Aplicación de técnicas de inteligencia de negocios para analizar el comportamiento de los niveles de inventarios de la bodega de insumos de producción en la empresa COFIMAR S.A.

MAESTRÍA EN SISTEMAS DE INFORMACIÓN CON MENCIÓN EN INTELIGENCIA DE NEGOCIOS

Trabajo de investigación bajo la modalidad proyecto de desarrollo, requisito previo para la obtención del título de:

Magíster en Sistemas de Información, Mención en Inteligencia de Negocios

Autor:

Luis Leandro Garzón Vélez

Tutor:

PhD. Sandra Lorena García Bustos

Samborondón - Ecuador

Agosto del 2019

DEDICATORIA

Al ser humano más bueno que pude haber conocido en esta vida (†) María Teresa, no hay palabras poder describir el papel que tuviste, aun tienes y seguirás teniendo en mi vida.

A mis tres querubines (†) Xavier, Jestyn y Lupe, por haberme permitido conocer y saber quién es Dios, a través de su ojitos y su inocencia de niños.

AGRADECIMIENTO

De manera infinita le agradezco a DIOS por darme todo salud, bienestar, tiempo, y muchas dadivas más, agradezco a mi señora madre Pilar Vélez Gonzáles por haberme dado la oportunidad de vivir, a mi abuelo sr. Segundo Garzón por haber asumido el rol y la toda responsabilidad de padre cuando fui un niño.

También quiero agradecer a la persona que luego de conocerla cambiaria mi vida para siempre, si existe la eternidad, sería una dicha y un placer compartirla contigo, muchas gracias Angélica Moreira.

CERTIFICACION DE REVISION FINAL

QUE EL PRESENTE PROYECTO DE DESARROLLO TITULADO:

Aplicación de técnicas de inteligencia de negocios para analizar el comportamiento de los niveles de inventarios de la bodega de insumos de producción en la empresa COFIMAR S.A.

, FUE REVISADO, SIENDO SU CONTENIDO ORIGINAL EN SU TOTALIDAD, ASÍ COMO EL CUMPLIMIENTO DE LOS REQUERIMIENTOS QUE SE DICTAN EN LA GUÍA DE ELABORACIÓN DEL TRABAJO DE TITULACIÓN, POR LO QUE SE AUTORIZA A: **Luis Leandro Garzón Vélez**, QUE PROCEDA A SU PRESENTACION.

Samborondón, 5 de Agosto del 2019

Atentamente



_____

PhD. Sandra Lorena García Bustos

C.I. # 0919742866

DIVISIÓN DE POSGRADO
PROCESO DE TITULACIÓN
CERTIFICADO DEL PORCENTAJE DE COINCIDENCIAS

Habiendo sido nombrado Sandra Lorena García Bustos, tutor del trabajo de titulación "Aplicación de técnicas de inteligencia de negocios para analizar el comportamiento de los niveles de inventarios de la bodega de insumos de producción en la empresa COFIMAR S.A." elaborado por Luis Leandro Garzón Vélez con mi respectiva supervisión como requerimiento parcial para la obtención del título de "Magíster en Sistemas de Información Mención Inteligencia de negocios".

Se informa que el mismo ha resultado tener un porcentaje de coincidencias 7% mismo que se puede verificar en el siguiente link: https://secure.urkund.com/view/53113327-224149-945842. Adicional se adjunta capturas de pantalla de dicho resultado.

URKUND

Urkund Analysis Result

Analysed Document:	1563904136_Trabajo final -TESIS-.docx (D54544133)
Submitted:	7/23/2019 7:49:00 PM
Submitted By:	slgarcia@espol.edu.ec
Significance:	7 %

URKUND 1563904136_Trabajo final -TESIS-.docx (D54544133)

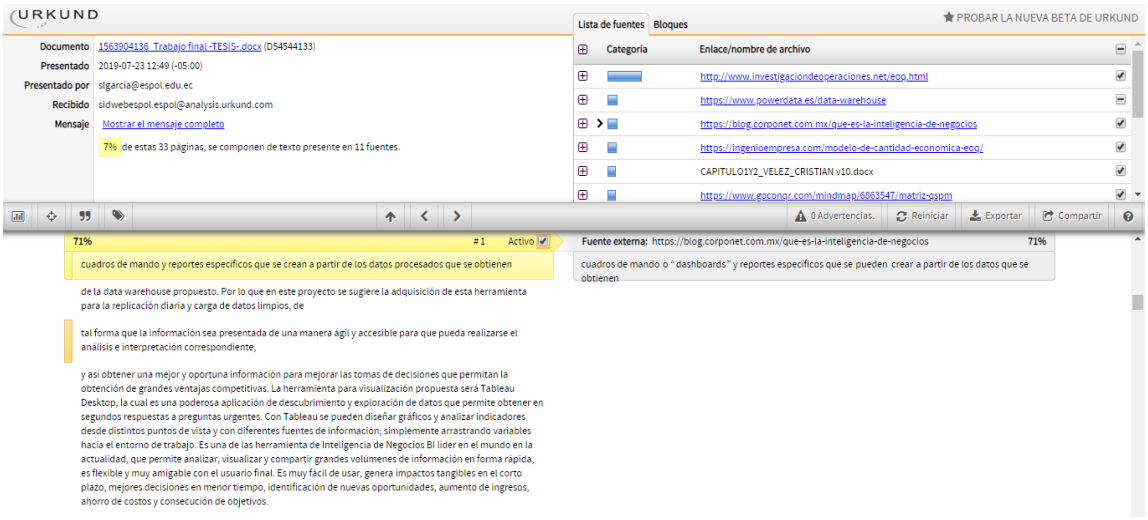
MAESTRÍA EN SISTEMAS DE INFORMACIÓN CON MENCIÓN EN INTELIGENCIA DE NEGOCIO

Aplicación de técnicas de inteligencia de negocios para analizar el comportamiento de los niveles de inventarios de la bodega de insumos de producción en la empresa COFIMAR S.A.

Trabajo de investigación requisito previo para la obtención del título de:

Magíster en Sistemas de Información Mención Inteligencia de negocios

Autor: Luis Leandro Garzón Vélez Tutor: PhD. Sandra Lorena García Bustos Guayaquil, Julio del 2019





_____

PhD. Sandra Lorena García Bustos

C.I. # 0919742866

RESUMEN

El principal objetivo de este trabajo fue plantear una solución al problema detectado en el área de bodega, altos costos, niveles de inventarios elevados en ciertos ítems y desabastecimiento en otros insumos durante los cierres de mes. La metodología que se utilizó en este trabajo fue la investigación de campo, científica y tecnológica. A través de técnicas para exploración de datos se consiguió identificar el producto del inventario con el mayor costo total, el cual para el cierre del año 2018 presento un valor de **$1.701.559,00** con una cantidad almacenada que ascendió a **3.267.075** kilogramos del producto, lo cual representaba dinero sin utilizar y una gran cantidad de espacio físico ocupado en las instalaciones de la organización.

Mediante la utilización de modernas técnicas para el procesamiento de datos, métodos estadísticos para predicción y un modelo para gestión de inventarios mediante lotes económicos de pedidos, se consiguió desarrollar un sistema de inteligencia de negocios el cual mediante su implementación podría dejar los siguientes beneficios: primero reducción de costos de inventarios hasta en un 75%, dejándole ahorros a la organización de más de un millón de dólares y segundo disminuir el nivel del inventario en más de dos millones de kilogramos de producto permitiendo tener un considerable espacio físico libre en el área de la bodega.

Palabras claves: inteligencia de negocios, métodos estadísticos, predicción, modelo para gestión de inventarios y lotes económicos de pedidos.

_____

Garzón Vélez Luis Leandro
C.I. # 0917972382

SUMMARY

The main objective of this work was to propose a solution to the problem detected in the warehouse area, high costs, high inventory levels in certain items and shortages in other supplies during the closing of the month. The methodology used in this work was field, scientific and technological research. Through data exploration techniques, it was possible to identify the product of the inventory with the highest total cost, which by the end of 2018 presented a value of $ 1,701,559.00 with a stored amount that amounted to 3,267,075 kilograms of product, which represented unused money and a large amount of physical space occupied in the organization's facilities.

Through the use of modern techniques for data processing, statistical methods for prediction and a model for inventory management through economic batches of orders, it was possible to develop a business intelligence system which through its implementation could leave the following benefits: first reduction of inventory costs up to 75%, leaving the organization savings of more than one million dollars and second decrease the inventory level by more than two million kilograms of product allowing to have a considerable free physical space in the area of the cellar

Keywords: business intelligence, statistical methods, prediction, inventory management model and economic order batches

CONTENIDO ANALITICO

1	INTRODUCCION.	1
1.1.	Antecedentes	3
1.2.	Descripción del problema	6
1.3.	Pregunta problemática	6
1.4.	Idea a defender	7
1.5.	Objetivo general	7
1.6.	Objetivos especificos	7
1.7.	Justificativos del Proyecto	8
1.8.	Metodologia	11
1.9.	Procedimiento metodologico, arquitectuta BI	12
2	MARCO TEORICO.	15
2.1.	Inteligencia de negocios.	15
2.2.	Datos.	16
2.3.	Información.	17
2.4.	Conocimiento.	17
2.5.	Data warehouse.	18
2.6.	Proceso ETL.	18
2.7.	Minería de datos.	19
2.8.	EOQ – Cantidad Economica de Pedido.	20
3	SOLUCION PROPUESTA.	24
3.1.	Analisis de la situacion actual.	24
3.2.	Analisis FODA.	26
3.3.	Definición de requerimientos funcionales de la propuesta.	31
3.4.	Definición de roles y responsabilidades del equipo de trabajo.	31
3.5.	Núcleo de la arquitectura tecnológica de la solución propuesta.	32
3.5.1	Datos: descripción.	33
3.5.2	Datos: limpieza.	33
3.5.3	Datos: formateo.	34
3.5.4	Integración de datos.	34
3.5.5	Estructura del sistema de la fuente de datos (ODS).	36
3.5.6	Estructura del sistema de almacenamiento datos (DW).	37
3.5.7	Diseño del proceso ETL.	39
3.5.8	Visualización y exploración de datos.	45
3.5.9	Selección de las técnicas de modelado predictivas.	49
3.5.9.1	Construcción del modelo ARIMA.	50
3.5.9.2	Construcción del modelo ARNN.	57
3.5.10	Evaluación de los modelos predictivos propuestos.	61
3.6.	Pronosticar la demanda anual y calcular la cantidad económica de pedido.	63

4	VALIDACION DE RESULTADOS.	65
4.1.	Esquema de validación.	65
4.2.	Revisión de resultados.	67
4.3.	Conclusiones.	67
4.2.	Recomendaciones propuestas.	68
5	BIBLIOGRAFIA.	69
6	ANEXOS.	70
Anexo #1,	Código en R del modelo de predicción ARIMA.	70
Anexo #2,	Código en R del modelo de predicción ARNN.	72
Anexo #3,	Código en R del algoritmo ARNN.	73
Anexo #4,	Código en PHP del aplicativo de carga de datos.	82

INDICE DE GRAFICOS.

Gráfico #1,	Captura de pantalla del sistema ERP de Cofimar S.A	10
Gráfico #2,	Arquitectura BI conceptual propuesta para este proyecto	13
Gráfico #3,	Pirámide de la información	17
Gráfico #4,	Proceso ETL	19
Gráfico #5,	Modelo de inventarios EOQ	22
Gráfico #6,	Módulo de inventarios que posee el sistema ERP de COFIMAR S.A	25
Gráfico #7,	Reporte actual que el módulo de inventarios del ERP ofrece	26
Gráfico #8,	Arquitectura hardware propuesta para implementar la solución BI del proyecto	30
Gráfico #9,	Arquitectura BI tecnológica propuesta para este proyecto	32
Gráfico #10,	Modelado CONCEPTUAL de datos propuesto para el proyecto, importación desde el almacén operacional de datos con nombre ODS_COFIMAR	35
Gráfico #11,	Modelado LOGICO de datos propuesto para el proyecto, variables que serán almacenadas al data warehouse de destino con nombre DW_COFIMAR	35
Gráfico #12,	Simulación del ODS actual, “ods_cofimar” de la bodega	37
Gráfico #13,	Estructura del DW propuesto, “dw_cofimar” de la bodega	38
Gráfico #14,	Arquitectura ETL propuesta para el proyecto	40
Gráfico #15,	Fuente de datos para la extracción, base ODS_COFIMAR	40
Gráfico #16,	Esquema del flujo de datos del proceso ETL	41
Gráfico #17,	Filtrado de datos realizado a través del proceso ETL	42
Gráfico #18,	Depuración de datos realizada a través del proceso ETL	42
Gráfico #19,	Aplicativo para carga de datos en el data warehouse	43
Gráfico #20,	Datos cargados en el data warehouse	43
Gráfico #21,	Consulta de datos con el botón “Consultar Registros”	44
Gráfico #22,	Consulta de datos cargados efectuada a la base	44
Gráfico #23,	Mapa de colores del inventario de bodega consolidado año 2018	47

Gráfico #24,	Histograma, mes de abril del 2018	47
Gráfico #25,	Gráfico de pastel, mes de abril del 2018	48
Gráfico #26,	Histograma de consumo histórico de metabisulfito.	48
Gráfico #27,	Data subida a R Studio con formato de series de tiempo	51
Gráfico #28,	Gráfica de tendencia acerca del consumo histórico de la variable	51
Gráfico #29,	Descomposición en serie de tiempos	52
Gráfico #30,	Prueba de la raíz unitaria	55
Gráfico #31,	Gráfica de tendencia sin estacionariedad en la variable	56
Gráfico #32,	Pronostico propuesto por el modelo ARIMA	57
Gráfico #33,	Valor real de la serie vs. Valor del modelo estimado	59
Gráfico #34,	Pronostico propuesto por el modelo ARNN	59
Gráfico #35,	Evaluación gráfica de los modelos predictivos	62
Gráfico #36,	Predicción de nivel de inventario, producto Metabisulfito	63
Gráfico #37,	Evaluación gráfica de la prueba de hipótesis t	66
Gráfico #38,	Resultados del análisis	67

INDICE DE TABLAS.

Tabla #1,	Análisis FODA para el proyecto	27
Tabla #2,	Tabla para ponderación del análisis FODA	29
Tabla #3,	Resultados obtenidos del FODA	29
Tabla #4,	Definición de requerimientos funcionales y roles de usuarios	31
Tabla #5,	Definición de roles y responsabilidades del equipo de trabajo	32
Tabla #6,	Modelo cantidad económica de pedido, producto Metabisulfito.	64
Tabla #7,	Resultados de la prueba de hipótesis t, aplicada a los procesos ARIMA y ARNN.	66
Tabla #8,	Datos finales del análisis.	68
Tabla #9,	Cronograma tentativo de implementación	68

Glosario de acrónimos y abreviaturas

CPK: Grupo empresarial Cofimar, Prilabsa y Krismare

TI: Tecnologías de Información

BI: Inteligencia de negocios, del inglés Business Intelligence

FODA: Fortaleza, Oportunidad, Debilidad y Amenaza

ETL: Extraer, transformar y cargar, del inglés Extract, Transform and Load

ERP: Planificación de recursos empresariales, del inglés Enterprise Resource Planning

RDBMS: Sistema de gestión de bases de datos relacionales, del inglés Relational database management system

ODS: Almacén operacional de datos, del inglés Operational Data Store

DW: Base de datos usada para análisis, del inglés Data Warehouse

ODBC: Conectividad abierta de bases de datos, del inglés Open Data Base Connectivity

EOQ: Cantidad económica de pedido, del inglés Economic Order Quantity

QSPM: Matriz de planificación estratégica cuantitativa, del inglés Quantitative Strategic Planning Matrix

SQL: Lenguaje estructurado de consulta, del inglés Structured Query Language

CSV: tipo de archivo de texto que almacena los datos en forma de columnas, separadas por coma y las filas se distinguen por saltos de línea del inglés comma-separated values

1°) INTRODUCCION

Aproximadamente a principios del siglo XX, se prestó mucha atención al desarrollo de métodos matemáticos analíticos destinados a brindar ayuda a quienes toman las decisiones en el establecimiento de los niveles óptimos de inventarios. Desde entonces se han aplicado a los problemas de administración de inventarios, instrumentos de análisis cada vez más precisos y refinados, la razón de que se haya dado mayor atención a los inventarios es que para muchas empresas, esa cifra es la partida de mayor valor que aparece del lado del activo en los balances generales. Los problemas del inventario relacionados con cantidades en existencia muy pequeñas o demasiado grandes pueden ser causa del fracaso de los negocios. Si un fabricante deja de tener existencias de un artículo considerado como crítico, esto podría dar como resultado una paralización no deseada del proceso productivo. Además, el comprador espera que el comerciante tenga en existencia el artículo que el necesita, si no hay existencias de algún artículo considerado como crítico cuando el cliente cree que debe haberlas, el comerciante pierde un cliente, no solo para este artículo, si no para muchos otros en el futuro.

Los costos en los que puede incurrir una empresa a consecuencia de los niveles de inventarios que establezca, podrían agruparse en tres categorías: costos de pedidos o de reabastecimiento, costos cargados al inventario o de almacenamiento y costos de pérdidas. El tercer costo hace referencia a la pérdida que sufre una organización o negocio si hay faltantes en sus existencias, pueden perderse ventas si los inventarios no son los adecuados para satisfacer la demanda de los consumidores, o la producción podría detenerse si los

inventarios considerados como críticos son insuficientes para cubrir sus necesidades, pueden requerirse inventarios de fluctuación si se quiere mantener un suministro adecuado de artículos, para cuando lo soliciten los consumidores si el objetivo es minimizar los faltantes de existencias, pero lo que en realidad se necesitan son existencias de seguridad a fin de poder satisfacer las fluctuaciones normales sobre el promedio de la demanda.

Una preocupación fundamental de la administración consiste en desarrollar políticas de inventarios que aminoren los costes totales de operación de las organizaciones. Para conseguir esta meta hay que tomar en consideración dos decisiones básicas de inventarios: la cantidad que hay solicitar y cuando se debe solicitar. Al considerar esas decisiones, una ruta consiste en pedir grandes cantidades, a fin de disminuir los costos de los pedidos. La otra alternativa consiste en pedir pequeñas cantidades para disminuir los costos cargados a los inventarios. Llevadas al extremo cualquiera de esas dos rutas tendrá un efecto desfavorable en las ganancias, y la mejor solución, en términos de ganancias e ingresos sobre los activos totales, es un compromiso entre los dos extremos.

La conclusión que se puede inferir de esta breve introducción es que una administración eficaz de los inventarios apoyándose en herramientas de actualidad como las técnicas de inteligencia de negocios, puede hacer una contribución muy significativa a las ganancias de una empresa, además de aumentar sus ingresos y optimizar el total de sus activos.

1.1 Antecedentes

COFIMAR S.A. es una empresa del grupo corporativo CPK, la cual se dedica a procesar camarón cultivado en piscinas con las siguientes presentaciones: camarón entero crudo congelados en bloque, camarón cola crudo congelados en bloque, camarón cola pelado y desvenado crudo congelados en bloque, los cuales cumplirán los lineamientos de calidad establecidos. COFIMAR S.A. es una empresa de reciente fundación con aproximadamente 8 años de funcionamiento, pero sus directivos son pioneros de la industria camaronera y han estado trabajando en el área de la acuacultura por más de 40 años, tanto en el área de producción como de exportación de camarón. Actualmente la empresa comercializa sus productos en varios países de Asia y Europa, y está logrando su objetivo principal, el cual es el de convertirse en una de las empresas con más éxito dentro del sector de la industria camaronera.

El grupo es dueño de más de 2000 hectáreas de camaroneras que han sido certificadas por el Instituto Nacional de Pesca, así como los laboratorios de donde provienen las larvas que se utilizan para la siembra, manejados bajo estándares orgánicos, utilizando probióticos e insumos biodegradables, para garantizar la calidad, y confiabilidad de sus camarones en toda la cadena de valor desde su inicio hasta llegar al consumidor final. Con ese propósito, COFIMAR S.A. se encuentra comprometida en un programa de expansión internacional con producción propia orientada a una filosofía de mejora continua a nivel organizacional.

Misión. _

COFIMAR S.A. tiene como misión producir, procesar y comercializar productos pesqueros y acuícolas que cumplan con los más altos estándares internacionales en lo que respecta a calidad e inocuidad y que satisfagan plenamente las expectativas de todos los clientes.

Esta misión se cumplirá mediante la capacitación constante del talento humano de la empresa para desarrollar habilidades en todos los procesos tecnológicos y de tal manera garantizar una adecuada relación con sus proveedores y clientes.

Visión. _

Ser una empresa líder a nivel nacional e internacional, reconocida y posicionada en el ámbito pesquero y acuícola, por su excelencia en la calidad e inocuidad de sus productos, generando en nuestros clientes un alto grado de satisfacción empresarial.

Valores. _

Su cultura organizacional está orientada con valores corporativos que enmarcan su labor empresarial. Todos los colaboradores que conforman COFIMAR interiorizan estos valores y lo hacen parte de sus acciones diarias para contribuir al desarrollo y éxito de la organización a través de:

- CALIDAD DE NUESTROS PRODUCTOS

- SUSTENTABILIDAD AMBIENTAL
- MEJORA CONTÍNUA
- INTEGRIDAD Y LIDERAZGO
- MOTIVACIÓN Y TRABAJO EN EQUIPO
- RESPONSABILIDAD SOCIAL

Objetivo organizacional. _

Actualmente está comercializando sus productos en varios países, y está logrando su objetivo principal:

"Convertirnos en una de las empresas con más éxito dentro del sector de la industria camaronera"

COFIMAR comparte la actual filosofía comercial mundial basada en la consideración de las cuestiones medioambientales en la producción de camarón y su comercialización. La evaluación técnica de productos está basada en sus normas de filosofía comercial y en la promoción de fuentes sostenibles dentro de los límites de convenios internacionales. La organización siempre está en busca de cumplir con los estándares que promuevan el bienestar de la industria, con el fin de que futuras generaciones puedan continuar con una cadena de comida sostenible.

Este proyecto de investigación tiene como finalidad analizar los flujos de mercancías de la bodega de COFIMAR S.A. para evitar el desabastecimiento involuntario y disminuir los altos niveles de stock que conllevan o se traducen en altos costes de almacenamiento,

para en lo posterior a esto sugerir la implantación de un sistema para la gestión de inventarios el cual permita identificar: la cantidad de unidades que deben solicitarse, en qué momento realizar el pedido, que artículos merecen atención especial, cuáles son las frecuencias de consumo de estos entre otras cosas.

1.2 Descripción del problema

Altos costos, niveles de inventarios elevados en ciertos ítems y desabastecimiento en otros insumos durante los cierres de mes en la bodega de insumos para producción de COFIMAR S.A. El problema se origina básicamente debido a que esta unidad de negocio no dispone de un sistema para gestión de inventarios, que les permita a los administradores de esta área obtener la información necesaria que pueda proporcionar una visión global de su panorama al momento de realizar la toma de decisiones, en lo que respecta o concierne a los niveles de compras y puntos de reabastecimiento.

1.3 Pregunta problemática

¿Cómo disminuir los costos de almacenamiento y el desabastecimiento de insumos en la bodega de la empresa COFIMAR S.A.?

1.4 Idea a defender

Para ofrecer una solución al problema anteriormente descrito, se propone utilizar técnicas de inteligencia de negocios que permitan identificar insumos que no se utilizan con mucha frecuencia, además de proyectar correctamente las compras para disminuir el costo de almacenamiento y evitar el desabastecimiento en la bodega de la empresa COFIMAR S.A.

1.5 Objetivo general

- Desarrollar un sistema de inteligencia de negocios aplicando un proceso ETL y técnicas de administración de inventarios para disminuir los costos de almacenamiento y el desabastecimiento de insumos en la bodega de la empresa COFIMAR S.A.

1.6 Objetivos específicos

- Analizar los referentes teóricos relacionados a la inteligencia de negocios, la aplicación de procesos de ETL y a la utilización de técnicas de administración de inventarios en bodegas.
- Modelar un proceso de extracción, transformación y carga de datos.

- Aplicar técnicas de administración de inventarios en el desarrollo de la plataforma de software de análisis y visualización de datos.
- Desplegar la solución propuesta en un escenario simulado de pruebas controlado.
- Validar la solución propuesta aplicando métodos cuantitativos.

1.7 Justificativos del proyecto

Criterio de selección de la unidad de análisis seleccionada. _

El inventario de mercancías garantiza la disponibilidad de los materiales requeridos, que aseguran la continuidad operativa de la organización. Por lo tanto, se deben emplear estrategias de control y análisis para reposición del inventario, estas estrategias deben consistir en hacer seguimiento de los bienes almacenados de un negocio, monitorear variables tales como el peso, las dimensiones, la cantidad y la ubicación, ayudará a los directivos de los negocios a saber cuándo es el momento de realizar una reposición de productos o comprar más material para el proceso de manufactura o servicio.

Si no se maneja correctamente el inventario de una bodega puede resultar que el negocio pierda dinero en ventas potenciales que no pueden satisfacerse o que malgaste dinero teniendo demasiado inventario (José Bardo, Marzo del 2018). Un sistema de gestión de inventario puede prevenir que ocurran este tipo de errores, además una adecuada gestión o administración de inventarios a más de ayudar a los negocios a mantenerse organizados, también puede ayudar a:

- Mantener el negocio rentable

- Reducir costos
- Alcanzar economías de escala
- Analizar patrones de venta para predecir ventas futuras
- Analizar el comportamiento contra los competidores
- Preparar el negocio para lo inesperado.

Por lo tanto, es de vital importancia para los negocios contar con un inventario adecuadamente administrado y controlado. El enfoque en esta área debe ser el de mantener un nivel óptimo de mercancías para no generar costes innecesarios ya que, además el tener niveles óptimos de inventario puede ayudar a liberar flujo de efectivo.

COFIMAR S.A. cuenta con un módulo de inventarios en su ERP para el almacenamiento diario de datos (insumos en general para fines de producción excluyendo los repuestos que utiliza el departamento de mantenimiento y la flota de vehículos) el cuál no está desarrollado para generar indicadores ni realizar operaciones de cálculo con los datos almacenados. Además, la empresa no posee un sistema informático automatizado para consulta de datos históricos y generación de indicadores acerca de las actividades de gestión de inventarios efectuadas en las transacciones diarias que realiza este departamento, es decir que estamos ante una evidente necesidad insatisfecha y una muy buena oportunidad de mejora.

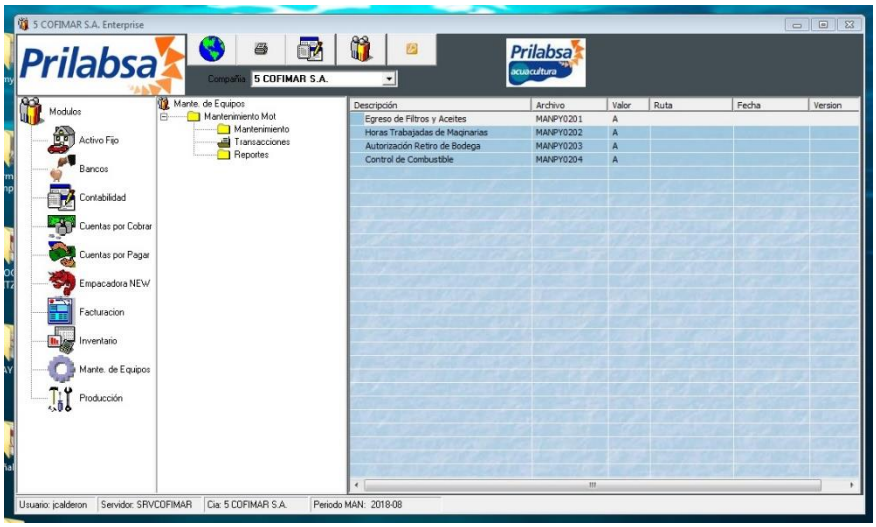


Gráfico #1, Captura de pantalla del sistema ERP de Cofimar S.A.
Fuente: Estación de trabajo departamento de TI del grupo corporativo CPK

Los procesos de gestión de inventarios determinarán qué información deberá estar registrada en el sistema y cuáles serán los reportes y documentos que este debe generar. El módulo para la Administración de inventarios del ERP empresarial no ha sido desarrollado por el departamento de TI en su totalidad y deberá cumplir con una serie de requisitos como mínimo a fin de satisfacer las necesidades del departamento de bodega de la empresa COFIMAR S.A. para temas de trazabilidad y toma de decisiones, el mismo debe ser evaluado por una persona afín al área en cuestión (bodega).

La ausencia de control y planificación estratégica en los inventarios, puede originar consecuencias negativas para los negocios o cualquier tipo de organización, como la pérdida de ventas, cuellos de botella en la producción, retrasos en temas logísticos, pérdidas financieras, máquinas fuera de servicio y demás eventualidades.

Por lo tanto, con el fin de prevenir este tipo de eventos negativos no deseados, es de suma importancia mejorar, organizar y proyectar los procesos de manejo de inventarios de una manera óptima y eficiente.

1.8 Metodología

Se procederá con la extracción de los datos del sistema ERP de la empresa, luego con la replicación de estos datos a un servidor equipado con un motor gestor de base de datos relacionales RDBMS se utilizará MySQL, en conjunto con su transformación en un formato de lectura compatible con el motor gestor de base de datos haciendo uso de herramienta software o aplicativo ETL para realizar este proceso, posterior a ello proceder con la carga de estos datos en un aplicativo de inteligencia de negocios donde serán transformados en la información que se necesita obtener y finalmente a esta información se le dará tratamiento en una herramienta software aplicando técnicas administrativas para la gestión inventarios que permitan a los administradores poder visualizar indicadores de una manera más ágil, práctica y sencilla.

Para realizar este proyecto se plantea el uso de una herramienta de inteligencia de negocios BI, la cual le permite al usuario poder observar la información en forma de tableros de control, cuadros de mando y reportes específicos que se crean a partir de los datos procesados que se obtienen de la data warehouse propuesto. Por lo que en este proyecto se sugiere la adquisición de esta herramienta para la replicación diaria y carga de datos limpios, de tal forma que la información sea presentada de una manera ágil y accesible para que pueda realizarse el análisis e interpretación correspondiente, y así obtener una mejor y oportuna información para mejorar las tomas de decisiones que permitan la obtención de grandes ventajas competitivas.

La herramienta para visualización propuesta será Tableau Desktop, la cual es una poderosa aplicación de descubrimiento y exploración de datos que permite obtener en segundos respuestas a preguntas urgentes. Con Tableau se pueden diseñar gráficos y analizar indicadores desde distintos puntos de vista y con diferentes fuentes de información; simplemente arrastrando variables hacia el entorno de trabajo. Es una de las herramientas de Inteligencia de Negocios BI líder en el mundo en la actualidad, que permite analizar, visualizar y compartir grandes volúmenes de información en forma rápida, es flexible y muy amigable con el usuario final. Es muy fácil de usar, genera impactos tangibles en el corto plazo, mejores decisiones en menor tiempo, identificación de nuevas oportunidades, aumento de ingresos, ahorro de costos y consecución de objetivos.

1.9 Procedimiento metodológico, arquitectura BI

Una solución de inteligencia de negocios BI, parte de los sistemas de origen de una organización, sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico. La arquitectura BI propuesta para esta investigación es la siguiente:

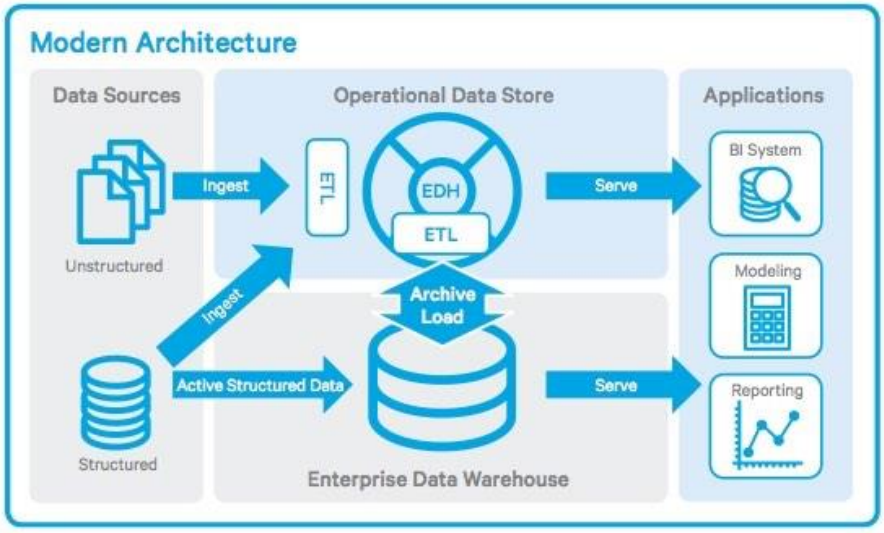


Gráfico #2, Arquitectura BI conceptual propuesta para este proyecto
Fuente: Estación de trabajo departamento de TI del grupo corporativo CPK

Data Sources: fuentes de datos, aquí se registran las transacciones diarias de egresos, ingresos y compras, físicamente son documentos en papel o archivos Microsoft Excel.

Operatinal Data Store: almacén transaccional de datos, físicamente es el modulo para gestión de inventarios del ERP de la empresa, que almacena sus datos en un RDBMS Microsoft SQL Server, es necesario tener acceso a la base de datos transaccional para extraer los datos históricos necesarios.

Enterprise Data Warehouse: almacén corporativo de datos, aquí los datos serán extraídos del almacén transaccional de datos ODS mediante un aplicativo ETL. Estos datos serán analizados en forma más detallada desde aquí, se procederá con la adquisición de un servidor para replicación de los datos transformados, el mismo deberá contener el

resumen consolidado de datos en un motor gestor de base de datos relacionales MySQL previo al tratamiento y técnicas de análisis.

Data Analisis: herramienta software para análisis de inventarios de los datos replicados en el almacén corporativo de datos DW expuesto en las líneas anteriores. Se utilizarán los datos del data warehouse en R Studio haciendo uso de un conector ODBC que servirá de enlace con el motor de base de datos, para en lo posterior aplicar en estos datos un modelo predictivo de minería de datos.

Applications: herramienta software BI para visualización de datos, se entregará la Visualización analítica de los datos con Tableau Desktop de una manera amigable y simple al usuario.

2°) MARCO TEORICO

2.1 Inteligencia de negocios

La inteligencia de negocios es el conjunto de tecnologías, herramientas, procesos y aplicaciones sinérgicas que permiten transformar datos en información, para posterior a esto convertir esta información en conocimiento y el conocimiento orientarlo por medio de acciones hacia un plan estratégico que facilite la toma de decisiones (Procibernetica, 2017).

Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir a la Inteligencia de negocios como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

Este conjunto de herramientas o tecnologías de inteligencia de negocios normalmente muestran sus productos finales de información en forma de cuadros de mando integral o "dashboards" y reportes específicos que se diseñan partiendo de los datos que se obtienen de una fuente ERP que la empresa utiliza para su gestión, de tal forma que la información es presentada al usuario de manera ágil y accesible para que pueda realizarse el análisis e interpretación correspondiente (Blog Corponet, Agosto del 2015).

2.2 Datos

Se conoce que la palabra dato etimológicamente proviene del latín “Dtum” cuyo significado es “lo que se da”.

Los datos son la representación simbólica mediante: números, letras, caracteres o incluso gráficos, tanto cualitativa o cuantitativa, de hechos o condiciones que a su vez describen situaciones que no tienen ningún tipo de información que facilite la deducción de una investigación o un hecho. Lo realmente importante de los datos es que están en la capacidad de asociarse dentro de un contexto el cual ofrece la posibilidad de que estos datos puedan convertirse en información.

2.3 Información

La información se puede definir como un conjunto de datos procesados y que tienen un significado relevancia, propósito y contexto, y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Está constituida por un grupo de datos ya supervisados y ordenados, que sirven para construir un mensaje basado en un cierto fenómeno o ente. La información permite resolver problemas, ya que su aprovechamiento racional es la base del conocimiento.

2.4 Conocimiento

El conocimiento es un conjunto de representaciones abstractas que se almacenan mediante la experiencia o el aprendizaje a través de la observación. En el sentido más extenso que se trata de la tenencia de variados datos interrelacionados que al ser tomados por sí solos, poseen un menor valor cualitativo.

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



Gráfico #3, Pirámide de la información
Fuente:http://www.seguridadinternacional.es/?q=es/content/introducción-la-inteligencia-en-el-ámbito-de-seguridad-y-defensa

2.5 Data warehouse

Es un almacén electrónico donde generalmente una empresa, organización o negocio mantiene una gran cantidad de información. Los datos de una data warehouse deben almacenarse de forma segura, fiable, fácil de recuperar y fácil de administrar (PowerData, referencia de fecha no disponible).

Una data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso, pese a las similitudes logísticas que puedan derivarse de la nomenclatura, se debe tener en claro que se trata de mucho más que un simple almacén de datos. Aunque podría definirse como un espacio de almacenamiento permanente de información, hay que puntualizar que en él se recogen los datos necesarios para apoyar la presentación de informes, el análisis y otras funciones de inteligencia de negocios, recoge todos aquellos que son realmente necesarios para la realización de análisis e informes relacionados con BI.

2.6 Proceso ETL

ETL es un término normalizado que se utiliza para referirse al movimiento y transformación de datos. Se trata del proceso que permite a las organizaciones, empresas o negocios, mover cualquier tipo de dato desde diversas fuentes, limpiarlos, reformatearlos y cargarlos en otro almacén de datos. En este estudio será un data

warehouse la fuente de destino, con el objeto de analizarlos. También pueden ser enviados a otro sistema operacional para apoyar un proceso de negocio, desglosando el acrónimo ETL podemos intuir su significado: Extract-Transform-Load, que significa Extraer-Transformar-Cargar. Las siglas ETL componen una parte muy importante en el mundo de la inteligencia de negocios o BI (BiGeek, Junio del 2018).

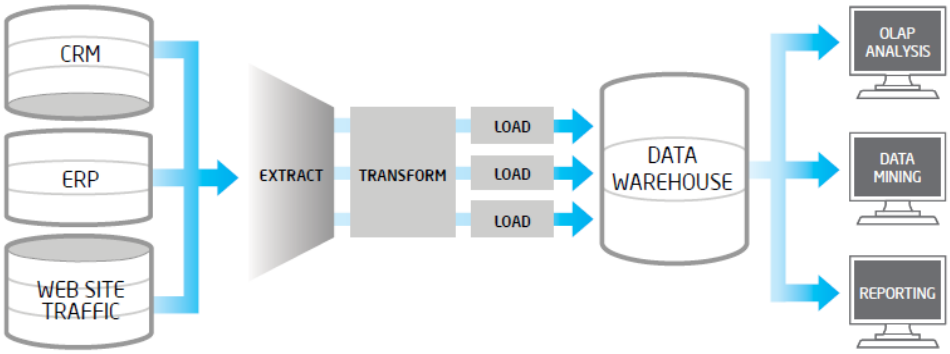


Gráfico #4, Proceso ETL
Fuente:http://www.dataprix.com/blog-it/big-data/consideraciones-procesos-etl-entornos-big-data-caso-hadoop

2.7 Minería de datos

Es la combinación de técnicas estadísticas y tecnologías de las ciencias de la computación, que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos que normalmente no se pueden detectar mediante la exploración tradicional de los datos

porque las relaciones son demasiado complejas o porque hay demasiado datos (Documentos Microsoft, 2017).

El software para minería de datos propuesto en esta investigación es R Studio. R es un entorno de programación libre que se utiliza para el procesamiento y análisis estadístico de datos implementado algoritmos computacionales, que permiten a los usuarios analizar los datos desde muchas dimensiones, ángulos o perspectivas además permite clasificar, y resumir las relaciones o patrones identificados en el análisis de datos.

2.8 EOQ - Cantidad Económica de Pedido

El modelo EOQ o Harris Wilson aborda la mencionada problemática a través del cálculo del tamaño de lote que minimiza los costos de mantenimiento de inventario y colocación de pedidos. Se caracteriza por generar un pedido justo cuando se llega a un nivel específico de inventario en el que es necesario hacer otro pedido. Esto ocurre según la demanda considerada (IngenioEmpresa, Septiembre del 2017).

Antes de desarrollar el modelo de inventarios de la cantidad económica de orden, hay que hacer ciertas suposiciones con respecto a la compra de un solo artículo de inventario. Primero, la demanda del artículo es de una tasa constante y quien es encargado de tomar decisiones deberá conocer esto de forma anticipada. Segundo, se conoce también el tiempo de adelanto, que es el tiempo transcurrido entre la colocación del pedido y su recibo en el inventario, o el tiempo necesario para adquirir el artículo. Aunque estos

supuestos rara vez son válidos en los problemas de inventarios del mundo de los negocios, nos permitirán el desarrollo de un modelo simplificado en el que puedan introducirse perturbaciones de complicación a sistemas más reales.

Existen varias razones para que una empresa mantenga productos terminados o insumos como inventario. El inventario permite enfrentar fluctuaciones de la demanda, evitar quiebres de stock, obtener economías de escala, permite una mayor flexibilidad productiva, se puede usar como un arma competitiva, etc.

Entonces, si mantener inventarios tiene importantes beneficios asociados ¿Por qué no llenamos nuestras bodegas de inventario? Las respuestas son múltiples, pero todas mantienen una base común que ya se la ha mencionado con antelación: **Costos**. Se afirma que mantener inventarios es un "mal necesario" dado los costos asociados a la gestión de inventarios. En este sentido podemos clasificar los costos de inventario en:

- **Costo de Órdenes**: costo que se incurre cada vez que se emite una orden.
- **Costo de mantener Inventario**: arriendo de bodegas, depreciación, costo de oportunidad, pérdidas, seguros, etc.
- **Costo de quiebre de stock:** es más difícil de estimar y está asociado al costo de la venta pérdida (perder un cliente, deterioro de imagen, multas, etc.).

Para el diseño de un modelo de inventarios EOQ se deben cumplir las siguientes premisas:

1. Demanda constante y conocida
2. Un solo producto
3. Los productos se producen o se compran en lotes
4. Cada lote u orden se recibe en un sólo envío
5. El costo fijo de emitir una orden es constante
6. El Lead Time (Tiempo de Espera) es conocido y constante
7. No hay quiebre de stock
8. No existen descuentos por volumen

Para un mejor entendimiento de la metodología se presenta el siguiente gráfico.

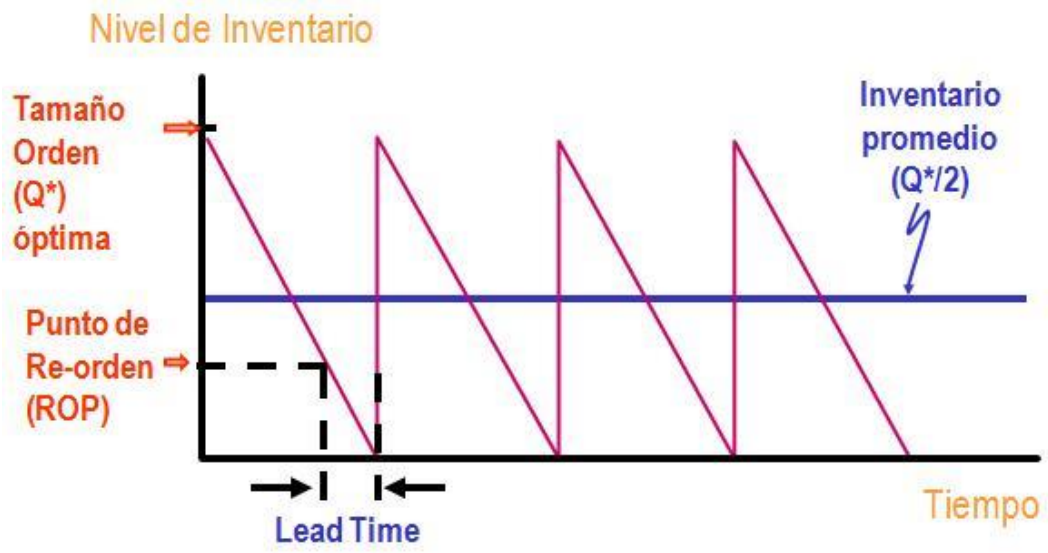


Gráfico #5, Modelo de inventarios EOQ
Fuente: http://www.investigaciondeoperaciones.net/eoq.html

La altura de cada triángulo representa el tamaño óptimo de pedido que minimiza la función de costos totales. La base del triángulo es el tiempo que pasa desde que se recibe

la orden hasta que se termina el lote (este tiempo se conoce como el tiempo de ciclo). Adicionalmente se puede identificar el punto de re-orden que es un nivel crítico de inventario de modo que cada vez que el inventario llegue a ese nivel se hace un pedido de Q unidades. Dado que existe un tiempo de espera (conocido) desde que se emite la orden hasta que se dispone del lote, una vez que se termina el inventario se dispone inmediatamente del nuevo lote y de esta forma no existe quiebre de stock (Investigación de Operaciones, referencia de fecha no disponible).

El modelo para efectos del cálculo matemático considera las siguientes variables:

D: Demanda de unidades por año

S: Costo de emitir una orden

H: Costo asociado a mantener una unidad en inventario en un año

Q: Cantidad a ordenar

$$Q = \sqrt{\frac{2DS}{H}}$$

3°) SOLUCION PROPUESTA

3.1 Análisis de la situación actual

Actualmente las bases de datos se encuentran en línea y funcionando bajo una aplicación de escritorio desarrollada en Visual Basic. Estos sistemas OLAP están siendo utilizados y alimentados diariamente. Se dispone de los siguientes sistemas para el desarrollo de este proyecto:

- Servidor de base de datos Microsoft SQL Server.
- Servidor de base de datos Oracle MySQL.
- Herramienta de inteligencia de negocios para visualización de datos Tableau Desktop, soporte con Microsoft Excel.
- Software R con interfaz gráfica de usuario GUI R Stuido para la Construcción de un Modelo de Predicción.
- IDE o entorno de desarrollo integrado Microsoft Visual Studio.

En los cuartos de servidores hay bastidores para alojar equipos tipo rack, que están sin utilizar y a disposición para instalar sistemas de hardware con las características que necesitamos para este proyecto. El equipamiento para la producción y almacenamiento de la base de datos procesados (servidor data warehouse), y el entrenamiento de los modelos, se deberán alojar en un servidor que disponga con las siguientes características:

Especificaciones técnicas:

Formato	Módulo 2U de ancho completo
Procesadores	Hasta 2 CPUs con procesadores de la familia Intel® Xeon® Scalable por nodo
Memoria	Hasta 768 GB con 24 módulos DIMM TruDDR4 a 2667 MHz por nodo
Expansión E/S	Hasta 3 adaptadores PCIe: 2 PCIe 3.0 x16 + 1 PCIe 3.0 x4 slots
Aceleración	Hasta 4 unidades GPU de doble slot, altura completa y longitud completa (cada PCIe 3.0 x16 slots), o bien hasta 8 unidades GPU de slot sencillo, altura completa y media longitud (cada PCIe 3.0 x8 slots)
Administrador de Interface de redes	1 RJ-45 para gestión BMC 1GbE
Almacenamiento interno	Hasta 8 unidades SSD o HDD SATA de 2,5" hot-swap en las bahías traseras Hasta 2 unidades SSD M.2 no hot-swap, SATA de 6 Gbps en bahías internas
Compatibilidad con RAID	RAID Direct PCH Access Intel RSTe / RAID por software
Administración del sistema	Gestión remota utilizando Lenovo XClarity Controller; NIC de 1 Gb dedicado para gestión
Compatibilidad con sistemas operativos	Red Hat Enterprise Linux 7.5; Windows Server.

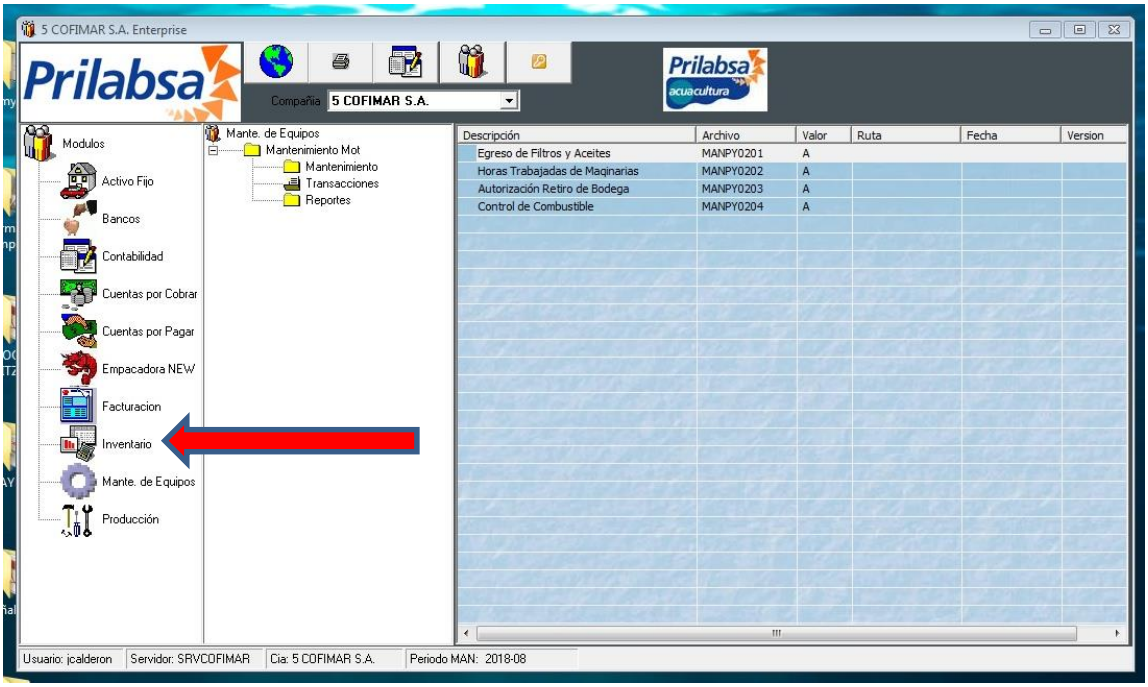


Gráfico #6, Modulo de inventarios que posee el sistema ERP de COFIMAR S.A.
Fuente: Estación de trabajo de COFIMAR S.A.

Usuario: jcalderon
C. de costos: Bodega
Mes: 4
Año: 2018

N° Articulo	Descripción	Saldo Final	Saldo Inicial	Total Egreso	Total Ingreso	Costo al cierre	Total Costo
1 504001	METABISULFITO	270575	372175	101600	0	0,48	128928,99
2 504002	CARNAL	47	62	15	0	86,25	4053,75
3 504003	CAMPHOS	28	36	8	0	70	1960
4 504006	ALTESA PS PLUS	43	6	13	50	66	2838
5 506001	Fundas Venta Local	33379	38081	19702	15000	0,04	1308,46
7 506003	Fundas Master	68122	50782	42660	60000	0,17	11253,75
8 506004	Fundas laminadas Santa Maria (1kilo)	40043	19216	13493	34320	0,15	5822,25
9 506005	Fundas laminadas natural con sellos (IQF transpare	11000	11000	0	0	0,24	2614,7
10 506006	Funda Poliolefina 26x14	33000	44800	11800	0	0,02	706,2
11 506007	Funda Poliolefina 24x14	12800	12800	0	0	0,02	264,96
12 506008	Funda nylon 350 mm x 420 mm x 80 natural(empaque	150	150	0	0	0,23	33,9
13 506009	Funda IQF Pigmentada Azul Baja Densidad Para Empaq	2698	2698	0	0	0,18	482,67
14 506010	Funda Global frozen (1 kilo)	33000	0	0	33000	0,05	1547,7
15 506011	Funda pigmentada azul 24 + FL8 X 3050 X 0.002 (EZE	3596	0	1824	5420	0,24	868,79
16 507011	Ribbon 110x450	216	99	228	345	12,05	2603,1
17 507012	Sellos Blancos	3000	2000	0	1000	0,19	555
18 507013	Sellos de Botella	0	0	0	0	0	0
19 507014	Sellos Plasticos	14000	5000	2000	11000	0,18	2573,2
20 507016	Sellos de cable para bines	1500	1500	0	0	0,23	345
21 507017	Zunchos naranjas	0	0	0	0	0	0
22 507018	Zunchos blancos	0	0	0	0	0	0
23 507019	Zunchos negros	0	0	0	0	0	0
24 507020	Zunchos amarillos	0	0	0	0	0	0
25 507021	Strech film	15	15	0	0	13,67	205

Gráfico #7, Reporte actual que el módulo de inventarios del ERP ofrece
Fuente: Estación de trabajo de COFIMAR S.A.

3.2 Análisis FODA

La Matriz de Planificación Estratégica Cuantitativa (QSPM) es un enfoque de gestión de alto nivel para evaluar posibles estrategias, las cuales permitirán hacer frente de manera anticipada a imprevistos. Esta matriz proporciona un método analítico para comparar alternativas de soluciones factibles, además, intenta seleccionar objetivamente la mejor estrategia utilizando información, técnicas administrativas y algunos cálculos fáciles. En otras palabras, la matriz QSPM utiliza datos de entrada, los compara con resultados obtenidos de un proceso y luego decide objetivamente entre estrategias alternativas.

Es importante tener en cuenta que las estrategias sujetas a comparación, de ser posible deberían ser eventos mutuamente excluyentes.

FACTORES INTERNOS				
Número	Fortaleza/Debilidad	Coeficiente	Ranking	Score
F1	FCPE1 Personal calificado	0.08	4	0.32
F2	FCTI1 Tareas programadas se cumplen al 100%	0.02	4	0.08
F3	FETI1 Sistema de evaluación, costeo y gestión de inventarios	0.05	3	0.15
D1	DDPE1 Una sola persona analiza los datos	0.05	2	0.10
D2	DDTI1 Los sistemas de despachos no están integrados	0.04	1	0.04
D3	DDIN1 No se aprovecha todos los datos históricos disponibles	0.12	1	0.12
D4	DDSE1 No existe actualmente sistema DW y BI	0.09	1	0.09
D5	DCPR1 La clasificación de inventarios se hace sin segmentar ítems	0.09	1	0.09
D6	DCIN1 Los datos pueden ser manipulados	0.09	1	0.09
D7	DCSE1 No existe actualmente sistema DW y BI	0.02	2	0.04
D8	DEPR1 Los procesos son lentos	0.15	1	0.15
D9	DEIN1 Mucho trabajo manual en Excel	0.06	2	0.12
D10	DESE1 No existe actualmente sistema DW y BI	0.02	1	0.02
D11	DIPR1 Hay resultados que no son coherente con los inventarios consolidados	0.03	2	0.06
D12	DISE1 No existe actualmente sistema DW y BI	0.09	2	0.18
	TOTAL	1		1.65

FACTORES EXTERNOS				
Número	Oportunidades/Amenazas	Coeficiente	Ranking	Score
O1	OPNC1 Mercado muy pequeño	0.03	3	0.09
O2	OECL1 Demanda baja de ciertos insumos utilizados en el proceso de producción	0.06	4	0.24
O3	OERI1 Empresa top 10 de exportaciones de camarón en el pais	0.09	4	0.36
O4	OESO1 Cambios en el estilo de vida y ocupacional.	0.06	4	0.24
O5	OSPR1 Asistencia técnica a sectores camaroneros	0.06	3	0.18
O6	OSSO1 Apoyo económico a fundaciones	0.02	3	0.06
O7	OTPR1 Mercados locales mal atendidos	0.09	3	0.27
O8	OTCL1 Uso de tecnología puede mejorar las eficiencias de las operaciones	0.20	4	0.80
O9	OTRI1 Mejor tecnología que la competencia	0.05	4	0.20
O10	OLPS1 Alianzas estratégicas con otras marcas	0.02	3	0.06
A1	APRI1 Cambio de las leyes tributarias y/o laborales	0.02	2	0.04
A2	AEPR1 Precios altos de los insumos	0.05	2	0.10
A3	AEPS1 Disminución de poder de compra de los consumidores	0.06	2	0.12
A4	ASCL1 Competencia agresiva de otras empresas exportadoras de camarón	0.12	1	0.12
A5	AAPR1 No tienen politicas verdes	0.02	1	0.02
A6	ALPR1 Normas IFS y BASC	0.02	2	0.04
A7	ALAM1 Regulaciones ambientales	0.03	1	0.03
	TOTAL	1		2.97

Tabla #1, Análisis FODA para el proyecto
Fuente: Elaboración del autor, basada en la metodología de la matriz QSPM

La importancia de esta matriz radica en que nos hace un diagnóstico real de la organización o negocio, nos dice cómo estamos actualmente, cómo vamos evolucionando y nos brinda el principal elemento requerido para tomar decisiones, conocer la realidad, pues cuando conocemos nuestra realidad, casi siempre sabemos qué hacer. El desarrollo de la matriz (QSPM) consiste en evaluar de manera subjetiva, en seis pasos él ponderado de los factores considerados como críticos del análisis interno y externo:

- Asignar los valores ponderados a cada uno de los valores críticos considerados según la importancia que tenga en la organización.
- Cada objetivo tendrá tres columnas, la primera columna es el peso asignado de forma porcentual (coeficiente), la segunda representa la calificación del impacto del factor (ranking) y en la tercera el valor resultante de la multiplicación del ponderado (score) también se la llama la columna de valor.
- El peso de cada factor va desde 0.0, menos importante a 1.0 muy importante, hay que recalcar que el peso indica la importancia del factor y su influencia para alcanzar el éxito en el sector. Además, a cada objetivo se le debe asigna una calificación de impacto de la siguiente manera: 0 no relevante, 1 bajo impacto, 2 moderado, 3 alto, 4 crítico.
- Para cada objetivo se tendrá que multiplicar el peso ponderado (coeficiente) por la calificación del impacto (ranking), hay que considerar que la suma de todos los coeficientes deberá dar como resultado 1.0
- Se suman las cantidades obtenidas de la columna de valor (score) para obtener una variable única de importancia del objetivo.

- Ordenar la variable única de importancia en cada objetivo de mayor a menor, lo de mayor valor representa mayor relevancia para la compañía y así se tomará la decisión de implementar (GoConqr, 2019).

Ponderación	Fortalezas	Debilidades	Factores Internos
Oportunidades	Conservadora	Agresiva	1 hasta 2.5
Amenazas	Defensiva	Competitiva	2.6 hasta 4
Factores externos	1 hasta 2.5	2.6 hasta 4	Ponderación

Tabla #2, Tabla para ponderación del análisis FODA
Fuente: Elaboración del autor, basada en la metodología de la matriz QSPM

Estrategia Conservadora		Estrategia Agresiva
Debilidades Oportunidades	 **(1,65 - 2,97)** **Cuadrante de estrategia defensiva**	Fortalezas Amenazas
Estrategia Defensiva		Estrategia Competitiva

Número	Estrategia sugeridas para implementación
ED3	Implementar modelo de predicción con R
ED5	Agrupar datos mediante replicación a un data warehouse para su posterior análisis
ED12	Implementar un sistema de soporte para toma de decisiones basado en BI
EO8	Aplicar técnicas de Análisis & Gestión de inventarios para minimizar costes

Tabla #3, Resultados obtenidos del FODA
Fuente: Elaboración del autor, basada en la metodología de la matriz QSPM

El departamento de bodega actualmente cuenta con tres usuarios los cuales disponen cada uno de un computador personal, pero como se mencionó en líneas anteriores no se disponen de herramientas de análisis que permitan una adecuada toma de decisiones. Para implementar las estrategias presentadas en la tabla #3 se propone el siguiente modelo de arquitectura:

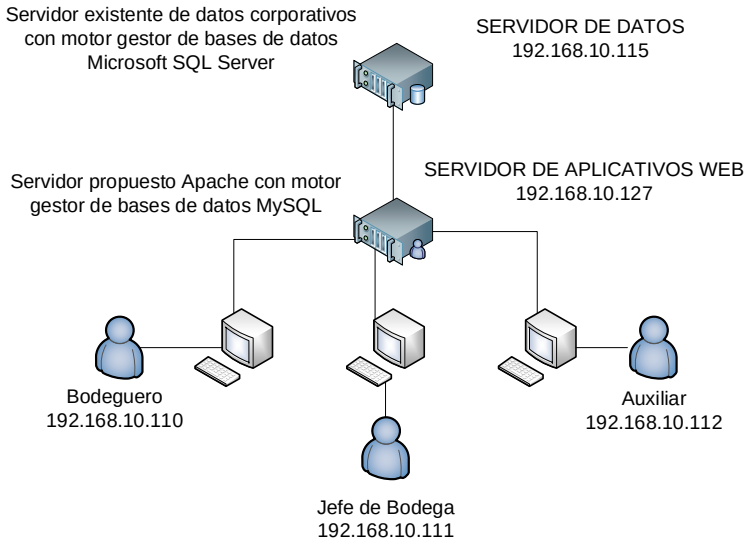


Gráfico #8, Arquitectura hardware propuesta para implementar la solución BI del proyecto
Fuente: Elaboración del autor, con la herramienta Microsoft Visio

3.3 Definición de requerimientos funcionales de la solución propuesta

Identificador (ID) del usuario	Rol	Característica / Funcionalidad	Razón / Resultado
000-001	"Jefe de bodega"	Necesita ver informes gráficos con información relevante de productos, realizar cálculos automáticos de indicadores y pronósticos.	Con la finalidad de evaluar resultados en el proceso de la toma acertada de decisiones.
000-002	"Bodeguero"	Necesita ver la lista con los ingresos, los egresos y los costos de los inventarios de la bodega.	Con la finalidad de tener la información necesaria para la recepción y despacho de materiales.
000-003	"Auxiliar"	Necesita ver un listado de categorías de productos y poder seleccionar una categoría.	Con la finalidad de realizar búsquedas de productos por categorías.

Tabla #4, Definición de requerimientos funcionales y roles de usuarios
Fuente: Elaboración del autor

3.4 Definición de roles y responsabilidades del equipo de trabajo

Persona	Contacto	Rol
Leandro Garzón	lgarzon@cofimar.com.ec	Líder de proyecto
Walter Moncayo	wmoncayo@cofimar.com.ec	Jefe de infraestructura
Anthony Navarrete	anavarrete@cofimar.com.ec	Asistente de hardware
Danilo Rendón	drendon@cofimar.com.ec	Desarrollador

Roles / Responsabilidades	Letra Código
Responsable	**R**
Aprobador	**A**
Consultado	**C**
Informado	**I**

Actividad		Roles / Responsabilidades			
ID Actividad	Actividad	Leandro	Walter	Anthony	Danilo
1	Entrevistas con los expertos	R-A	I	I	I
2	Análisis de datos	C-I	A	I	R
3	Análisis de requisitos	C-I	A	I	R
4	Ingeniería de Sistemas	C-I	A	I	R
5	Programación de los modelos	C-I	R-A	I	C-I
6	Construcción del data warehouse	C-I	I	R	C-I
7	Desarrollo de pantallas de visuales	C-I	I	R	C-I
8	Informes de finalización por fase	C-I	R-A	I	C-I
9	Manuales de usuarios	R	C-I	I	C-I
10	Guías de administración	R	C-I	I	C-I
11	Seguridad informática	I	R-A	I	C-I
12	Análisis de riesgo informático	I	C-I	I	I
13	Acta de entrega & satisfacción	R-A	C-I	I	I

Tabla #5, Definición de roles y responsabilidades del equipo de trabajo
Fuente: Elaboración del autor

3.5 Núcleo de la arquitectura tecnológica de la solución propuesta

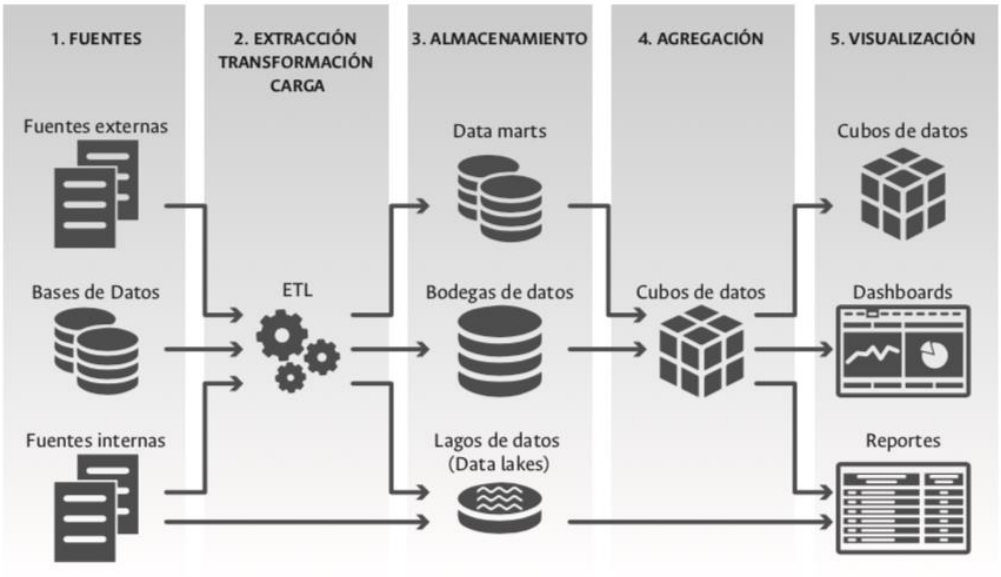


Gráfico #9, Arquitectura BI tecnológica propuesta para este proyecto
Fuente: Estación de trabajo departamento de TI del grupo corporativo CPK

3.5.1 Datos: descripción

Los datos obtenidos del almacén operacional de datos, ODS son los siguientes:

VARIABLES	TIPO DE DATO
Articulo (código)	Numérico
Descripción	Carácter
Saldo Final	Numérico
Saldo Inicial	Numérico
Total Egreso	Numérico
Total Ingreso	Numérico
Costo al cierre	Numérico
Total Costo	Numérico
Mes	Carácter
Año	Numérico

3.5.2 Datos: limpieza

El objetivo de esta tarea es mejorar la calidad de los datos para alcanzar el nivel de calidad requerido por las técnicas de análisis o inferencia que se pretendan implementar en la siguiente fase del proyecto. No podría ser posible lograr un resultado final satisfactorio en un proceso ETL, acorde a los objetos planteados, sino se realiza previamente una buena limpieza de los datos.

Solo se utilizarán las variables o campos contienen información relevante y representativa por su cantidad y diversidad de movimientos. Por este motivo se procedió a eliminar del presente proyecto las demás variables o campos los cuales no agregan valor al proyecto:

- **“N° de registro”** (generado de manera automática por SQL Server)
- **“Costo al cierre”**

3.5.3 Datos: formateo

La tarea de formateado de datos se refiere principalmente a modificaciones sintácticas hechas a los datos que no cambian su significado, pero podría ser requerido por la herramienta de modelado.

La data source o fuente de datos de la que disponemos para el estudio en este proyecto contiene la opción de bajar los datos en varios formatos digitales. Las respuestas están representadas mediante una codificación numérica. Debido a eso no ha sido necesario formatear los datos del almacén de datos operacional el cual nos proporciona números reales para la inserción a la base de datos. Esta actividad también facilitará el modelado de los datos

3.5.4 Integración de datos

La integración de los datos, involucra la creación de nuevas estructuras, a partir de los datos seleccionados, por ejemplo, generación de nuevos campos a partir de otros existentes, creación de nuevos registros, fusión de tablas campos o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen.

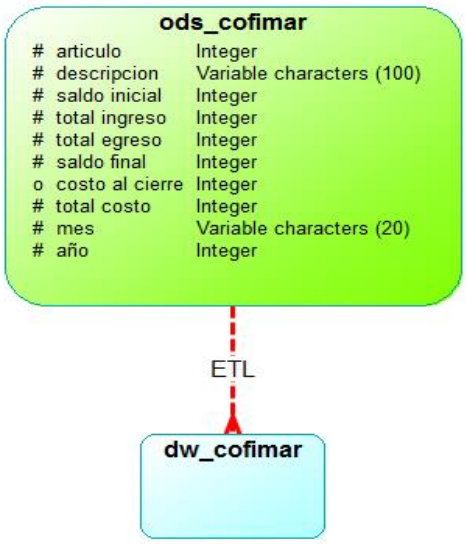


Gráfico #10, Modelado CONCEPTUAL de datos propuesto para el proyecto, importación desde el almacén operacional de datos con nombre ODS_COFIMAR
Fuente: Elaboración del autor, con la herramienta PowerDesigner

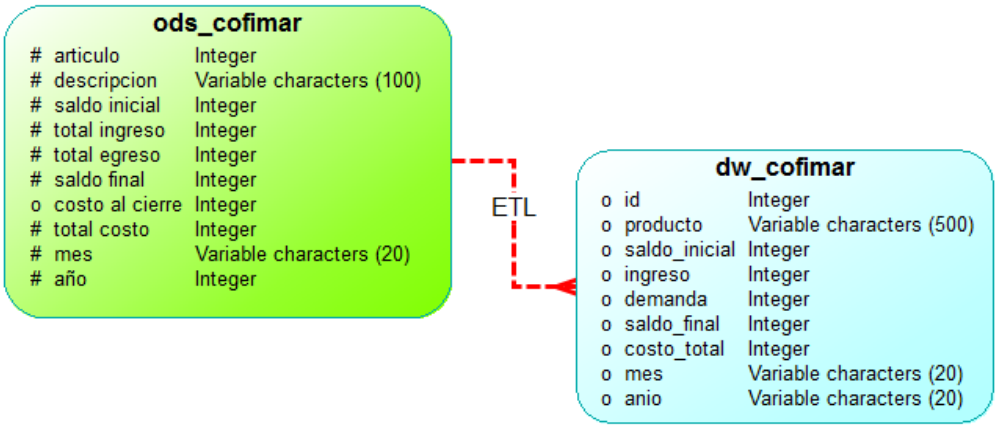


Gráfico #11, Modelado LOGICO de datos propuesto para el proyecto, variables que serán almacenadas al data warehouse de destino con nombre DW_COFIMAR
Fuente: Elaboración del autor, con la herramienta PowerDesigner

3.5.5 Estructura del sistema de la fuente de datos, (ODS)

El origen de los datos es un almacén operacional que contiene el repositorio de las transacciones diarias y lleva el nombre de "ods_cofimar", cuenta con un motor gestor de bases de datos relacionales Microsoft SQL Server.

A continuación, se detalla la codificación SQL necesaria para la creación de la base de datos de la fuente:

```
CREATE TABLE "inv_mes_año" (
  "Articulo" INT,
  "Descripción" VARCHAR (100),
  "Saldo Inicial" INT,
  "Total Ingreso" INT,
  "Total Egreso" INT,
  "Saldo Final" INT,
  "Costo al cierre" INT,
  "Total Costo" INT,
  "Mes" VARCHAR (20),
  "Año" INT
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

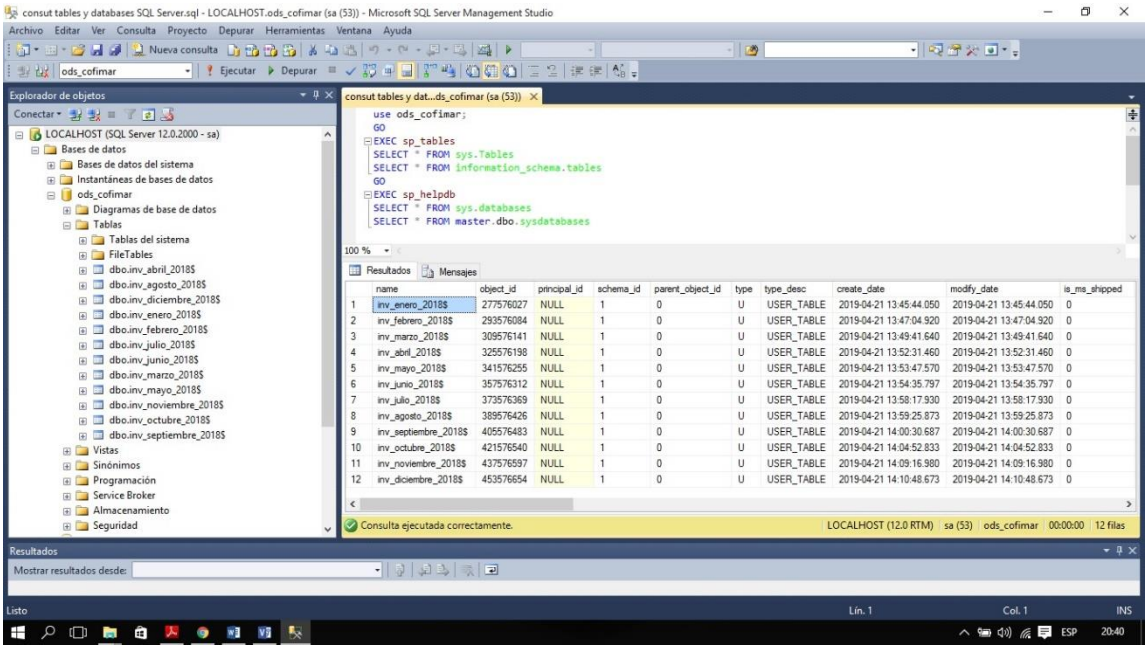


Gráfico #12, Simulación del ODS actual, “ods_cofimar” de la bodega de COFIMAR S.A. Fuente: Elaboración del autor, con la herramienta Microsoft SQL Server

3.5.6 Estructura del sistema de almacenamiento de datos, (DW)

El almacén destino propuesto, es un data warehouse para replicación de datos limpios que deberá disponer de una base de datos de nombre “dw_cofimar”, además deberá poseer un motor gestor de bases de datos relacionales Oracle MySQL Server.

A continuación, se detalla la codificación SQL necesaria para la creación de la base de datos del almacén data warehouse:

```
CREATE TABLE `datos_bodega` (
  `id` int(10) NOT NULL,
  `producto` varchar(500) CHARACTER SET utf8 DEFAULT NULL,
  `saldo_inicial` int(20) NOT NULL,
  `ingreso` int(20) NOT NULL,
  `demanda` int(20) NOT NULL,
  `saldo_final` int(20) NOT NULL,
  `costo_total` int(20) NOT NULL,
  `mes` varchar(20) COLLATE latin1_spanish_ci NOT NULL,
  `anio` varchar(20) COLLATE latin1_spanish_ci NOT NULL
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_spanish_ci;
```

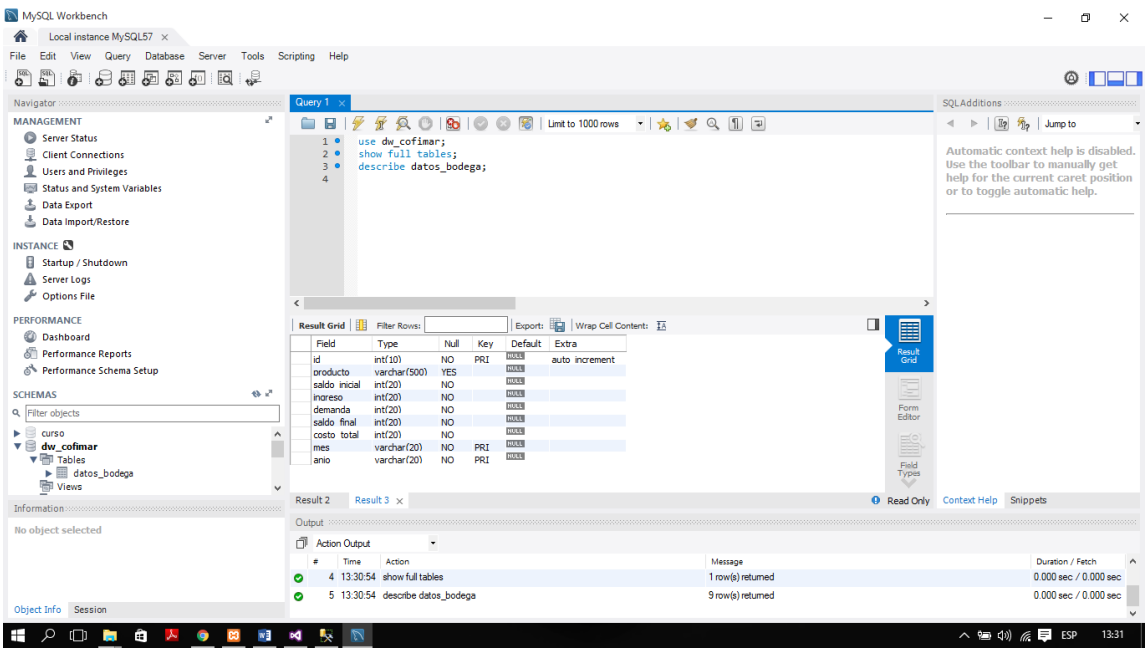


Gráfico #13, Estructura del DW propuesto, “dw_cofimar” para el presente proyecto
Fuente: Elaboración del autor, con la herramienta MySQL Workbench

3.5.7 Diseño del proceso ETL

Para esta parte del proyecto es necesario diseñar correctamente un ETL el cual va a extraer los datos de una fuente de origen en este caso la base de datos “ods_cofimar” SQL Server, para transformarlos y llevarlos a la data warehouse de destino que será la base de datos “dw_cofimar” MySQL, el término ETL como se definió en el apartado 2.5 de este proyecto, es un estándar normalizado que se utiliza para referirse al movimiento y transformación de datos. Se trata del proceso que permite mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos (denominada data mart o data warehouse) con el objeto de analizarlos.

La nueva estructura que se obtendrá producto del proceso ETL, tendrá que ser una colección de datos normalmente de tipo simple y que estos se caractericen por su sencilla organización y la facilidad con que se puedan ejecutar las operaciones que se realicen con ellos. Esta tarea consiste principalmente, en la realización de transformaciones sintácticas de los datos sin modificar su significado, esto, con la idea de permitir o facilitar en lo posterior el empleo de alguna técnica de DM (minería de datos) en particular, como por ejemplo la reordenación de los campos y/o registros de la tabla o el ajuste de los valores de los campos a las limitaciones de las herramientas de modelación; en este proyecto se utilizarán modelos predictivos.

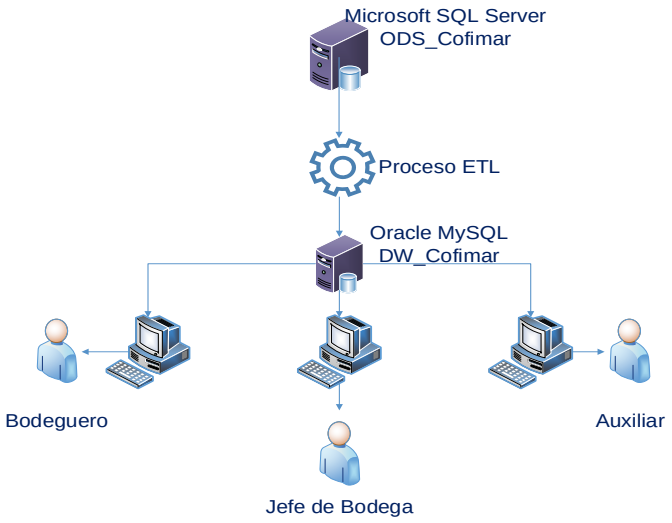


Gráfico #14, Arquitectura ETL propuesta para implementar la solución BI del proyecto
Fuente: Elaboración del autor, con la herramienta Microsoft Visio

Extracción. _

Se procederá a realizar la extracción de la fuente, la cual es el almacén operacional de datos donde se registran las transacciones diarias a través del ERP de la organización.

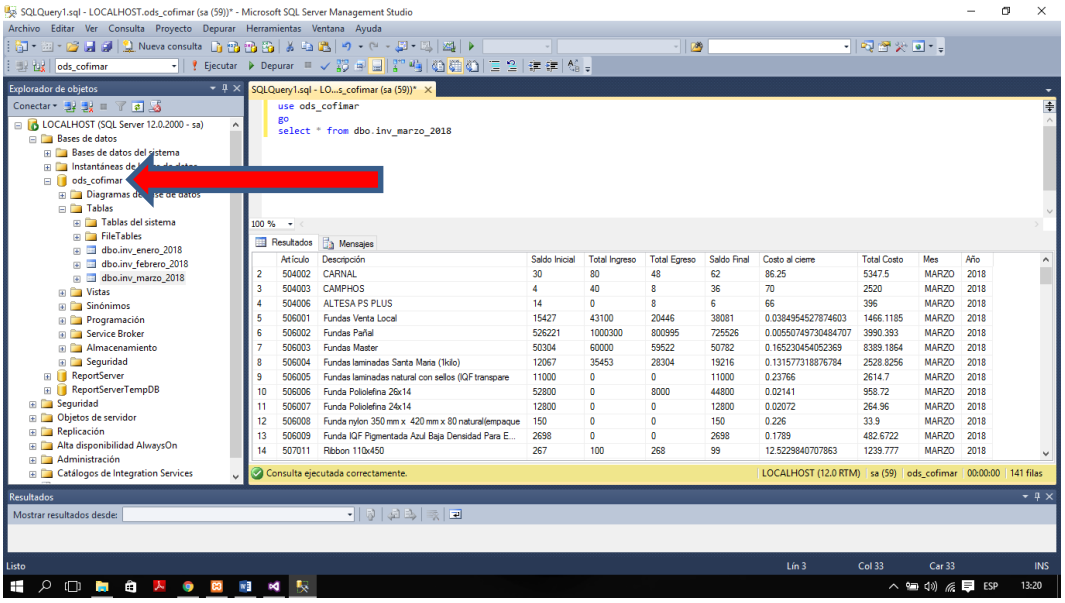


Gráfico #15, Fuente de datos para la extracción, base ODS_COFIMAR
Fuente: Elaboración del autor, con la herramienta Microsoft SQL Server

Transformación. _

Esta será efectuada mediante un proyecto de integración de servicios hecho con la herramienta flujo de datos disponible en el paquete Microsoft Visual Studio.

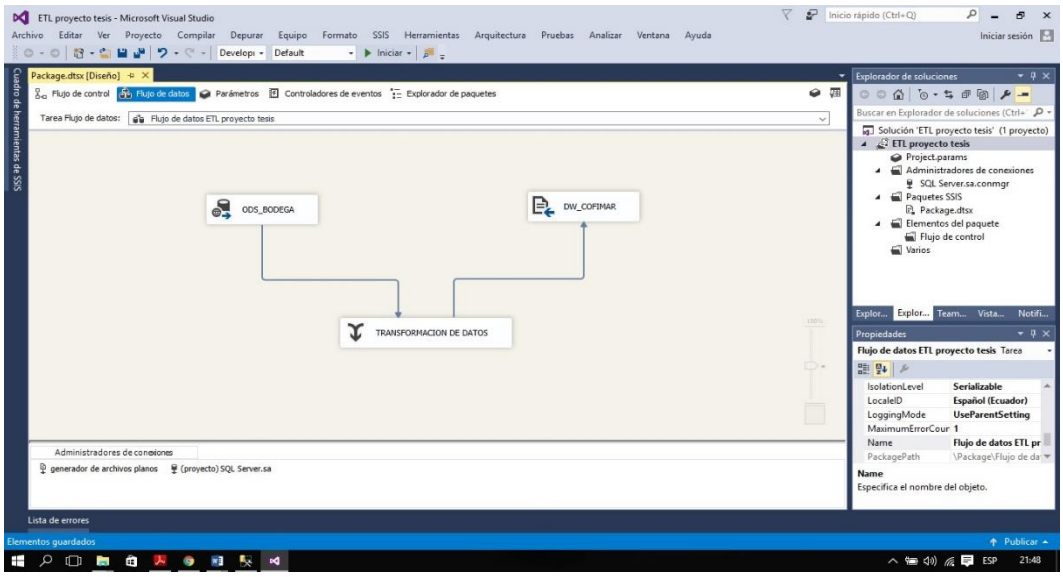


Gráfico #16, Esquema del flujo de datos del proceso ETL
Fuente: Elaboración del autor, con integración de servicios de Microsoft Visual Studio

Este proceso ETL consiste en filtrar los datos de la base de origen SQL Server "ODS_COFIMAR", extrayendo los campos que se van a utilizar, para llevarlos a la base de destino "DW_COFIMAR" desde donde serán procesados y analizados. En el proceso de transformación obtenemos como resultado un archivo plano con los datos limpios y procesados, la carga de estos datos en el servidor data warehouse será efectuada por medio de una aplicación que funciona en ambiente web, en los gráficos 19 y 20 se puede observar el ambiente de trabajo con el logo corporativo de la organización que el usuario podrá visualizar previo a la carga de datos.

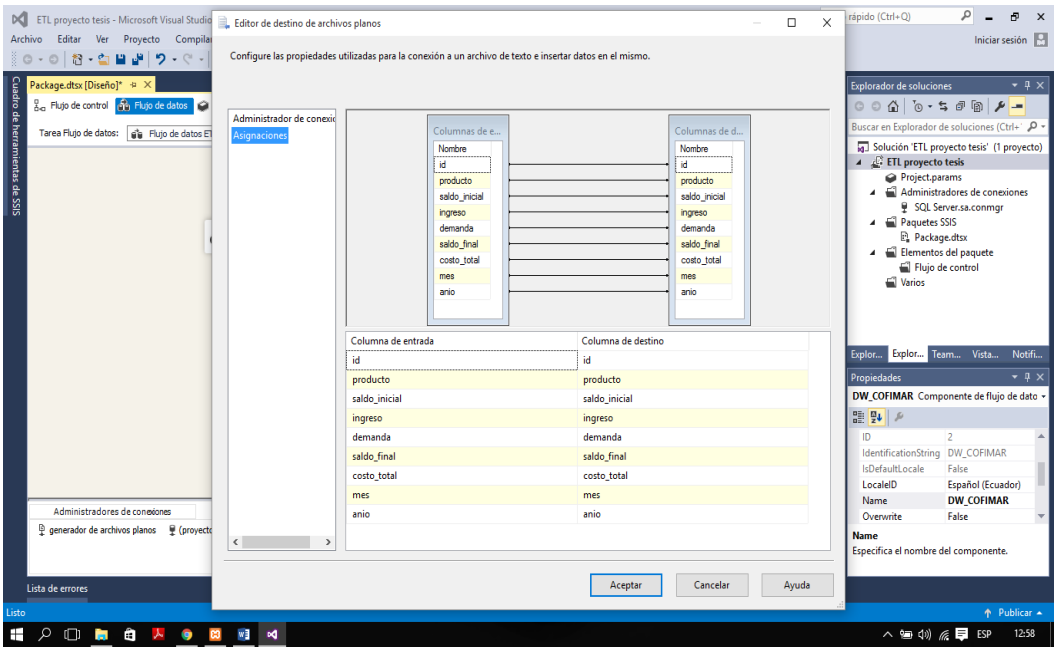


Gráfico #17, Filtrado de datos realizado a través del proceso ETL
Fuente: Elaboración del autor, con integración de servicios de Microsoft Visual Studio

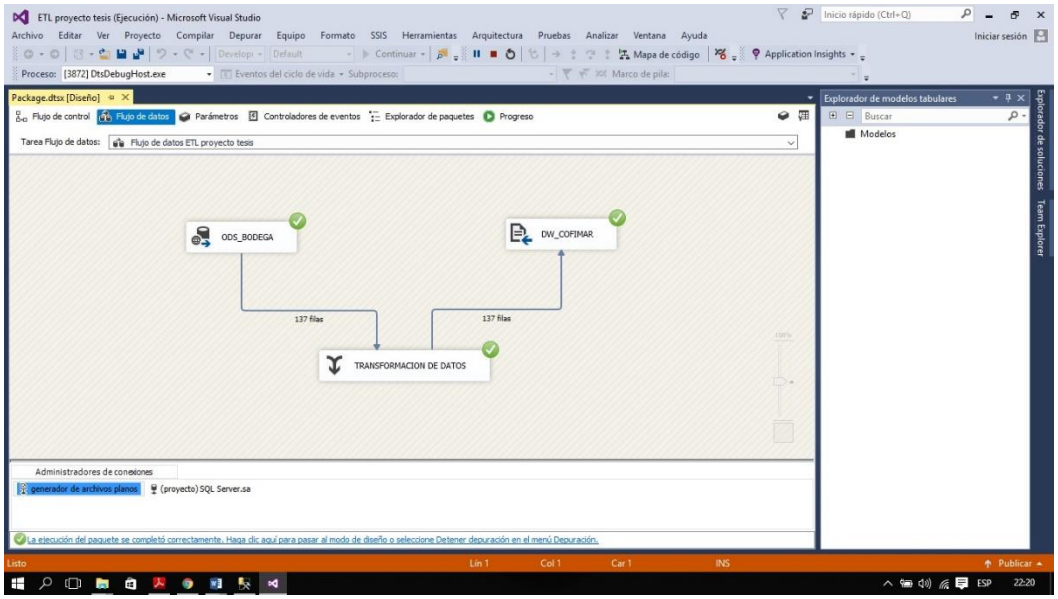


Gráfico #18, Depuración de datos realizada a través del proceso ETL
Fuente: Elaboración del autor, con integración de servicios de Microsoft Visual Studio

Carga._



Gráfico #19, Aplicativo para carga de datos en el data warehouse
Fuente: Elaboración del autor, código desarrollado en php.

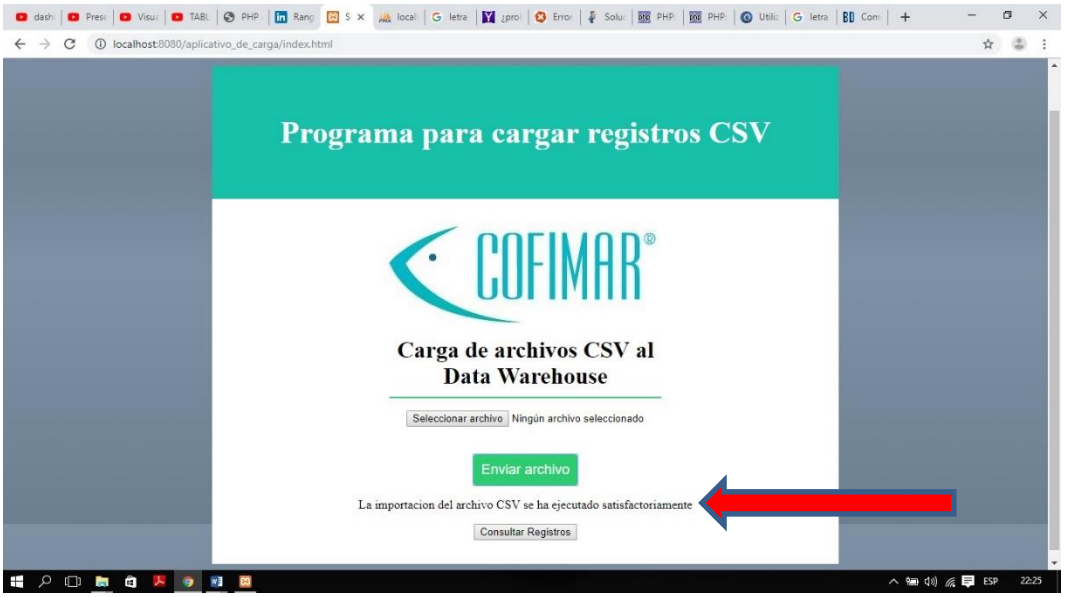


Gráfico #20, Datos cargados en el data warehouse
Fuente: Elaboración del autor, código desarrollado en php

Una vez que el usuario ha seleccionado el archivo plano y ha enviado los datos al servidor de almacenamiento data warehouse, la pantalla del aplicativo mostrara este mensaje de confirmación “La importación del archivo CSV se ha ejecutado satisfactoriamente”, los datos se pueden consultar de dos formas: desde el botón “Consultar Registros” del aplicativo o mediante consulta directa a la base de datos.

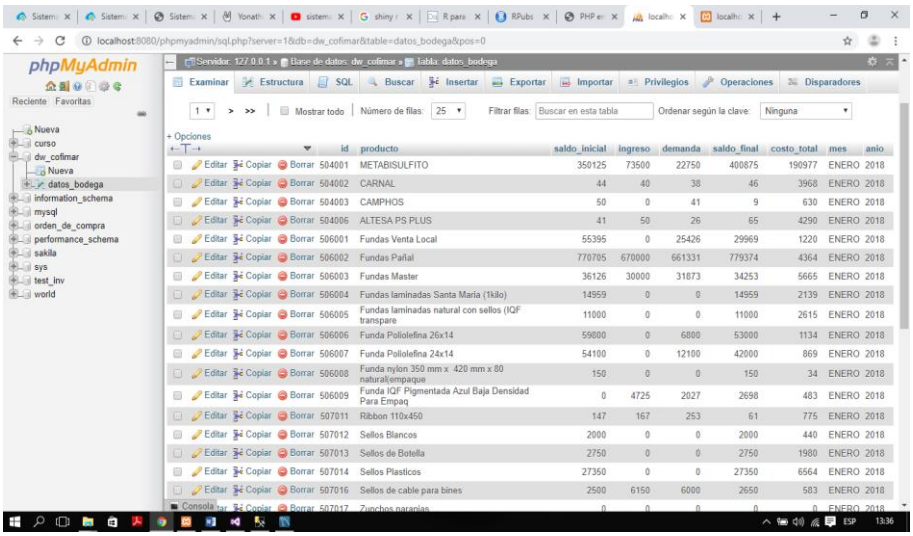
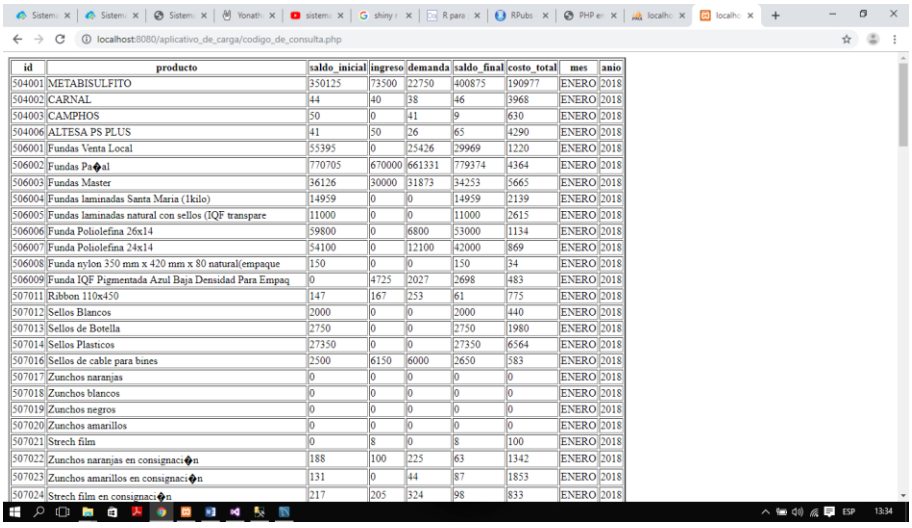


Gráfico #21, Consulta de datos cargados efectuada con el botón “Consultar Registros”
Fuente: Elaboración del autor, código desarrollado en php



id	producto	saldo_inicial	ingreso	demanda	saldo_final	costo_total	mes	anio
504001	METABISULFITO	350125	73500	22750	400875	190977	ENERO	2018
504002	CARNAL	44	40	38	46	3968	ENERO	2018
504003	CAMPHOS	50	0	41	9	630	ENERO	2018
504006	ALTESA PS PLUS	41	50	26	65	4290	ENERO	2018
506001	Fundas Venta Local	55395	0	25426	29969	1220	ENERO	2018
506002	Fundas Pa&#65533;al	770705	670000	661331	779374	4364	ENERO	2018
506003	Fundas Master	36126	30000	31873	34253	5665	ENERO	2018
506004	Fundas laminadas Santa Maria (1kilo)	14959	0	0	14959	2139	ENERO	2018
506005	Fundas laminadas natural con sellos (IQF transpare	11000	0	0	11000	2615	ENERO	2018
506006	Funda Polilefina 26x14	59800	0	6800	53000	1134	ENERO	2018
506007	Funda Polilefina 24x14	54100	0	12100	42000	869	ENERO	2018
506008	Funda nylon 350 mm x 420 mm x 80 natural(empaque	150	0	0	150	34	ENERO	2018
506009	Funda IQF Pigmentada Azul Baja Densidad Para Empaq	0	4725	2027	2698	483	ENERO	2018
507011	Ribbon 110x450	147	167	253	61	775	ENERO	2018
507012	Sellos Blancos	2000	0	0	2000	440	ENERO	2018
507013	Sellos de Botella	2750	0	0	2750	1980	ENERO	2018
507014	Sellos Plasticos	27350	0	0	27350	6564	ENERO	2018
507016	Sellos de cable para bines	2500	6150	6000	2650	583	ENERO	2018
507017	Zunchos naranjas	0	0	0	0	0	ENERO	2018
507018	Zunchos blancos	0	0	0	0	0	ENERO	2018
507019	Zunchos negros	0	0	0	0	0	ENERO	2018
507020	Zunchos amarillos	0	0	0	0	0	ENERO	2018
507021	Strech film	0	8	0	8	100	ENERO	2018
507022	Zunchos naranjas en consignaci&#65533;n	188	100	225	63	1342	ENERO	2018
507023	Zunchos amarillos en consignaci&#65533;n	131	0	44	87	1853	ENERO	2018
507024	Strech film en consignaci&#65533;n	217	205	324	98	833	ENERO	2018

Gráfico #22, Consulta de datos cargados efectuada a la base
Fuente: Elaboración del autor, herramienta para administrar MySQL phpMyAdmin

3.5.8 Visualización y Exploración de datos

La visualización de datos es la representación gráfica de información y datos. Al utilizar elementos visuales como cuadros, gráficos y mapas, las herramientas de visualización de datos proporcionan una manera accesible de ver y comprender tendencias, valores atípicos y patrones en los datos (Tableau, 2019). Esto involucra la aplicación de pruebas de estadísticas descriptivas básicas, que revelen propiedades en los datos recién adquiridos, se crean tablas de frecuencia y se construyen gráficos de distribución. La salida de esta tarea es un informe de exploración de los datos, el análisis exploratorio de datos es el paso previo a la aplicación de cualquier método de análisis inferencial.

Una primera aproximación consiste en reducir toda la información de la muestra a un reducido grupo de valores numéricos, los parámetros estadísticos estimados, cuyo conocimiento puede dar pie a la detección de algún patrón de comportamiento que aporte información sobre la distribución de frecuencias bajo la que se ha obtenido la muestra, es casi preceptivo que en la fase exploratoria las confecciones de gráficos permitan comprender mejor algunos comportamientos de los datos. Se utilizará una visualización para explorar los datos la cual es comúnmente un gráfico, tales como: un gráfico de barras, de anillos, una cronología o un mapa de calor. También puede ser datos en forma tabular, como una tabla de comparación o una tabla dinámica. Cada visualización tiene una consulta subyacente, que es el método que utiliza Analytics para recuperar información a partir de los datos de origen.

El Gráfico 23 muestra todos los registros en función del costo total, de este gráfico se puede extraer la información de que el metabisulfito es el ítem de mayor valor monetario con el que actualmente cuenta la bodega de insumos de producción de COFIMAR S.A.

El gráfico 24 muestra varios registros de ítems de la bodega combinados con algunas variables, específicamente total de egreso, total costo, saldo inicial y total ingreso, de aquí podemos inferir por observación que el ítem fundas pañal es el que mayor cantidad de unidades registra, más aún así se sigue observando que la columna color naranja de coste total es casi imperceptible a la vista siendo notorio el coste total en otra variable.

El gráfico 25 muestra una vista desde otro ángulo todos los registros de la bodega en función del coste total, siendo el 28% la mayor porción de la torta ocupada por el producto metabisulfito, lo cual refrenda la información del mapa de colores del gráfico número cinco de que este producto es el más costoso en esta bodega de insumos.

Finalmente, la gráfica 26 nos muestra un histograma del consumo histórico de varios años de este producto con un consumo mínimo de 14850 Kg, un promedio de consumo de 38982 Kg y un consumo máximo 101600 Kg registrado en el mes de abril del año 2018.

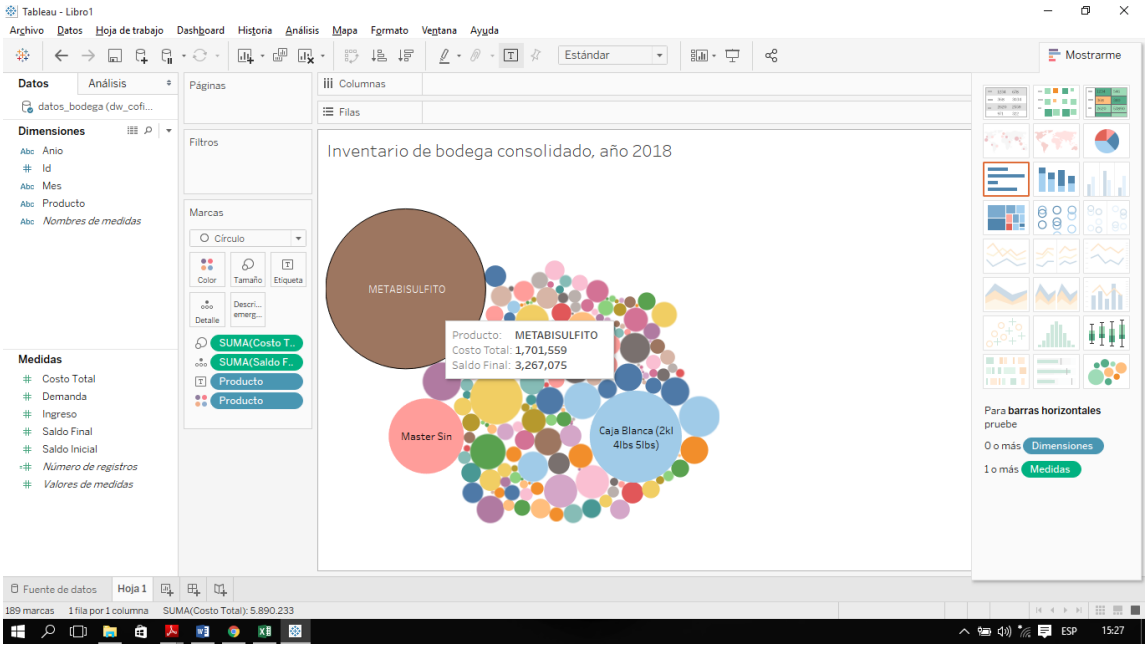


Gráfico #23, Mapa de colores del inventario de bodega consolidado año 2018
Variables: productos, costo total y demanda
Fuente: Elaboración del autor, herramienta Tableau Desktop

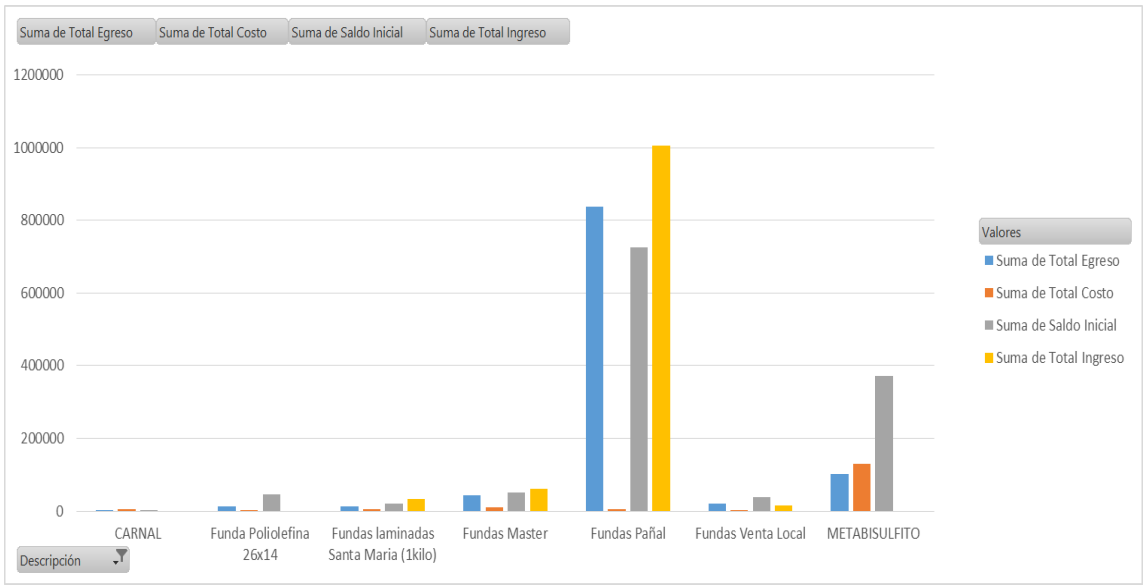


Gráfico #24, Histograma, mes de abril del 2018
Variables: Descripción, Total Egreso, Costo Total, Saldo inicial y total de ingresos
Fuente: Elaboración del autor, herramienta Microsoft Excel

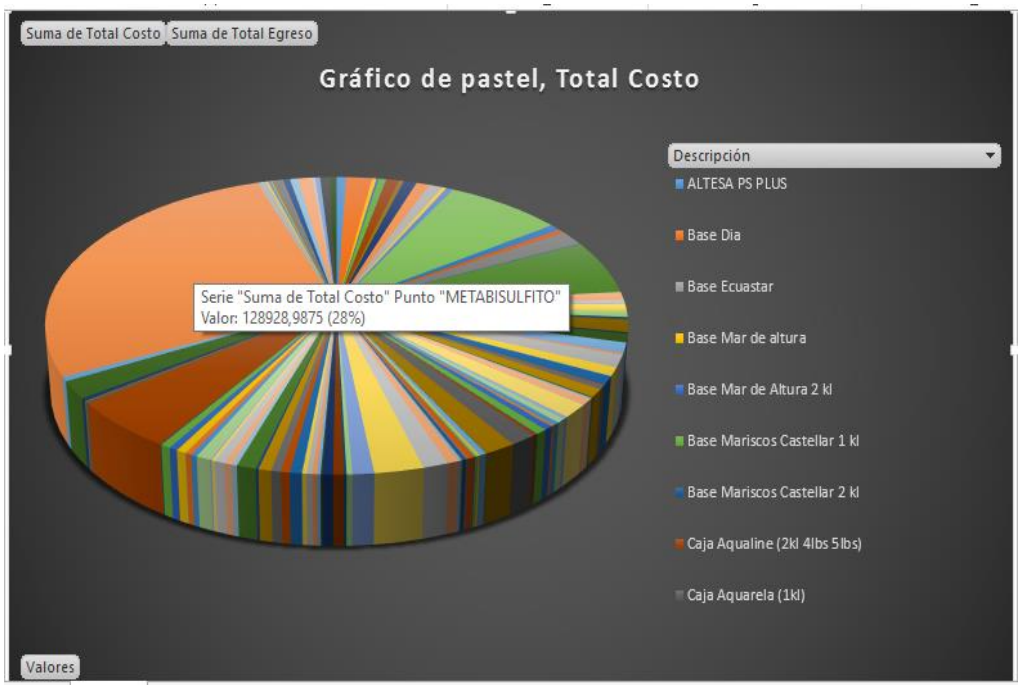


Gráfico #25, Gráfico de pastel, mes de abril del 2018
Variables: Descripción, Total Egreso, Costo Total, Saldo inicial y total de ingresos
Fuente: Elaboración del autor, herramienta Microsoft Excel

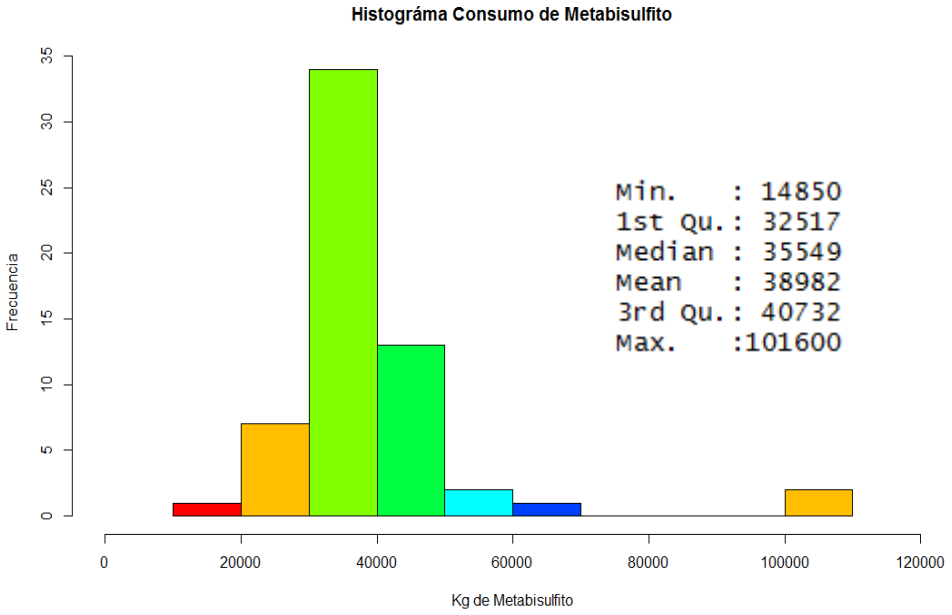


Gráfico #26, Histograma de consumo histórico de metabisulfito.
Variables utilizadas: Total de Egreso de los registros de metabisulfito.
Fuente: Elaboración del autor, herramienta RStudio.

3.5.9 Selección de las técnicas de modelado predictivas

ARIMA._

La palabra ARIMA significa Modelos Auto Regresivos Integrados de Medias Móviles. ARIMA (acrónimo del inglés Auto-Regressive Integrated Moving Average) es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

ARNN._

Redes neuronales autorregresivas ARNN (acrónimo del inglés Auto-Regressive Neural Network) se obtienen al considerar la fusión de un modelo lineal autorregresivo con un modelo MLP (multi linear perceptor). Su desarrollo conceptual inicial está basado en el desarrollo de un contraste estadístico de no linealidad en el que se comparan los dos modelos anteriores. No obstante, la ARNN es una importante alternativa al uso de los MLP en la predicción de series de tiempo debido a la incorporación de la componente lineal auto regresiva, una de las principales herramientas para realizar el análisis y la predicción de series de tiempo es el lenguaje de programación R ya que posee un amplio repertorio de funciones para ese fin; una recopilación de las principales funciones implementadas y paquetes disponibles es presentada.

Para definirlo de una manera muy sencilla, estos modelos de predicción hacen uso de los datos del pasado reciente o más distante para modelar los datos existentes, así como para hacer predicciones adecuadas del comportamiento futuro. El objetivo es identificar un modelo subyacente que explique el cambio en el proceso. Cualquier punto que se desvíe de este comportamiento pronosticado podría considerarse una causa especial, puesto que no sigue los movimientos generales de los datos.

Para el modelado vamos a utilizar el artículo de mayor valor que se almacena en la bodega de esta planta empacadora de camarón, el metabisulfito de sodio el cual es un preservante que se utiliza en los productos alimenticios para evitar la melanosis, que es una reacción enzimática la cual produce oscurecimiento y degradación en la piel y todo el tejido del camarón.

3.5.9.1 Construcción del modelo ARIMA

Procedemos con la exportación de 60 registros (correspondientes a 5 periodos anuales de 12 meses) al ambiente de trabajo en R Studio para aplicar los métodos predictivos de análisis. Lo primero que se realizará previo al análisis será aplicar a la data el formato de serie de tiempos, lo cual no es otra cosa que re organizar la información en meses y años con la función TS del motor R.

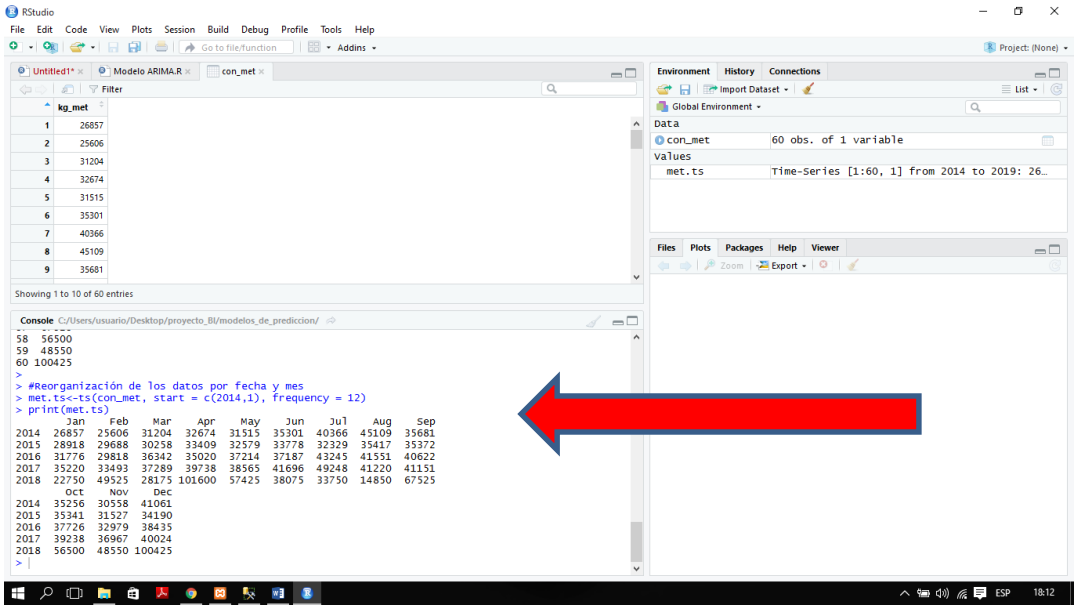


Gráfico #27, Data subida a R Studio con formato de series de tiempo
Fuente: Elaboración del autor, herramienta RStudio

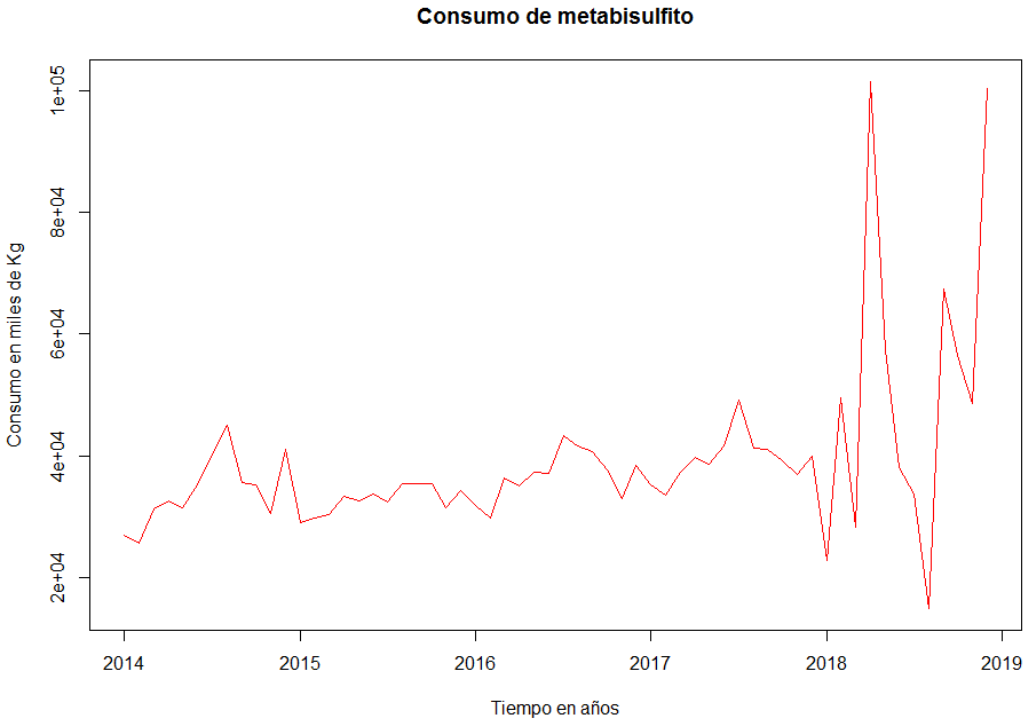


Gráfico #28, Gráfica de tendencia acerca del consumo histórico de la variable
Fuente: Elaboración del autor, herramienta RStudio

Ahora debemos comprender los tres componentes de los datos de una serie temporal:

Tendencia: un aumento o disminución a largo plazo de los datos se denomina tendencia. No tiene por qué ser necesariamente lineal. Es el patrón subyacente en los datos a lo largo del tiempo.

Estacional o Periódico: cuando una serie está influenciada por factores estacionales, es decir, un trimestre del año, mes o días de una semana, la estacionalidad existe en la serie. Siempre es de un período fijo y conocido. P.ej. – Un aumento repentino en las ventas durante el verano, etc.

Cíclico: cuando los datos muestran subidas y caídas que no son del período fijo, lo llamamos patrón cíclico. Por ejemplo – La duración de estas fluctuaciones suele ser de al menos 2 años.

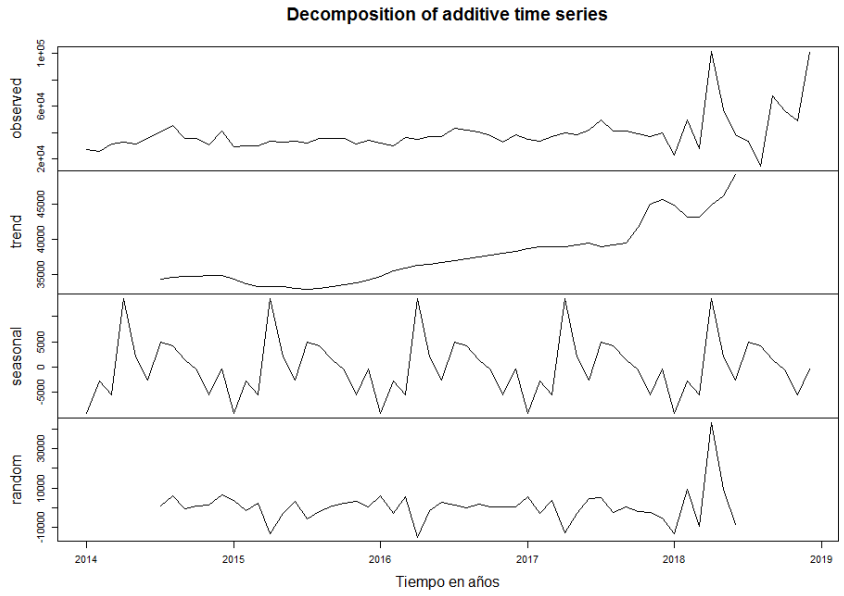


Gráfico #29, Descomposición en serie de tiempos
Fuente: Elaboración del autor, herramienta RStudio

En este gráfico podemos observar 4 sub-gráficos:

Observed: los datos actuales

Trend: el movimiento general hacia arriba o hacia abajo de los puntos de datos

Seasonal: cualquier patrón mensual / anual de los puntos de los datos

Random – parte inexplicable de los datos

Observando de cerca cada uno de estos gráficos, podemos descubrir si los datos satisfacen todos los supuestos del modelado ARIMA, principalmente la estacionalidad y la estacionariedad.

A continuación, debemos eliminar la parte no estacionaria de ARIMA. Para facilitar el análisis aquí, también eliminaremos la parte estacional de los datos. La parte estacional puede eliminarse del análisis y agregarse más tarde, o puede tratarse en el propio modelo ARIMA.

Para conseguir estacionariedad se aplica prueba de raíz unitaria: esta prueba se usa para descubrir la primera diferencia o regresión que se debe usar en los datos de tendencias para hacerla estacionaria. En la prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS), los valores pequeños de la variable sugieren que se requiera una diferenciación.

¿Qué es raíz unitaria?

Una raíz unitaria, es una tendencia estocástica en la serie temporal. Algunas veces se le llama “paseo aleatorio con deriva”. Por tanto, si la serie tiene una raíz unitaria, ésta presenta un patrón sistemático que es impredecible.

Entonces, una serie temporal es estacionaria si un cambio en el tiempo no cambia la forma de la distribución; y las raíces unitarias son una causa de no estacionariedad.

¿Qué es la estacionariedad?

Una serie temporal es estacionaria cuando la media y la variabilidad se mantienen constantes a lo largo del tiempo, es decir, no es en función del tiempo; y además, no presenta tendencia, aquí nos centramos en la media, varianza y auto correlación.

Para eliminar la estacionalidad de los datos, restamos el componente estacional de la serie original (media cero) y luego lo diferenciamos para que sea estacionario (varianza constante e independiente).

Una vez que tenemos los datos listos y han satisfecho todas las suposiciones del modelo, para determinar el orden del modelo que se ajustará a los datos, necesitamos tres variables: p, d y q que son enteros positivos que se refieren al orden de las partes medias auto regresivas, integradas y móviles del modelo, respectivamente.

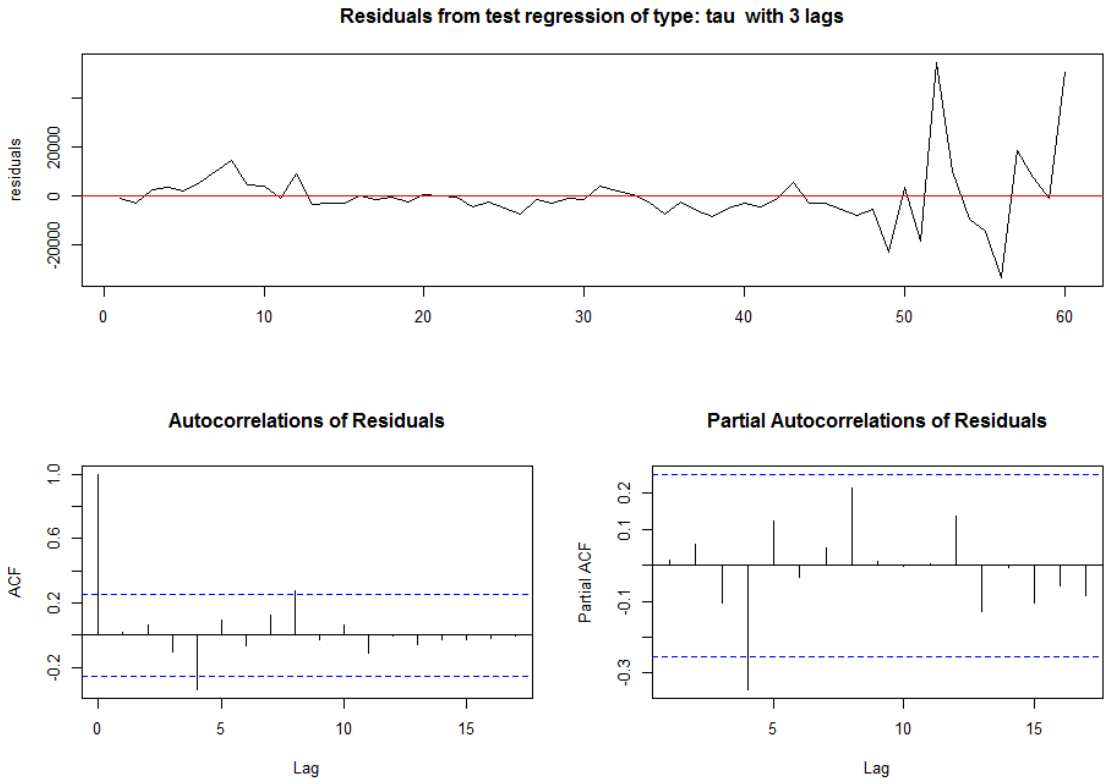


Gráfico #30, Prueba de la raíz unitaria
Fuente: Elaboración del autor, herramienta RStudio

Para identificar que valores de p y q serán apropiados, necesitamos ejecutar las funciones acf() y pacf().

Los parámetros del modelo ARIMA se pueden usar como un modelo predictivo para hacer pronósticos de valores futuros de las series temporales, una vez que se selecciona el modelo más adecuado para los datos de la serie temporal. El valor d afecta a los intervalos de predicción; los intervalos de predicción aumentan de tamaño con valores más altos de "d". Todos los intervalos de predicción serán esencialmente los mismos cuando d = 0

porque la desviación estándar de la previsión a largo plazo será la desviación estándar de los datos históricos, (Diego Calvo, Noviembre del 2018). En este caso el mejor modelo fue un ARIMA$(1,1,1)(1,1,0)_{12}$.

Existe una función llamada predict () que se usa para predicciones a partir de los resultados de varias funciones de ajuste de modelos. Se necesita un argumento n.ahead () que especifica la cantidad de tiempo que se adelanta para predecir.

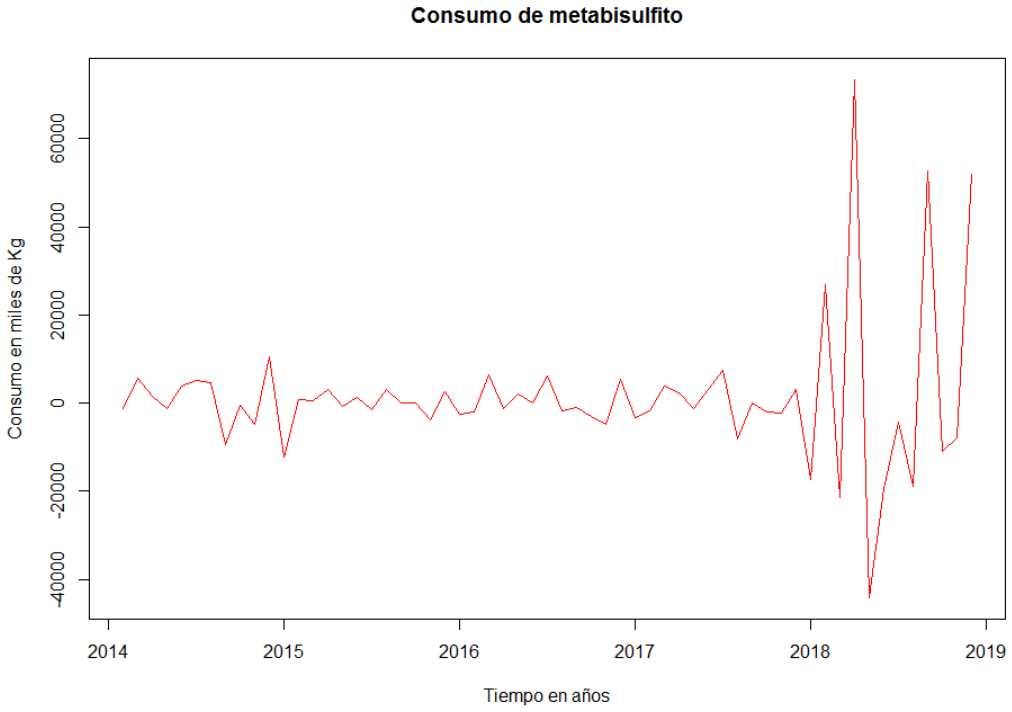


Gráfico #31, Gráfica de tendencia sin estacionariedad en la variable
Fuente: Elaboración del autor, herramienta RStudio

Resultados del Modelo ARIMA

	Jan	Feb	Mar	Apr	May	Jun
2019	53577.74	55718.24	52540.08	63227.65	56795.87	53978.60
	Jul	Aug	Sep	Oct	Nov	Dec
2019	53348.90	50597.15	58266.38	56661.19	55503.71	63056.46

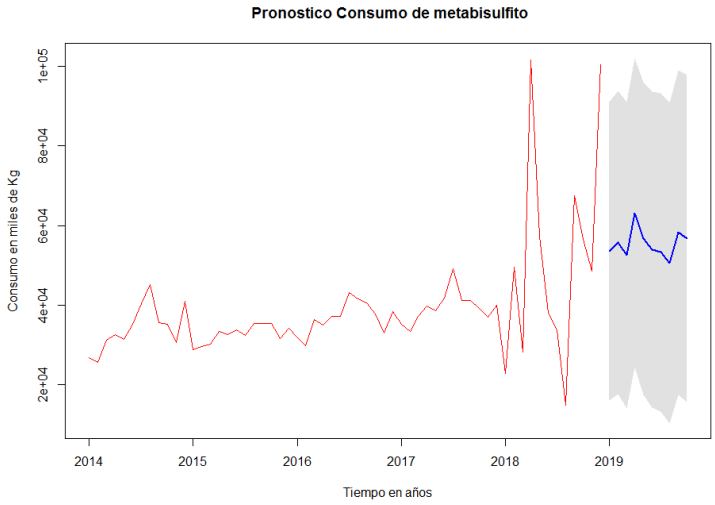


Gráfico #32, Pronostico propuesto por el modelo ARIMA
Fuente: Elaboración del autor

3.5.9.2 Construcción del modelo ARNN

Se procede con la carga de datos en serie de tiempos:

x <- ts(con_met, s = 1, f = 1)

Time Series:

Start = 1

End = 144

Frequency = 1

kg_met

"DATA CONSUMO DE METABISULFITO" 60 registros

Ajuste de una red neuronal MLP._ la red neuronal elegida usa los primeros cuatro rezagos de la serie como entrada y dos neuronas en su capa oculta. La creamos mediante una simple llamada a la función 'arnn' en R, (Juan D. Velásquez PhD, 2011).

```
x <- ts(con_met, s = 1, f = 1)
fit <- arnn(x=x, lags=1:4, isMLP=FALSE, H=2, w.max=1e-3,
  restarts=10, seed = 1234, lambda=0, optim.control=list(maxit=200))
```

Información sobre el modelo ajustado._ la estimación numérica de los parámetros es realizada usando el algoritmo BFGS implementado en la función 'optim' del paquete 'stats', cuya ejecución es controlada con el argumento 'optim.control' de la función 'arnn'. La implementación realizada permite que el algoritmo sea reiniciado varias veces de forma automática, de tal manera que 'arnn' devuelve el mejor modelo encontrado.

Method: arnn

Call:

arnn(x = x, lags = 1:4, isMLP = FALSE, H = 2, w.max = 0.001,

Call:

restarts = 10, seed = 1234, lambda = 0, optim.control = list(maxit = 200))

Sigma^2 estimated as: 168145829.477087

LogL: -567.73370075626

Information Criteria:

AK AKc HQ SC

1171.467 1192.195 1184.932 1206.590

Ahora se procederá a trazar el valor real de la serie (en color negro) y el pronóstico un periodo adelante (en color rojo) en un mismo gráfico, visualizando como se ajusta el modelo estimado a los datos reales.

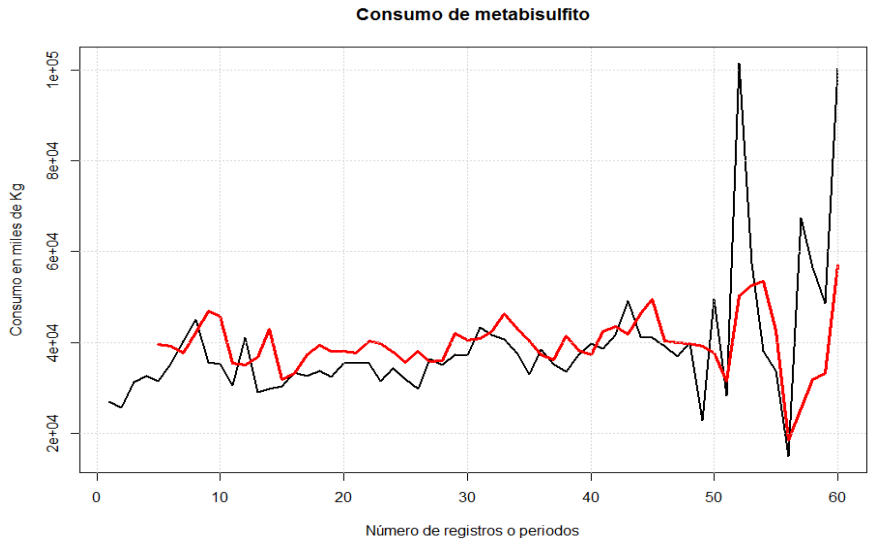


Gráfico #33, Valor real de la serie ━━ Valor del modelo estimado ━━
Fuente: Elaboración del autor, herramienta RStudio

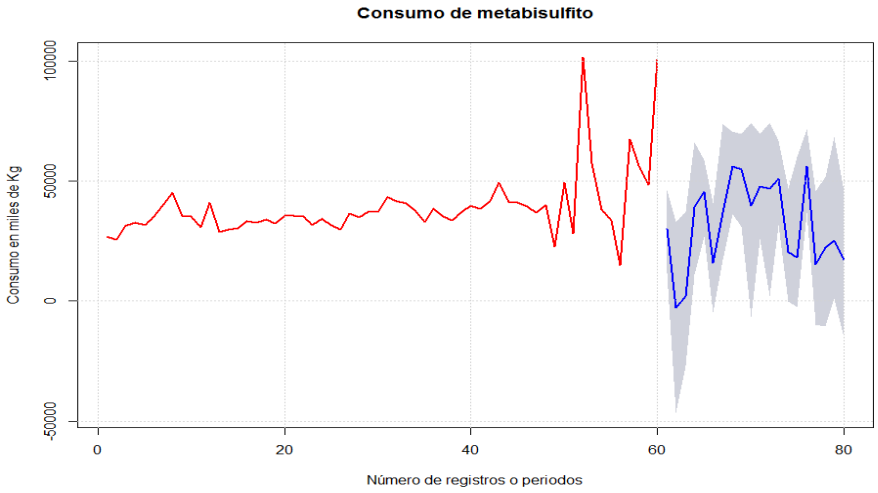


Gráfico #34, Pronostico propuesto por el modelo ARNN
Fuente: Elaboración del autor, herramienta RStudio

A continuación, se procederá con la estimación del pronóstico de consumo de metabisulfito para los siguientes 24 periodos adelante a partir de 1000 series simuladas con intervalos de confianza del 90%. Estos son los resultados del modelo ARNN:

forecast(fit, h=24, level=90, fan=FALSE, bootstrap=FALSE, npaths=1000)

Point	Forecast	Lo 90	Hi 90
61	25187.822	-1985.5107	45354.49
62	14265.543	-29853.0941	45371.80
63	26641.914	-148.0445	48747.89
64	25481.696	3215.1158	68711.12
65	41895.412	27881.3240	71917.13
66	34734.955	15476.3557	73189.14
67	44140.152	24808.5173	62384.89
68	31035.344	-8464.7435	70945.67
69	43908.090	16679.7617	73492.97
70	51632.777	6778.3134	79432.08
71	50966.358	28859.6074	66151.34
72	32222.823	18821.9379	58623.55
73	42154.470	35338.6447	68730.43
74	33915.389	8433.5369	70762.22
75	28059.851	-5092.0201	68058.84
76	19090.628	-17409.2233	64272.00
77	26623.069	10529.9356	60099.48
78	9324.770	-23620.1059	44097.25
79	42177.388	14292.2290	68513.32
80	18014.690	-3923.7406	51817.11
81	22217.238	-11260.6064	73658.10
82	20123.468	-9408.6419	43120.02
83	42781.626	22580.2300	56749.80
84	7362.891	-18160.9258	39785.20

3.5.10 Evaluación de los modelos predictivos propuestos

Una vez ejecutadas las dos técnicas de modelado ARIMA & ARNN, pasamos a la selección del modelo generado que utilizaremos para encontrar y definir el patrón con mayor precisión que sea el más adecuado en la estimación del pronóstico.

Modelos DM	Pronóstico Enero 2019	Error
ARIMA	53577,74	947,26
ARNN	45354,49	9170,51
Enero 2019, Real	54525,00	

En este cuadro podemos observar los resultados para el periodo de enero del 2019, producto del proceso de minería de datos ejecutados, se cuenta con el dato real de consumo de metabisulfito de enero del 2019 y se procedió a calcular el sesgo, desviación o error que ofrece la aplicación de ambos métodos.

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse, además que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se haya posiblemente cometido algún error.

En la gráfica 35 podemos observar que el modelo Auto Regresivo de Redes Neuronales presenta una cantidad de error muy superior al modelo Auto Regresivo Integrado de Medias Móviles, concluyendo que el modelo DM que se debe utilizar para este proyecto es el ARIMA.

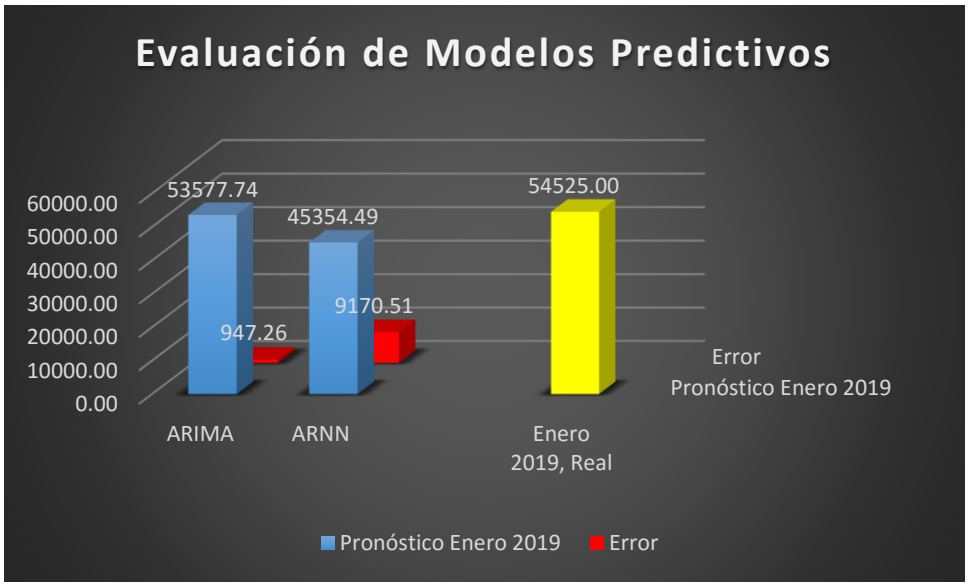


Gráfico #35, Evaluación gráfica de los modelos predictivos
Fuente: Elaboración del autor, herramienta Microsoft Excel

En este punto, el modelo resultante aceptado provee resultados muy adecuados y parece satisfacer necesidades de negocio. Es ahora apropiado hacer una revisión más cuidadosa de las promesas de minería de datos para determinar si hay algún factor importante o tarea que de algún modo ha sido pasada por alto. En esta etapa del ejercicio de minería de datos, el proceso de revisión toma la forma de una Revisión de Garantía de Calidad.

Con un sesgo de error del 1,74% con respecto al valor real el modelo ARIMA nos brinda la garantía en cuanto a calidad de datos para considerar las futuras cantidades de consumo de materiales en la bodega.

3.6 Pronosticar la demanda anual y calcular la cantidad económica de pedido

Utilizando la técnica ARIMA se procederá a realizar el pronóstico para todos los doce meses del año 2019 y así poder efectuar el modelo de cálculo de la cantidad económica de pedido sobre el producto que se escogió para este estudio, el metabisulfito.

Enero	53577.74
Febrero	55718.24
Marzo	52540.08
Abril	63227.65
Mayo	56795.87
Junio	53978.60
Julio	53348.90
Agosto	50597.15
Septiembre	58266.38
Octubre	56661.19
Noviembre	55503.71
Diciembre	63056.46
TOTAL	673271.97

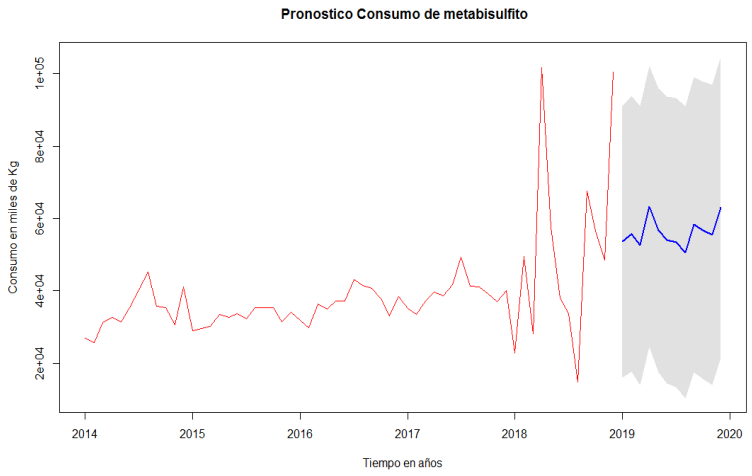


Gráfico #36, Predicción de nivel de inventario, producto Metabisulfito
Fuente: Elaboración del autor, herramienta RStudio

Ítem	Datos	Detalle	Interpretación
Demanda (D)	673272	anual	
Costo de ordenar (S)	500	por orden	
Costo de mantener (H)	0.25	anual por unidad	
Número de días de trabajo (X)	365	por año	
Costo © (C)	0.61	por unidad	
Tiempo de entrega en días	15	por pedido	
Cantidad óptima de pedido Q	51894.97	unidades	Se deben pedir 51894.97 unidades por orden
Número de ordenes (N)	12.97	órdenes	Se realizan 12.97 órdenes de pedido al año
Tiempo entre órdenes (L)	28.133	días	El tiempo entre órdenes es de 28.13 días
Punto de re-orden ® (PR)	27668.71	unidades	Colocar un nuevo pedido en 27668.71 unidades
Costo total (CT)	423669.66	anual	El costo total de inventario es de $423669.66
Costo de ordenar (CO)	6486.87	anual total	El costo anual de ordenar es de $6486.87
Costo de mantener (CM)	6486.87	anual total	El costo anual de mantener es de $6486.87

Tabla #6, Modelo cantidad económica de pedido, producto Metabisulfito
Fuente: Elaboración del autor, datos de costos de inventarios obtenidos de la bodega de COFIMAR S.A., facilitados por el auxiliar de bodega.

$$\boldsymbol{Q} = \sqrt{\frac{\boldsymbol{2DS}}{\boldsymbol{H}}} = \sqrt{\frac{2(673272)(500)}{0.25}} = 51894.97$$

$$\boldsymbol{N} = \frac{\boldsymbol{D}}{\boldsymbol{Q}} = \frac{673272}{51894.97} = 12.97$$

$$\boldsymbol{L} = \frac{\boldsymbol{X}}{\boldsymbol{N}} = \frac{365}{12.97} = 28.133$$

$$\boldsymbol{PR} = \frac{\boldsymbol{D}}{\boldsymbol{X}} * \boldsymbol{tiempo\ de\ entrega\ en\ días} = \frac{673272}{365} * 15 = 27668.71$$

$$\boldsymbol{CT} = \boldsymbol{CO} + \boldsymbol{CM} + (\boldsymbol{D} * \boldsymbol{C}) = 6486.87 + 6486.87 + (673272*0.61) = 423669.66$$

4°) VALIDACION DE RESULTADOS

4.1 Esquema de validación

Se puede concluir que el método ARIMA seleccionado para calcular el pronóstico de la demanda anual de metabisulfito es el más adecuado, se ha constatado que la diferencia entre ambos métodos comparado con un valor real, deja constancia de que menor margen de error presenta el método seleccionado comparado con el ARNN, pero debido a que ambos métodos presentan resultados con valores que tienen similitud podría surgir el planteamiento de la siguiente premisa ¿el producto de ambos métodos es igual?

Para despejar esta interrogante y concluir que son diferentes, el resultado de ambos métodos será sometido a una prueba de hipótesis para probar que no son iguales y que se ha elegido el método predictivo más adecuado. Una prueba de hipótesis es un proceso para determinar la validez de una aseveración hecha sobre una población, es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos. Evalúa dos hipótesis opuestas sobre una población: la hipótesis nula y la hipótesis alternativa, siendo la hipótesis nula la afirmación que se está comprobando.

H_0: Hipótesis nula, métodos ARIMA y ARNN son iguales

H_1: Hipótesis alternativa, métodos ARIMA y ARNN son diferentes

Año 2019	ARIMA	ARNN	Prueba t para dos muestras		
Enero	53577.74	45354.49			
Febrero	55718.24	45371.80		*ARIMA*	*ARNN*
Marzo	52540.08	48747.89	Media	56105.9975	63693.50583
Abril	63227.65	68711.12	Varianza	15143095.08	137102848.2
Mayo	56795.87	71917.13	Observaciones	12	12
Junio	53978.60	73189.14	Varianza agrupada	76122971.62	
Julio	53348.90	62384.89	Diferencia hipotética de las medias	0	
Agosto	50597.15	70945.67	Grados de libertad	22	
Septiembre	58266.38	73492.97	Estadístico t	-2.130183123	
Octubre	56661.19	79432.08	P(T<=t) una cola	0.02229209	
Noviembre	55503.71	66151.34	Valor crítico de t o z (una cola)	1.717144374	
Diciembre	63056.46	58623.55	P(T<=t) dos colas	0.044584179	
TOTAL	673271.97	764322.07	Valor crítico de t o z (dos colas)	±2.073873068	

Tabla #7, Resultados de la prueba de hipótesis t, aplicada a los procesos ARIMA y ARNN
Fuente: Elaboración del autor, herramienta Microsoft Excel

Siendo el valor obtenido del estadístico de prueba t igual a 2.13018 mayor que el valor critico de dos colas 2.07387, hace que este quede fuera de la región de aceptación y se concluye que la hipótesis nula es falsa por consiguiente se acepta la hipótesis alternativa la cual afirma que ambos métodos son completamente diferentes y es un acierto utilizar el método ARIMA para proyección de la demanda futura, para contrastar y probar la normalidad del conjunto de datos se utilizó el test de Shapiro–Wilk.

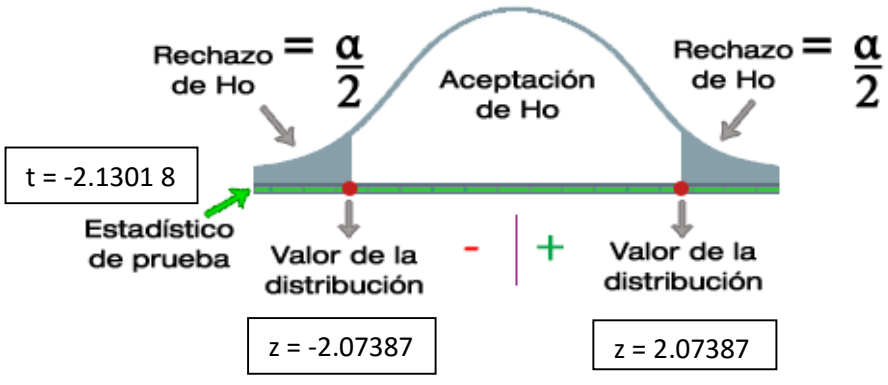


Gráfico #37, Evaluación gráfica de la prueba de hipótesis t
Fuente: Elaboración del autor, herramienta Microsoft Excel

4.2 Revisión de resultados

La exploración y visualización de datos en Tableau Desktop permite referenciar de una manera más rápida cuales son los productos que tienen poca o ninguna rotación en la bodega, para tomar una decisión con esta mercancía que no tiene ningún uso debido a que un inventario sin rotación es dinero que no está utilizando. Además, por medio de la exploración de datos a través de Tableau Desktop podemos inferir la información necesaria para hacer la evaluación final y comprar valores del último año con los valores propuestos con el modelo de predicción ARIMA y las cantidades económicas de pedido.



Gráfico #38, Resultados del análisis
Fuente: Elaboración del autor, herramienta Microsoft Excel

4.3 Conclusiones

Las diferencias entre el escenario real y el propuesto son muy evidentes según lo que se puede observar en la gráfica #38. El costo total de almacenamiento se ve disminuido en un 75% y los niveles de inventarios podrían disminuir en un 79%

METABISULFITO	REAL	PROPUESTO	Disminución
	2018	2019	%
Costo Total en $	$1,701,559.00	$423,669.00	75%
Saldo Final en Kg	3267075.00	673272.00	79%

Tabla #8, Datos finales del análisis
Fuente: Elaboración del autor, herramienta Microsoft Excel

4.4 Recomendaciones propuestas

Se recomienda a la alta dirección de la empresa evaluar el presente estudio en coordinación con el departamento de TI para llevar a cabo la implementación definitiva del mismo, los beneficios económicos que podría dejar ejecutar esta propuesta fueron claramente demostrados con fundamentos técnicos.

Producto Esperado	**Fase**	**Actividad**	**Duración Estimada de Actividad**
Constitución del proyecto	1	Entrevistas con los expertos	4 días
	1	Análisis de datos	7 días
Análisis del producto	1	Análisis de requisitos	3 días
	1	Ingeniería de Sistemas	5 días
Desarrollo DW y BI	2	Programación de los modelos predictivos	2 días
	2	Construcción de los DataMarts	3 días
	2	Desarrollo de las pantallas de visualización	1 día
Documentación Entregables	3	Informes de finalización de cada fase	5 días
	3	Manuales de usuario	
	3	Guías de administración	
	3	Seguridad Informática	
	3	Análisis de riesgo informático	
	3	Acta de entrega y satisfacción	
		TOTAL	**30 días**

Tabla #9, Cronograma tentativo de implementación
Fuente: Elaboración del autor, herramienta Microsoft Excel

5°) BIBLIOGRAFIA

BiGeek, (Junio del 2018): ¿Qué es una ETL?
Disponible en: https://blog.bi-geek.com/que-es-una-etl/

Blog Corponet, (Agosto del 2015): ¿Qué es la inteligencia de negocios y como beneficia a tu empresa? Disponible en: https://blog.corponet.com.mx/que-es-la-inteligencia-de-negocios

Diego Calvo, (Noviembre del 2018): Análisis de series temporales en R. ARIMA.
Disponible en: http://www.diegocalvo.es/analisis-de-series-temporales-en-r-arima/

Documentos Microsoft, (2017): Conceptos de minería de datos.
Disponible en: https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/

GoConqr,(2019): Matriz QSPM.
Disponible en: https://www.goconqr.com/mindmap/6863547/matriz-qspm

IngenioEmpresa, (Septiembre del 2017): Modelo de cantidad económica de pedido (EOQ): Qué es y cómo se hace. Disponible en: https://ingenioempresa.com/modelo-de-cantidad-economica-eoq/

Investigación de Operaciones, (referencia de fecha no disponible): EOQ - Cantidad Económica de Pedido (Economic Order Quantity). Disponible en: http://www.investigaciondeoperaciones.net/eoq.html

José Bardo, (Marzo del 2018): Como realizar un análisis de stock.
Disponible en: https://meetlogistics.com/inventario-almacen/como-realizar-un-analisis-del-stock/

Juan D. Velásquez PhD, (Julio del 2011): ARNN: Un paquete para la predicción de series de tiempo usando redes neuronales autorregresivas. Artículo científico de la Facultad de Minas, Universidad Nacional de Colombia

PowerData, (referencia de fecha no disponible): Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos. Disponible en: https://www.powerdata.es/data-warehouse

Procibernetica, (2017): Soluciones de inteligencia de negocios. Disponible en: https://www.procibernetica.com/soluciones-de-inteligencia-de-negocios/

Tableau, (2019): Guía de visualización de datos para principiantes: definición, ejemplos y recursos de aprendizaje. Disponible en: https://www.tableau.com/es-mx/learn/articles/data-visualization

6°) ANEXOS

Anexo #1, Código en R del modelo de predicción ARIMA

```
#Carga de datos
con_met<-
read.table("C:/Users/usuario/Desktop/proyecto_BI/modelos_de_prediccion/consumo de metabisulfito.txt", quote="\"", comment.char="")
View(con_met)
colnames(con_met)<-c("kg_met")
print(con_met)

#Histograma y resumen de datos de la variable
hist(con_met$kg_met,col = rainbow(8),xlab="Kg de Metabisulfito", ylab="Frecuencia",main = "Histográma Consumo de Metabisulfito",nclass = 10,xlim=c(0,120000))
summary(con_met)

#Reorganización de los datos por fecha y mes
met.ts<-ts(con_met, start = c(2014,1), frequency = 12)
print(met.ts)
plot(met.ts,col="red",main="Consumo de metabisulfito",xlab="Tiempo en años",ylab="Consumo en miles de Kg")

#Análisis de cajas por meses
boxplot(met.ts ~ cycle(met.ts),main="Consumo de metabisulfito",xlab="Meses del año",ylab="Consumo en miles de Kg")
cycle(met.ts)

#Descomponer la serie temporal en sus componentes = Tendencia, Senoidal y Aleatoria
#En este gráfico podemos observar 4 sub-graficos:

#Observed: los datos actuales
#Trend: el movimiento general hacia arriba o hacia abajo de los puntos de datos
#Seasonal: cualquier patrón mensual / anual de los puntos de los datos
#Random: parte inexplicable de los datos
met.ts.desc<-decompose(met.ts)
plot(met.ts.desc, xlab='Tiempo en años')

#Raiz de prueba unitaria
library("fUnitRoots")
urkpssTest(met.ts, type = c("tau"), lags = c("short"),use.lag = NULL, doplot = TRUE)

#Codigo R para calcular la autocorrelación:
acf(met.ts,lag.max=140)
```

```
pacf(met.ts,lag.max=140)

#Después eliminamos la no estacionariedad:
tsstationary = diff(met.ts, differences=1)
plot(tsstationary,col="red",main="Consumo de metabisulfito",xlab="Tiempo en años",ylab="Consumo en miles de Kg")

fitARIMA <- arima(met.ts, order=c(1,1,1),seasonal = list(order = c(1,0,0), period = 12),method="ML")
library(lmtest)
coeftest(fitARIMA)
```

```
z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1   0.039651   0.165605   0.2394   0.8108
ma1  -0.861064   0.079680 -10.8066   <2e-16 ***
sar1  0.145595   0.266470   0.5464   0.5848
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
confint(fitARIMA)
```

```
2.5 %     97.5 %
ar1  -0.2849294  0.3642314
ma1  -1.0172326 -0.7048946
sar1 -0.3766765  0.6678670
```

```
#Prueba de Box-Ljung
acf(fitARIMA$residuals,lag.max=140)
library(FitAR)
boxresult=LjungBoxTest (fitARIMA$residuals,k=2,StartLag=1)
plot(boxresult[,3],main= "Ljung-Box Q Test", ylab= "P-values", xlab= "Lag")
qqnorm(fitARIMA$residuals)
qqline(fitARIMA$residuals)onfint(fitARIMA)

#Ahora, este proceso se repite hasta que no se pueda encontrar un AIC inferior
library(forecast)
auto.arima(met.ts, trace=TRUE)

#Predecir utilizando un modelo ARIMA
predict(fitARIMA,n.ahead = 12)
futurVal <- forecast(fitARIMA,h=12, level=c(99.5))
plot(futurVal,col="red",main=" Pronostico Consumo de metabisulfito",xlab="Tiempo en años",ylab="Consumo en miles de Kg")
```

Anexo #2, Código en R del modelo de predicción ARNN

```
#Carga de datos
con_met<-
read.table("C:/Users/usuario/Desktop/proyecto_BI/modelos_de_prediccion/consumo de metabisulfito.txt", quote="\"", comment.char="")
View(con_met)
colnames(con_met)<-c("kg_met")
print(con_met)

###load function
source("arnn.R")
library(forecast)

#### First 60 points are used to fit the model
x <- ts(con_met, s = 1, f = 1)

#### A mlp neural network is fitted
fit <- arnn(x=x, lags=1:4, isMLP=FALSE, H=2, w.max=1e-3,
  restarts=10, seed = 1234, lambda=0, optim.control=list(maxit=200))

#### information about the fitted model
fit
summary(as.ts(fit))

#### in-sample errors
forecast::accuracy(as.ts(fit))

#### out-of-sample errors
fit1 <- arnn(x, model = fit)

forecast::accuracy( fitted(as.ts(fit1))[76:96], con_met$kg_met[81:100] )

#### one-step forecasts plot
plot(x, lwd = 2,main="Consumo de metabisulfito",xlab="Número de registros o periodos",ylab="Consumo en miles de Kg")
lines(fitted(fit1), col = 'red', lwd=3)
grid()

#### multi-step forecast plot
plot(forecast(fit, h=20, level=90, fan=FALSE, bootstrap=FALSE,
  seed=1234, npaths=1000), col="red",lwd=2,main="Consumo de metabisulfito",xlab="Número de registros o periodos",ylab="Consumo en miles de Kg")
grid()
###prediticon
forecast(fit, h=24, level=90, fan=FALSE, bootstrap=FALSE, npaths=1000)
```

Anexo #3, Código en R del algoritmo ARNN

```
# Autoregressive and multilayer perceptron artificial neural networks

.onAttach <-
function(...)
{
   version = library(help = arnn)$info[[1]]
   version = version[pmatch("Version",version)]
   um = strsplit(version, " ")[[1]]
   version = um[nchar(um) > 0][2]
   #
   cat(paste("This is arnn package", version, "\n"))
}
#
.onLoad <- function(lib,pkg){ library.dynam("arnn","arnn") }
#
#############################################################################
#############
#
coef.arnn      <- function(object, ...) { object$par       }
fitted.arnn    <- function(object, ...) { object$fitted    }
residuals.arnn <- function(object, ...) { object$residuals }
logLik.arnn                    <-        function(object,        ...)        {        structure(
object$loglik,df=length(object$par),class="logLik") }
#
summary.arnn   <- function(object, ...)
{
   require(forecast)
   print(object)
   cat("\nIn-sample error measures:\n")
   print(accuracy(object))
}
#
print.arnn <-
function (x, ...)
{
   cat(paste("\nMethod:", x$method, "\n\n"))
   cat(paste("Call:\n", deparse(x$call), "\n\n"))
   cat("Parameters:\n\n")
   print(coef(x))
   cat("\n\n")
   cat("Sigma^2 estimated as: ", as.character(x$sigma^2))
   cat("\n\n")
   cat("LogL: ", as.character(x$loglik))
   cat("\n\n")
```

```
   cat("Information Criteria:\n")
   print(x$IC)
   cat("\n\n")
}
#---------------------------------------------------------------------------#
arnn <-
function(x, lags = NULL, isMLP = FALSE, H = 1, w.max = 1.0, restarts = 1, seed =
NULL,
   lambda = 0, model = NULL,  optim.control = list())
{
 #

#############################################################################
#########
 #                                                    #
 #                    Funciones auxiliares                    #
 #                                                    #
                                  ##################################################
###################################
 #
 lagvector <- function(x, lags)
 {
                                        z = embed(x, max(lags)+1)
                                        return (z[,lags+1])
 }
                                  #
 #---------------------------------------------------------------------------
 #
                                  object2par <- function(object)
 {
                                        if (isMLP == TRUE)
                                        {
                                              return(c(       object$M,       c(object$Wih),
c(object$Wbh), c(object$Who), object$Wbo))
                                        }
                                        return(c( object$M, c(object$Wio), c(object$Wih),
c(object$Wbh), c(object$Who), object$Wbo))
                                  }
                                  #
 #---------------------------------------------------------------------------
 #
                                  par2object <- function(object, par)
 {
                                        k = 2
  ###
   object$M     = par[1]
```

```
  ###
  if (isMLP == FALSE)
  {
    object$Wio = matrix( data = par[k:(k + object$nlags - 1)],
                         nrow = object$nlags,
                         ncol = 1)
    k          = k + object$nlags
  }
  else
  {
    object$Wio = matrix( 0, nrow = object$nlags, ncol = 1)
  }
  ###
  if( object$H > 0)
  {
    ###
                                          object$Wih = matrix( data = par[k:(k + object$H *
object$nlags - 1)],
                       nrow = object$nlags,
                                                ncol = object$H)
                                          k          = k + object$H * object$nlags
    ###
                                          object$Wbh = matrix( data = par[k:(k+object$H-
1)],
                         nrow = object$H,
                         ncol = 1)
                                          k          = k + object$H
                                          ###
    object$Who = matrix( data = par[k:(k+object$H-1)],
                         nrow = object$H,
                         ncol = 1)
                                          k          = k + object$H
  }
                                        ###
  object$Wbo = par[k]
  ###
                                              return(object)
                                    }
                                    #
#-------------------------------------------------------------------------
#
                                    fn.foptim = function(w)
{
                                              object = par2object(object, w)
                                              object = arnn(x = object$x, model = object)
```

```
                                    return((1 - object$lambda) * object$sigma ^ 2 +
object$lambda * sum( abs(w[-1])))
                                  }
                                  #

#####################################################################
#########
                                  #                                          #
 #                  Cuerpo de la funcion                    #
 #                                              #
                                  ##############################################
###################################
 #
 if (!exists(".Random.seed", envir = .GlobalEnv, inherits = FALSE)) { runif(1) }
 if (is.null(seed))
 {
  RNGstate <- get(".Random.seed", envir = .GlobalEnv)
 }
 else
 {
  R.seed <- get(".Random.seed", envir = .GlobalEnv)
  set.seed(seed)
  RNGstate <- structure(seed, kind = as.list(RNGkind()))
  on.exit(assign(".Random.seed", R.seed, envir = .GlobalEnv))
 }
 #
 #
 #
                                  if (is.null(model))
 {
                                    object = list( x = x, lags = lags, nlags = length(lags),
            maxlag = max(lags), call = match.call(),
            method   = "arnn",
            restarts = restarts,
            H        = H,               # nÃºmero de neuronas en la capa oculta

            lambda   = lambda,            # parÃ¡metro de regularizaciÃ³n
            numpar   = 1 + 1 + 2 * H + length(lags) * H)
                                    #
  if( isMLP == FALSE )
  {
   object$numpar = object$numpar + object$nlags
  }
  #
                                    object = structure(object, class = "arnn")
```

```
                                        runopt = TRUE

                                        #
                                    }
 else
 {
                                        object      = model
                                        object$x    = x
                                        object$call = match.call()
                                        runopt      = FALSE
                                    }
                                    #
 #-------------------------------------------------------------------------
 #
                                    X.lagged = lagvector(object$x, object$lags)
 y = x[(max(object$lags)+1):length(x)]
 L = nrow(X.lagged)
                                    #
                                    #-------------------------------------------------------------------
-------------
                                    #
                                    if (runopt == TRUE)
 {
  for( irestart in 1:restarts)
  {
   wopt = optim( par     = runif(object$numpar, min = -w.max, max = w.max),
                 fn      = fn.foptim,
                 method  = "BFGS",
                 control = optim.control )

   u =  fn.foptim(wopt$par)

   if (irestart == 1 || u < u.opt )
   {
    w.gopt = wopt
    u.opt  = u
   }
  }
  object = par2object(object, w.gopt$par)
                                        #
  #-------------------------------------------------------------------------
  #
  names(object$M   ) = "M"
  names(object$Wbo ) = "Wbo"
  names(object$Wio ) = paste("Wio[", object$lags, "]", sep = "")
  if(object$H > 0)
```

```
   {
    names(object$Wbh ) = paste("Wbh[", 1:object$H,  "]", sep = "")
    names(object$Who ) = paste("Who[", 1:object$H,  "]", sep = "")
    names(object$Wih ) = paste("Wih[",
                        matrix( rep(1:object$nlags, object$H),  object$nlags, object$H),
                        ",",
                        t(matrix( rep(1:object$H, object$nlags), object$H, object$nlags)),
                        "]", sep = "" )

    #
    object$par = c( object$M, c(object$Wio), c(object$Wih), c(object$Wbh),
                c(object$Who), object$Wbo)
   }
   else
   {
    object$par = c( object$M, c(object$Wio), object$Wbo)
   }
                                       }
                                       #
                                       #-----------------------------------------------------------------
-------------
                                       #
  if( object$H > 0)
  {
                                        f       =       X.lagged       %*%       object$Wih       +
t(matrix(rep(object$Wbh, L), nrow = object$H, ncol = L))
                                        f = (1 / (1 + exp(-f))) ^ object$M
                                        f = f %*% object$Who + object$Wbo + X.lagged %*%
object$Wio
  }
  else
  {
   f = object$Wbo + X.lagged %*% object$Wio
  }
                                       #
                                       #-----------------------------------------------------------------
-------------
                                       #
  response        = y
  residuals       = y - f
  object$sigma     = sqrt(sum(residuals ^ 2) / (length(residuals) - 1))
  if(L >= object$maxlag)
  {
   object$loglik    = -0.5 * (L - object$maxlag) * (log(2 * pi) + log(object$sigma^2)) -
                0.5 * sum(residuals^2) / object$sigma^2
   p               = 1 + object$nlags
```

```
                                    N               = L - object$maxlag
   AK              = -2 * object$loglik  + 2 * object$numpar
   AKc             = -2 * object$loglik  + 2 * object$numpar * (N / (N - object$numpar - 1))
   HQ              = -2 * object$loglik  + 2 * object$numpar * log(log(L - object$maxlag))
   SC             = -2 * object$loglik  +     object$numpar * log(N)
   object$IC       = c(AK, AKc, HQ, SC)
   names(object$IC) = c("AK", "AKc", "HQ", "SC")
  }
  else
  {
   object$loglik = AK = AKc = HQ = SC = NULL
  }
  #
  #-------------------------------------------------------------------------------
  #
  frequency       = attributes(object$x)$tsp[3]
                                    start           = attributes(object$x)$tsp[1] + object$maxlag
/ frequency
  object$response  = ts(response,  start = start, frequency = frequency)
  object$fitted    = ts(f,         start = start, frequency = frequency)
  object$residuals = ts(residuals, start = start, frequency = frequency)
  #
                                    return(object)
}
#-------------------------------------------------------------------------------#
simulate.arnn <-
function (object, nsim = 1000, seed = NULL, h = length(object$x),
   bootstrap = FALSE, ...)
{
   #-------------------------------------------------------------------------
   if (!exists(".Random.seed", envir = .GlobalEnv, inherits = FALSE)) { runif(1) }
   if (is.null(seed))
   {
      RNGstate <- get(".Random.seed", envir = .GlobalEnv)
   }
   else
   {
      R.seed <- get(".Random.seed", envir = .GlobalEnv)
      set.seed(seed)
      RNGstate <- structure(seed, kind = as.list(RNGkind()))
      on.exit(assign(".Random.seed", R.seed, envir = .GlobalEnv))
   }
   #-------------------------------------------------------------------------
   if (bootstrap)
   {
      e = matrix(sample(object$residuals, h * nsim, replace = TRUE),
```

```
        nrow = h, ncol = nsim)
  }
  else
  {
    e = matrix(rnorm(h * nsim, 0, object$sigma),
        nrow = h, ncol = nsim)
  }
  #-------------------------------------------------------------------------
  To = length(object$x)
  s = matrix(NA, nrow = h, ncol = nsim)
  #
                                for (iserie in 1:nsim)
  {
                                        x.sim = c(object$x, rep(NULL, times = h))
    #
                                        for (k in 1:h)
    {
                                                X = x.sim[(To + k) - object$lags]

                                ##############################################
############################
                                                f = matrix(NA, nrow = 1, ncol = object$H)
                                                for(i in 1:object$H) {
                                                        f[,i] = X %*% object$Wih[,i]+
object$Wbh[i]
                                                        f[,i] = (1 / (1 + exp(-
f[,i])))^object$M
                                                }
                                                f = f %*% object$Who + object$Wbo + X
%*% object$Wio

                                ##############################################
############################
                                                x.sim[To + k] = f + e[k, iserie]
                                        }
                                        s[, iserie] = x.sim[(To + 1):(To + h)]
                                }
                                return(s)
}
#-------------------------------------------------------------------------------#
forecast.arnn <-
function (object, h = 10, level = c(80, 95), fan = FALSE, bootstrap = FALSE,
  seed = 1234, npaths = 5000, ...)
{
  if (fan)
    level <- seq(51, 99, by = 3)
```

```
   else
   {
      if (min(level) > 0 & max(level) < 1)
         level <- 100 * level
      else if (min(level) < 0 | max(level) > 99.99)
         stop("Confidence limit out of range")
   }
   nconf = length(level)
   lower = upper = matrix(NA, nrow = h, ncol = nconf)
   m = rep(0, h)
                                 ##############################################
##########################
   paths = simulate(object = object, npaths = npaths, seed = seed, h = h, bootstrap =
bootstrap)
                                 ##############################################
##########################
   for (k in 1:h) {
      m[k] = quantile(paths[, k], 0.5)
      lower[k, ] = quantile(paths[, k], (1 - level/100))
      upper[k, ] = quantile(paths[, k], level/100)
   }
   colnames(lower) = colnames(upper) = paste(level, "%", sep = "")
   #
   #
   #
   result          = list()
   result$model     = object
   result$method    = object$method
   result$level     = level
   result$x         = object$x
   result$residuals = residuals(object)
   result$fitted    = fitted(object)
   #
   frequency       = attributes(object$x)$tsp[3]
   start           = attributes(object$x)$tsp[1] + length(object$x)/frequency
   result$mean      = ts(data = m, start = start, frequency = frequency)
   result$lower     = ts(data = lower, start = start, frequency = frequency)
   result$upper     = ts(data = upper, start = start, frequency = frequency)
   #
   result = structure(result, class = "forecast")
   return(result)
}
```

Anexo #4, Código en PHP del aplicativo de carga de datos

//conexión

```
<?php

define("SERVIDOR","localhost");
define("USUARIO","root");
define("CLAVE","1234");
define("BASE_DATOS","dw_cofimar");

////////////////////////////////////////////////////////////////////////////////////////////////////////////

function Conexion()
{
   if (!($cn=mysqli_connect(SERVIDOR,USUARIO,CLAVE,BASE_DATOS)))
   {
    echo "Error conectando a la base de datos.";
    exit();
   }
   if (!mysqli_select_db($cn,BASE_DATOS))
   {
    echo "Hay un error en la conexión a la base de datos.";
    exit();
   }

   return $cn;

}
```

//código de consulta

```
<?php

  $servername = "localhost";
  $username   = "root";
  $password   = "1234";
  $dbname     = "dw_cofimar";

// crear conexion con new mysqli
  $conn = new mysqli($servername, $username, $password, $dbname);

// Chequear la conexion
  if ($conn->connect_error) {
      die("Connection failed: " . $conn->connect_error);
  }
```

```
 $sql = "SELECT * FROM datos_bodega ORDER BY id";
 $result = $conn->query($sql);

 if ($result->num_rows > 0) {
    // preparamos nuestra tabla
    echo "<table border=2>";
    echo "<tr>";
    echo "<th>id</td> <th>producto</td> <th>saldo_inicial</td> <th>ingreso</td>
<th>demanda</td> <th>saldo_final</td> <th>costo_total</td> <th>mes</td>
<th>anio</td>";
    echo "</tr>";

    // recorrer los registros del select cargando la consulta en un array asociativo
    while( $row = $result->fetch_assoc() ) {

      // descripcion de las variables;
      echo "<tr>";
      echo "<td>" .  $row["id"] . "</td>";
      echo "<td>" .  $row["producto"] . "</td>";
      echo "<td>" .  $row["saldo_inicial"] . "</td>";
      echo "<td>" .  $row["ingreso"] . "</td>";
      echo "<td>" .  $row["demanda"] . "</td>";
      echo "<td>" .  $row["saldo_final"] . "</td>";
      echo "<td>" .  $row["costo_total"] . "</td>";
      echo "<td>" .  $row["mes"] . "</td>";
      echo "<td>" .  $row["anio"] . "</td>";
      echo "</tr>" ;
    }
    echo "</table>";
 } else {
   echo "0 results";
}
$conn->close();
?>
<html>
<body>
<form action="index.html" method="post">
   <p><input type="submit" name="Boton" VALUE="Regresar">
</form>
</body>
</html>
```

//formulario, front end

```
<!DOCTYPE html>
<html lang="es">
<head>
                                        <meta charset="UTF-8">
                                        <link rel="stylesheet" href="css/style.css" />
                                        <title>Subir archivo CSV en PHP</title>
                                        <!-- libreria jquery -->
                                        <script
src="https://ajax.googleapis.com/ajax/libs/jquery/3.3.1/jquery.min.js"></script>
                                        <!-- cargar js -->
                                        <script src="js/cargar.js"></script>
</head>
<body>

                                        <!-- contenedor -->
                                        <main id="contenedor">
                                                <!-- formulario -->
                                                <section class="formulario">

                                                        <!-- -->
                                                        <header>
                                                <h1>Programa para cargar registros CSV</h1>
                                                        </header>
                                                        <!-- -->

                                                        <!-- -->
                                                        <section>
                                                                <form name="frmSubirCSV"
id="frmSubirCSV" method="post" enctype="multipart/form-data">
                                                                        <p><img
src="images/logo.jpg" alt="" class="imagen"></p>
                                                                        <h2>Carga de archivos CSV
al Data Warehouse</h2>
                                                                        <p><input type="file"
name="archivo_csv" id="archivo_csv"></p>
                                                                        <p><input type="submit"
class="enviar_archivo separar_boton" value="Enviar archivo"></p>
                                                                        <div id="estado"></div>
                                                                </form>
                                                <form action="codigo_de_consulta.php" method="post">
                                                  <p><input type="submit" name="Boton" VALUE="Consultar Registros">
                                                                </form>
                                                        </section>
                                                        <!-- -->
```

```
                </section>
                <!-- ./formulario -->

            </main>
            <!-- ./contenedor -->

</body>
```

//procesador de archivos planos y generador de respaldos

```
<?php

            include("includes/conexion.php");

            $cn = Conexion();

            if(substr($_FILES['archivo_csv']['name'], -3)=="csv")
            {
                $fecha          = date('Y-m-d');
                $carpeta        = "tmp_excel/";
                $n_archivo  = $fecha.'-
'.$_FILES['archivo_csv']['name'];

                $row = 1;


            move_uploaded_file($_FILES['archivo_csv']['tmp_name'
], $carpeta.$n_archivo);

                $fp = fopen($carpeta.$n_archivo, "r");

                while($data = fgetcsv($fp, 1000, ","))
                {

                    // Si la variable $row es diferente a 1, que
no registre los titulos en la tabla
            if($row!=1)
            {
            $sql_archivo  = "INSERT INTO
datos_bodega(id,producto,saldo_inicial,ingreso,demanda,saldo_final,costo_total,mes,ani
o)";
                                $sql_archivo .=
"VALUES('$data[0]','$data[1]','$data[2]','$data[3]','$data[4]','$data[5]','$data[6]','$data[7
```

```
]','$data[8]')";

                                                $rpta_archivo =
mysqli_query($cn,$sql_archivo) or die(mysqli_error($cn));

                                                if(!$sql_archivo)
                                                {
echo "<p>Hubo un problema al momento de importar el archivo. Por favor vuelva a
intentarlo</p>";
                                                        exit;

                                                }

                                        }

                                $row++;

                                }

                                fclose($fp);

echo "<p>La importacion del archivo CSV se ha ejecutado satisfactoriamente</p>";

                                exit();

                        }
```