart como subconjunto de los datos del data warehouse cuya meta Es responder a un minucioso análisis, función o necesidad, contando con una población de clientes específicos. Esta función es igual que la Data warehouse, pues los datos están organizados en modelos de estrella o copos de nieve; el autor indica que Data Mart, fue creado para cubrir los requerimientos de las necesidades de un equipo de trabajo o de una determinada área dentro de la empresa.

Así mismo, Kroenke [26] expresa que el Data Mart es parecido a un Data warehouse, sin embargo muestra un dominio más pequeño. Data Mart puede restringirse a un tipo en particular de administración de datos, a una función determinada de la empresa, una unidad de negocio concreto o un área geográfica.

Según Ramos [25], lo que distingue a un Data Mart de un Data Warehouse se trataría básicamente a su alcance, en cambio el Data Warehouse se trata de un sistema basado en los datos generales de la organización y de todos los procesos operacionales. El Data Mart se orienta a un proceso de negocio específico, el cual consta de una estructura especializada que verifica las perspectivas que afectan los procesos de dicha área. Así mismo, Ralph Kimball, menciona que un Data Mart tiene que estar encaminado a un proceso determinado dentro de la empresa, en cuanto a los pedidos que hacen clientes, compras, inventarios de almacén, envíos de materiales, entre otros; para el autor en mención, todo lo que engloba el Data Mart origina el Data Warehouse.

2.2.9. Proceso ETL:

De acuerdo a Ramos [27], un Data Warehouse o Data Mart, se basan en reunir información de diversas áreas, por lo que debe cargarse de manera frecuente, a fin de crear una base de datos que tenga los requisitos que se han descrito anteriormente, lo que involucra la existencia de una lista de etapas que puedan leer los datos de diversas áreas, transformándolos y adaptándolos al modelo que se haya propuesto, de esta manera se introducirán en la base de datos de destino, a este proceso se le denomina proceso ETL (Load).

Respecto al proceso anteriormente descrito, es fundamental diseñarlo de manera adecuada, pues este debe reunir todos los elementos de diversas áreas, a fin de ejecutar cálculos propios para la mejora de la calidad de datos, a fin de que se puedan adaptar y almacenar en el nuevo modelo físico.

Por otro lado, Trujillo y Manzon [28], indicaron El proceso ETL son muy importantes en la construcción del almacenamiento de datos, debido a que este proceso se encargan de obtener datos de las diversas áreas que contienen datos transaccionales, además de ejecutar las transformaciones que son requeridas, los cuales son cargados en el almacén cuando se hayan tratado, ejecutando las cargas de datos mientras dure la vida del almacén de datos.

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

- Extracción de los datos.- En este proceso, se debe tener en cuenta la decisión de trabajar con una actualización completa, la cual es más sencilla o con una actualización incremental, la cual es la más idónea. Es necesario mencionar que si elegimos la primera opción, lo que haremos serpa eliminar los datos y volver a cargarlos, y si escogemos la segunda opción, se generara una serie de técnicas y controles.

- Transformación de los datos.- en este proceso es fundamental cerciorarse de la validez de los datos, la integridad y el uso que le demos, lo que involucra ejecutar cálculos, a fin de crear nuevos valores. Deberán filtrarse los datos, para cercenar inconsistencias y duplicidades, lo que trae cambios en relación a la estructura de origen a fin de amoldarla al destino, a las modificaciones en el contenido de valores iniciales y generación de valores nuevos en las filas de destino.

- Carga en el Data Warehouse.- Como último paso debemos realizar el proceso de incorporar los datos al Data Warehouse y/o a los diferentes Data Marts, y a los cubos OLAP. Todo ello según la presentación, formatos definidos y periodicidad de actualización propuesta.

2.2.10. Algoritmo de serie temporal:

Este modelo permite la predicción de tendencias, que se basan en todos los datos originales empleados para generar el modelo. Así mismo, existe la posibilidad de generar datos nuevos a este modelo, cuando se ejecute la predicción, a fin de integrar de manera rápida los datos nuevos para el análisis de tendencias.

En la figura N^o4, se puede observar el modelo común que se emplea para la previsión en el tiempo que se requiere en ventas de un producto en cuatro regiones de ventas distintas. Este modelo evidencia las ventas de cada localización en líneas de los colores amarillo, rojo, purpura y azul. Cada línea de la región presenta dos partes:

- La línea vertical simboliza la información histórica, representando los datos que el algoritmo emplea para generar dicho modelo.

- A la derecha de la línea vertical, se encuentra la información de la predicción, representando la previsión que ejecuta el modelo.

Se le denomina serie, cuando se combinan los datos de inicio y los datos de predicción.

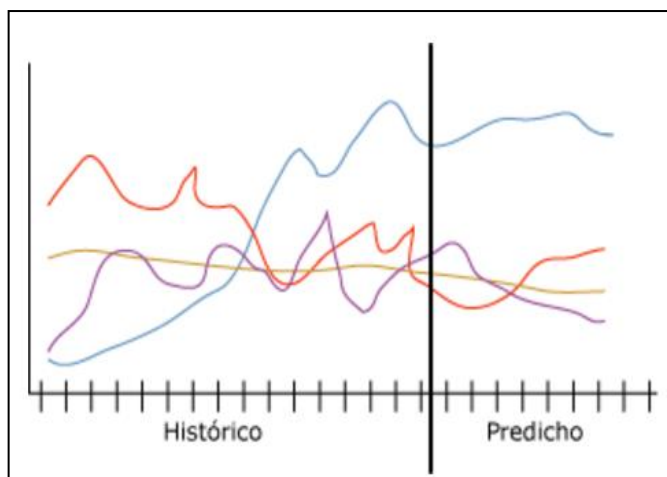


Fig. 2. Modelo típico de previsión en el tiempo de las ventas de un producto

El algoritmo de serie temporal de Microsoft, presenta una característica fundamental, la cual trata de la capacidad para mostrar predicciones de forma cruzada. Al emplear el modelo que se crea para predecir el resultado de una serie que se basa en el comportamiento de otra, deberá entrenarse el algoritmo que presenta dos series independientes. Tomando como ejemplo, las ventas que se observan en un producto puede involucrarse en las ventas que se prevee en otro producto.

La utilidad de la predicción cruzada, se basa en que es posible la creación de un modelo general, al que se le puede añadir múltiples series. Por ejemplo, la inestabilidad de las predicciones para una determinada región, suceden porque la serie no cuenta con datos de calidad adecuada. Así mismo, se podría aplicar el modelo a las series individuales, cuando se entrene un modelo general sobre las medidas

de las cuatro regiones, a fin de generar predicciones confiables para cada región.

- Algoritmo ARTXP: Tipo de serie temporal para predicciones a corto plazo.
- Algoritmo ARIMA: Tipo de serie temporal para predicciones a largo plazo.
- Algoritmo MIXED: Es la combinación de los dos mencionados anteriormente.

2.2.11. SQL Server:

De acuerdo a Microsoft se trata de un sistema que gestiona la base de datos relacionales (RDBMS) de Microsoft, el cual se ha creado a fin de ser de utilidad para la organización. Este sistema funciona en T-SQL (Transact-SQL), el cual está compuesto por extensiones de programación de Sybase y Microsoft, las cuales le proporcionar características a SQL estándar, incorporando el control de transacciones, manejo de errores y su filtro, procesamiento fila, y variables declaradas.

III. METODOLOGÍA

3.1. Tipo y nivel de investigación

3.1.1. Tipo de investigación

La investigación del presente proyecto es de tipo Pre-experimental. Donde tendremos un control sobre cada variable interviniente, haciendo una contrastación de la situación actual (Pre-test) con el resultado obtenido. Es decir, después de implementar el sistema (Post-Test).

3.1.2. Nivel de investigación

Nivel de investigación.

3.2. Diseño de investigación

TABLA I.
DISEÑO DE LA INVESTIGACIÓN

Variable independiente	Aplicación	Variable dependiente
O1	X	O2

Donde:

O1 = Gestión de ventas en la empresa MotoFuerza S.A.C. antes del sistema.

X = Implementación del Sistema BI con predicción de ventas basado en el algoritmo de series temporales en la empresa MotoFuerza S.A.C.

O2 = Gestión de ventas en la empresa MotoFuerza S.A.C. después del sistema.

3.3. Población, muestra y muestreo

3.3.1. Población

La población objeto del estudio estuvo constituida por:

Todas las personas que son responsables para la toma de decisiones ante las situaciones ocurridas en el proceso de ventas en la empresa MotoFuerza S.A.C.

3.3.2. Muestra

La muestra está conformada por el gerente de venta ya que él toma las decisiones de lo que a ventas se refiere puesto que se le ejecuto la entrevista.

La muestra del trabajo de tesis es muestreo no probabilístico intencionado; es decir se ha escogido como muestra a la población de la Empresa MotoFuerza S.A.C.

3.3.3. Muestreo

La técnica de muestreo que se aplicó ha sido en formato de encuesta y entrevista.

3.4. Criterios de selección

TABLA II.
MÉTODOS, TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Método	Técnica e Instrumento	Población
Entrevista y encuesta	Entrevista (Ver Anexo N° 3) Encuesta (Ver Anexo N 4)	Gerente de Ventas

3.5. Operacionalización de variables

Las variables que se han utilizado como elementos básicos en el desarrollo de la hipótesis están identificadas de la siguiente manera:

3.5.1. Variables

3.5.1.1. Variable independiente

Sistema BI basado en el Algoritmo de Series Temporales.

3.5.1.2. Variable dependiente

Gestión en la empresa MOTOFUERZA S.A.C

3.5.2. Indicadores (Operacionalización de variables)

**TABLA III.
INDICADORES**

Objetivo específico	Indicador(es)	Definición conceptual	Unidad de medida	Instrumento	Definición operacional
Incrementar los reportes sobre predicción de productos a vender	Cantidad de reportes sobre predicción de productos a vender	Reportes que ayudaran a la gerencia a tomar decisiones a largo o corto plazo	Unidad	Cuestionario Entrevista (Ver Anexo N° 3)	N° reportes con sistema BI - N° reportes sin el sistema BI
Incrementar el número de reportes para el análisis en la toma de decisiones	Numero de reportes analizando las ventas a través de las dimensiones	Reportes que se utilizaran para llevar un mejor manejo en la gestión de ventas	Unidad	Cuestionario Entrevista (Ver Anexo N° 3)	N° reportes con sistema BI - N° reportes sin el sistema BI
Incrementar el grado de satisfacción sobre la información solicitada	Grado de satisfacción del gerente de ventas sobre la información solicitado	Porcentaje de satisfacción por parte del gerente de ventas con respecto a la información solicitada	Porcentaje	Encuesta a los miembros de la gerencia (Ver Anexo N° 4)	(N° miembros satisfechos y muy satisfechos / total) *100
Incrementar la eficiencia de un modelo de predicción optimo	Eficacia de los modelos de proyección de ventas	Identificar el modelo óptimo para la futura proyección	Porcentaje	Hoja de Calculo	Porcentaje de error MAPE

3.6. Técnicas e instrumentos de recolección de datos

A continuación, en la siguiente tabla se muestra las técnicas e instrumentos que fueron útiles para la recolección de datos.

TABLA IV.
TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Técnicas	Instrumentos	Elementos de la población	Propósito
Entrevista y encuesta	Cuestionario Entrevista (Ver Anexo N° 3)	Gerente de ventas	Saber las dificultades y carencias al no tener un sistema BI
	Encuesta (Ver Anexo N° 4)		

3.7. Procedimientos

3.7.1. Metodología de desarrollo

La metodología seleccionada para el desarrollo de la investigación fue una híbrida que combina las metodologías de Crisp DM con la de Ralph Kimball ya que se utilizó la Minería de Datos, así como el procedimiento de Business Intelligence con las etapas descritas a continuación:

- Planificación del proyecto
- Definición de requerimientos de Minería de Datos y BI
- Comprensión de los Datos
- Modelo Dimensional
- Preparación de los datos de Minería de Datos
- Modelado de Minería de Datos
- Diseño Físico
- ETL
- Diseño de la arquitectura técnica
- Implementación
- Evaluación

3.7.2. Análisis de riesgos

El análisis de riesgos en el desarrollo de la presente tesis se efectuó con la finalidad de identificar las fases, entregables y objetivos afectados durante desarrollo de la presente tesis, las mismas se detallan en el Anexo N° 02.

3.7.3. Producto acreditable

1. Interfaces

Se construyeron las interfaces del sistema... haciendo uso del lenguaje... las mismas que se presentan en el *Capítulo IV. Resultados.*

2. Arquitectura

De diseñó una arquitectura idónea para el funcionamiento del sistema de negocios inteligentes basado en el algoritmo de series temporales, el cual se detalla en el *Capítulo IV. Resultados.*

3. Infraestructura tecnológica

Considerando la arquitectura anteriormente descrita, se definen las características de cada uno de sus componentes en el *Capítulo IV. Resultados.*

3.7.4. Manual de usuario

Se elaboró un manual de usuario con la finalidad de ayudar a los usuarios en el uso del sistema que se implementó.

3.8. Plan de procesamiento y análisis de datos

- ✓ Se utilizó instrumentos de recolección y procesamiento de datos tales como encuestas y entrevistas detalladas en el Anexo 3 y 4.
- ✓ La evaluación de estos resultados fueron sencillos ya que solo se le encuestó al gerente de ventas posteriormente él toma decisiones estratégicas con respecto al proceso de ventas en la empresa.

3.9. Matriz de consistencia

TABLA V.
MATRIZ DE CONSISTENCIA

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES
<u>FORMULACIÓN DEL PROBLEMA</u>	<u>OBJETIVO GENERAL</u>	<u>HIPÓTESIS</u>	<u>VARIABLES DE ESTUDIO</u>
De qué manera se puede apoyar la gestión en la empresa Motofuerza S.A.C.	Apoyar la gestión en la empresa Motofuerza S.A.C. mediante la implementación de un sistema BI con predicción de ventas basado en el algoritmo de series temporales.	La implementación de un sistema BI con predicción de ventas basado en el algoritmo de series temporales permitirá apoyar la gestión en la empresa Motofuerza S.A.C.	VARIABLE INDEPENDIENTE Sistema BI basado en el algoritmo de series temporales VARIABLE DEPENDIENTE Gestión de inventarios y predicción de ventas
<u>OBJETIVOS ESPECÍFICOS</u>	<u>DESCRIPCIÓN DEL LOGRO DE LOS OBJETIVOS ESPECÍFICOS</u>		<u>INDICADORES</u>
Incrementar los reportes sobre predicción de productos a vender	Se realizó una solución BI para crear reportes de predicción de ventas.		Cantidad de reportes sobre predicción de productos a vender
Incrementar el número de reportes para el análisis en la toma de decisiones	Se realizó una solución de BI para crear reportes que analizan las ventas a través de las dimensiones Se realizó una encuesta antes y después de realizar la solución		Numero de reportes analizando las ventas a través de las dimensiones Grado de satisfacción del gerente de ventas sobre la información que solicito
Incrementar el grado de satisfacción sobre la información solicitada	Usando métodos de margen de error porcentual se obtuvo el modelo optimo a utilizar		Eficacia de modelos de proyección de ventas

3.10. Consideraciones éticas

A continuación, se listan los aspectos que se han considerado para la protección y bienestar de los participantes de esta investigación, en este caso el gerente de ventas de la empresa Motofuerza S.A.C, así como de la seguridad (resguardo) de los datos:

- ✓ Aplicación de técnicas de recolección de datos: Encuestas, entrevistas, etc.
- ✓ Seguridad de la información (Por medio de backups y copias de respaldo tanto en la nube como en almacenamiento físico).
- ✓ Protección de contraseñas y datos registrados (Encriptando información).
- ✓ Resguardo de los datos y secreto de la información.

IV. RESULTADOS

4.1. En base a la metodología utilizada

4.1.1. Iteración #1: Planificación del proyecto (Plan de sistemas)

Para el desarrollo del presente proyecto, se incluirán área de ventas, en el cual es necesario plantearse como punto de partida los objetivos a conseguir con cualquier solución BI y tener en cuenta los fallos que se han presentado a lo largo del tiempo a la hora de llevar a cabo un proyecto de BI para no caer en ellos. El objetivo que se tiene en el presente proyecto es automatizar cada uno de los procedimientos que realizan los usuarios mediante reportes de forma precisa y oportuna y con ello permita una adecuada toma de decisiones en lo que respecta a las ventas.

Item	Descripción / Fase	Duración (Días)	Comienzo	Fin	% Avance
1	Planificación del proyecto	7	21/03/2018	27/03/2018	20%
2	Definición de requerimientos de Minería de Datos y BI	8	28/03/2018	4/04/2018	
3	Comprensión de los datos	6	5/04/2018	10/04/2018	
4	Modelado Dimensional	6	11/04/2018	16/04/2018	
5	Preparación de los datos de Minería de Datos	4	17/04/2017	20/04/2018	
6	Modelado de Minería de Datos	4	21/04/2018	24/04/2018	
7	AVANCE AL 20%	1	25/04/2018	25/04/2018	
8	Diseño Físico	6	26/04/2018	1/05/2018	50%
9	ETL	3	2/05/2018	4/05/2018	
10	Diseño de la arquitectura técnica	4	5/05/2018	8/05/2018	
11	AVANCE AL 50%	1	9/05/2018	9/05/2018	
12	Implementación	20	10/05/2018	29/05/2018	80%
13	AVANCE AL 80%	1	30/05/2018	30/05/2018	
14	Evaluación	20	31/05/2018	29/06/2018	100%
15	AVANCE AL 100%	1	30/06/2018	30/06/2018	
16	SUSTENTACION FINAL	1	6/07/2018	6/07/2018	100%

4.1.2. Iteración #2: Definición de los requerimientos de Minería de Datos y BI

En el trabajo realizado con respecto al proceso de recolección de requerimientos se realizaron entrevistas al personal encargado del área de ventas, en este caso a la Gerencia general de ventas, además de ello se recopiló documentos, bases de datos operacionales y reportes estadísticos de la organización.

- Se realizará el proceso de carga en el Data Mart con la información extraída de las hojas de cálculo XLS brindadas por la empresa
- Se creará una secuencia de rutinas de limpieza de datos ya que se verificará la validez y fiabilidad de los datos. Se verán los valores nulos y se verificara que la data sea consistente
- Permitir mostrar el análisis de proyección de datos ya que se utilizará para la toma de decisiones por parte del gerente de ventas.
- Evaluar el NTOP de los clientes, cantidad de ventas, unidades vendidas monto facturado en el tiempo por año.
- Predicción de los productos que se venden anualmente.

El objetivo principal:

Así mismo, dichos requerimientos permitieron identificar:

- ❖ Medidas
- ❖ Dimensiones

Con la información brindada a través de entrevistas y contrastando dicha información con la Base de Datos Operacional se definieron los siguientes requerimientos de BI:

- ❖ Cuál es el modelo de artículo más vendido.
- ❖ Cuál es el modelo que genera mayor rentabilidad al negocio.
- ❖ Cuál es la cantidad de productos vendidos en un determinado periodo de tiempo
- ❖ Cuál es el producto que genera mayor rentabilidad al negocio
- ❖ Cuáles son los clientes que generan mayor ingreso al negocio.
- ❖ Cuál es el monto de ingreso que ha sido generado por las ventas realizadas en un determinado periodo de tiempo.
- ❖ Cuál es la cantidad de ingresos que se han generado por cada empleado por medio de sus ventas en un determinado periodo de tiempo.
- ❖ Cuál será la cantidad de productos vendidos para los siguientes 3 meses.
- ❖ Participación de ingresos por producto

4.1.2.1. Requerimientos de Minería de Datos:

Los objetivos de Minería de datos suelen relacionarse mucho con los objetivos del negocio; sin embargo, estos son más técnicos. A continuación, se lista los requerimientos de minería de datos:

- ❖ Predecir la venta de motocicletas en un determinado lapso de tiempo, esto permitirá una planificación de las compras, mejorar la toma de decisiones estratégicas y minimizar los costos de inventario.

4.1.3. Iteración #3: Comprensión de Datos

En esta fase de la metodología híbrida se utilizó CRISP-DM, se realizó la recopilación de datos iniciales para establecer un primer contacto con el problema y realidad de la data, luego se procedió con la descripción de los datos obtenidos con el fin de familiarizarse con los datos y se concluye con la verificación de la calidad de datos; lo que ha permitido averiguar su calidad, así como identificar las relaciones más evidentes para formular las primeras hipótesis.

4.1.3.1. Recopilación de datos iniciales

Los registros transaccionales en bruto proporcionados por la empresa Motofuerza S.A.C. contienen toda la información del negocio a partir del año 2014 hasta el mes de agosto de 2017, a continuación, se detalla el estado de la recopilación inicial de los datos:

➤ Ventas por artículo detalle motos:

Contiene información en forma de columnas de la fecha de compra, el tipo de documento de compra el cual puede ser factura (FAC) o boleta (BOL), el documento que es un número que el sistema transaccional asigna; este dato debe analizarse y de ser necesario se debe reemplazar con un código; siendo que el contenido del dato es muy variado, la columna cliente contiene el nombre y apellido del cliente que realizó la compra el dato es consistente, la condición de pago; es la forma de pago como se realizó la compra; el dato es necesario excluirlo, la columna chasis es un dato que indica la serie del chasis del artículo; el dato muestra muchas filas con registros erróneos; se considera necesario excluirlo, la columna número de motor; este dato indica la serie del motor este dato es consistente e indispensable, color esta columna indica la color del artículo es necesario para establecer preferencias de los clientes, la columna cantidad; contiene las unidades de artículos vendidas al cliente este dato necesita un análisis exhaustivo por el hecho que se observa

cantidades con el símbolo negativo(-) y porque no coinciden con la sumatoria de las cantidades vendidas, la columna precio indica el valor del artículo, la columna importe indica el monto de dinero pagado por el artículo vendido y adicionalmente en este registro se puede encontrar debajo de la columna fecha el modelo del artículo comprado; el cual está en forma aislada e indica que los registros correspondientes de esa página son de los artículos vendidos que corresponden al del modelo que indica esta celda.

La estructura de la tabla no es la adecuada por lo que se ha tenido que reestructurar las tablas, utilizando para ello unos scripts con sentencias SQL y pasar el archivo a texto plano, para hacer una primera limpieza antes de pasar al ETL. A continuación, la Fig. 3 Ventas por artículo detalle motos, muestra parte del registro Ventas por artículo detalle motos descrito líneas arriba. Luego la Fig. 4 y 5, contiguas muestran la primera limpieza de los archivos obtenidos del Excel.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	T	U	V	W
1	FECHA	TIP.	DOCUMENTO	CLIENTE	CON.PAGO	CHASIS	MOTOR	COLOR	CANT	PRECIO	IMPORTE											
1541	07/05/14	BOL	00060022694	CAJUSOL CHAPOÑAN, EDUARDO JHOAN	CONTADO	EBFC-1011398	GCAEH-0028219	ROJO	1.00	5,099.00	5,099.00											
1544	30/04/14	FAC	00060008178	MUNICIPALIDAD DISTRITAL DE TUNAN	CRED 7 DIAS	-----	EBFC-1013176	ROJO	1.00	5,100.00	5,100.00											
1545	22/05/14	FAC	00060008377	MUNICIPALIDAD DISTRITAL DE ILLIMO	CRED (NO USAR)	3,196.00	EBFC-1013196	ROJO	1.00	5,100.00	5,100.00											
1546	16/06/14	FAC	00060008659	EDPYME ALTERNATIVA	CRED (NO USAR)	---	EBFC-1013229	ROJO	1.00	5,300.00	5,300.00											
1547	10/10/14	FAC	00060009666	SERVI-SAP SRL	CRED 15 DIAS	***	EBFC-1013203	ROJO	1.00	5,300.42	5,300.42											
1548	13/10/14	BOL	00060026762	VASQUEZ COEÑA, MILTON OMBR	CONTADO	---	EBFC-1013330	ROJO	1.00	5,099.00	5,099.00											
1549	09/12/14	FAC	00060010248	MUNICIPALIDAD DISTRITAL DE IMAZA	CONTADO	---	EBFC-1013204	ROJO	1.00	5,099.00	5,099.00											
1550	06/03/15	BOL	00060029497	BURGA CHONLON, TIMOTEO	CONTADO	361.*	EBFC-1014531	ROJO	1.00	5,299.00	5,299.00											
1551	26/03/15	FAC	00060011071	EDPYME ALTERNATIVA	CRED 30 DIAS	5.00	EBFC-1015042	ROJO	1.00	5,300.00	5,300.00											
1552	26/03/15	FAC	00060011070	EDPYME ALTERNATIVA	CRED (NO USAR)	0.00	EBFC-1015041	ROJO	1.00	5,300.00	5,300.00											
1553	26/03/15	FAC	00060011069	EDPYME ALTERNATIVA	CRED (NO USAR)	0.90	EBFC-1015039	ROJO	1.00	5,300.00	5,300.00											
1554	26/03/15	FAC	00060011068	EDPYME ALTERNATIVA	CRED 30 DIAS	---	EBFC-1015040	ROJO	1.00	5,300.00	5,300.00											
1555	26/03/15	FAC	00060011072	EDPYME ALTERNATIVA	CRED (NO USAR)	*6	EBFC-1015051	ROJO	1.00	5,300.00	5,300.00											
1556	24/04/15	NOT	00060001443	EDPYME ALTERNATIVA	CRED (NO USAR)	0.90	EBFC-1015039	ROJO	-1.00	5,300.00	-5,300.00											
1557	24/04/15	FAC	00060011642	EDPYME ALTERNATIVA	CRED (NO USAR)	0.90	EBFC-1015039	ROJO	1.00	5,300.00	5,300.00											
1558	23/10/15	FAC	00060013124	INVERSIONES AGRICOLAS OLMOS S.A.C.	CRED 30 DIAS	*	EBFC-1015050	ROJO	1.00	5,869.47	5,869.47											
1559	07/12/15	FAC	00060013600	SERVICIO NACIONAL DE SANIDAD AGROPEC	CRED 2 DIAS	..*8	EBFC-1014557	ROJO	1.00	5,780.00	5,780.00											
1560	14/12/15	FAC	00060013678	GUIVAR SILVA MARIA CONSUELO	CONTADO	++98	GCAEH-0034197	ROJO	1.00	6,585.67	6,585.67											
1561	22/01/16	NOT	00060001495	GUIVAR SILVA MARIA CONSUELO	CONTADO	++98	GCAEH-0034197	ROJO	-1.00	6,585.67	-6,585.67											
1562	28/01/16	FAC	00060014028	EDPYME ALTERNATIVA	CRED (NO USAR)	++9369	GCAEH-0034108	ROJO	1.00	5,848.15	5,848.15											
1563	05/02/16	FAC	00060014099	EDPYME ALTERNATIVA	CRED (NO USAR)	++98	GCAEH-0034197	ROJO	1.00	6,199.26	6,199.26											
1564	27/02/16	FAC	00130000278	AGRICOLA PANPA BAJA S.A.C.	CRED 15 DIAS	12,563.00	GCAEH-0034098	ROJO	1.00	6,035.70	6,035.70											
1565	22/06/16	FAC	00130000307	EDPYME ALTERNATIVA	CRED 7 DIAS	5,552.00	GCAEH-0035475	ROJO	1.00	5,780.00	5,780.00											
1566	22/06/16	FAC	00060015212	MOTO NORTE S.A.C.	CRED 60 DIAS	666,662.00	GCAEH-0035462	ROJO	1.00	5,606.04	5,606.04											
1567				MOTOFUERZA S.A.C.																		
1568	20478217898																					
1569																						
1570																						
1571																						

Fig. 3 Ventas por artículo detalle motos
Fuente: empresa Motofuerza S.A.C.

Número	Fecha	TipoDoc	ConPago	Cantidad	Monto	CódigoCliente
00060023371	16/04/14	BOL	CONTADO	1.00	4690	CASTAÑEDA SALDAÑA, ROBERTO CRHISTIAN
60023421	21/04/14	BOL	CONTADO	1.00	4690	BARBOZA SOSA, SEGUNDO ROLANDO
60023980	22/05/14	BOL	CONTADO	1.00	4690	CRUZ CALVANAPON, NILSER ONAN
00060024055	27/05/14	BOL	CONTADO	1.00	4690	TALLEDO LARRAIN, EDGAR NOLBERTO
00060024233	07/06/14	BOL	CRED SCOTIA	1.00	4690	SERRATE CARRASCO, HENRRY SAMUEL
00060024328	09/06/14	BOL	CRED IDIA	1.00	4690.34	CAMPOS TAVARA, JOSE CRISTIAN
00060024537	19/06/14	BOL	CRED SCOTIA	1.00	4690.0	DE LA CRUZ ZAPATA, CESAR MIGUEL
00060024891	07/07/14	BOL	CONTADO	1.00	4790	TIGRE ACOSTA, EDWIN ALEXANDER
00060024930	08/07/14	BOL	CONTADO	1.00	4790	RUJZ REQUEJO, JULIO CESAR
00060025072	15/07/14	BOL	CONTADO	1.00	4790	ARMAS REAÑO, LILIANA YSABEL
00060025572	08/08/14	BOL	CRED SCOTIA	1.00	4790	VENTURA MENESES, FRANCO
00060026005	29/08/14	BOL	CRED IDIA	1.00	4790	GONZALES MESTANZA, LUIS ELVIS
00060009427	11/09/14	FAC	CRED IDIA	1.00	4790.72	CALDERON AGUIRRE JOSE LUIS
00060026287	15/09/14	BOL	CRED SCOTIA	1.00	4790.05	BAUTISTA DIAZ, SEGUNDO AMADOR
00060009489	18/09/14	FAC	CRED IDIA	1.00	4790	BURGA RUJZ ANITA
00060026426	23/09/14	BOL	CREDITO DIRECTO	1.00	4790	MOZOMBITE SHUÑA, AMADEO
00060009541	24/09/14	FAC	CRED 60 DIAS	1.00	4790	MOTO NORTE S.A.C.
00060009542	24/09/14	FAC	CRED 60 DIAS	1.00	4790	MOTO NORTE S.A.C.
00060026525	30/09/14	BOL	CONTADO	1.00	4790	BRAVO ROJAS, WALTER ISAIAS

Fig. 4 Primera limpieza de la tabla ventas

IDArtículo	Artículo	Modelo	Chasis	Motor	Color	Precio
1	MOTOCICLETA CB110	CB110	ME4JC47J4E7002564	JC47E-7-1026480	NEGRO	4690
2	MOTOCICLETA CB110	CB110	ME4JC47J9E7002561	JC47E-7-1026508	NEGRO	4690
3	MOTOCICLETA CB110	CB110	ME4JC47J0E7002528	JC47E71025409	NEGRO	4690
4	MOTOCICLETA CB110	CB110	ME4JC47J2E7002563	JC47E-7-1026510	NEGRO	4690
5	MOTOCICLETA CB110	CB110	ME4JC47J8E7001952	JC47E-7-1019671	ROJO	4690
6	MOTOCICLETA CB110	CB110	ME4JC47JXE7002570	JC47E-7-1026517	NEGRO	4690.34
7	MOTOCICLETA CB110	CB110	ME4JC47J5E7002525	JC47E71025435	NEGRO	4690
8	MOTOCICLETA CB110	CB110	ME4JC47J9E7001670	JC47E71019278	NEGRO	4790
9	MOTOCICLETA CB110	CB110	ME4JC47J3E7001650	JC47E71019164	NEGRO	4790
10	MOTOCICLETA CB110	CB110	ME4JC47J4E7001656	JC47E71019220	NEGRO	4790
11	MOTOCICLETA CB110	CB110	ME4JC47J7E7001683	JC47E71019375	NEGRO	4790
12	MOTOCICLETA CB110	CB110	ME4JC47J3E7001681	JC47E-7-1019341	NEGRO	4790
13	MOTOCICLETA CB110	CB110	ME4JC47J5E7001746	JC47E-7-1019424	NEGRO	4790.72
14	MOTOCICLETA CB110	CB110	ME4JC47J7E7001621	JC47E71019133	NEGRO	4790.05
15	MOTOCICLETA CB110	CB110	ME4JC47J1E7001680	JC47E71019340	NEGRO	4790
16	MOTOCICLETA CB110	CB110	ME4JC47J1E7001954	JC47E-7-1019661	ROJO	4790
17	MOTOCICLETA CB110	CB110	ME4JC47J9E7002706	JC47E-7-1030698	NEGRO	4790
18	MOTOCICLETA CB110	CB110	ME4JC47JXE7002763	JC47E-7-1030758	NEGRO	4790
19	MOTOCICLETA CB110	CB110	ME4JC47JXE7001712	JC47E-7-1019408	NEGRO	4790
20	MOTOCICLETA CB110	CB110	ME4JC47J6F7003085	JC47E-7-1037912	NEGRO	4790

Fig. 5 Primera limpieza de la tabla artículo

- **Compras por artículo detalle:** la tabla contiene los registros de las compras de artículos realizadas al proveedor, contiene en forma de columnas los campos: MODELO, CHASIS, MOTOR, COLOR y PRECIO. La estructura de la tabla es adecuada. A continuación, la Figura n°6 Compras por artículo detalle, muestra lo detallado líneas arriba.

MODELO	CHASIS	MOTOR	COLOR	PRECIO
XL200	9C2MD2891DR203124	MD28E9D203124	ROJO	9,263.16
XL200	9C2MD2892DR203374	MD28E9D203374	BLANCO	9,263.16
XL200	9C2MD2899ER200246	MD28E9E200246	NEGRO	9,263.16
XL200	9C2MD2899DR203050	MD28E9D203050	ROJO	9,263.16
XL200	9C2MD2899DR203534	MD28E9D203534	BLANCO	9,263.16
XL200	9C2MD289XDR203316	MD28E9D203316	BLANCO	9,263.16
XL200	9C2MD2895DR203255	MD28E9D203255	BLANCO	9,263.16
XL200	9C2MD2890DR203342	MD28E9D203342	BLANCO	9,263.16
XL200	9C2MD289XDR203526	MD28E9D203526	BLANCO	9,263.16
XL200	9C2MD2892DR203519	MD28E9D203519	BLANCO	9,263.16

Fig.6 Compras por artículo detalle
Fuente: empresa Motofuerza S.A.C.

Se tiene una dificultad para conocer la dirección de los clientes y proveedores puesto que por motivos de seguridad personal de datos la empresa no facilita esa información. Para concluir se hace presente que no se considera la adquisición de datos de alguna otra fuente.

4.1.3.2. Descripción de los datos

En la data de la empresa existe multitud de registros y atributos para procesar en una aplicación de minería de datos. En el proyecto de minería de datos se ha limitado su estudio inicial a unos 4.754 registros de venta de artículos aproximadamente, aunque se hayan registrado en data bruta, 15, 000 datos.

Los tipos de valor de estos orígenes de datos son simbólicos, ya sean las fechas, categórico ya sean chasis, motor, modelo del artículo, y numéricos ya sean el precio, cantidad del artículo y, el importe.

La forma de codificación de los diferentes orígenes de datos causa dificultad para procesarlos, porque los orígenes de datos contienen atributos muy diferentes y difíciles de relacionar entre sí. Por ejemplo las variables que aparecen como “claves”, como los códigos de clientes, códigos de vendedor, códigos de productos y código de venta. Estas variables deben tener esquemas de codificación idénticos desde un origen de los datos a otro; de otro modo no será imposible relacionar los orígenes de datos. Se deberá realizar una preparación adicional de los datos para volver a codificar estos campos clave para fusionar.

4.1.3.3. Verificación de la calidad de datos

La verificación de la calidad de los datos realizada durante estos procesos de descripción y exploración ha permitido detectar algunos de los problemas los cuales suelen incluir:

Datos perdidos. Los datos perdidos conocidos incluyen campos vacíos, campos con caracteres especiales y datos numéricos, en los registros de chasis. Sin la información necesaria que proporciona este dato, estos registros de chasis se pueden omitir en algunos de los siguientes modelos.

Errores de datos. La mayoría de los orígenes de datos se generan automáticamente, por lo que no es un problema grave, sin embargo se observa que la variable “Numero” tiene errores tipográficos por lo que es recomendable generarle un valor para identificar al cliente y relacionarlo con el artículo.

Errores de mediciones. El origen principal de los errores de mediciones es el campo “CAN” (cantidad) en la tabla Ventas por artículo detalle motos. Se observa que existen elementos que expresan cantidad negativa, lo que está dificultando la sumatoria de productos y el importe pagado. Si alguno de los elementos no está cumplimentado correctamente, es posible que no proporcione la información que se espera obtener. De nuevo, durante el proceso de exploración, es importante prestar una especial atención a los elementos que expresan la cantidad del artículo vendido.

A continuación la tabla VI, muestra el resumen de los datos con errores de medición, valores perdidos, valores nulos; además indica la cantidad de muestras que contiene los registros, el porcentaje de aprobación del dato que contiene la variable y el procedimiento que se realizara para cada caso.

TABLA VI.
CUADRO DE RESULTADO DE VALORES CON ERRORES DE MEDICIÓN, VALORES PERDIDOS Y VALORES NULOS, (ELABORACIÓN PROPIA)

Variable	Total errores medición	Total perdidos	Total nulos	Total muestras	Porcentaje de aprobación	Procedimiento
CAN	68	0	0	4754	98.56%	Se corregirá manualmente
chasis	0	0	362	4754	92.38%	Eliminar variable no relevante
motor	0	0	1	4754	99.77	Eliminar Variable
Modelo	0	0	0	4754	100%	Dato correcto
artículo	0	0	0	4754	100%	Asignar un número como identificador
Precio	0	0	0	4754	100%	Dato correcto
color	0	0	0	4754	100%	Dato correcto
Numero	4754	0	0	4754	0%	Asignar un número como identificador
Monto	0	0	0	4754		Dato correcto

Los datos utilizados en este proyecto son datos referentes a empleados y clientes que incluyen información personal como son únicamente, sus nombres y apellidos.

A continuación, listamos los datos adquiridos:

❖ **Cliente**

Cada cliente está identificado por su código de cliente que es un valor numérico. Todo cliente está relacionado con una venta y un pago que éste realizó.

❖ **Vendedor**

Cada vendedor está identificado por su código de vendedor el cual es un valor numérico. Todo empleado está relacionado con una venta que éste realizó durante el día.

❖ **Artículo**

Cada artículo está identificado con un iD artículo el cual es un valor numérico. Todo artículo está relacionado con una localidad en la cual se encuentra.

❖ **Forma de Pago**

Cada venta está identificada con su código de forma de pago el cual es un valor numérico. La forma de pago está relacionada con la venta directamente realizada en el transcurso del día a día.

❖ **Tipo de Documento**

Cada venta está identificada con su tipo de documento el cual es un valor numérico. El tipo de documento está relacionado con la venta directamente realizada en el transcurso del día a día.

❖ **Tiempo**

El tiempo es extraído en formato numérico.

Los atributos específicos que serán útiles a la hora de hacer la minería de datos son:

- Identificador de Cliente
- Identificador de Vendedor
- Identificador de Artículo
- Ventas realizadas anualmente.
- Tiempo de ventas realizadas por los Empleados.

4.1.4. Iteración #4: Preparación de los datos de Minería de Datos

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

❖ **Descripción de Tablas Fuentes**

Las tablas a continuación muestran los datos que contienen las tablas fuentes, se toma el total de 4754 registros, se considera las variables más representativas que permitan analizar las ventas y posteriormente predecir el comportamiento de las mismas, a través de los modelos de predicción que se ajusten al objetivo del proyecto. Los datos que se considera son: fecha de venta, artículo, cantidad y monto por mes.

**TABLA VII.
VENTAS DE MOTOCICLETAS POR MESES EN EL AÑO 2014.**

	FECHA	ARTICULO	CANTIDAD	MONTO
Año 2014	Enero	motocicletas	100	535,468.68
	Febrero	motocicletas	93	471739.85
	Marzo	Motocicletas	138	669,259.17
	Abril	Motocicletas	92	467,966.26
	Mayo	Motocicletas	124	683,944.09
	Junio	Motocicletas	208	951,855.16
	Julio	Motocicletas	93	496,100.41
	Agosto	Motocicletas	104	751910.62
	Septiembre	Motocicletas	89	533860.55
	Octubre	Motocicletas	116	617269.65
	Noviembre	Motocicletas	91	457690.15
	Diciembre	Motocicletas	91	448427.46

TABLA VIII.
VENTAS DE MOTOCICLETAS POR CADA MES EN EL AÑO 2015

FECHA	ARTICULO	CANTIDAD	MONTO	
Año 2015	Enero	Motocicletas	91	448427.46
	Febrero	Motocicletas	92	467,966.26
	Marzo	Motocicletas	124	683,944.09
	Abril	Motocicletas	120	630,900.09
	Mayo	Motocicletas	91	448427.46
	Junio	Motocicletas	93	471739.85
	Julio	Motocicletas	138	669,259.17
	Agosto	Motocicletas	208	951,855.16
	Septiembre	Motocicletas	93	496,100.41
	Octubre	Motocicletas	104	751910.62
	Noviembre	Motocicletas	89	533860.55
	Diciembre	Motocicletas	116	617269.65

TABLA IX.
VENTAS DE MOTOCICLETAS POR CADA MES EN EL AÑO 2016

FECHA	ARTICULO	CANTIDAD	MONTO	
Año 2016	Enero	Motocicletas	81	469555.80
	Febrero	Motocicletas	102	708338.53
	Marzo	Motocicletas	110	749877.42
	Abril	Motocicletas	117	789355.68
	Mayo	Motocicletas	103	692158.61
	Junio	Motocicletas	87	507659.11
	Julio	Motocicletas	100	592107.43
	Agosto	Motocicletas	101	660180.63
	Septiembre	Motocicletas	95	561613.05
	Octubre	Motocicletas	69	444864.77
	Noviembre	Motocicletas	80	487874.37
	Diciembre	Motocicletas	100	742872.00

**TABLA X.
VENTAS DE MOTOCICLETAS POR CADA MES EN EL AÑO 2017**

FECHA		ARTICULO	CANTIDAD	MONTO
Año 2017	Enero	Motocicletas	99	639784.74
	Febrero	Motocicletas	133	723460.18
	Marzo	Motocicletas	111	597755.75
	Abril	Motocicletas	96	560085.76
	Mayo	Motocicletas	75	482806.49
	Junio	Motocicletas	115	816024.44
	Julio	Motocicletas	90	638424.20
	Agosto	Motocicletas	99	640721.67
	Septiembre	Motocicletas	110	905952.76
	Octubre	Motocicletas	94	899083.86
	Noviembre	Motocicletas	97	856535.53
	Diciembre	Motocicletas	102	900267.56

**TABLA XI.
VENTAS DE MOTOCICLETAS DESDE ENERO HASTA MAYO EN EL AÑO 2018**

FECHA		ARTICULO	CANTIDAD	MONTO
Año 2018	Enero	Motocicletas	121	1166585.31
	Febrero	Motocicletas	118	1101831.29
	Marzo	Motocicletas	124	1132184.70
	Abril	Motocicletas	89	835397.09
	Mayo	Motocicletas	87	808399.32

A continuación, la tabla XII muestra los diversos modelos de los artículos que contienen las tablas fuentes y a su vez la cantidad que se ha vendido durante este periodo de tiempo (2014-2018).

**TABLA XII.
MODELO DE ARTÍCULO Y CANTIDAD VENDIDA**

MODELO DEL ARTICULO	CANTIDAD VENDIDA
CB110	144
CB125F	32
CB150 INVICTA	164
CB160F	72
CB190R	283
CBF110MC	1
CBF110ME	20
CBF125MC	7
CBF125 STUNNER	15
CBF150 INVICTA	39
CBF150MA	63
CBF150MC	1
CBR250R	22
CBR250RAD	5
CBR250RC	6
CBX250	3
CCG125	214
CG 150	16
CRF230F	2
CRF250LF	1
DIO	35
EG 1000 TIPO S	41
EG4000CX	7
EG5000CX	33
EG6500 CX S	16
EG6500 CX S S	23
ELITE 125	101
EP2500C SC	9
EP2500CX	18
EU10-I	1
EU20-I	3
G200 QAMD	11
GL125	897
GL150	328
GX120T1 QH	1
GX160T1QAB3	2
GX270 T2 QP	1
GX270 T2QH	2
GX35T-SD	4
GX390H1 QH1	6
GX390T1QH	2

HJ / 4 / 13 C V	45
HJ / 4 / 9 C	22
HJ/2/9 MP	2
HJ/3/13AP	4
HJ/3/7 C1	32
HJ/3/9 S	3
HJ/4/13 A	30
HJ/4/13 C	84
HJ/4/9/A	3
HV/28/5.5 QAB31	14
NF100 WAVE	5
SHADOW 150	4
STORM	60
TRX250TEE	8
TRX420FE1E	9
TRX420FE1F	12
TRX420FE1G	4
TRX420FE1H	12
TRX420FED	1
TRX420TE1E	1
UMK435	1
UMK435T-ACC01	13
UMR435T	15
V-MEN	3
WAVE 110	162
WB20XHDR	64
WB30XHDR	189
WB30XTDR	1
WJR2525T	17
WJR4025T	31
X190CT	4
XL200	36
XR125L	142
XR150L	689
XR190L	106
XR250-TORNADO	262
TOTAL VENTAS	4754

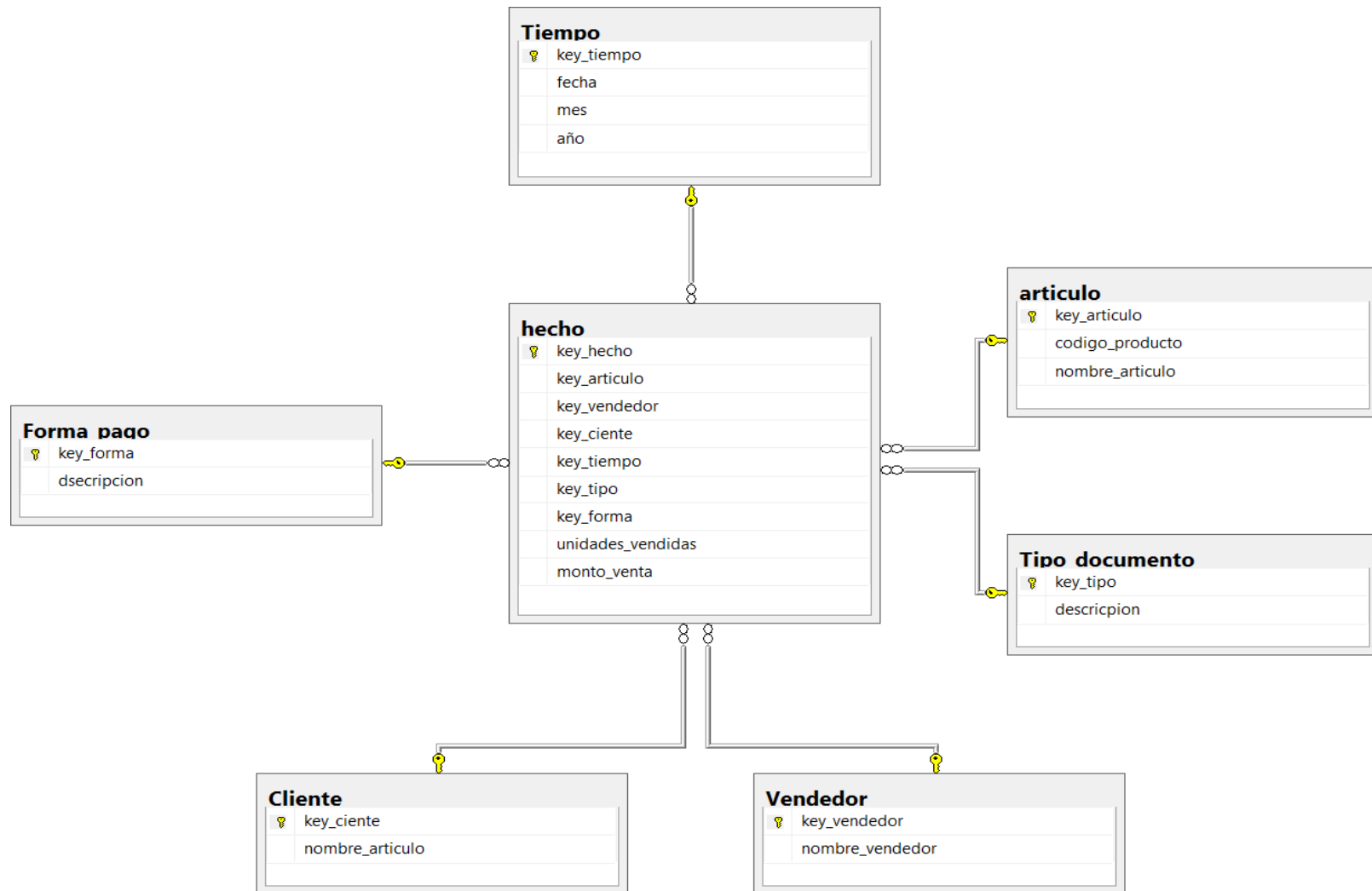


Fig. 7. Modelado Dimensional

4.1.5. Iteración #5: Modelado Dimensional

Para los modelos dimensionales que se han realizado mediante un Esquema Estrella. Dicho Esquema Estrella se realiza mediante la elaboración de Tabla o tablas de Hechos o Fact Table, donde la tabla contiene los datos para el análisis y está rodeada de Tablas de Dimensiones que son tablas más pequeñas y con datos necesarios para la tabla hechos.

A continuación, en la Tabla XIII se presentan las consideraciones de la creación de los modelos dimensionales respecto a los estándares de nombres usados:

4.1.5.1. Dimensiones VS Hechos – Matriz Bus

En la siguiente TABLA XIII se muestra la matriz bus con las dimensiones que utiliza cada tabla de hechos.

**TABLA XIII.
MATRIZ BUS**

Hechos	Matrices	Dimensiones				
		Tiempo	Cliente	Artículo	Tipo Pago	Vendedor
Hecho Ventas	Unidades vendidas	X	X	X	X	X
	Monto soles ventas	X	X	X	X	X

4.1.5.2. Start Mart

La siguiente Fig. 8 muestra el hecho ventas con sus respectivas dimensiones, así como también los atributos de cada una de las dimensiones.

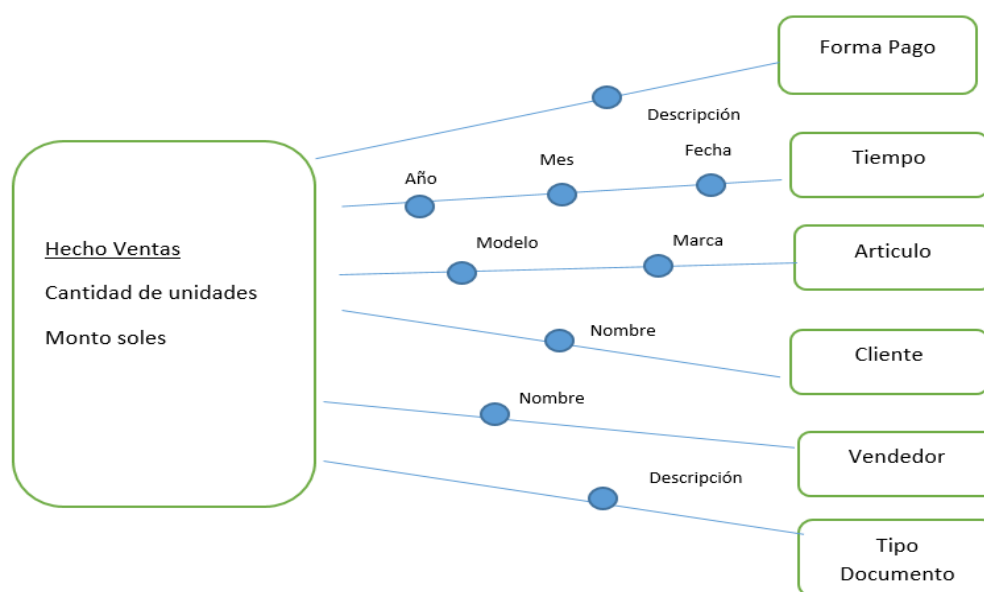


Fig. 8. Start Mart Ventas

4.1.5.3. Análisis de Procesos

En Este apartado se explica los procesos cubiertos por el datamart, y además se mostrará el nivel de granularidad de cada uno de ellos.

4.1.5.3.1. Procesos de Ventas

Este proceso tiene por objetivo analizar las ventas de la organización.

Los objetivos que busca cumplir este tema son:

- Analizar la Cantidad de Unidades de artículos vendidos
- Analizar el Costo de los artículos vendidos
- Analizar el margen de utilidad de ventas
- Analizar el monto en soles ingresado por las ventas
- Analizar el Precio Unitario de los artículos vendidos
- Analizar la Utilidad Bruta de las ventas.

Los resultados de este tema se miden por:

- Periodos de tiempo, año y mes.
- Campañas, que son conjuntos de fechas en las que se ofrecen artículos de temporada.
- Empleados.
- Clientes.
- Ventas

4.1.5.4. Dimensiones y atributos

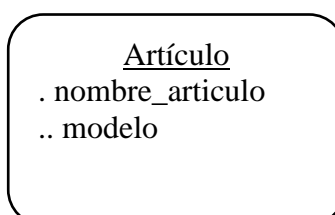
A continuación, en la TABLA XIV se representan las dimensiones que son utilizadas en los temas que son presentados en el punto anterior.

**TABLA XIV.
DIMENSIONES**

Nº	Nombre
1	Dimensión Artículo
2	Dimensión Tiempo
3	Dimensión Cliente
4	Dimensión Vendedor
5	Dimensión Forma de Pago
6	Dimensión Tipo de Documento

4.1.5.4.1. Dimensión Artículo

Esta dimensión contiene los datos de los artículos para los cuales se registran las compras y las ventas. Esta dimensión presenta en su estructura una jerarquía, la cual es detallada en la Tabla



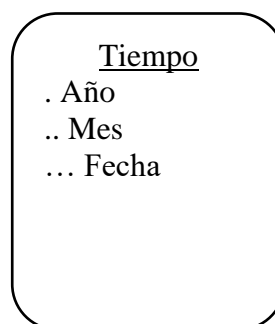
Se puede verificar que la TABLA XV muestra la lista de atributos que componen la Dimensión Artículo

**TABLA XV.
DIMENSION ARTÍCULO**

Atributo	Contenido	
	Descripción	Formato
key_artículo	Identificador de Artículo	Entero correlativo
código_artículo	Código de artículo de BD origen	Cadena
precio_artículo	Precio unitario	Decimal

4.1.5.4.2. Dimensión Tiempo

Esta dimensión contiene los periodos de tiempo por el cual se obtendrá información.



A continuación, en la TABLA XVI se muestra la lista de atributos que componen la Dimensión Tiempo

**TABLA XVI.
DIMENSIÓN TIEMPO**

Atributo	Contenido	
	Descripción	Formato
key_tiempo	Identificador de Tiempo	Entero correlativo
año	Número de año	Entero
mes	Número de mes	Entero
fecha	Fecha	Fecha

4.1.5.4.3. Dimensión Cliente

En esta dimensión se van a contener los datos de los clientes de la organización en los que se registra las ventas.

En la TABLA XVII se muestra la lista de atributos que componen la Dimensión Cliente

**TABLA XVII.
DIMENSION CLIENTE**

Atributo	Contenido	
	Descripción	Formato
key_cliente	Identificador de Cliente	Entero correlativo
nombre_cliente	Número de Cliente	Cadena
direcciónCliente	Dirección de Cliente	Cadena

4.1.5.4.4. Dimensión Vendedor

En esta dimensión se van a contener los datos de los empleados de la organización quienes registran las ventas.

En la TABLA XVIII se muestra la lista de atributos que componen la Dimensión Vendedor

**TABLA XVIII.
DIMENSIÓN VENDEDOR**

Atributo	Contenido	
	Descripción	Formato
key_vendedor	Identificador de vendedor	Entero correlativo
código_vendedor	Código de vendedor	Entero
nombre_vendedor	Nombre de vendedor	Cadena

4.1.5.4.5. Dimensión Tipo Pago

En esta dimensión se van a contener los datos de los tipos de pago que la organización maneja.

En la TABLA XIX se muestra la lista de atributos que componen la Dimensión Vendedor

**TABLA XIX.
DIMENSIÓN FORMA PAGO**

Atributo	Contenido	
	Descripción	Formato
key_forma_pago	Identificador de forma pago	Entero correlativo
código_forma_pago	Código forma pago	Entero
descripcion	Descripción forma pago	Cadena

4.1.5.4.6. Tabla de Hechos

En la TABLA XIX se presenta la lista de la tabla de hechos (fact table) que conforman el modelo dimensional estrella

**TABLA XX.
LISTA DE HECHOS**

Nº	Tema	Nombre
1	Ventas	HechoVentas

En el siguiente apartado se muestra la descripción de cada una de las tablas de hechos:

Hecho Ventas

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

4.1.6. Iteración #6: Modelado de Minería de Datos

Cuando se utiliza una metodología híbrida, en el cual se ha integrado la metodología de Ralph Kimbal – Bottom- Down para el DataWarehouse y la metodología CRIPS (para minería), pues en esta parte de la metodología se desarrolla la fase de la metodología de la minería.

Para ello, al modelar la minería en Microsoft Analysis Services se va a crear una vista con cada uno de los datos preparados, y es a partir de ello que se van a seleccionar nuestros datos de aprendizaje, tal como se muestra en la siguiente figura:

```
-- CREACION DE LA VISTA SEGUN LA SITAXIS QUE SERA TOMADA POR EL ALGORITMO DE SERIES TEMPORALES DE MICROSOFT
create view v_tp_articulos as
select cast('1/'+right('00'+cast(mes as varchar),2)+'/'+cast(año as char(4)) as date) as fecha, articulo, unidades_vendidas from tp_articulos

select * from v_tp_articulos where year(fecha) = 2017 order by 2,1

--CONSULTA
delete from v_tp_articulos where year(fecha) =2017 and month(fecha)>=10
delete from v_tp_articulos where year(fecha) =2017
```

Fig. 9 Moldeamiento de los datos para adecuar al modelo de predicción

4.1.7. Iteración #7: Diseño Físico

**TABLA XXI.
MODELADO FÍSICO DE LA TABLA
DIMENSIÓN CLIENTE**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
key_cliente	Int		No	Si	PK
nombre_cliente	varchar	300	Si	No	
Origen Datos	Tabla de Cliente en el archivo XLS				

**TABLA XXII.
MODELADO FÍSICO DE LA TABLA
DIMENSIÓN ARTÍCULO**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
key_artículo	Int		No	Si	PK
codigo_artículo	Int		No	No	
nombre_artículo	Varchar	200	No	No	
precio_artículo	Float		No	No	
Origen Datos	Tabla de Articulo en el archivo XLS.				

**TABLA XXIII.
MODELADO FÍSICO DE LA TABLA FORMA
DE PAGO**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
key_forma_pago	Int		No	Si	PK
descripcion	Varchar	200	No	No	
Origen Datos	Tabla de Venta en el archivo XLS				

**TABLA XXIV.
 MODELADO FÍSICO DE LA TABLA
 DIMENSIÓN TIEMPO**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
key_tiempo	Int		No	Si	PK
fecha	smalldatetime		No	No	
mes	int		No	No	
año	int		No	No	
Origen Datos	Tabla generada por una consulta sql observada en la fig. 6				

**TABLA XXV.
 MODELADO FÍSICO DE LA TABLA DIMENSIÓN TIPO DE
 DOCUMENTO**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
key_tipo_documento	Int		No	Si	PK
descripcion	Varchar	200	No	No	
Origen Datos	Tabla de Venta en el archivo XLS				

**TABLA XXVI.
 MODELADO FÍSICO DE LA TABLA
 DIMENSIÓN VENDEDOR**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restricción
Key_vendedor	Int		No	Si	PK
Código_vendedor	Int		No	No	
Nombre_vendedor	Varchar	200	No	No	
Origen Datos	Tabla de Vendedor en el archivo XLS				

**TABLA XXVII.
HECHO VENTA**

Atributo	Tipo Dato	Longitud	Nulo	Autoincremental	Restriccion
key_hecho	Int		No	Si	PK
Key_artículo	Int		No	No	FK
key_vendedor	Int		No	No	FK
key_cliente	Int		No	No	FK
key_tiempo	Int		No	No	FK
key_tipo_documento	Int		No	No	FK
key_forma_pago	Int		No	No	FK
unidades_vendidas	Int		No	No	
monto_venta	numeric	10,2	No	No	
Origen Datos	Tabla de Ventas en el archivo XLS				

4.1.8. Iteración #8: ETL (Extraer, transformar y cargar)

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

Se realizó de la siguiente forma:

- La herramienta Microsoft SQL Server Management Studio como una herramienta de administración de base de datos.
- Con esta herramienta creamos la base de datos DM_venta_motos.

Por consiguiente, realizamos las siguientes conexiones:

- Conectamos el origen de datos denominado origen_datos_2017 es un archivo en formato XLS.
- Conectamos el destino denominado DESKTOP-4G1TUJJ\SQLEXPRESS.DM_ventas_motos
- Se realiza el proceso de carga de los datos desde la base de datos fuente transaccional hacia la base de datos recientemente creada.
- En la Figura 10 y 11. Se visualizan los flujos de control y las conexiones al origen y al destino.

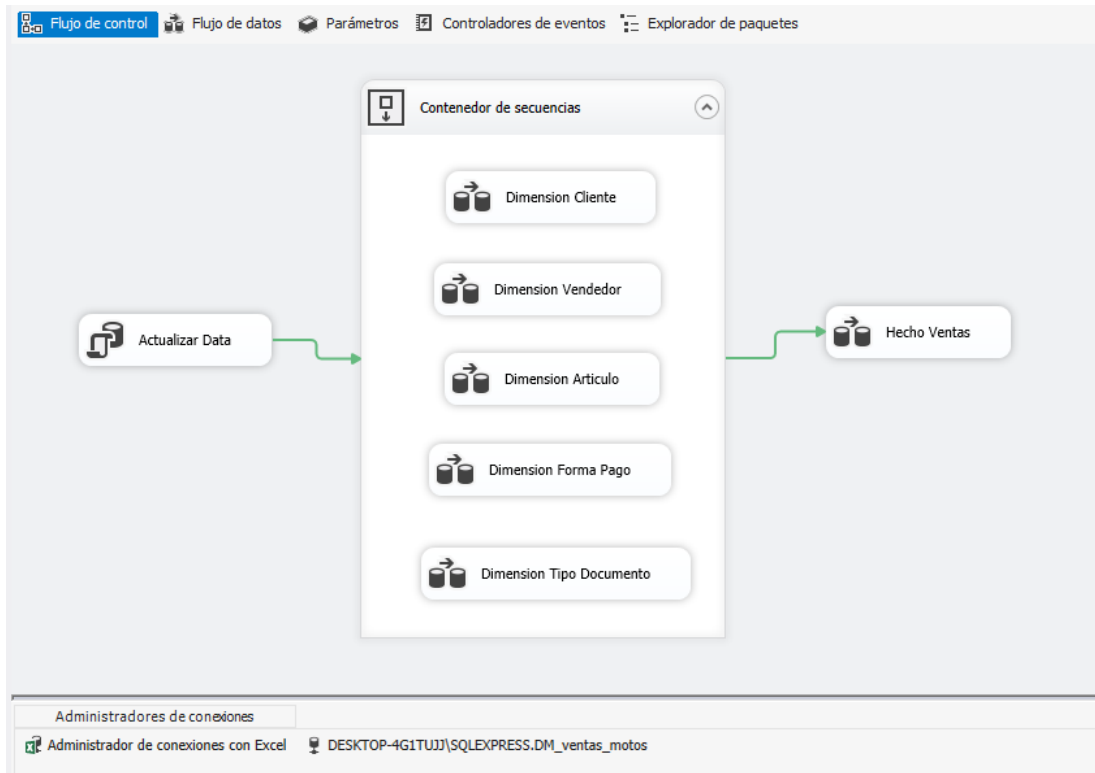


Fig. 10 Conexión al origen y destino

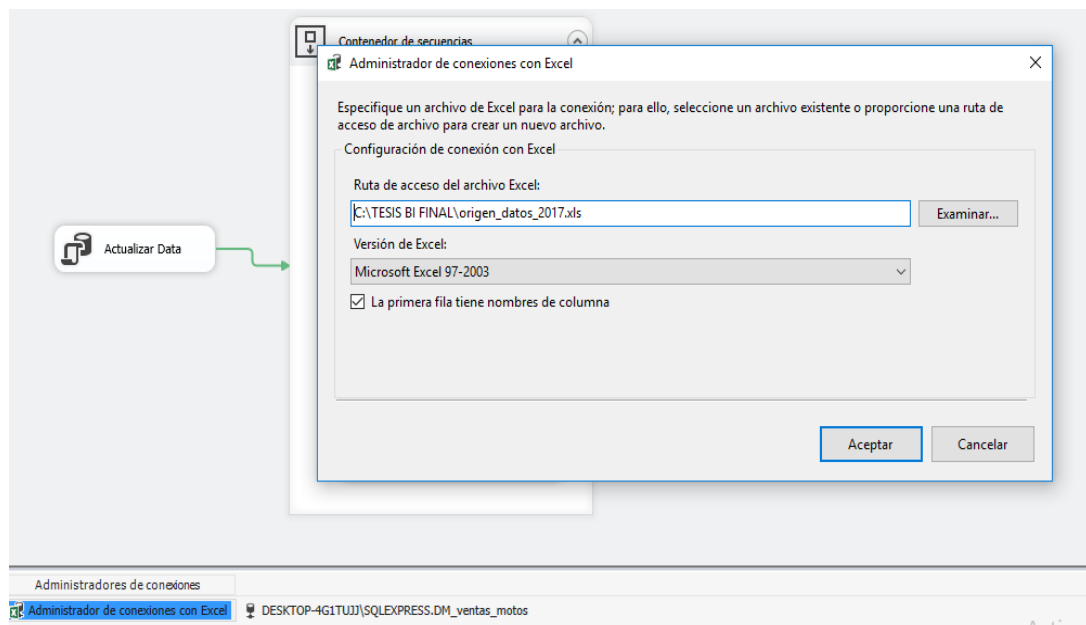


Fig. 11 Conexión al origen y destino

➤ Poblamiento de la dimensión Vendedor:

Para llevar a cabo el proceso ETL de los productos se utilizó un origen de Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento en XLS. La información necesaria se extrajo a través de los nombres de columnas que en la hoja de Excel estaban por ejemplo el código y nombre del vendedor.

La figura 12 muestra el proceso de ETL para la dimensión Vendedor

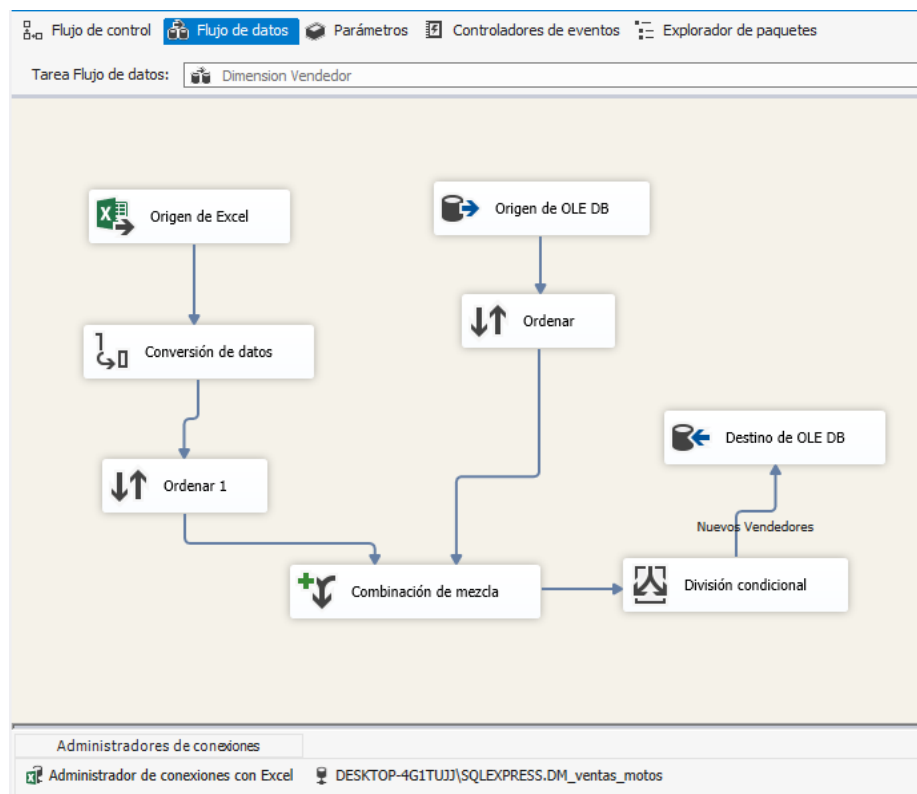


Fig. 12 Flujo de datos Vendedor

➤ Poblamiento de la dimensión Artículo:

Para llevar a cabo el proceso ETL de los productos se utilizó un origen de Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento de hoja de cálculo, exactamente de la Tabla Artículo. La información necesaria se extrajo a través de los nombres de columnas que en la hoja de Excel estaban por ejemplo el código y nombre del artículo.

La figura 13 muestra el proceso de ETL para la dimensión artículo.

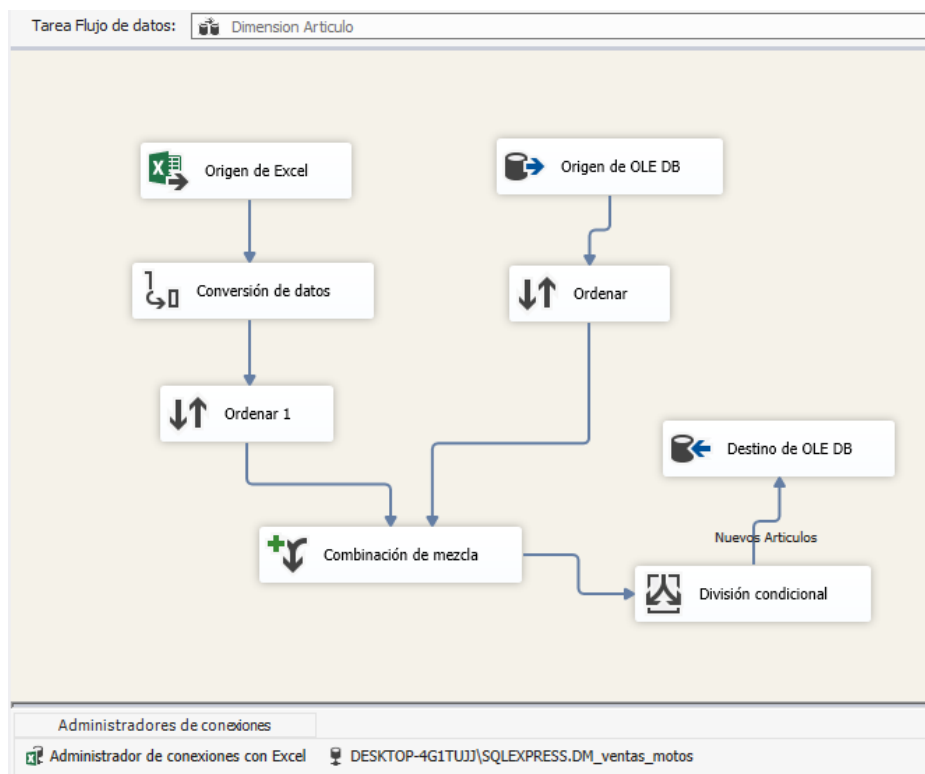


Fig. 13 Flujo de datos Artículo

➤ **Poblamiento de la dimensión Tipo de Documento:**

Para llevar a cabo el proceso ETL del tipo de documento se utilizó un origen Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento en formato XLS específicamente se trabajó con la tabla ventas posteriormente se realizó el siguiente procedimiento:

- Se convirtieron los datos de entrada de la tabla ventas ya que no eran compatibles con los datos del destino. La información necesaria se extrajo a través de los nombres de columnas que en la hoja de Excel estaban por ejemplo el tipo de documento.
- Luego que los datos hayan sido compatibles y ordenados, se procedía con la combinación de mezcla usando una función de left join (Combinación externa izquierda).
- Finalmente los datos limpios pasaban al Destino OLE DB.

La figura 14 muestra el proceso de ETL para la dimensión tipo documento.

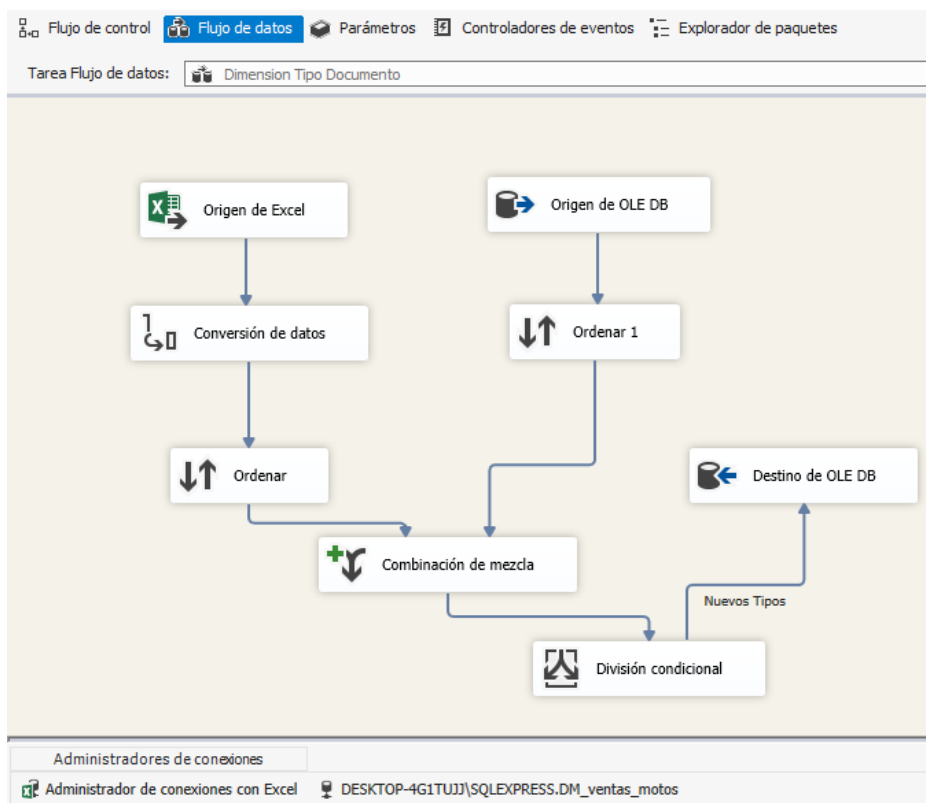


Fig. 14 Flujo de datos Tipo Documento

➤ Poblamiento de la dimensión Forma de Pago:
 Para llevar a cabo el proceso ETL del tipo de documento se utilizó un origen de Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento de hoja de cálculo, específicamente se trabajó con la tabla ventas posteriormente se realizó el siguiente procedimiento:

- Se convirtieron los datos de entrada de la tabla ventas ya que no eran compatibles con los datos del destino. La información necesaria se extrajo a través de los nombres de columnas que en la hoja de Excel estaban en este caso forma de pago.
- Luego que los datos hayan sido compatibles y ordenados, se procedía con la combinación de mezcla usando una función de left join (Combinación externa izquierda).
- Finalmente por medio de la división condicional solo los datos limpios pasaban al Destino OLE DB.

La figura 15 muestra el proceso de ETL para la dimensión forma pago.

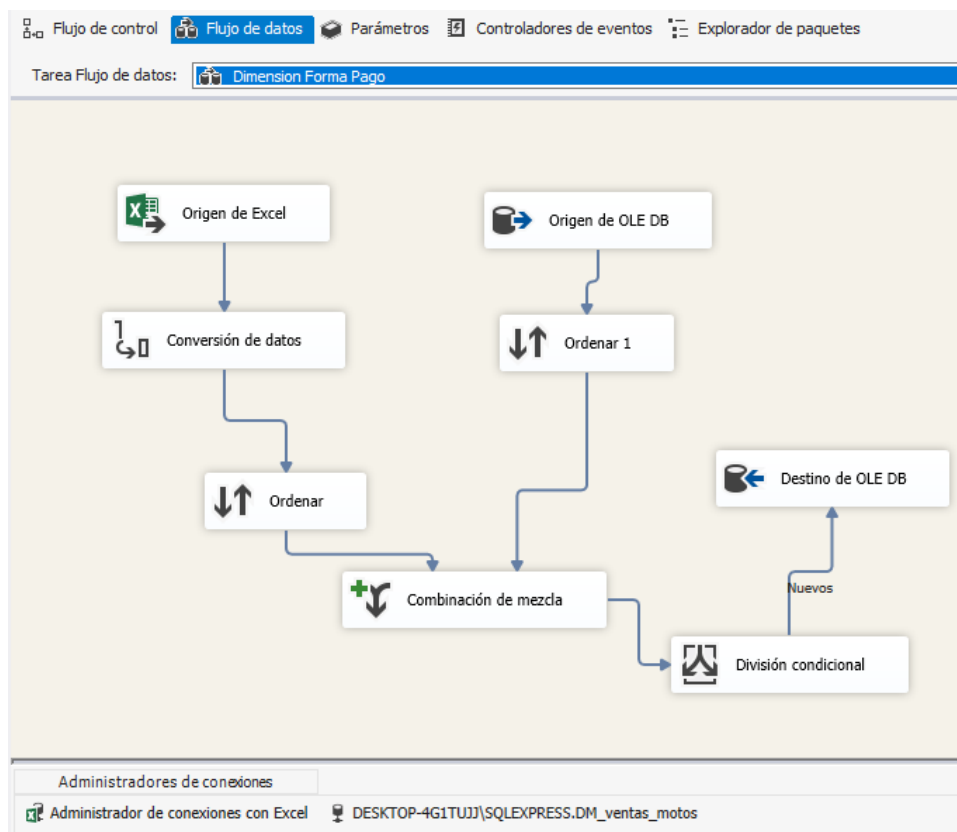


Fig. 15 Flujo de datos Forma de Pago

➤ **Poblamiento de la dimensión Cliente:**

Para llevar a cabo el proceso ETL del tipo de documento se utilizó un origen de Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento con hojas de cálculo específicamente de la tabla cliente posteriormente se realizó el siguiente procedimiento:

- Se convirtieron los datos de entrada de la tabla ventas ya que no eran compatibles con los datos del destino. La información necesaria se extrajo a través de los nombres de columnas que en la hoja de Excel estaban en este caso nombre cliente.
- Luego que los datos hayan sido compatibles y ordenados, se procedía con la combinación de mezcla usando una función de left join (Combinación externa izquierda).
- Finalmente, por medio de la división condicional solo los datos limpios pasaban al Destino OLE DB.

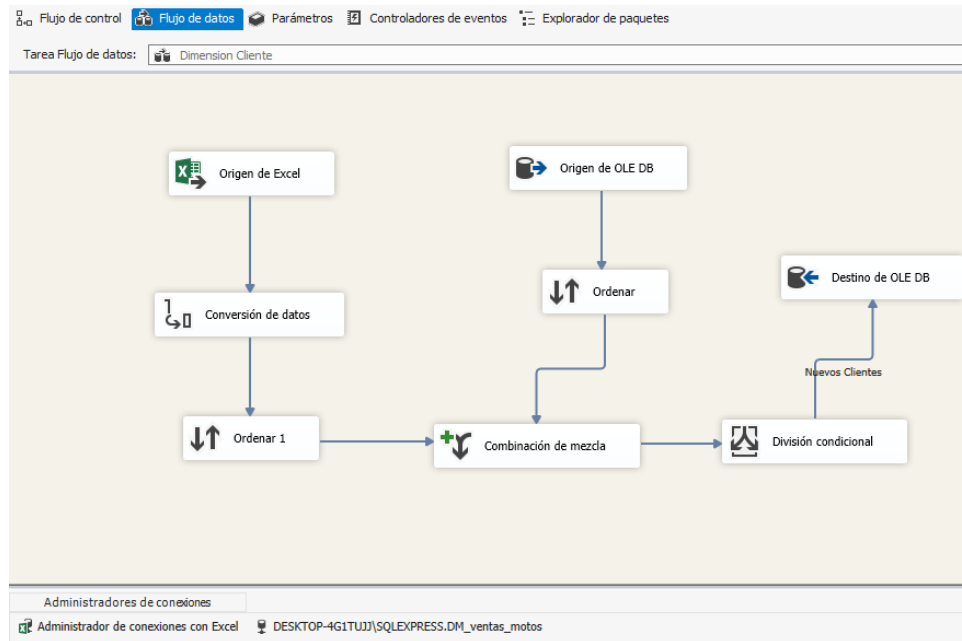


Fig. 16 Flujo de datos Cliente

➤ Poblamiento de la dimensión Tiempo:

Se extrajo la información por medio de una consulta acorde con la fecha inicial y final según la tabla de Ventas en el XLS.

En la Fig. 17 observamos la consulta que autogenera la dimensión Tiempo en el DataMart

```
alter table DTiempo add unique (fecha) declare @f1 smalldatetime='01/01/2014' declare @f2 smalldatetime='30/06/2018'
while @f1<=@f2 begin insert into DTiempo values (@f1,MONTH(@f1), datepart(quarter,@f1),
(case when MONTH(@f1) <7 then 1 else 2 end), YEAR(@f1)) set @f1=@f1+1 end

select * from DTiempo
```

Fig. 17 Consulta para el poblamiento de la dimensión Tiempo

➤ Poblamiento del Hecho de ventas:

Finalmente, cuando se haya terminado la población de las dimensiones se empieza a poblar la dimensión Hecho.

Para llevar a cabo el proceso ETL del Hecho se utilizó un origen de Excel lo cual se extrajo la información necesaria desde una base de datos en este caso un documento en formato XLS, específicamente se trabajó con la tabla ventas posteriormente se realizó el siguiente procedimiento:

- Se convirtieron los datos de entrada de la tabla ventas ya que deben de ser compatibles con los campos de las tablas destino.
- Se buscará el key de cada una de las dimensiones (en este caso por id artículo, tipo de documento, forma de pago, nombre cliente, nombre vendedor, fecha, cantidad y monto)
- Luego estos datos se almacenarán en la tabla Hecho como clave foránea, el principal propósito de esto es que los datos que espera la tabla Hecho sean limpios e iguales en tipo de dato por eso se lleva a cabo este poblamiento que termina en el Destino OLE DB.

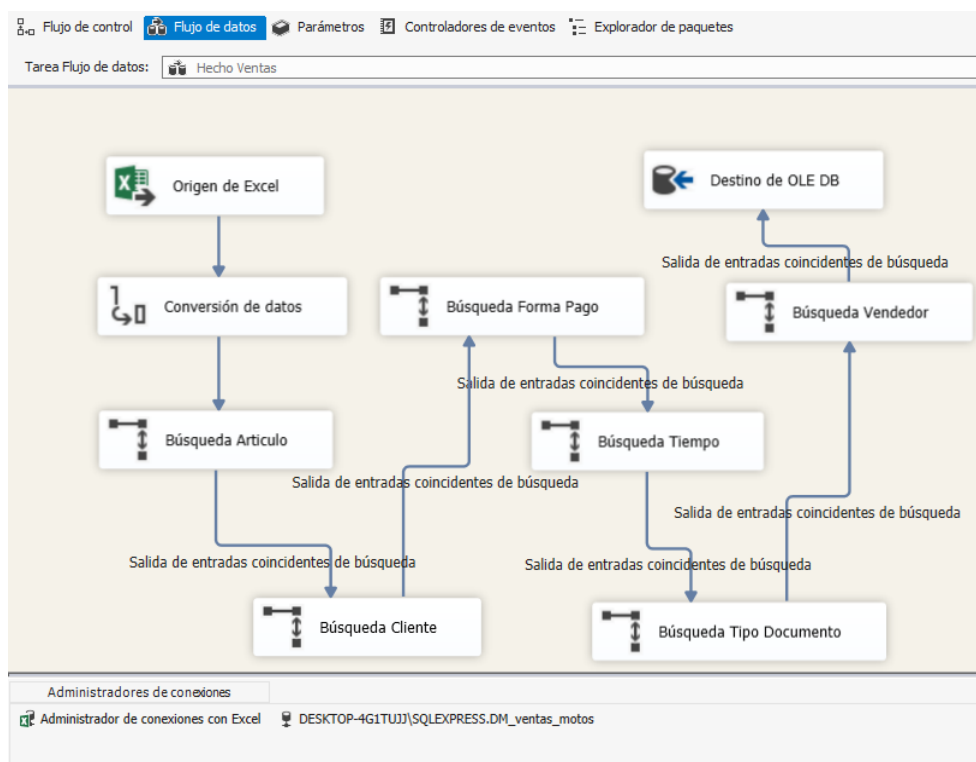


Fig. 18 Flujo de datos Hecho Ventas

➤ **Elaboración del cubo OLAP:**

Se utilizó SQL Server Analysis Services, se lleva a cabo después de haber concluido con el poblamiento del Data Mart por medio del proceso ETL. Por consiguiente, se genera una vista origen de datos:

En la Fig. 20 se muestra la vista de los datos denominada Vista_DM Ventas Motos la cual se generó de la conexión al Data Mart denominado DM_ventas_motos.

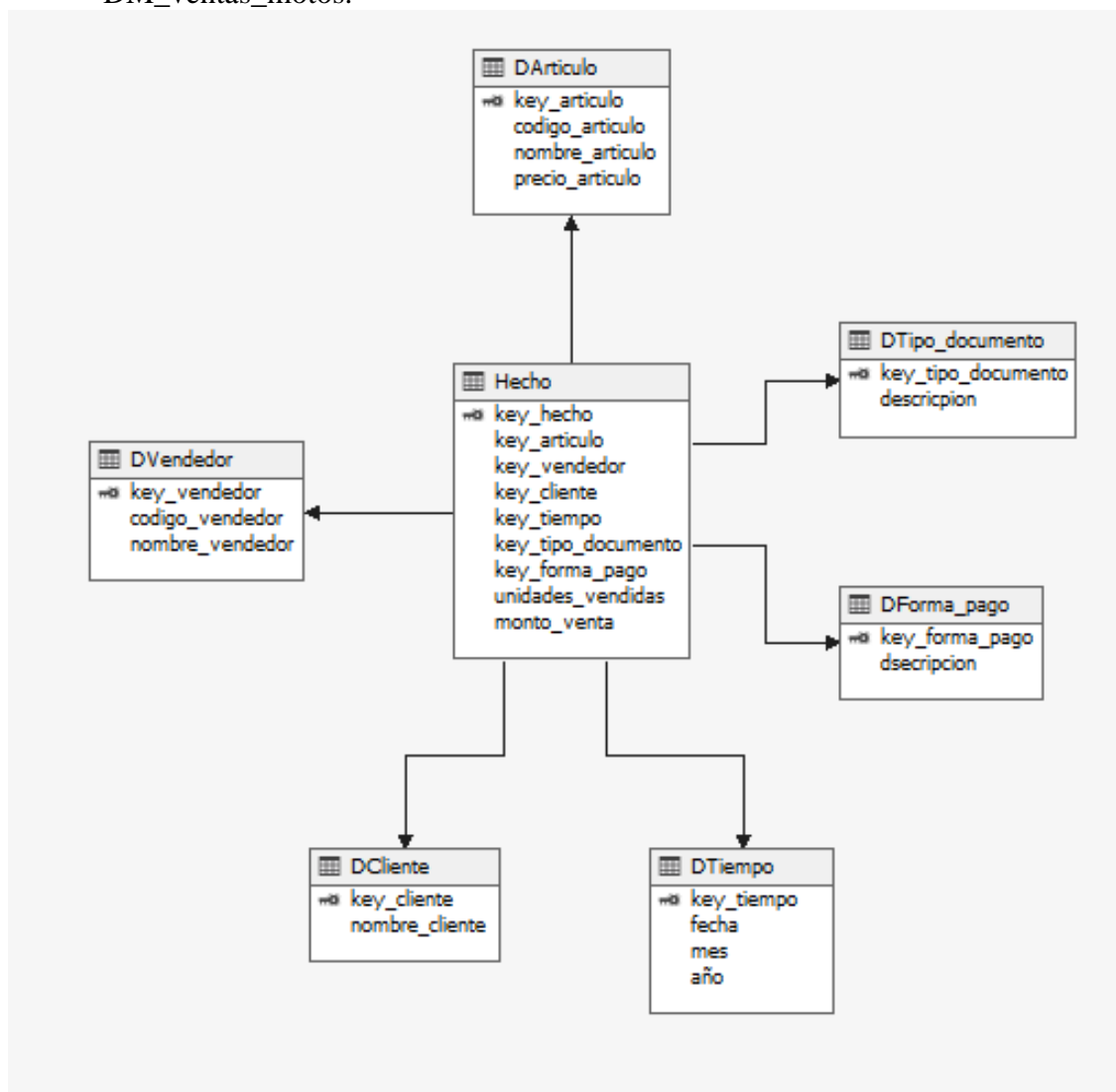


Fig. 19 Vista del origen de datos Vista_DM Venta Motos

Posteriormente de generar la vista se procede a crear las dimensiones jerarquizadas y sus unidades de medida. En este caso el cubo se llamara Cubo_DM Ventas Motos

Las figuras 20 y 21 muestran los pasos de este proceso y en la figura 22 se muestra el cubo ya implementado.

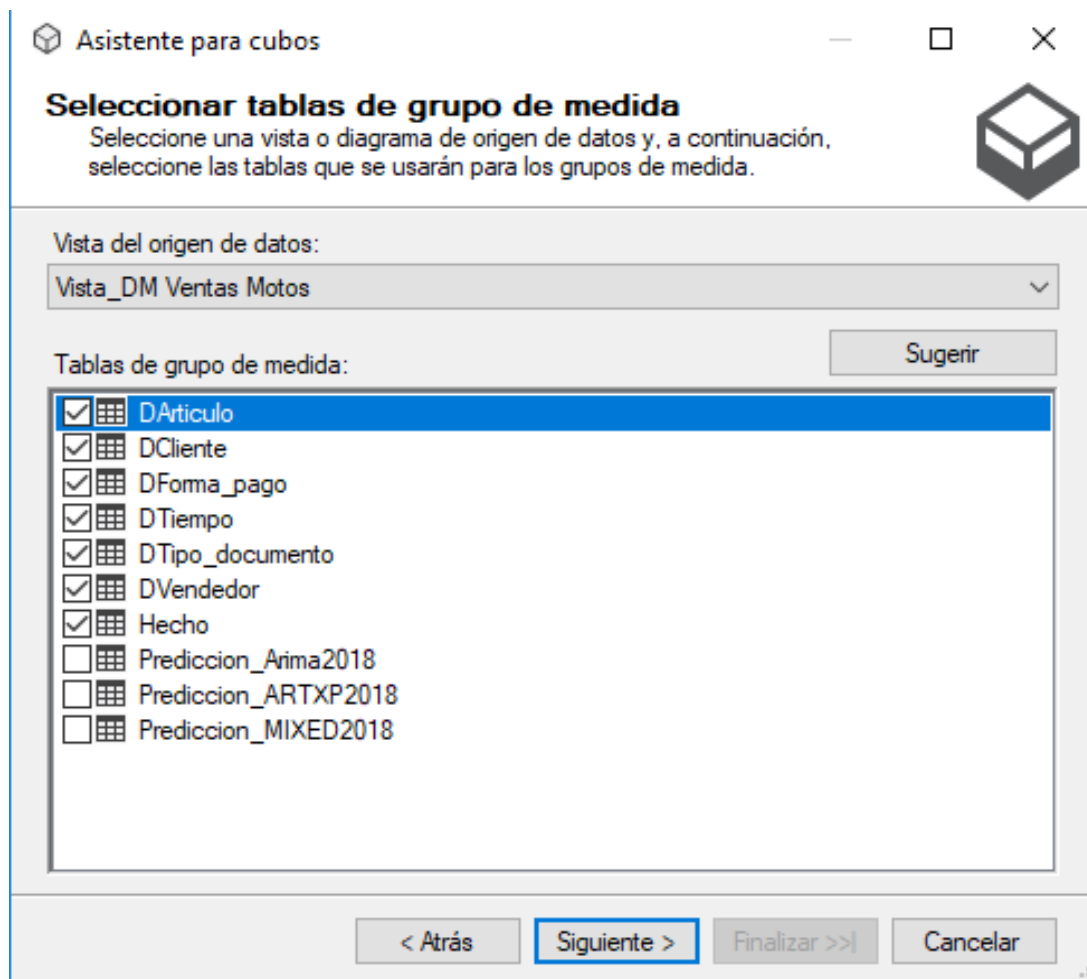


Fig. 20 Creacion del Cubo

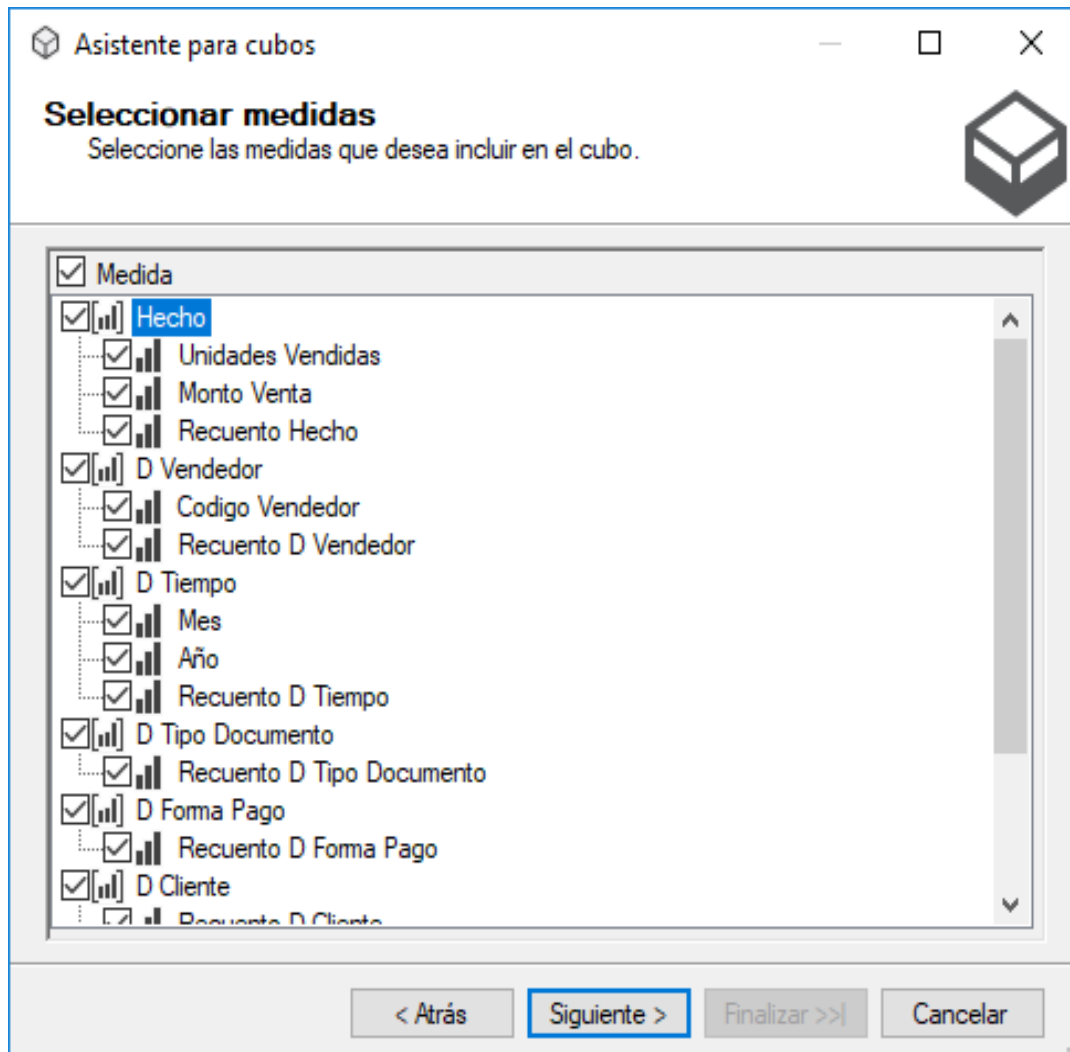


Fig. 21 Eleccion de Dimensiones y unidades de medidas para el CUBO

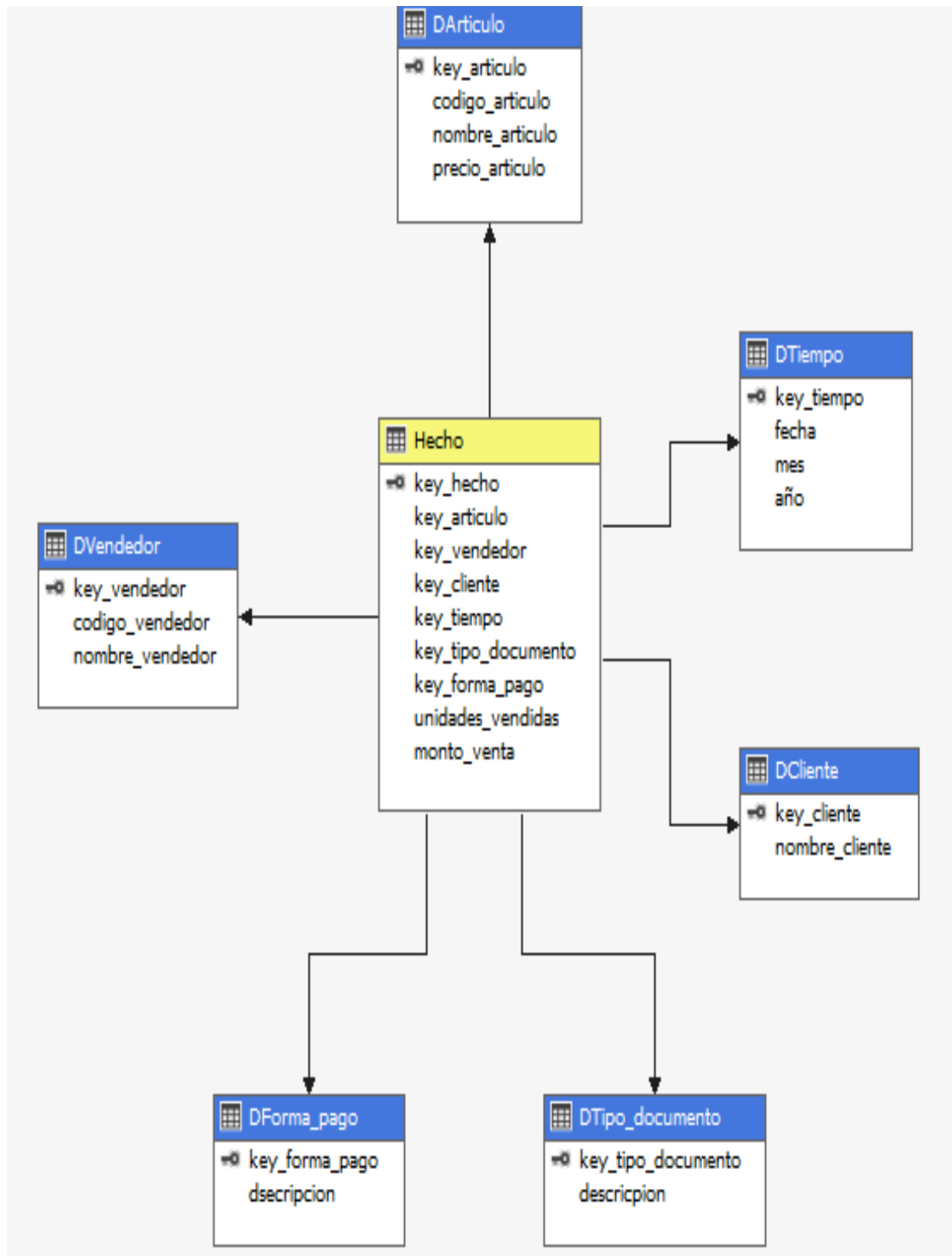


Fig. 22 Creación del cubo Cubo_DM Ventas Motos

4.1.9. Iteración #9: Diseño de la arquitectura técnica

En la siguiente figura se muestra la arquitectura que se ha realizado en el presente proyecto de investigación. El cual constituye el diseño de una aplicación con alto nivel, ya que se ha realizado el estudio de la estructura de la aplicación desde el punto de vista de componentes que interactúan entre sí.

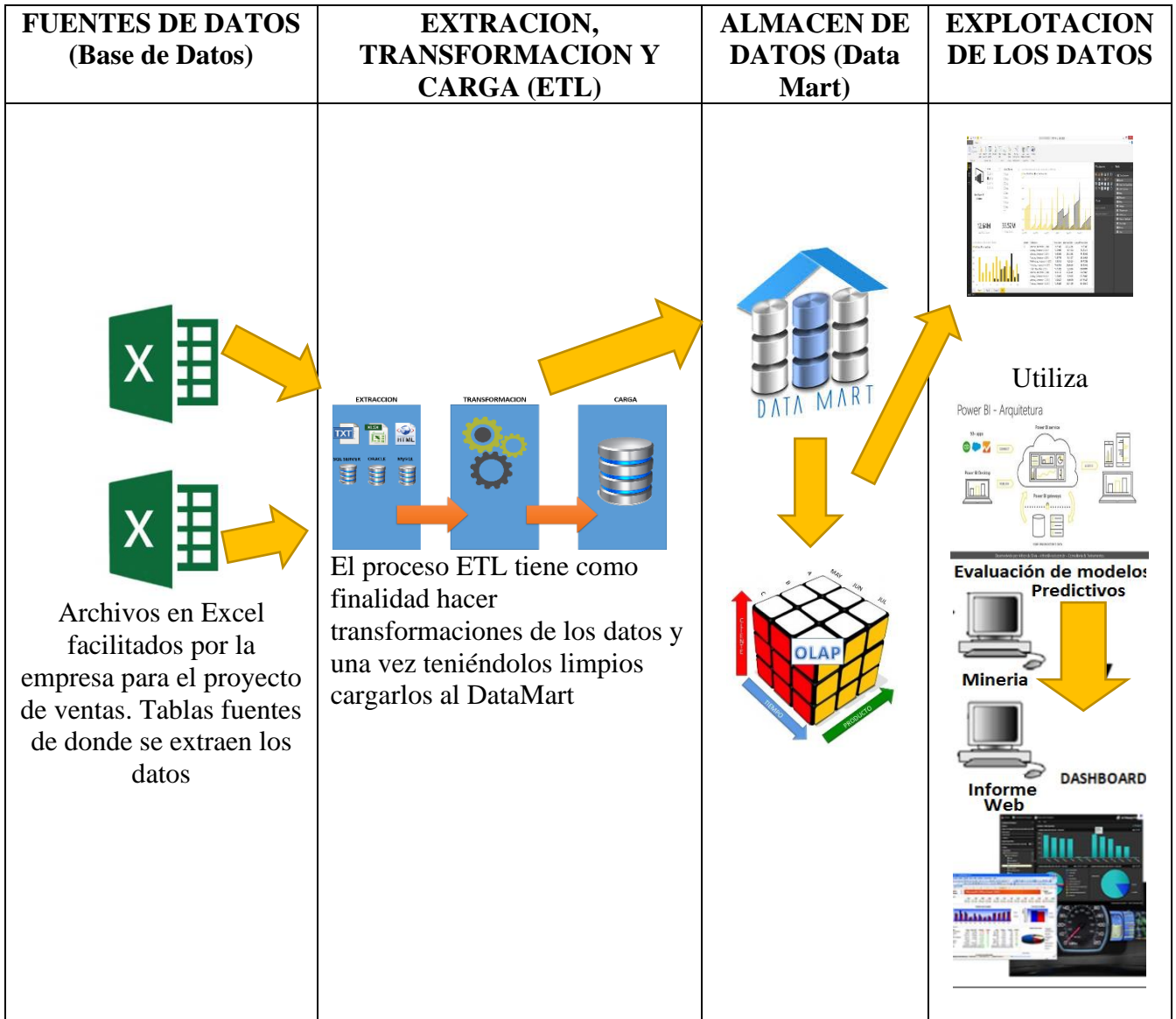


Fig. 23. Arquitectura del Proyecto

Ahora detallaremos cada componente de la Arquitectura del Proyecto:

- **Base de datos transaccional OLTP**

La base de datos transaccional que es usada en el proceso ETL es la base de datos del sistema transaccional de la empresa, cuyos datos son soportados por una base de datos de Origen XLS (Excel)

- **Proceso ETL+**

El proceso ETL (Extracción, Transformación y Carga) del datamart ha sido desarrollado usando la herramienta Microsoft SQL Server Integration Services 2012. Mediante esta herramienta se creó un paquete ETL para el modelo dimensional creado, para que de esta forma accedan a la fuente de datos mencionada anteriormente.

- **Datamart**

Mediante el uso de la herramienta administradora de base de datos Microsoft SQL Server 2012 se crea una base de datos que contiene las tablas que conforman la estructura del modelo dimensional diseñado. En esta base de datos se depositaran toda la data de la base de datos XLS (Excel) los cuales serán cargados mediante el proceso ETL

- **Cubo OLAP**

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

- **Proceso Explotación de la Información**

Para la explotación de datos primero vamos a utilizar la herramienta Analisis Service para el modelado de series temporales de Microsoft este algoritmo nos permitirá proyectar las ventas en lapsos de tiempo. Con respecto al análisis de datos se crearan Dashboard y reportes según los requerimientos con la herramienta de Power BI desktop y Web.

4.1.10. Iteración #10: Implementación

Al utilizar una metodología híbrida, en que la integro la metodología de Ralph Kimbal - Bottom-Down (para el DataWarehouse) y la metodología CRIPS (para la minera), en esta parte de la metodología se desarrolla la fase de Modelado de la Minería. Para poder modelar la minería en Microsoft Analysis Services se crea una vista con los datos preparados, y a partir de ello seleccionar nuestros datos de aprendizaje, tal como se aprecia en la Fig. 28.

Con respecto a la minería de datos la cual sería la proyección de ventas. Primero debemos de tener en cuenta los datos que necesitamos y que estos formen parte del modelo del algoritmo de series temporales.

En este caso el pronóstico se realizara para las ventas por artículo en los diferentes años, se realizó la consulta a diferentes tablas (DArticulo, Hecho y DTiempo).

Se considerara el año, mes, nombre_artículo y unidades_vendidas por cada producto

```

select t.año, t.mes, a.nombre_articulo, sum(h.unidades_vendidas) as unidades_vendida
from DArticulo a
inner join Hecho h on a.key_articulo=h.key_articulo
inner join DTiempo t on t.key_tiempo=h.key_tiempo
where t.año = 2017
group by a.nombre_articulo,t.año,t.mes
order by 2

```

	año	mes	nombre_articulo	unidades_vendidas
1	2017	1	MOTOCICLETA CB110	3
2	2017	1	MOTOCICLETA CB125F	2
3	2017	1	MOTOCICLETA CB160F	7
4	2017	1	MOTOCICLETA CB190R	19
5	2017	1	MOTOCICLETA CCG125	5
6	2017	1	MOTOCICLETA CRF250LF	1
7	2017	1	MOTOCICLETA DIO	1
8	2017	1	MOTOCICLETA EG 1000 TIPO S	1
9	2017	1	MOTOCICLETA EG5000CX	1
10	2017	1	MOTOCICLETA GL125	18
11	2017	1	MOTOCICLETA GL150	1
12	2017	1	MOTOCICLETA HJ / 4 / 13 C V	1
13	2017	1	MOTOCICLETA HJ / 4 / 9 C	1
14	2017	1	MOTOCICLETA HJ/4/13 C	3
15	2017	1	MOTOCICLETA UMK435	1
16	2017	1	MOTOCICLETA WAVE 110	1
17	2017	1	MOTOCICLETA WB30XHDR	7
18	2017	1	MOTOCICLETA XR150L	18
19	2017	1	MOTOCICLETA XR190L	8

Consulta ejecutada correctamente. | DESKTOP-4G1TUJJ\SQLEXPRESS ... | DESKTOP-4G1TUJJ\LUF

Fig. 24. Consulta para pronóstico

Luego de tener la consulta que se implementara como parte del modelo y así alimentar el algoritmo se creó una tabla temporal con la misma estructura de datos que necesitaremos: año, mes, artículo y unidades_vendidas.

En la Fig. 25 podemos observar la tabla temporal creada como tp_artículos.

```

create table tp_articulos
(
  articulo varchar(100) not null,
  año int not null,
  mes int not null,
  unidades_vendidas int not null
)

select * from tp_articulos

```

	articulo	año	mes	unidades_vendidas
550	MOTOCICLETA GX270 T2...	2014	6	0
551	MOTOCICLETA GX270 T2...	2014	7	0
552	MOTOCICLETA GX270 T2...	2014	8	0
553	MOTOCICLETA GX270 T2...	2014	9	0
554	MOTOCICLETA GX270 T2...	2014	10	0
555	MOTOCICLETA GX270 T2...	2014	11	0
556	MOTOCICLETA GX270 T2...	2014	12	0
557	MOTOCICLETA GX270 T2...	2015	1	0
558	MOTOCICLETA GX270 T2...	2015	2	0
559	MOTOCICLETA GX270 T2...	2015	3	0
560	MOTOCICLETA GX270 T2...	2015	4	0
561	MOTOCICLETA GX270 T2...	2015	5	0
562	MOTOCICLETA GX270 T2...	2015	6	0
563	MOTOCICLETA GX270 T2...	2015	7	0
564	MOTOCICLETA GX270 T2...	2015	8	0
565	MOTOCICLETA GX270 T2...	2015	9	0
566	MOTOCICLETA GX270 T2...	2015	10	0
567	MOTOCICLETA GX270 T2...	2015	11	0
568	MOTOCICLETA GX270 T2...	2015	12	0

Consulta ejecutada correctamente. DESKTOP-4G1TU.

Fig. 25. Creacion de tabla temporal tp_artículos

Visualizando no todos los artículos tenían monto para todos los meses, lo cual no tenían registros en todos los meses de todos los años. Para esto se creó un cursor el cual llenó la tabla temporal tp_artículos con los artículos 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 26 muestra el primer cursor creado, en este caso llamado artículos

```

--CREACION DE UN CURSOR PARA COMPLETAR LAS COLUMNAS DE LOS MESES Q NO SE VENDIO ALGUN PRODUCTO ESPECIFICO EN SI
--Y EN LA COLUMNA NO SE MUESTRA
|
declare @articulo varchar(100), @año int, @mes int
| declare articulos cursor for
| select distinct a.nombre_articulo
| from DArticulo a
| inner join Hecho h on a.key_articulo=h.key_articulo
| open articulos;
| fetch next from articulos into @articulo
| set @mes=1
| while @@FETCH_STATUS=0
| begin
|     set @año=2014
|     while @año<=2017
|     begin
|         set @mes=1
|         while @mes<=12
|         begin
|             insert into tp_articulos values(@articulo,@año,@mes,0)
|             set @mes=@mes+1
|         end
|         set @año=@año+1
|     end
|     fetch next from articulos into @articulo
| end
| close articulos;
| deallocate articulos;

```

Fig. 26. Creacion del cursor denominado artículos

Seguidamente se creó otro cursor para actualizar el monto que el primer cursor le registró a todos los artículos con el valor “0” por el monto real de cada uno de los artículos. De esta manera siempre que un artículo no tenga monto en un mes determinado vaya una fila pero con valor “0”.

Como se aprecia en la fig. 27 se muestra el segundo cursor creado, en este caso llamado artículos2.

```

declare @articulo_ varchar(100), @año_ int, @mes_ int, @unidades_vendidas_ int
| declare articulos2 cursor for
| select t.año, t.mes, a.nombre_articulo, sum(h.unidades_vendidas) as unidades_vendidas
| from DArticulo a
| inner join Hecho h on a.key_articulo=h.key_articulo
| inner join DTiempo t on t.key_tiempo=h.key_tiempo
| group by a.nombre_articulo, t.año, t.mes order by 1,2,3;
| open articulos2;
| fetch next from articulos2 into @año_,@mes_,@articulo_,@unidades_vendidas_
| while @@FETCH_STATUS=0
| begin
|     update tp_articulos set unidades_vendidas=@unidades_vendidas_
|     where articulo=@articulo_ and año=@año_ and mes=@mes_
|     fetch next from articulos2 into @año_, @mes_, @articulo_, @unidades_vendidas_
| end
| close articulos2
| deallocate articulos2;

```

Fig. 27. Creacion del segundo cursor denominado artículos2

Continuando con el proceso del modelado de datos para alimentar el algoritmo se genero una vista por lo cual se realizo una consulta respectiva de la tabla temporal que se poblo.

La Fig. 28 se visualiza la vista creada.

```
-- CREACION DE LA VISTA SEGUN LA SITAXIS QUE SERA TOMADA POR EL ALGORITMO DE SERIES TEMPORALES DE MICROSOFT
create view v_tp_articulos as
select cast('1/'+right('00'+cast(mes as varchar),2)+'/'+cast(año as char(4)) as date) as fecha, articulo, unidades_vendidas from tp_articulos

select * from v_tp_articulos where year(fecha) = 2017 order by 2,1
```

Fig. 28 Creacion de la vista denominada v_tp_artículos

Ahora que tenemos todos los datos moldeados para alimentar al algoritmo debemos de crear la estructura de minería de datos y como primer punto debemos de generar una vista de acuerdo a la coneccion de origen que es DM_Ventas Motos.ds

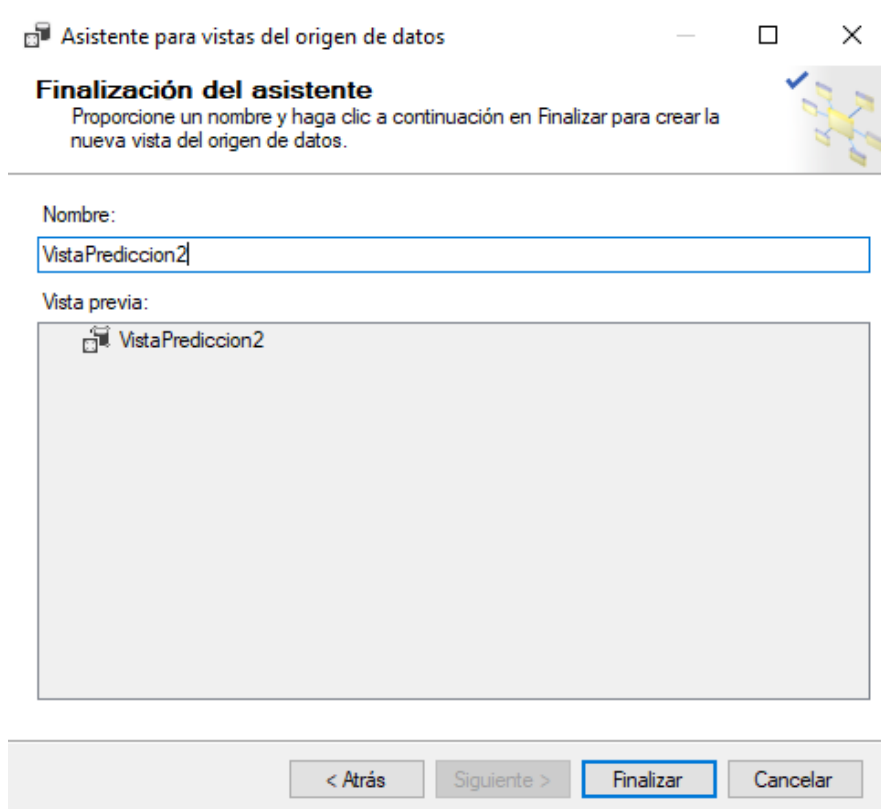


Fig. 29. Creación de la vista denominada VistaPrediccion por medio del origen de datos DM_Ventas Motos.ds

Como segundo punto debemos de crear la estructura de minería de datos con respecto al algoritmo que utilizaremos en este caso series temporales.

En la Fig. 30. Observamos la elección de la técnica.

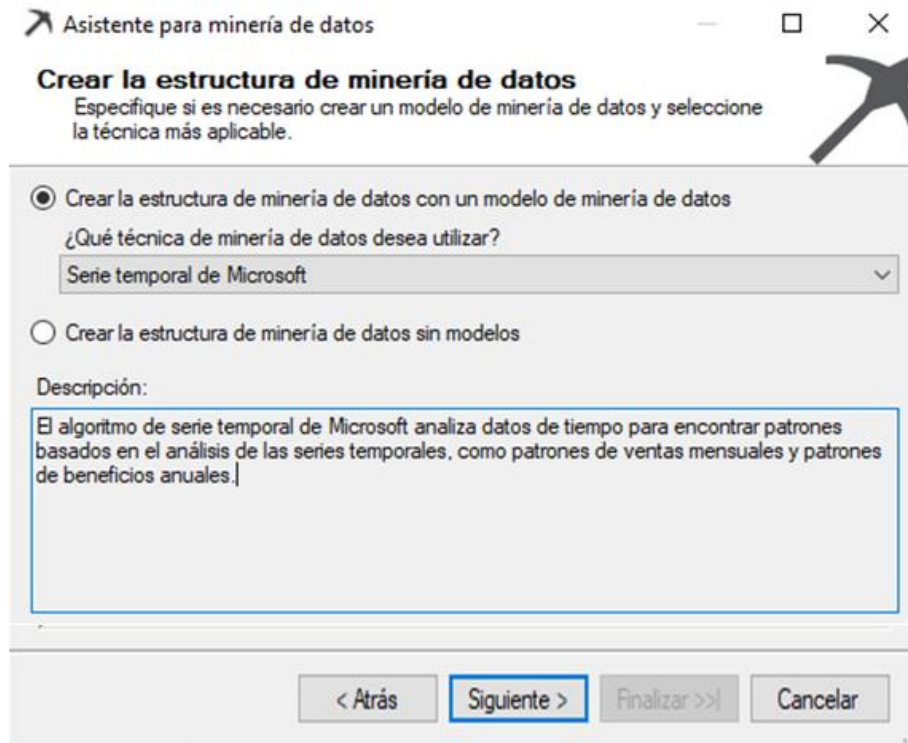


Fig. 30 Selección de la tecnica de minería de datos

En la Fig. 31 veremos que seleccionamos la vista que creamos de acuerdo al origen de datos DM_Ventas Motos

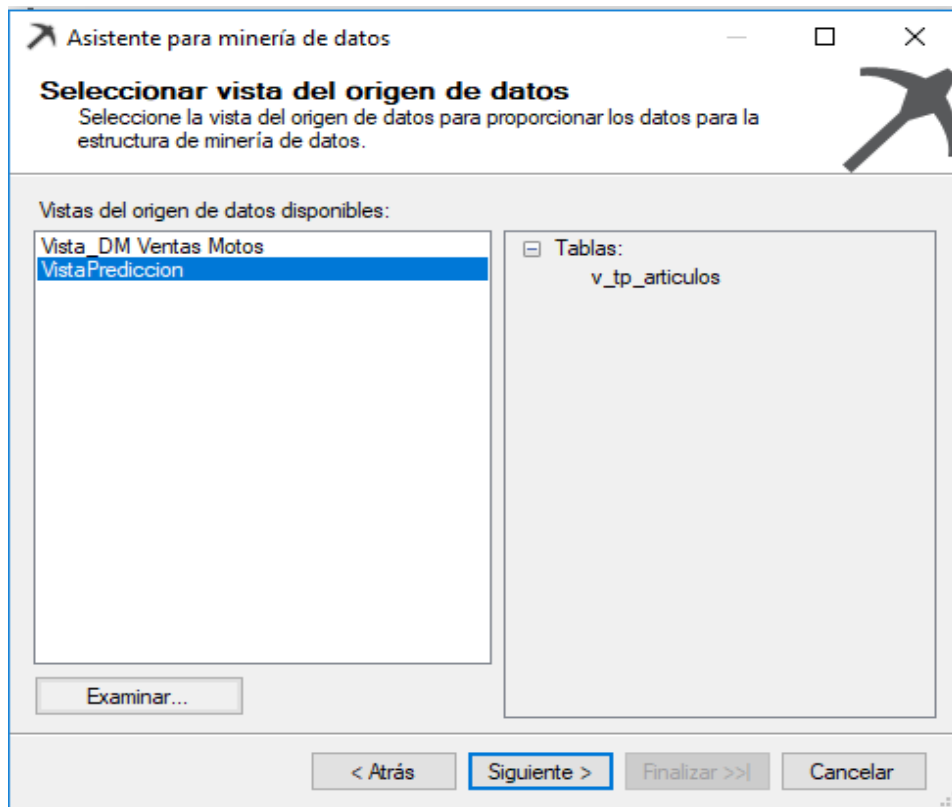


Fig. 31 Selección de la VistaPrediccion

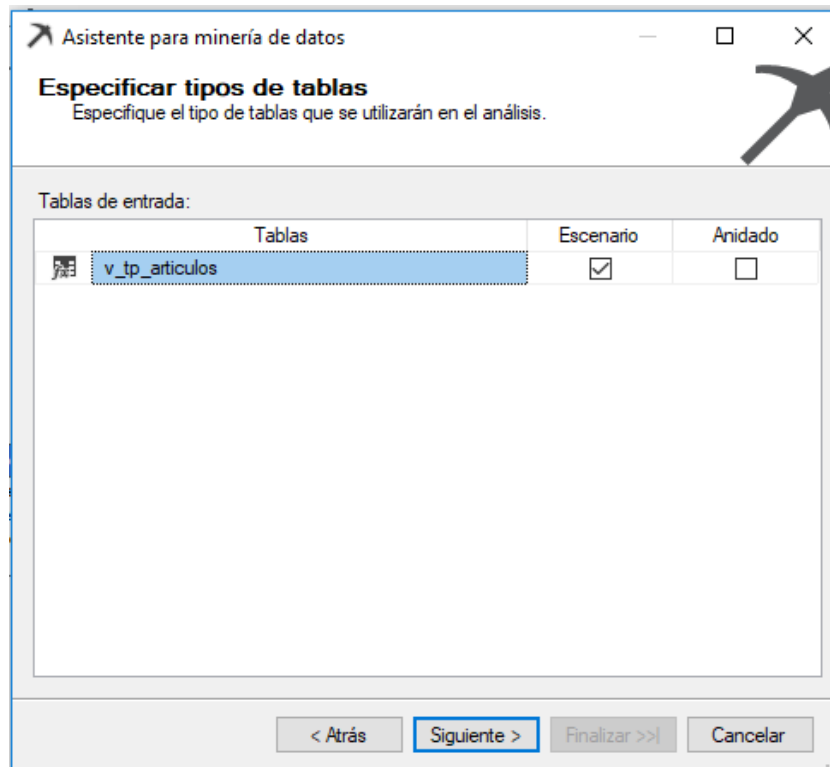


Fig. 32 Selección de la tabla modelada para el algoritmo de series temporales

A continuación escogeremos los datos de aprendizaje.

La Fig. 33 muestra la selección de los artículos, fechas como clave y unidades vendidas como entrada y predicción.

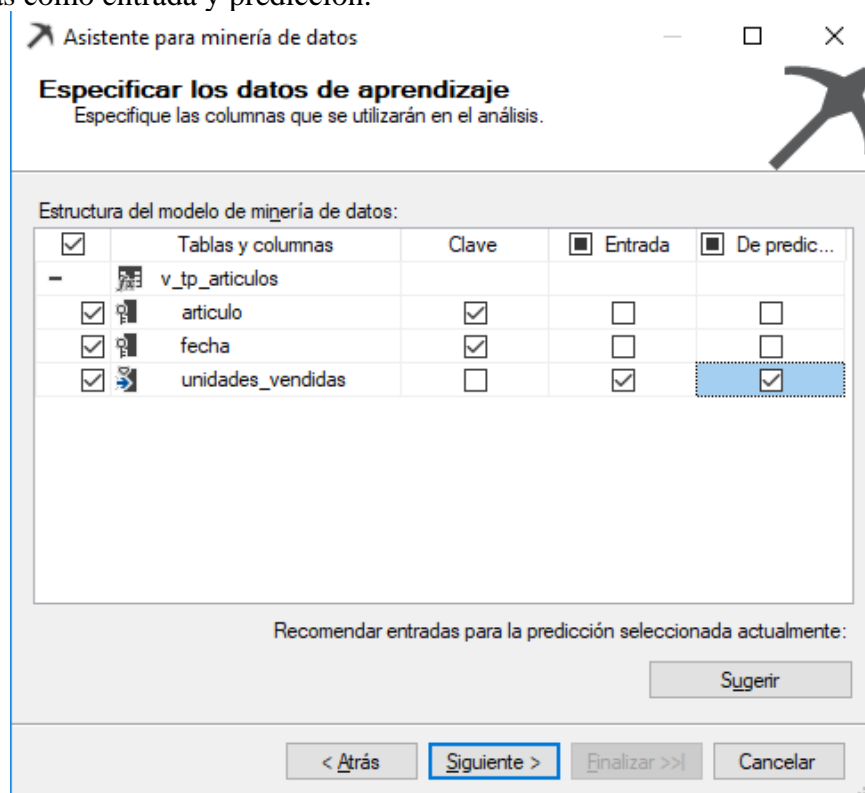


Fig. 33 Selección de datos de aprendizaje

Luego automáticamente se asignan tipos de datos a los datos que el algoritmo necesita.

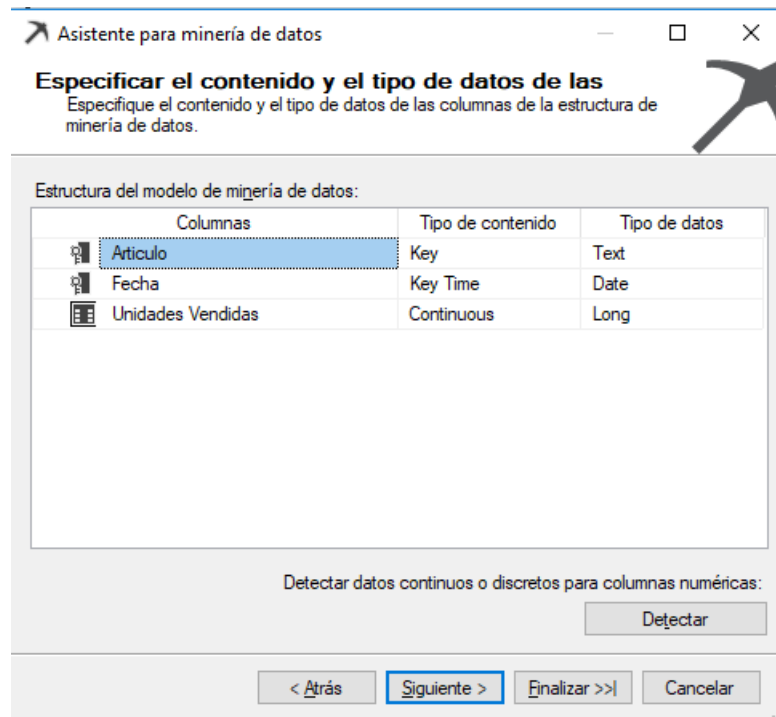


Fig. 34 Especificación de tipo de dato y contenido

Por último, se le brinda un nombre a la estructura y al modelo de minería de datos. La Fig. 35 muestra la culminación de la estructura de minería de datos.

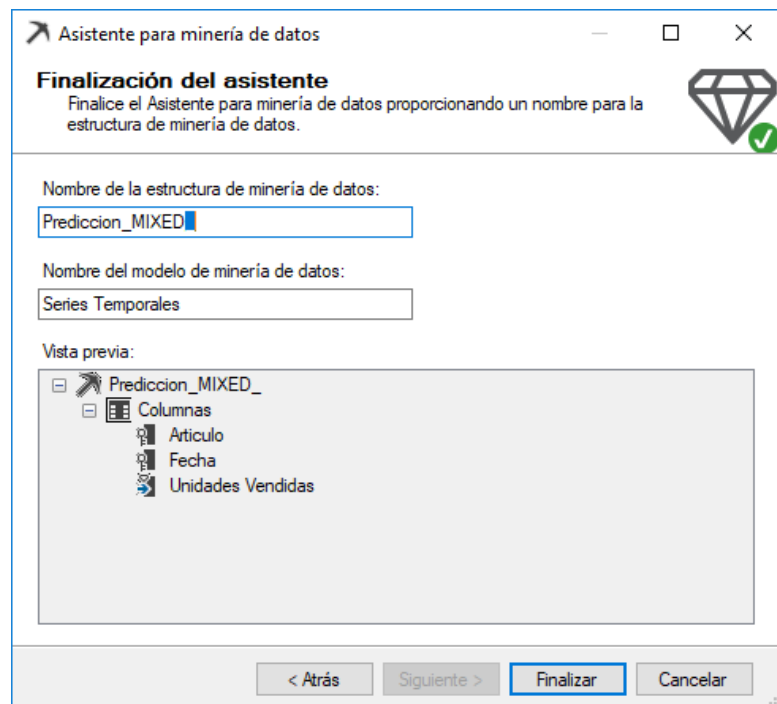


Fig. 35 Finalización de la estructura de minería de datos

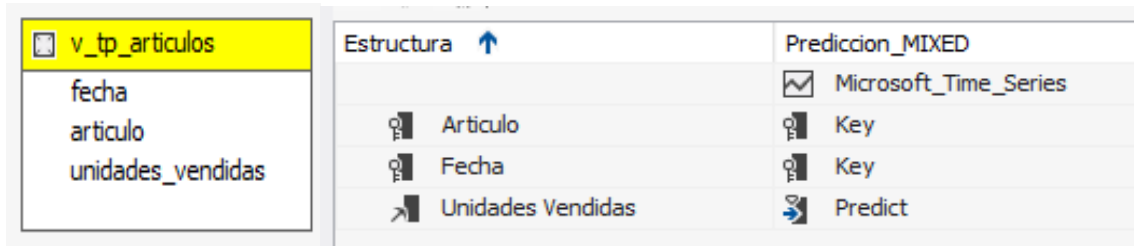


Fig. 36 Estructura de Minería de Datos y Modelado de Minería de Datos

Por consiguiente procedemos a integrarle parámetros al algoritmo ya que se va añadir en FORECAST_METHOD el valor de MIXED en este parámetro podemos elegir el modelo de serie temporal que utilizaremos y por otro lado en el parámetro PERIODICITY_HINT se le añade {12} ya que la predicción la haremos anualmente por 12 meses.

En la Fig. 37 observaremos los parámetros del algoritmo y los que utilizaremos.

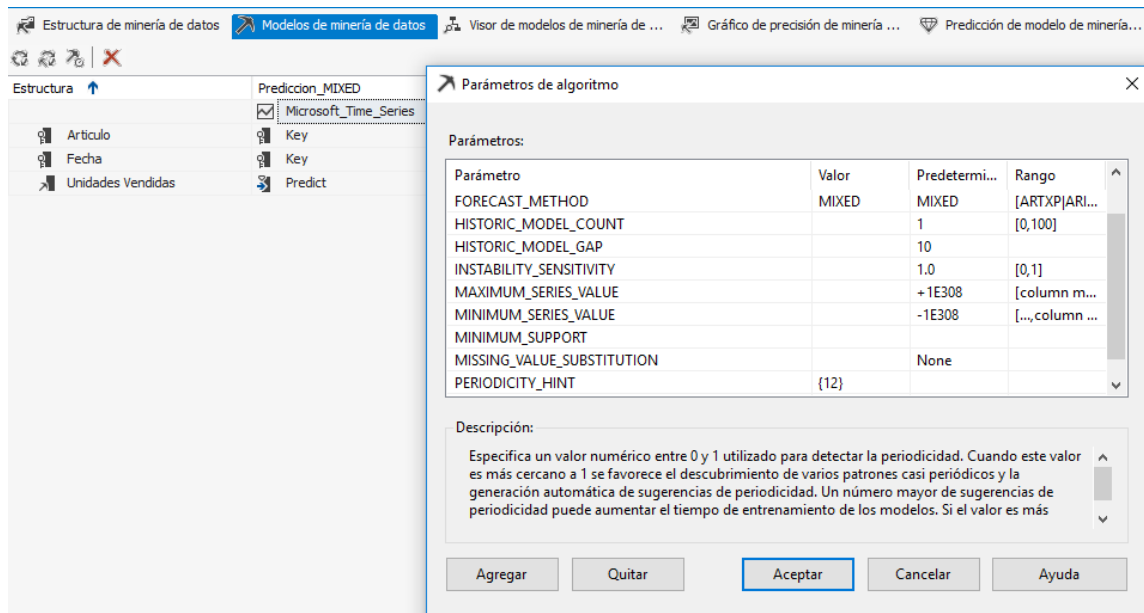


Fig. 37 Parámetros para el algoritmo con el metodo MIXED

A continuación veremos la grafica que nos brinda el modelo de minería de datos MIXED de los artículos mas vendidos en el transcurso de los años iniciando desde el 2014.

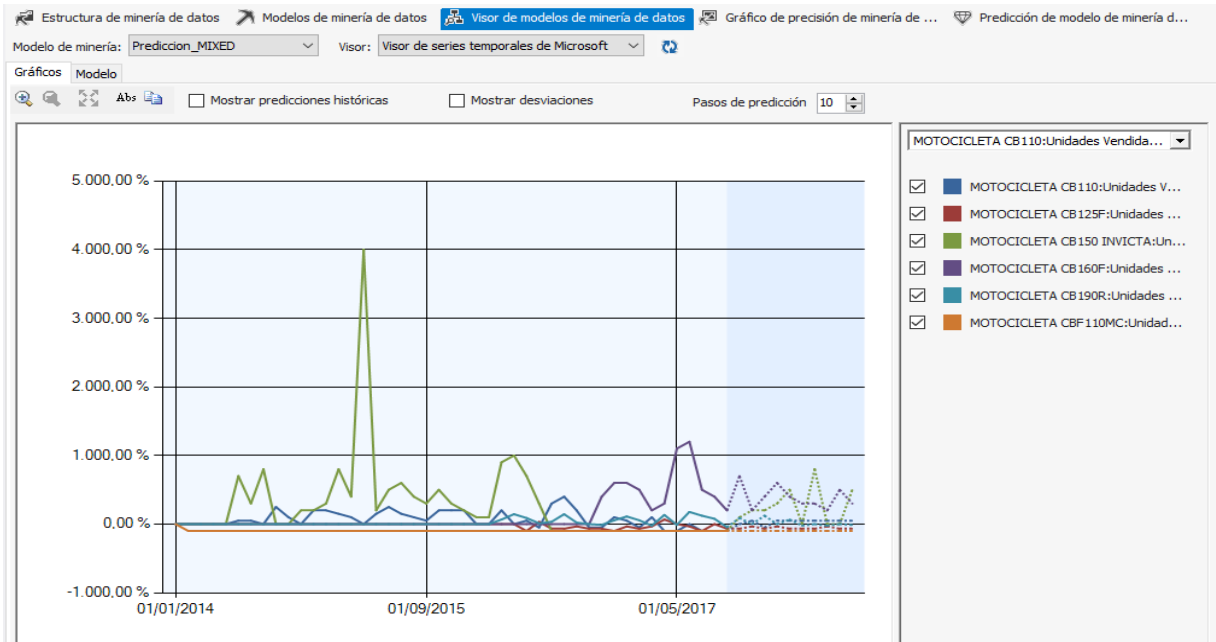


Fig. 38 Visor del Modelo de minería de datos MIXED

Ahora se evaluará un modelo con el método ARTXP para su posterior fiabilidad y ver cual se adecua o tiene menos margen de error.

Procedemos a integrarle parametros al algoritmo ya que se va añadir en FORESCAST_METHOD el valor de ARTXP en este parametro podemos elegir el modelo de serie temporal que utilizaremos y por otro lado en el parametro PERIODICITY_HINT se le añade {12} ya que la prediccion la haremos anualmente por 12 meses.

En la Fig. 39 observaremos los parametros del algoritmo y los que utilizaremos.

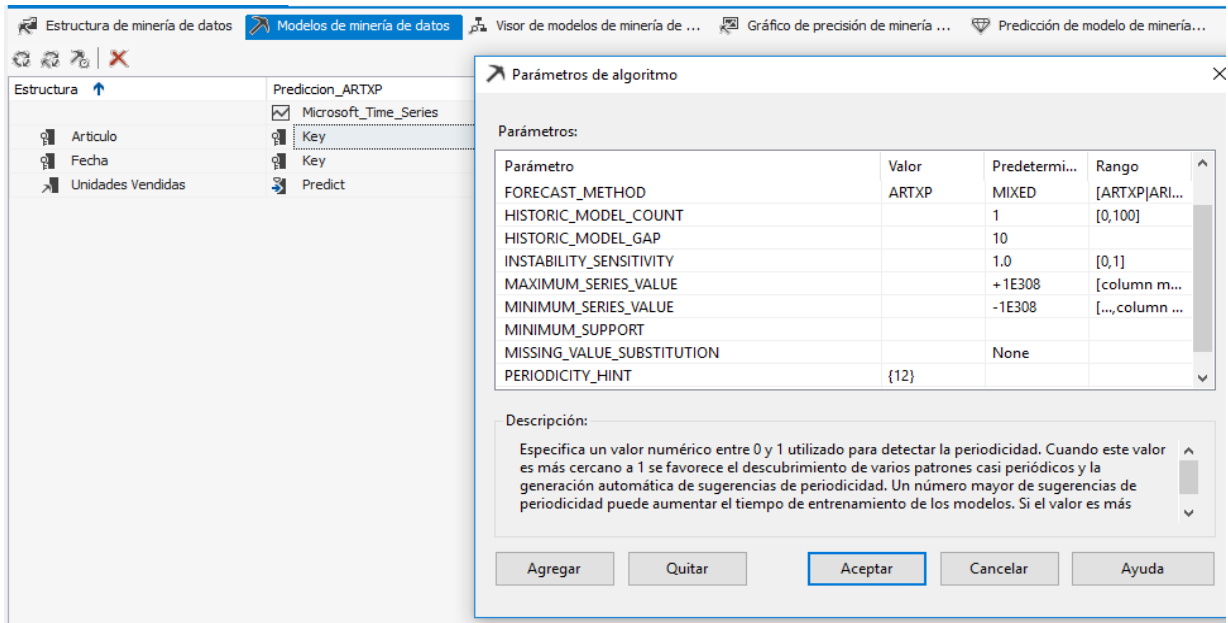


Fig. 39 Parametros del algoritmo con el metodo ARTXP

A continuacion veremos la grafica que nos brinda el modelo de minería de datos ARTXP de los 5 artículos mas vendidos en el transcurso de los años iniciando desde el 2014.

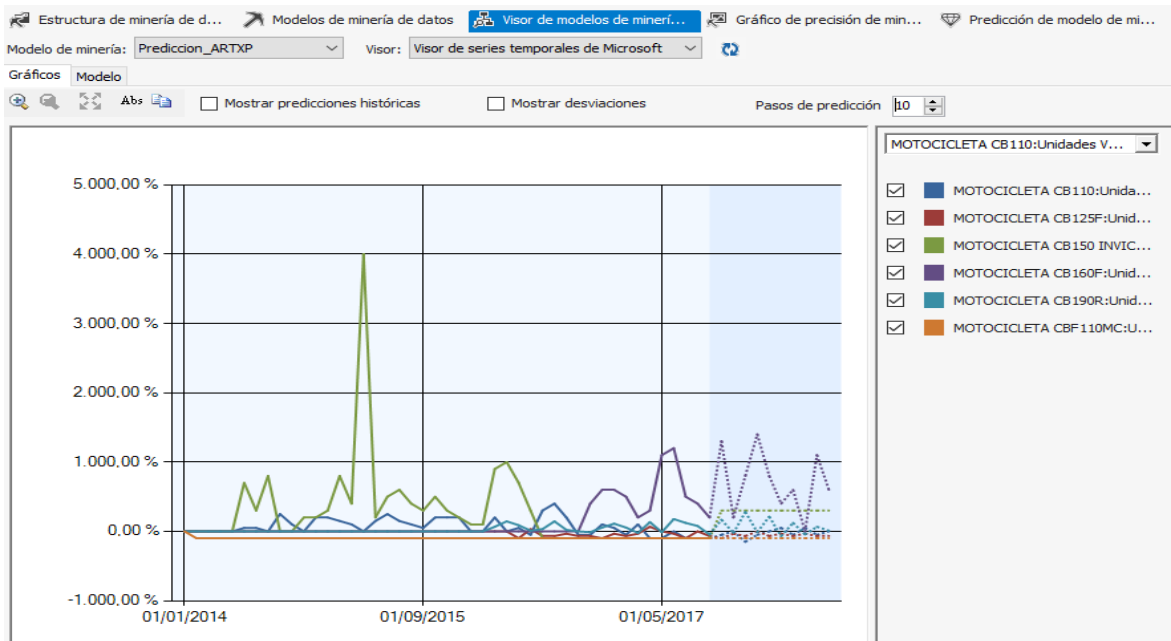


Fig. 40 Visor de modelo de minería de datos ARTXP

Por último, se evaluó un modelo con el método ARIMA para su posterior fiabilidad y ver cual se adecua o tiene menos margen de error.

Procedemos a integrarle parametros al algoritmo ya que se va añadir en FORESCAST_METHOD el valor de ARIMA en este parametro podemos elegir el modelo de serie temporal que utilizaremos y por otro lado en el parametro PERIODICITY_HINT se le añade {12} ya que la prediccion la haremos anualmente por 12 meses.

En la Fig. 41 observaremos los parametros del algoritmo y los que utilizaremos.

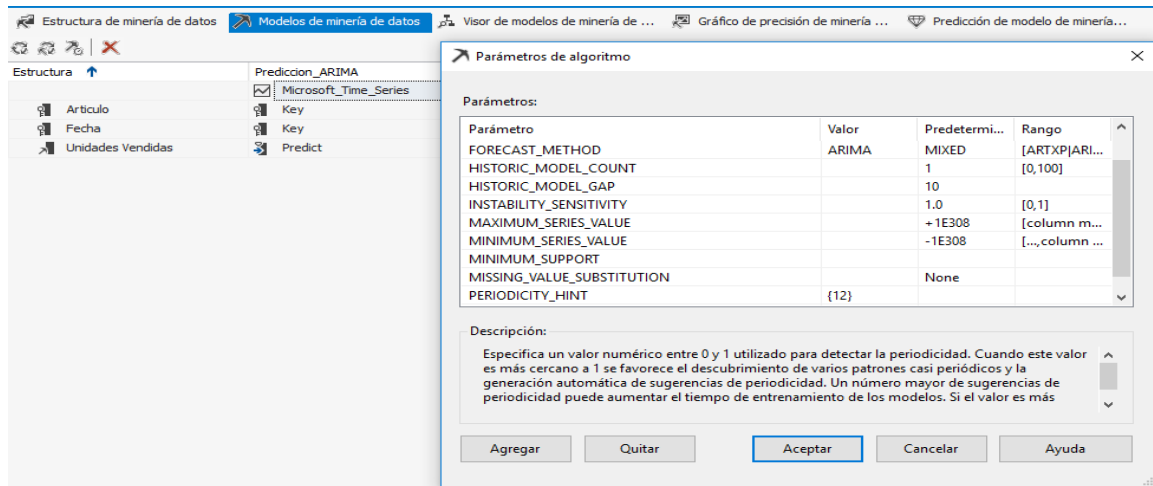


Fig. 41 Parametros del algoritmo con el metodo ARIMA

A continuacion veremos la grafica que nos brinda el modelo de minería de datos ARTXP de los 5 artículos mas vendidos en el transcurso de los años iniciando desde el 2014.

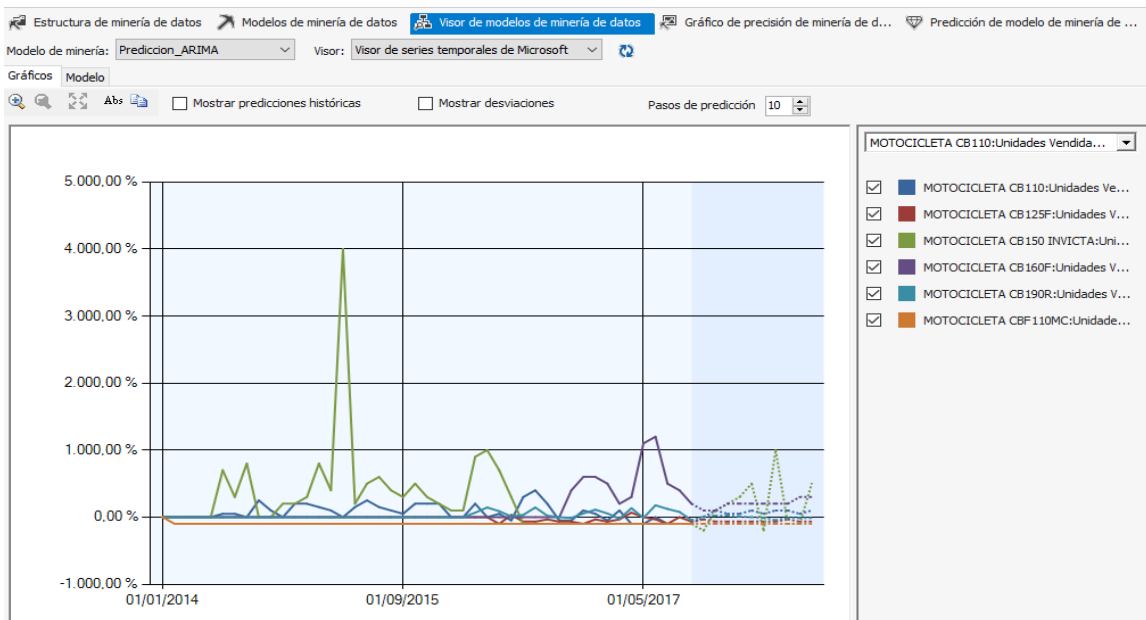


Fig. 42 Visor de modelo de minería de datos ARIMA

Estos modelos de minería de datos también se pueden crear directamente desde el Analysis Services con el lenguaje DMX. Al igual que con la herramienta para iniciar se creaba el modelo, acá hacemos lo mismo solo que con el lenguaje apropiado ya que los metadatos se consultan en DMX.

En la Fig. 43 observamos la creación de los modelos de minería de datos

```
--CREANDO MODELO DE PREDICCIÓN DE SERIES DE SERIES TEMPORALES ARIMA
CREATE MINING MODEL [Prediccion_ARIMA](
    fecha date KEY TIME,
    articulo text key,
    unidades long CONTINUOUS PREDICT
)
USING Microsoft_Times_Series(AUTO_DETECT_PERIODICITY = 0.8, FORECAST_METHOD = 'ARIMA')

--CREANDO MODELO DE PREDICCIÓN DE SERIES DE SERIES TEMPORALES ARTXP
CREATE MINING MODEL [Prediccion_ARTXP](
    fecha date KEY TIME,
    articulo text key,
    unidades long CONTINUOUS PREDICT
)
USING Microsoft_Times_Series(AUTO_DETECT_PERIODICITY = 0.8, FORECAST_METHOD = 'ARTXP')

--CREANDO MODELO DE PREDICCIÓN DE SERIES DE SERIES TEMPORALES MIXED
CREATE MINING MODEL [Prediccion_MIXED](
    fecha date KEY TIME,
    articulo text key,
    unidades long CONTINUOUS PREDICT
)
USING Microsoft_Times_Series(AUTO_DETECT_PERIODICITY = 0.8, FORECAST_METHOD = 'MIXED')
```

Fig. 43 Creacion de modelos de minería de datos basado en el algoritmo de series temporales

Al tener creados los modelos ARIMA, ARTXP y MIXED procedemos a hacer uso de estos. Usando sentencias en DMX consultando predicciones en el tiempo de 12 meses por cada producto vendido específicamente.

```
--UTILIZANDO EXTENCIONES DE MINERIA DE DATOS(DMX) PARA CREAR LOS DIFERENTES TIPOS DE
--PREDICIONES BASADOS EN MODELOS DE SERIES TEMPORALES
select FLATTENED
[Prediccion_MIXED].[Articulo],
PredictTimeSeries([Prediccion_MIXED].[Unidades Vendidas],12) as PrediccionCantidad
from [Prediccion_MIXED]
where [Articulo] = 'MOTOCICLETA CB110'

select
[Prediccion_ARIMA].[Articulo],
PredictTimeSeries([Prediccion_ARIMA].[Unidades Vendidas],12) as PrediccionCantidad
from [Prediccion_ARIMA]
where [Articulo] = 'MOTOCICLETA CB110'

select
[Prediccion_ARTXP].[Articulo],
PredictTimeSeries([Prediccion_ARTXP].[Unidades Vendidas],12) as PrediccionCantidad
from [Prediccion_ARTXP]
where [Articulo] = 'MOTOCICLETA CB110'
```

Fig. 44 Consultas DMX para crear los diferentes tipos de predicciones de un artículo específico

4.1.11. Iteración #11: Evaluación

Para comprobar la exactitud de un modelo de predicción es dividirlo en formación del modelo y otro validándolo.

Para la siguiente validación los valores pronosticados se comparan con los valores reales. Puesto que al final los modelos deben ser reconstruidos ya que nuevos datos se integrarían y esto no debe afectar a predicciones futuras con datos recientes que se haga en la empresa.

Existe un método que mide las desviaciones y valores exactos de la serie de tiempo en porcentajes este método se denomina MAPE (Porcentaje promedio absoluto de error).

De acuerdo a los algoritmos implementados en la investigación ARIMA, ARTXP Y MIXED para el análisis del proyecto se procedió a realizar la predicción del año 2017 para sus 3 primeros meses (enero, febrero y marzo) de los productos que más se vendieron según el histórico, puesto que se compararían con los valores reales que se tienen respecto a las ventas.

A continuación, explicaremos el procedimiento con la técnica MAPE:

- **ERROR = DATO REAL – DATO PREDICHO**
- **ERROR ABSOLUTO = |ERROR|**
- **ERROR ABSOLUTO PORCENTUAL = (ERROR ABSOLUTO / DATO REAL)*100**
- **MAPE = (Suma de ERROR ABSOLUTO %) / CANTIDAD ARTICULOS**

En la Fig. 45 se hace la evaluación del algoritmo ARIMA con el método MAPE. Según este argumento indica que se haya un error de un 115% en el pronóstico.

Artículo	Fecha	ARIMA	DATOS REALES	ALGORITMO ARIMA		
		Predicción Unidades	Ventas en el 2017	ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
MOTOCICLETA CB110	012017 0:00:00	3	3	0	0	0
MOTOCICLETA CB110	022017 0:00:00	5	1	-4	4	400
MOTOCICLETA CB110	032017 0:00:00	4	4	0	0	0
MOTOCICLETA CB125F	012017 0:00:00	0	2	2	2	100
MOTOCICLETA CB125F	022017 0:00:00	-1	1	2	2	200
MOTOCICLETA CB125F	032017 0:00:00	0	2	2	2	100
MOTOCICLETA CB160F	012017 0:00:00	1	7	6	6	86
MOTOCICLETA CB160F	022017 0:00:00	-13	6	19	19	317
MOTOCICLETA CB160F	032017 0:00:00	-35	3	38	38	1267
MOTOCICLETA CB190R	012017 0:00:00	14	19	5	5	26
MOTOCICLETA CB190R	022017 0:00:00	15	14	-1	1	7
MOTOCICLETA CB190R	032017 0:00:00	15	7	-8	8	114
MOTOCICLETA CCG 125	012017 0:00:00	6	5	-1	1	20
MOTOCICLETA CCG 125	022017 0:00:00	5	4	-1	1	25
MOTOCICLETA CCG 125	032017 0:00:00	5	3	-2	2	67
MOTOCICLETA CRF250LF	012017 0:00:00	0	1	1	1	100
MOTOCICLETA CRF250LF	022017 0:00:00	0	0	0	0	0
MOTOCICLETA CRF250LF	032017 0:00:00	0	0	0	0	0
MOTOCICLETA DIO	012017 0:00:00	1	1	0	0	0
MOTOCICLETA DIO	022017 0:00:00	1	1	0	0	0
MOTOCICLETA DIO	032017 0:00:00	1	2	1	1	50
MOTOCICLETA EG 1000 TIPO S	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	022017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	032017 0:00:00	1	2	1	1	50
MOTOCICLETA EG 5000CX	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 5000CX	022017 0:00:00	1	0	-1	1	0
MOTOCICLETA EG 5000CX	032017 0:00:00	1	2	1	1	50
MOTOCICLETA GL 125	012017 0:00:00	28	18	-10	10	56
MOTOCICLETA GL 125	022017 0:00:00	22	12	-10	10	83
MOTOCICLETA GL 125	032017 0:00:00	22	5	-17	17	340
				MAPE	115	

Fig. 45 Evaluación de algoritmo ARIMA con el método MAPE

En la Fig. 46 se hace la evaluación del algoritmo MIXED (Mezcla de ARIMA y ARTXP) con el método MAPE. Según este argumento indica que se haya un error de un 89% en el pronóstico.

Artículo	Fecha	MIXED	DATOS REALES	ALGORITMO MIXED		
		Prediccion Unidades	Ventas en el 2017	ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
MOTOCICLETA CB110	012017 0:00:00	4	3	-1	1	33
MOTOCICLETA CB110	022017 0:00:00	5	1	-4	4	400
MOTOCICLETA CB110	032017 0:00:00	5	4	-1	1	25
MOTOCICLETA CB125F	012017 0:00:00	0	2	2	2	100
MOTOCICLETA CB125F	022017 0:00:00	0	1	1	1	100
MOTOCICLETA CB125F	032017 0:00:00	0	2	2	2	100
MOTOCICLETA CB160F	012017 0:00:00	3	7	4	4	57
MOTOCICLETA CB160F	022017 0:00:00	-6	6	12	12	200
MOTOCICLETA CB160F	032017 0:00:00	-20	3	23	23	767
MOTOCICLETA CB190R	012017 0:00:00	12	19	7	7	37
MOTOCICLETA CB190R	022017 0:00:00	12	14	2	2	14
MOTOCICLETA CB190R	032017 0:00:00	13	7	-6	6	86
MOTOCICLETA CCG 125	012017 0:00:00	5	5	0	0	0
MOTOCICLETA CCG 125	022017 0:00:00	3	4	1	1	25
MOTOCICLETA CCG 125	032017 0:00:00	6	3	-3	3	100
MOTOCICLETA CRF250LF	012017 0:00:00	0	1	1	1	100
MOTOCICLETA CRF250LF	022017 0:00:00	0	0	0	0	0
MOTOCICLETA CRF250LF	032017 0:00:00	0	0	0	0	0
MOTOCICLETA DIO	012017 0:00:00	1	1	0	0	0
MOTOCICLETA DIO	022017 0:00:00	1	1	0	0	0
MOTOCICLETA DIO	032017 0:00:00	2	2	0	0	0
MOTOCICLETA EG 1000 TIPO S	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	022017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	032017 0:00:00	1	2	1	1	50
MOTOCICLETA EG 5000CX	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 5000CX	022017 0:00:00	1	0	-1	1	0
MOTOCICLETA EG 5000CX	032017 0:00:00	1	2	1	1	50
MOTOCICLETA GL 125	012017 0:00:00	24	18	-6	6	33
MOTOCICLETA GL 125	022017 0:00:00	20	12	-8	8	67
MOTOCICLETA GL 125	032017 0:00:00	21	5	-16	16	320
MAPE						89

Fig. 46 Evaluación de algoritmo MIXED con el método MAPE

En la Fig. 47 se hace la evaluación del algoritmo ARTXP con el método MAPE. Según este argumento indica que se haya un error de un 62% en el pronóstico.

Artículo	Fecha	ARTXP	DATOS REALES	ALGORITMO ARTXP		
		Prediccion Unidades	Ventas en el 2017	ERROR	ERROR ABSOLUTO	ERROR ABSOLUTO %
MOTOCICLETA CB110	012017 0:00:00	4	3	-1	1	33
MOTOCICLETA CB110	022017 0:00:00	5	1	-4	4	400
MOTOCICLETA CB110	032017 0:00:00	5	4	-1	1	25
MOTOCICLETA CB125F	012017 0:00:00	1	2	1	1	50
MOTOCICLETA CB125F	022017 0:00:00	1	1	0	0	0
MOTOCICLETA CB125F	032017 0:00:00	0	2	2	2	100
MOTOCICLETA CB160F	012017 0:00:00	4	7	3	3	43
MOTOCICLETA CB160F	022017 0:00:00	2	6	4	4	67
MOTOCICLETA CB160F	032017 0:00:00	2	3	1	1	33
MOTOCICLETA CB190R	012017 0:00:00	10	19	9	9	47
MOTOCICLETA CB190R	022017 0:00:00	9	14	5	5	36
MOTOCICLETA CB190R	032017 0:00:00	10	7	-3	3	43
MOTOCICLETA CCG 125	012017 0:00:00	5	5	0	0	0
MOTOCICLETA CCG 125	022017 0:00:00	1	4	3	3	75
MOTOCICLETA CCG 125	032017 0:00:00	8	3	-5	5	167
MOTOCICLETA CRF250LF	012017 0:00:00	0	1	1	1	100
MOTOCICLETA CRF250LF	022017 0:00:00	0	0	0	0	0
MOTOCICLETA CRF250LF	032017 0:00:00	0	0	0	0	0
MOTOCICLETA DIO	012017 0:00:00	2	1	-1	1	100
MOTOCICLETA DIO	022017 0:00:00	2	1	-1	1	100
MOTOCICLETA DIO	032017 0:00:00	2	2	0	0	0
MOTOCICLETA EG 1000 TIPO S	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	022017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 1000 TIPO S	032017 0:00:00	1	2	1	1	50
MOTOCICLETA EG 5000CX	012017 0:00:00	1	1	0	0	0
MOTOCICLETA EG 5000CX	022017 0:00:00	1	0	-1	1	0
MOTOCICLETA EG 5000CX	032017 0:00:00	1	2	1	1	50
MOTOCICLETA GL 125	012017 0:00:00	19	18	-1	1	6
MOTOCICLETA GL 125	022017 0:00:00	19	12	-7	7	58
MOTOCICLETA GL 125	032017 0:00:00	19	5	-14	14	280
MAPE						62

Fig. 47 Evaluación de algoritmo ARTXP con el método MAPE

El algoritmo basado en series temporales que fue seleccionado para las predicciones es el algoritmo ARTXP, ya que basado en los resultados mostrados con fundamento, el algoritmo ARTXP presenta un porcentaje de error (MAPE) menor con respecto a los demás.

4.2. En base a los objetivos de la investigación

4.2.5. Descripción del objetivo 1:

Incrementar los reportes sobre predicción de productos a vender

Con ayuda de la Herramienta SQL Server logramos llevar la predicción del año 2018 por unidades vendidas en el transcurso de los meses realizada en esta herramienta hasta POWER BI donde plasmamos nuestros reportes incluso minería de datos. Esto nos permite tener un mejor entendimiento de la data histórica ya que al saber que voy a vender, sabré lo que voy a comprar puesto que ayudara a la gestión de inventarios.

Esto también ayudara a tomar decisiones estratégicas con respecto al proceso de ventas

Articulo	Year	Month	Day	Prediccion Unidades.	Unidades Vendidas
MOTOCICLETA CB110	2018	January	1		3
MOTOCICLETA CB110	2018	February	1		3
MOTOCICLETA CB110	2018	March	1		3
MOTOCICLETA CB110	2018	April	1		3
MOTOCICLETA CB110	2018	May	1		3
MOTOCICLETA CB110	2018	June	1		3
MOTOCICLETA CB110	2018	July	1		3
MOTOCICLETA CB110	2018	August	1		3
MOTOCICLETA CB110	2018	September	1		4
MOTOCICLETA CB110	2018	October	1		3
MOTOCICLETA CB110	2018	November	1		3
MOTOCICLETA CB110	2018	December	1		3
MOTOCICLETA CB125F	2018	January	1		2
MOTOCICLETA CB125F	2018	February	1		1
MOTOCICLETA CB125F	2018	March	1		1
MOTOCICLETA CB125F	2018	April	1		1
MOTOCICLETA CB125F	2018	May	1		2
MOTOCICLETA CB125F	2018	June	1		1
MOTOCICLETA CB125F	2018	July	1		1
MOTOCICLETA CB125F	2018	August	1		1
MOTOCICLETA CB125F	2018	September	1		1
MOTOCICLETA CB125F	2018	October	1		1
MOTOCICLETA CB125F	2018	November	1		1
MOTOCICLETA CB125F	2018	December	1		1
MOTOCICLETA CB150 INVICTA	2018	January	1		4
MOTOCICLETA CB150 INVICTA	2018	February	1		6
Total					1271

Fig. 48 Predicción de unidades vendidas en el 2018 por meses.

4.2.6. Descripción del objetivo 2:

Incrementar el número de reportes para el análisis en la toma de decisiones

➤ Reporte Top clientes por año:

El siguiente reporte muestra el top de los clientes por año indicando las unidades y los montos de venta.

Cabe recalcar que el top de clientes se filtra por unidades vendidas y el costo de estas en los diferentes años.

La Fig. 49 muestra el reporte del top de clientes por año

nombre_cliente	monto_venta	Year
BANCES SUCLUPE, NESTOR RAUL	27.984,00	2017
CHAVARRY AREVALO, ROBERTH	9.315,00	2016
CIEZA PEREZ, VICTOR YAMPIER	9.290,00	2015
SALDAÑA SANCHEZ, PACO	9.290,00	2014
BECERRA QUIROZ, OSCAR FERNANDO	9.135,00	
GARCIA NEYRA, DENNIS ALONSO	9.091,00	
CHIRA DIAZ, ELVIS ROLLER	9.080,00	
ESPEJO VARGAS, JOSE ANDRES	9.080,00	
DIAZ QUISPE, VICTOR RAUL	9.069,00	
PAISIG MONTENEGRO, EMILIA	9.069,00	
SAMAME SUCLUPE, JOSE ALONSO	9.063,00	
NOMBERA VEGA, MARIA EUGENIA	9.052,00	
NUÑEZ GUERRERO, JHON CARLOS	9.052,00	
VILLALOBOS ARIZOLA, LEODAN	9.052,00	
FERROÑAN VIERA, JUAN JOSE	9.047,00	
CERDAN MARIN, FELIX HERMOGENES	9.041,00	
LLANCA CHOCACA, MANUEL HUMBERTO	9.041,00	
BERRU HERNANDEZ, JOHNNY ALEXANDER	9.033,00	
ESTACION DE SERVICIOS EL CARIBE S.A.C	9.033,00	
VELEZ ANACLETO, JORGE EMERSON	9.033,00	
Total	200.850,00	

nombre articulo
<input type="checkbox"/> MOTOCICLETA CB110
<input type="checkbox"/> MOTOCICLETA CB125F
<input type="checkbox"/> MOTOCICLETA CB160F
<input checked="" type="checkbox"/> MOTOCICLETA CB190R
<input type="checkbox"/> MOTOCICLETA CBX250
<input type="checkbox"/> MOTOCICLETA CCG125
<input type="checkbox"/> MOTOCICLETA CG 150
<input type="checkbox"/> MOTOCICLETA CRF250LF
<input type="checkbox"/> MOTOCICLETA DIO
<input type="checkbox"/> MOTOCICLETA EG 1000 TIPO S
<input type="checkbox"/> MOTOCICLETA EG5000CX
<input type="checkbox"/> MOTOCICLETA EG6500 CX S
<input type="checkbox"/> MOTOCICLETA EG6500 CX S S
<input type="checkbox"/> MOTOCICLETA ELITE 125
<input type="checkbox"/> MOTOCICLETA EP2500C SC
<input type="checkbox"/> MOTOCICLETA EP2500CX
<input type="checkbox"/> MOTOCICLETA G200 QAMD
<input type="checkbox"/> MOTOCICLETA GL125
<input type="checkbox"/> MOTOCICLETA GL150
<input type="checkbox"/> MOTOCICLETA GX270 T2 QP
<input type="checkbox"/> MOTOCICLETA GX390T1QH
<input type="checkbox"/> MOTOCICLETA HJ / 4 / 13 C V
<input type="checkbox"/> MOTOCICLETA HJ / 4 / 9 C
<input type="checkbox"/> MOTOCICLETA HJ/2/9 MP
<input type="checkbox"/> MOTOCICLETA HJ/3/13AP
<input type="checkbox"/> MOTOCICLETA HJ/4/13 A
<input type="checkbox"/> MOTOCICLETA HJ/4/13 C
<input type="checkbox"/> MOTOCICLETA SHADOW 150
<input type="checkbox"/> MOTOCICLETA STORM

Fig. 49 Reporte top de clientes por año.

➤ **Reporte Top de unidades vendidas por año:**

En el presente reporte nos muestra las unidades en este caso son los artículos (Motocicletas), nos muestra información de los clientes y por último de los años.

Veremos que unidades se vendieron más en el año 2015 al igual que a quien se las vendieron. Puesto que en este año se vendieron más al Gobierno Regional Lambayeque según el grafico expuesto

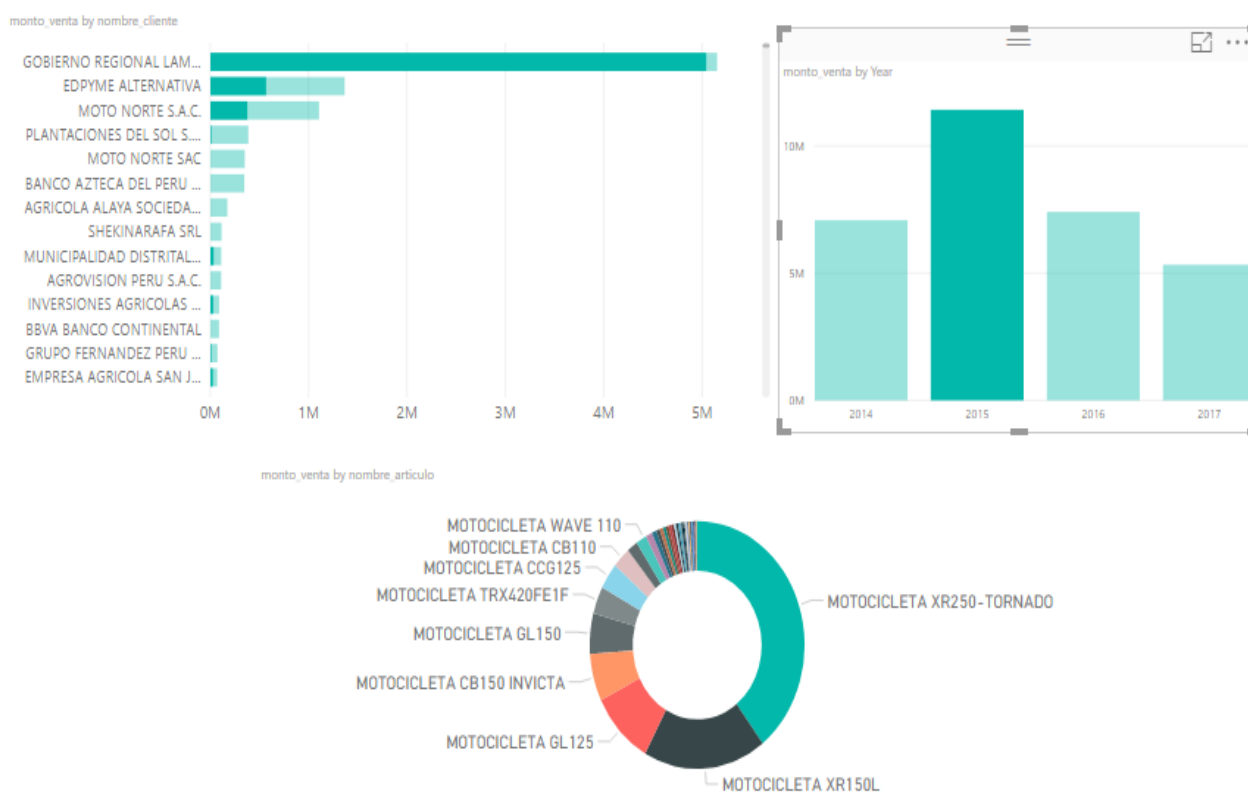


Fig. 50 Top unidades vendidas por año

4.2.7. Descripción del objetivo 3:

Incrementar el grado de satisfacción sobre la información solicitada

El grado de satisfacción es muy importante, principalmente por el gerente de ventas ya que él está a cargo de las estrategias del área de ventas al cual le realizaremos una encuesta para saber qué tan satisfecho está con la información que solicita después de aplicar el sistema. Anexo N° 5

4.2.8. Descripción del objetivo 4:

Incrementar la eficiencia de un modelo de predicción óptimo

Es de suma importancia saber qué modelo de predicción de ventas utilizar. Se empleó una técnica de procesamiento de error porcentual absoluto medio llamado MAPE, esta fue realizada y elaborada en la Iteración #11 lo cual en conclusión nos resultó más favorable trabajar con el modelo de ARTXP por su menos porcentaje de error absoluto.

4.3. Impactos esperados

4.3.5. Impactos económicos

El impacto que mi proyecto brinda es que ayudara a las empresas a reducir costes de análisis y evaluación esto quiere decir que apoyan a los gastos de gestión ya que le simplifica a los gerentes el proceso de supervisión respecto a las ventas

4.3.6. Impactos sociales

En este punto podemos recalcar que al momento de hacer una proyección de ventas se podrá saber cuánto comprar y por lo mismo se mantendrá el stock en los productos de más demanda esto evitara la deserción de clientes así mismo tendrá un impacto positivo a la sociedad.

4.3.7. Impactos en tecnología

Las grandes empresas se desarrollaron antes de la era de los sistemas de información y tecnologías de información, estas suelen ser ineficientes y lentas puesto que son menos competitivas. Al igual que el proyecto que desarrollamos la TI dirigen la capacidad de decisión a los niveles gerenciales, en conclusión la solución genera un impacto positivo en la tecnología

4.3.8. Impactos ambientales

Con respecto a los impactos ambientales el desarrollo del producto no apoya a la degradación. Negatividad o dificultad ambiental.

4.3.9. Impactos en la formación de cadenas productivas:

A través del uso de las redes globales ayudan a las empresas a reducir los costes de su participación en el mercado, haciendo que sea muy valioso hacer contratos con proveedores externos.

V. DISCUSIÓN

5.1. Indicador 1 Y 2:

Estos indicadores se relacionan ya que se evaluaron conjuntamente

5.1.1. Indicador 1: Cantidad de reportes sobre predicción de productos a vender

5.1.2. Indicador 2: Numero de reportes analizando las ventas a través de las dimensiones.

TABLA XXVIII.
INDICADOR 1 – INDICADOR 2

Indicador	O1	O2	Diferencia
Cantidad de reportes sobre predicción de productos a vender	0	8	8
Numero de reportes analizando las ventas a través de dimensiones			

- ✓ **O1:** Es la cantidad de reportes sobre la predicción de productos a vender y el número de reportes analizando las ventas a través de las dimensiones sin el uso del sistema de BI.
- ✓ **O2:** Es la cantidad de reportes sobre la predicción de productos a vender y el número de reportes analizando las ventas a través de las dimensiones con el uso del sistema de BI.

Diferencia entre (O2 – O1): En el transcurso del análisis de la situación problemática en la empresa MotoFuerza S.A.C se identificó que se tenía 0 reportes respecto a la predicción de productos a vender y al análisis de las ventas a través de las dimensiones. Por consiguiente, gracias al sistema de negocios inteligentes construido, la empresa puede analizar la realidad en base a su giro de negocio en cuanto a indicadores de monto total y unidades vendidas en el transcurso de un tiempo establecido al igual que filtrar datos top ya sea de un artículo, cliente en particular.

Se generaron 8 reportes puesto que 4 fueron de predicción de productos y 4 sobre el análisis de las ventas a través de las dimensiones. Se observó

un incremento del 100% del total de reportes. Por ende, se cumplió con los objetivos:

- Incrementar la cantidad de reportes sobre predicción de productos a vender
- Incrementar el número de reportes analizando las ventas a través de las dimensiones

5.2. Indicador 3: Grado de satisfacción del gerente de ventas sobre la información que solicito

Para la evaluación de este indicador se han tomado los datos de la tabla en el Anexo 5 ya que está el formulario aplicado para evaluar la satisfacción del gerente de ventas después de implementarse el sistema de negocios inteligentes.

- ✓ **O1:** Grado de satisfacción del gerente de ventas sobre la información solicitada sin el uso del sistema de BI.
- ✓ **O2:** Grado de satisfacción del gerente de ventas sobre la información solicitada con el uso del sistema de BI.

Ahora se muestra el cálculo de la respectiva fórmula para comparar si la hipótesis dada es aceptada o denegada.

S = Grado de satisfacción por parte de la gerencia sobre la información solicitada (gerente de ventas)

$S = (\text{N}^\circ \text{ miembros satisfechos y muy satisfechos} / \text{Total de encuestados}) * 100$

$S = (1/1) * 100$

$S = 100$

Diferencia entre (O2 – O1): En el transcurso del análisis de la situación problemática en la empresa MotoFuerza S.A.C se identificó que el grado de satisfacción de la gerencia sobre la información solicitada era de un 20%. Ahora gracias al desarrollo del sistema de negocios inteligentes el grado de satisfacción se incrementó en una 80% lo cual significa que el 100% de los miembros están satisfechos o muy satisfechos con la información solicitada y obtenida

En conclusión, se cumplió el objetivo:

- Incrementar el grado de satisfacción del gerente de ventas sobre la información que solicito

TABLA XXIX.

INDICADOR 3

Indicador	O1	O2	Diferencia
Grado de satisfacción del gerente de ventas sobre la información que solicito	20%	100%	80%

5.3. Indicador4: Eficacia de modelos de proyección de ventas

TABLA XXX.

INDICADOR 4

Indicador	O1	O2	Diferencia
Eficacia de modelos de proyección de ventas	0	3	3

- ✓ **O1:** Eficacia de modelos de proyección de ventas sin el uso del sistema de BI
- ✓ **O2:** Eficacia de modelos de proyección de ventas con el uso del sistema de BI

Diferencia entre (O2 – O1): En el transcurso del análisis de la situación problemática en la empresa MotoFuerza S.A.C se identificó que se tenía 0 reportes respecto a la predicción de productos a vender ahora para obtenerlos después de crear el sistema BI necesitamos que el modelo de predicción sea fiable por lo cual se implementó un método para hallar el menos margen de error por cada modelo. Por consiguiente, gracias al sistema de negocios inteligentes construido, la empresa puede analizar las proyecciones según las ventas teniendo una validación concreta por parte de los diferentes modelos utilizados.

Se observó un incremento del 62% del total de confianza ante la predicción de ventas gracias al modelo ARTXP de series temporales. Por ende, se cumplió con los objetivos:

- Incrementar la eficacia de modelos de proyección de ventas.

VI. CONCLUSIONES

- 6.1. Se incrementó en un 100% el número de reportes sobre proyección de los productos a vender logrando que el gerente de ventas pueda aproximarse al número de motocicletas se vendieron en el tiempo con el propósito de preveer la compra de productos y así mejorar su stock.
- 6.2. Se incrementó a un 100% el número de reportes analizando las ventas a través de dimensiones logrando así que el gerente de ventas pueda examinar en distintas situaciones a partir de distintos puntos de vista considerados importantes por la empresa en la toma de decisiones.
- 6.3. Se incrementó en un 80% el grado de satisfacción por parte de la gerencia sobre la información solicitada, demostrando que al emplear el sistema de BI este brinde información analítica importante, resumida, concisa y basada en las necesidades del usuario final para apoyar la toma de decisiones estratégicas.
- 6.4. Se incrementó la eficacia de modelos de proyección de ventas por medio del método de error porcentual, esto ayudará a que las decisiones y estrategias sean más concretas y así se tendrá un apoyo fiable al momento de analizar y evaluar las proyecciones de ventas.

VII. RECOMENDACIONES

- 1.** Utilizando una plataforma de programación Android se podría plasmar llevar toda esta data y mostrarla en dispositivos más portátiles que podría ser en los smartphones, tablets, etc.
- 2.** Integrar un sistema distribuido de registro de artículos y desglosarlo en más campos tal como modelo, marca, chasis, motor, color para su posterior análisis, siendo cada vez más fiable y consistente la data obtenida

VIII. LISTA DE REFERENCIAS

8.1. Bibliografía:

- [1] J. Salkind, de *Metodos de investigación*, México, Prentice Hall, 1999.
- [2] C. Cano, *Business Intelligence, decisiones del negocio basadas en tecnología: ruca crítica del negocio moderno*, Soluciones Avanzadas, 1999.
- [3] G. Correa, *Desarrollo de algoritmo y sus aplicaciones en Basic, Pascal, Cobol y C*, McGraw-Hill, 1992.
- [4] Hackney, *Your Business Intelligence arsenal. Telephony*, Soluciones Avanzadas, 2000.
- [5] Microsoft, «www.microsoft.com,» junio 2016. [En línea]. Available: <https://msdn.microsoft.com/es-es/library/ms174923.aspx> . [Último acceso: 2016].
- [6] C. Koch, *Nike Rebounds*, Copyright, 2004, pp. 511-516.
- [7] S. Nahamias, *Analisis de la produccion y las operaciones*, McGraw-Hill, 2007.
- [8] Noori y Radford, *Administracion de operaciones y produccion: calidad total y respuesta sensible rapida*, 1997.
- [9] J. Supulveda, *Inteligencia de Negocios*, 2002.
- [10] H. Taha, *Investigaciones de Operaciones*, Alfaomega, 1991.
- [11] Anonymus, *The critical role of business intelligence in e-business*, New York: Soluciones Avanzadas, 1999.

IX. ANEXOS

ANEXO N° 01.**CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS**

Motofuerza S.A.C
 Av. Felipe Santiago Salaverry 585
 Urb. Los Parques – Chiclayo – Perú
 Teléfono: (074) 224831
 Email: f.vidaurre@moto-fuerza.com

“AÑO DE LA LUCHA CONTRA LA CORRUPCION E IMPUNIDAD”

EMPRESA MOTOFUERZA S.A.C

CHICLAYO , 25 DE SETIEMBRE DE 2019

ASUNTO: CULMINACION DE PROYECTO DE TESIS SATISFACTORIAMENTE

De mi especial consideracion:

Reciba un saludo cordial a nombre de todos los integrantes de nuestra empresa Honda que se dedica al rubro de comercialización de motocicletas.

Sirva la presente para hacer de su conocimiento que hemos aceptado la culminación eficiente de la Tesis de título: "Sistema BI con predicción de ventas basado en el algoritmo de series temporales para apoyar la gestión en la empresa MOTOFUERZA S.A.C "cuyo alumno es responsable es:

- **Luis Fernando Juniors Santa Cruz Montaña**

Estudiante del X ciclo de la carrera profesional de Ingenieria de Sistemas y Computacion De la Institucion que usted dignamente representa.

Sin otro particular me despido y que Dios bendiga su favorable gestion.

Atentamente,

Fernando Vidaurre De La Cruz
 Gerente de Ventas

COPIA DE LA CARTA DE ACEPTACION DE LA EMPRESA

ANEXO N° 02.**ANÁLISIS DE RIESGOS****1. Datos generales**

- **Tesista** : Luis Fernando Juniors Santa Cruz Montaña
- **Fecha inicial** : 19 de agosto de 2019
- **Fecha final** : 10 de diciembre de 2019

2. Alcance del proyecto

Se desarrollará una solución de inteligencia de negocios para apoyar en diferentes aspectos a la empresa sea en la gestión de inventarios como en la toma de decisiones, con la finalidad de tener mejor análisis de los históricos con respecto al alto mando del área de ventas haciendo uso de herramientas, métodos, algoritmos de predicción de ventas.

El sistema implementado permite analizar, evaluar y predicción de ventas

La información será presentada mediante gráficos, tablas, dashboard para ello será necesario tener las herramientas necesarias y tener conocimiento sobre los diferentes modelos de proyección.

3. Interesados (Stakeholders)

Durante el desarrollo de la presente tesis se ha identificado a los siguientes interesados:

- **Internos**

TABLA XXXI**INTERESADOS INTERNOS**

Interesado	Participación
Interesado 1	Falta de disponibilidad de herramientas necesarias para el desarrollo.
Interesado 2	Carencia de investigación de los temas en mi proyecto de tesis
Interesado N	Mala planificación de las etapas e iteraciones a desarrollar

- **Externos**

TABLA XXXII**INTERESADOS EXTERNOS**

Interesado	Participación
Interesado 1	Infección de virus que afecte al sistema en desarrollo
Interesado 2	Carencia de internet y comunicación por motivos de lluvias
Interesado N	Falta de respaldo del desarrollo en un backup

4. Beneficios

Los beneficios que se van a obtener con el producto que se ha desarrollado son:

- Apoyar a la gestión de ventas con respecto a tomar decisiones estratégicas
- Apoyar en la gestión de ventas con respecto a pronósticos en lapsos de tiempo
- Incrementar el grado de satisfacción del gerente de ventas respecto a la información que solicite
- Incrementar la eficacia de un modelo óptimo para predecir

5. Etapas de desarrollo

Para el desarrollo del producto de la presente tesis se ha realizado considerando las etapas de la Metodología Ralph Kimball y Crisp DM, tomaremos en cuenta las etapas más importantes y delicadas de la metodología:

- **Iteración #1: Planificación del proyecto**

- **Matriz de riesgos**

Entre los riesgos identificados en esta etapa se mencionan:

TABLA XXXIII
MATRIZ DE RIESGOS DE PLANIFICACION DEL PROYECTO

Código del riesgo	Descripción del riesgo	Fase afectada	Causa raíz	Entregables afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad por impacto	Nivel de riesgo
RE1 – 001	Requerimientos incompletos o ambiguos	Análisis	Los requerimientos no se definieron de forma clara y completa	Documento de análisis de requerimientos	5	Alcance	4	20	MUY ALTO
						Tiempo	5	25	
						Costo	5	25	
						Calidad	4	20	
						Total probabilidad por impacto		90	
RE1 – 002	Retraso en la especificación de requerimientos	Análisis	Las reuniones con el gerente para el levantamiento de requerimientos se posponen.	Documento de análisis de requerimientos	4	Alcance	5	20	ALTO
						Tiempo	4	16	
						Costo	4	16	
						Calidad	4	16	
						Total probabilidad por impacto		68	
RE1 – 004	Aumentar requerimientos seguidamente	Análisis	El gerente no está claro en lo que requiere.	Documento de análisis de requerimientos	3	Alcance	3	9	MEDIO
						Tiempo	4	12	
						Costo	4	12	
						Calidad	4	12	
						Total probabilidad por impacto		42	
RE1 – 003	Falta de compromiso del responsable	Análisis	El gerente evade preguntas puntuales.	Documento de análisis de requerimientos	2	Alcance	3	6	BAJO
						Tiempo	3	6	
						Costo	3	6	
						Calidad	3	6	
						Total probabilidad por impacto		24	

– **Matriz salvaguarda de riesgos**

Entre los planes de mitigación para superar riesgos identificados en esta etapa se mencionan:

**TABLA XXXIV
MATRIZ DE SALVAGUARDA DE RIESGOS DE LA PLANIFICACION DEL PROYECTO**

Código del riesgo	Amenaza / Oportunidad	Descripción del riesgo	Fase	Nivel de riesgo	Tipo de respuesta	Responsable	Plan de mitigación
RE1 – 001	Amenaza / Oportunidad	Requerimientos incompletos o ambiguos	Análisis	MUY ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Tener preguntas puntuales al gerente ✓ Usar herramientas de extracción de datos ✓ Sintetizar mejor la información obtenida
RE1 – 002	Amenaza / Oportunidad	Retraso en la especificación de requerimientos	Análisis	ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Acordar horas donde el gerente este libre ✓ Organizar y planificar tiempos del tesista ✓ Ser claro y conciso
RE1 – 004	Amenaza / Oportunidad	Aumentar requerimientos seguidamente	Análisis	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Elaborar clausulas ✓ Elaborar restricciones en estos casos ✓ Llegar a un acuerdo con el cliente
RE1 – 003	Amenaza / Oportunidad	Falta de compromiso del responsable	Análisis	BAJO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Utilizar técnicas de motivación ✓ Utilizar técnicas de aprendizaje ✓ Utilizar métodos y planes para beneficio

- **Iteración #2: Definición de los requerimientos de Minería de Datos y BI**

- **Matriz de riesgos**

Entre los riesgos identificados en esta etapa se mencionan:

TABLA XXXV
RIESGOS IDENTIFICADOS EN LA DEFINICIÓN DE LOS REQUERIMIENTOS DE MINERÍA DE DATOS Y BI

Código del riesgo	Descripción del riesgo	Fase afectada	Causa raíz	Entregables afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad por impacto	Nivel de riesgo
RE1 – 001	Requerimientos de minería de datos incompletos	Análisis	Los requerimientos de minería de datos no se definieron de forma clara y completa	Documento de análisis de requerimientos de minería de datos	5	Alcance	5	25	MUY ALTO
						Tiempo	5	25	
						Costo	4	20	
						Calidad	5	25	
						Total probabilidad por impacto		95	
RE1 – 002	Retraso en la especificación de requerimientos de minería de dato	Análisis	Las reuniones con el gerente para el levantamiento de requerimientos de minería se posponen.	Documento de análisis de requerimientos de minería de datos	4	Alcance	4	16	ALTO
						Tiempo	4	16	
						Costo	5	25	
						Calidad	5	25	
						Total probabilidad por impacto		82	
RE1 – 004	Aumentar seguidamente requerimientos de minería de datos	Análisis	El gerente no está claro en lo que requiere.	Documento de análisis de requerimientos de minería de datos	4	Alcance	4	16	MEDIO
						Tiempo	4	16	
						Costo	3	12	
						Calidad	4	16	
						Total probabilidad por impacto		60	
RE1 – 003	Falta de compromiso	Análisis	El gerente evade preguntas puntuales sobre minería.	Documento de análisis de requerimientos de minería de datos	2	Alcance	3	6	BAJO
						Tiempo	3	6	
						Costo	3	6	
						Calidad	2	4	
						Total probabilidad por impacto		22	

– **Matriz salvaguarda de riesgos**

Entre los planes de mitigación para superar riesgos identificados en esta etapa se mencionan:

**TABLA XXXVI
MATRIZ DE SALVAGUARDA DE RIESGOS EN LA DEFINICIÓN DE REQUERIMIENTOS DE MINERÍA DE DATOS Y BI**

Código del riesgo	Amenaza / Oportunidad	Descripción del riesgo	Fase	Nivel de riesgo	Tipo de respuesta	Responsable	Plan de mitigación
RE1 – 001	Amenaza / Oportunidad	Requerimientos de minería de datos incompletos	Análisis	MUY ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Tener preguntas puntuales de req .minería de datos al gerente ✓ Usar herramientas de extracción de req. minería de datos ✓ Sintetizar mejor la información obtenida
RE1 – 002	Amenaza / Oportunidad	Retraso en la especificación de requerimientos de minería de dato	Análisis	ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Acordar horas donde el gerente este libre ✓ Organizar y planificar tiempos del tesista ✓ Ser claro y conciso
RE1 – 004	Amenaza / Oportunidad	Aumentar seguidamente requerimientos de minería de datos	Análisis	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Elaborar clausulas ✓ Elaborar restricciones en estos casos ✓ Llegar a un acuerdo con el cliente
RE1 – 003	Amenaza / Oportunidad	Falta de compromiso	Análisis	BAJO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Utilizar técnicas de motivación ✓ Utilizar técnicas de aprendizaje ✓ Utilizar métodos y planes para beneficio

- **Iteración #5: Modelado Dimensional**

- **Matriz de riesgos**

Entre los riesgos identificados en esta etapa se mencionan:

TABLA XXXVII
RIESGOS IDENTIFICADOS EN EL MODELADO DIMENSIONAL

Código del riesgo	Descripción del riesgo	Fase afectada	Causa raíz	Entregables afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad por impacto	Nivel de riesgo
RE1 – 001	Definición incoherente de las dimensiones	Modelado	Las dimensiones no se definieron de acuerdo a la data que se tiene	Documento de modelado dimensional	5	Alcance	5	25	MUY ALTO
						Tiempo	5	25	
						Costo	4	20	
						Calidad	4	20	
						Total probabilidad por impacto		90	
RE1 – 002	Granularidad de dimensiones mal definidas	Modelado	No se evaluaron correctamente los campos según su importancia	Documento de modelado dimensional	4	Alcance	5	20	ALTO
						Tiempo	4	16	
						Costo	4	16	
						Calidad	4	16	
						Total probabilidad por impacto		68	
RE1 – 004	Definición errónea de los atributos por dimensión	Modelado	No se definieron bien los atributos puesto que se tenían datos sucios	Documento de modelado dimensional	3	Alcance	4	16	MEDIO
						Tiempo	3	16	
						Costo	3	12	
						Calidad	3	16	
						Total probabilidad por impacto		60	
RE1 – 003	Tipos de datos incompatibles en el poblamiento de dimensiones	Modelado	No se realizó una limpieza correcta al poblar las dimensiones	Documento de modelado dimensional	2	Alcance	3	6	BAJO
						Tiempo	3	6	
						Costo	2	4	
						Calidad	2	4	
						Total probabilidad por impacto		20	

– **Matriz salvaguarda de riesgos**

Entre los planes de mitigación para superar riesgos identificados en esta etapa se mencionan:

TABLA XXXVIII
MATRIZ DE SALVAGUARDA DE RIESGOS EN EL MODELADO DIMENSIONAL

Código del riesgo	Amenaza / Oportunidad	Descripción del riesgo	Fase	Nivel de riesgo	Tipo de respuesta	Responsable	Plan de mitigación
RE1 – 001	Amenaza / Oportunidad	Definición incoherente de las dimensiones	Análisis	MUY ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Conocer bien el proceso o área ✓ Evaluar según requerimientos ✓ Sintetizar mejor la información obtenida
RE1 – 002	Amenaza / Oportunidad	Granularidad de dimensiones mal definidas	Análisis	ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Evaluar la importancia de cada campo ✓ Desmenuzar los campos de una dimensión
RE1 – 004	Amenaza / Oportunidad	Definición errónea de los atributos por dimensión	Análisis	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Evaluar los atributos según la data que se tiene ✓ Administrar y analizar bien la data histórica
RE1 – 003	Amenaza / Oportunidad	Tipos de datos incompatibles en el poblamiento de dimensiones	Análisis	BAJO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Limpiar la data de origen ✓ Verificar tipos de datos en el destino ✓ Usar un convertidor de datos

- **Iteración #9: Diseño de la arquitectura técnica**

- **Matriz de riesgos**

Entre los riesgos identificados en esta etapa se mencionan:

**TABLA XXXIX.
MATRIZ DE RIESGOS PARA EL DISEÑO DE LA ARQUITECTURA TÉCNICA**

Código del riesgo	Descripción del riesgo	Fase afectada	Causa raíz	Entregables afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad por impacto	Nivel de riesgo
RE1 – 001	Insuficiente conocimiento de lo que se va a desarrollar	Diseño	Falta de investigación	Documento de diseño	5	Alcance	5	25	MUY ALTO
						Tiempo	5	25	
						Costo	5	20	
						Calidad	5	25	
						Total probabilidad por impacto		100	
RE1 – 002	Desorden al plasmar las etapas a desarrollar	Diseño	No está equilibrado con el proceso	Documento de diseño	4	Alcance	5	20	ALTO
						Tiempo	4	16	
						Costo	5	20	
						Calidad	5	20	
						Total probabilidad por impacto		66	
RE1 – 004	Deficiencia de herramientas a implementar	Diseño	Presupuesto o costos	Documento de diseño	4	Tiempo	3	12	MEDIO
						Costo	3	12	
						Calidad	3	12	
						Total probabilidad por impacto		54	
						Alcance	3	6	
RE1 – 003	Información innecesaria	Diseño	Falta de dominio en el tema	Documento de diseño	2	Tiempo	3	6	BAJO
						Costo	2	4	
						Calidad	2	4	
						Total probabilidad por impacto		20	
						Alcance	3	6	

– **Matriz salvaguarda de riesgos**

Entre los planes de mitigación para superar riesgos identificados en esta etapa se mencionan:

**TABLA XL.
MATRIZ DE SALVAGUARDA DE RIESGOS PARA EL DISEÑO DE ARQUITECTURA TÉCNICA**

Código del riesgo	Amenaza / Oportunidad	Descripción del riesgo	Fase	Nivel de riesgo	Tipo de respuesta	Responsable	Plan de mitigación
RE1 – 001	Amenaza / Oportunidad	Insuficiente conocimiento de lo que se va a desarrollar	Diseño	MUY ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Investigar sobre el tema a desarrollar ✓ Empaparse con conocimiento específico
RE1 – 002	Amenaza / Oportunidad	Desorden al plasmar las etapas a desarrollar	Diseño	ALTO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Adquirir conocimiento puntual y riguroso ✓ Analizar bien el contexto del proceso que estamos llevando a cabo
RE1 – 004	Amenaza / Oportunidad	Deficiencia de herramientas a implementar	Diseño	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Apoyarse en herramientas gratuitas ✓ Buscar herramientas el cual podremos desarrollar el proceso
RE1 – 003	Amenaza / Oportunidad	Información innecesaria	Diseño	BAJO	Salvaguarda	Tesista	<ul style="list-style-type: none"> ✓ Descartar información que no sume a la investigación ✓ Definir clara y concisa la información puntual

ANEXO N° 03. INSTRUMENTOS DE RECOLECCIÓN DE DATOS**ANEXO N° 3**

Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y
Computación

ENTREVISTA AL GERENTE DE VENTAS:

Nombre: Fernando Vidaurre De La Cruz
Cargo: Gerente general de ventas
Fecha: 25/04/2017

1. **¿Cuál es el giro de su negocio?**
Venta de motocicletas y motokar.
2. **¿Con cuántos trabajadores cuenta la empresa?**
35 trabajadores
3. **¿Cuál es la cantidad promedio de productos que vende la empresa por mes?**
De 100 a 130 motocicletas
4. **En cuanto al producto que el cliente le solicita, ¿Tienen todos los productos en su stock?**
No, hay veces que se ofrecen motos opcionales
5. **En caso de no tener el producto solicitado por el cliente, ¿qué tiempo demora en atenderse o entregar el pedido de los productos solicitados?**
Si no hay el producto solicitado mientras verificamos el stock se demora 15 min y por otro lado si hago pedido se demoran 2 días.
6. **¿El cliente que no encuentra el producto que busca, vuelve otra vez?**
No vuelve, ya que piensan que no tenemos su producto en stock
7. **¿Qué factores consideran para hacer el pedido para surtir su inventario?**
Nos fijamos en las ventas que se realizaron en un corto plazo a lo mucho 2 a 3 meses de antigüedad.
8. **¿Cuál es costo del producto que más vende la empresa?**
No le podría asegurar, ya que no tenemos posibilidades de hacer un reporte de tal magnitud con la data que manejamos.
9. **¿Cuenta con algún software de predicción de ventas que le ayude a tomar decisiones en la empresa?**

No, solo contamos con información detallada de las ventas en el transcurso del tiempo. Esta información no está ordenada ni moldeada como para un análisis concreto.

10. ¿Quién es el encargado del análisis, evaluación y toma de decisiones estratégicas en el proceso de ventas?

Gerente de Ventas, en este caso yo quien le habla.

11. ¿Las estrategias que usted toma ante el mercado se ajustan con alguna proyección de ventas de motos a corto, mediano o largo plazo?

No, la toma de decisiones es algo básico porque nos encargamos de comparar las ventas realizadas a corto plazo, no tenemos una fuente fiable por el cual nos apoyemos y podamos tomar decisiones estratégicas más maduras y proyectivas.

12. ¿Usted analiza la información histórica respecto a las ventas? Si su respuesta es no ¿Cree usted que es necesario analizar esa información?

No analizo información muy antigua por lo cual la data no está apta para este tipo de procesos. Creo que si es necesario para saber el comportamiento a largo plazo de las ventas que se realizan.

13. ¿Cuál es el grado de satisfacción que usted tiene ante el análisis, procesamiento y proyección de ventas actualmente en el proceso de ventas?

- A. Totalmente satisfecho
- B. Muy Satisfecho
- C. Satisfecho
- D. Poco Satisfecho
- E. Nada Satisfecho

ANEXO N° 4

Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y
Computación

Encuesta para verificar el grado de satisfacción por parte del gerente de ventas sobre la información que solicita con respecto a realizar una venta luego de aplicar el Sistema BI.

Indicar con una X su grado de satisfacción:

Pregunta	Puntuación
¿Cuál es el grado de satisfacción sobre la información solicitada con respecto a realizar una venta?	5

Donde:

- 1: Muy Insatisfecho
- 2: Insatisfecho
- 3: Ni Satisfecho ni insatisfecho
- 4: Satisfecho
- 5: Muy satisfecho