

ESCUELA SUPERIOR POLITECNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICAS**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN DE
LA CALIDAD Y PRODUCTIVIDAD”**

TEMA:

APLICACIÓN DE UN MODELO ESTADÍSTICO PARA EL
PRONÓSTICO DE LA DEMANDA DE PRODUCTOS DE UNA
EMPRESA COMERCIALIZADORA DE ÍTEMS DE FERRETERÍA DE
LA CIUDAD DE GUAYAQUIL.

AUTOR:

STEFANIE DESIRÉ VILLANUEVA SAMPÍN

Guayaquil - Ecuador

2021

RESUMEN

La complejidad en la que está envuelto el mundo de los negocios ha aumentado con el tiempo generando que las organizaciones operen en una atmósfera de incertidumbre; a pesar de esto, deben tomar decisiones que conllevan a grandes riesgos y que terminan afectando su futuro. Un aspecto importante que ha destacado es la elaboración de pronósticos permitiendo a las organizaciones toma de decisiones más oportunas y confiables a diferencia de las estimaciones realizadas de manera subjetiva las cuales no suelen ser efectivas y seguras.

Por tal motivo, el presente trabajo propone la aplicación de un Modelo Estadístico para el pronóstico de la demanda de productos de una empresa comercializadora de ítems de ferretería de la ciudad de Guayaquil, que permita la disminución del error al pronosticar y la optimización de recursos.

La metodología empleada consiste en la aplicación y comparación del modelo no estacionario ARIMA, de Regresión Múltiple y Regresión con errores en una serie de datos históricos de ventas, mediante el software RStudio. Se seleccionará aquel modelo que describa de forma plausible el comportamiento de la demanda. Para la evaluación del desempeño, se realizaron pronósticos de períodos de prueba para compararlos con los datos reales y evaluar el error de estimación.

Como resultado de la investigación se obtuvo que los modelos ARIMA y de Regresión con errores presentaron mejores ajustes del comportamiento de los datos a diferencia del modelo de Regresión Múltiple que no tuvo buenos resultados. Los modelos de Regresión con errores mostraron mayor precisión en los pronósticos.

Palabras clave: pronóstico, demanda, ARIMA, regresión

ABSTRACT

The complexity in which the business world is involved has increased over time, causing organizations to operate in an atmosphere of uncertainty; in spite of this, they must make decisions that entail great risks and end up affecting their future. An important aspect that has been highlighted is the elaboration of forecasts allowing organizations to make more timely and reliable decisions as opposed to subjective estimates, which are not usually effective and safe.

For this reason, the present work proposes the application of a Statistical Model for the forecast of the demand of products of a company that sells hardware items in the city of Guayaquil, which allows the reduction of the error in forecasting and the optimization of resources.

The methodology used consists of the application and comparison of the non-stationary ARIMA model, Multiple Regression and Regression with errors in a series of historical sales data, using RStudio software. The model that plausibly describes the demand behavior will be selected. For the performance evaluation, forecasts of test periods were made to compare them with the real data and to evaluate the estimation error.

As a result of the research, the ARIMA and Regression with errors models showed better fits to the behavior of the data, unlike the Multiple Regression model, which did not perform well. The Regression models with errors showed better forecasting accuracy.

Keywords: forecast, demand, ARIMA, regression

DEDICATORIA

A Dios por guiarme en cada paso y por las bendiciones recibidas a lo largo de mi vida.

A mi madre Elba Sampín Candelario por ser mi mayor inspiración y quien estoy segura desde el cielo estará muy orgullosa por el cumplimiento de esta nueva meta.

A mi padre y hermana por su apoyo y aliento constante e incondicional en todos los momentos de mi vida.

A mi novio por su comprensión, paciencia y por ser quien ha estado presente en cada uno de los momentos vividos.

AGRADECIMIENTO

A Dios, por darme la sabiduría y fortaleza para cumplir esta meta propuesta.

A mi tutora Mgs. Nadia Cárdenas Escobar quien con su paciencia y dedicación contribuyeron a la realización de este proyecto.

A mi papá Pedro Villanueva y a mi hermana Yara, quienes han sido un pilar importante a lo largo de mi vida y me impulsan a afrontar todo reto que me proponga.

A Argeni Vera R. por su amor incondicional, su apoyo constante y por brindarme sus palabras de aliento en cada ocasión que las necesité.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.



Stefanie Desiré Villanueva Sampín

TRIBUNAL DE GRADUACIÓN



Sergio Bauz Olvera, Ph.D.
PRESIDENTE



Mgs. Nadia Cárdenas Escobar
DIRECTOR



Mgs. Mariela González Narváez
VOCAL 1



Mgs. Sandra González Camba
VOCAL 2

ABREVIATURAS O SIGLAS

ACF	AutoCorrelation Function (Función de autocorrelación)
ADF	Augmented Dickey–Fuller (Prueba Dickey–Fuller Aumentado)
AIC	Akaike information criterion (Criterio de información Akaike)
AICc	Akaike information criterion (Criterio de información Akaike Corregido)
AR	AutoRegressive (Autoregresivos)
ARIMA	Autoregressive Integrated Moving Average (Modelo Autorregresivo Integrado de Promedio Móvil)
BCE	Banco Central del Ecuador
BIESS	Banco del Instituto Ecuatoriano de Seguridad Social
DF	Prueba de Dickey–Fuller
EMV	Estimación de máxima verosimilitud (MLE, por sus siglas en inglés)
ENEMDU	Encuesta Nacional de Empleo, Desempleo y Subempleo
EOQ	Economic Order Quantity (Cantidad Económica de Pedido)
INEC	Instituto Nacional de Estadísticas y Censos
KPSS	Kwiatkowski-Phillips-Smichdt-Shin Test (Prueba de Kwiatkowski-Phillips-Smichdt-Shin)
MA	Moving Average (Medias móviles)
MAE	Mean Absolute Error (Error absoluto de la Media)
MAPE	Mean Absolute Percentage Error (Error absoluto porcentual de la media)
MSE	Mean Squared Error (Error cuadrático de la media)
OMS	Organización Mundial de Salud
PACF	Partial AutoCorrelation Function (Función de Autocorrelación Parcial)
PEA	Población Económicamente Activa
PIB	Producto Interno Bruto
PIBC	Producto Interno Bruto en Construcción
PP	Prueba de Phillips– Perron
RMSE	Root Mean Squared Error (Raíz del error cuadrático de la media)
SKU	Stock Keeping Unit (Unidad de Mantenimiento en Almacén)
SRI	Servicio de Rentas Internas

TABLA DE CONTENIDO

CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1. Antecedentes	4
1.2. Descripción del problema.....	6
1.3. Objetivos	8
1.3.1. Objetivo General	8
1.3.2. Objetivos Específicos.....	8
1.4. Hipótesis	8
1.5. Alcance.....	9
CAPÍTULO 2	10
2. MARCO TEÓRICO Y ESTADO DEL ARTE.....	10
2.1. Marco Teórico	10
2.1.1. Generalidades de los pronósticos.....	10
2.1.2. Clasificación de los pronósticos	11
2.1.3. Modelo Estadístico	12
2.1.3.1. Generalidades	12
2.1.3.2. Métodos de clasificación Paramétricos	14
2.1.4. Procesos estocásticos	15
2.1.4.1. Definición.....	16
2.1.4.2. Procesos estocásticos estacionarios	16
2.1.4.3. Primer y Segundo Momento de un Proceso Estocástico	17
2.1.5. Series de Tiempo.....	18
2.1.5.1. Objetivos del análisis de series de tiempo.....	18
2.1.5.2. Componentes de una serie de tiempo.....	19
2.1.6. Modelos de ajustes para Series de Tiempo	19
2.1.6.1. Modelo Autoregresivos AR (p).....	21
2.1.6.2. Modelo de Medias móviles MA (q)	22
2.1.6.3. Modelo Autorregresivo de Medias Móviles ARMA (p,q)	22
2.1.6.4. Modelo ARIMA.....	23
2.1.6.5. Modelo de Regresión Múltiple.....	25
2.1.6.6. Modelo de Regresión con errores ARIMA	27
2.1.7. Raíz Unitaria	29
2.1.8. Error del pronóstico	30

2.1.8.1. Error Absoluto de la Media (MAE).....	31
2.1.8.2. Error Absoluto porcentual de la Media (MAPE)	31
2.1.8.3. Error cuadrático de la media (MSE).....	31
2.1.8.4. Raíz del Error cuadrático de la media (RMSE)	31
2.2. Estado del Arte	32
CAPÍTULO 3	37
3. METODOLOGÍA.....	37
3.1. Los datos	37
3.2. Métodos estadísticos utilizados	41
3.3. Análisis de datos.....	43
3.4. Obtención del Modelo estadístico.....	43
CAPÍTULO 4	44
4. RESULTADOS	44
4.1. Comportamiento de los datos.....	44
4.1.1. Análisis de la Variable Ventas Totales	44
4.1.2. Análisis de Variables Macroeconómicas	47
4.2. Construcción del Modelo.....	54
4.2.1. Modelo ARIMA no estacionario	55
4.2.2. Modelo de Regresión Múltiple	65
4.2.3. Modelo de Regresión con errores ARIMA.....	78
4.3. Selección del Modelo.....	83
4.4. Validación del Modelo.....	89
4.4.1. Autocorrelación de los residuos	89
4.4.2. Distribución normal de los residuos	90
4.4.3. Gráfico de residuos vs ajustes.....	91
4.4.4. Predicción del Modelo seleccionado	92
CAPÍTULO 5	94
5. CONCLUSIONES Y RECOMENDACIONES	94
5.1. Conclusiones.....	94
5.2. Recomendaciones.....	96
6. Bibliografía	1
7. Apéndices y anexos	4

LISTADO DE FIGURAS

Figura 1. Elementos inherentes en pronósticos	11
Figura 2. Clasificación de los pronósticos	12
Figura 3. Gráfico de Dispersión Tipo Lineal de la variable respuesta pérdida de peso y variable predictora contenido de Hierro	15
Figura 4. Ventas de los Productos de Ferretería	44
Figura 5. Ventas estacionales de los Productos de Ferretería.....	45
Figura 6. Ventas estacionales de Productos de Ferretería – Diagrama de Cajas .	46
Figura 7. Top Ten de Sku's más vendidos	47
Figura 8. Producto Interno Bruto Construcción	48
Figura 9. Índice de Precios de Tuberías	49
Figura 10. Tasa de Variación de Precios de Tuberías.....	50
Figura 11. Tasa de Desempleo Nacional.....	51
Figura 12. Gasto Público del Estado	52
Figura 13. Operaciones de Créditos Hipotecarios en el Biess	53
Figura 14. Operaciones de Crédito Vivienda en Sector Privado.....	54
Figura 15. Demanda del SKU 17050	55
Figura 16. Transformación logarítmica del SKU 17050.....	56
Figura 17. Primera Diferencia de los datos.....	57
Figura 18. Funciones de Autocorrelación ACF y PACF	58
Figura 19. Autocorrelación ACF de los Residuos.....	61
Figura 20. Valores p del Estadístico Ljung-Box	61
Figura 21. Curva ajustada del Modelo ARIMA (1,1,3)	62
Figura 22. Autocorrelación ACF de los Residuos.....	64
Figura 23. Curva ajustada del Modelo ARIMA (0,1,3)	65
Figura 24. Ajuste de Modelo Inicial (Todas las Variables)	69
Figura 25. Ajuste de Modelo # 2.....	71
Figura 26. Ajuste de Modelo # 3.....	72
Figura 27. Ajuste de Modelo # 4.....	74
Figura 28. Ajuste de Modelo # 5.....	76
Figura 29. Ajuste de Modelo # 6.....	77
Figura 30. Modelo # 1 Regresión con errores ARIMA (0, 1, 3)	79

Figura 31. Modelo Regresión con errores ARIMA (PPT, GP)	82
Figura 32. Modelo Regresión con errores ARIMA (PIBC, PPT y DN).....	82
Figura 33. Modelo Regresión con errores ARIMA (PIBC y PPT).....	83
Figura 34. Gráfica de ACF de los residuos	90
Figura 35. Histograma de los residuos	90
Figura 36. Gráfico de dispersión Residuos vs Ajuste	91
Figura 37. Pronóstico de la demanda del modelo seleccionado.....	93

LISTADO DE TABLAS

Tabla 1. Coeficientes del modelo ARIMA (1, 1, 3).....	59
Tabla 2. Medidas de Error del modelo ARIMA (1, 1, 3).....	60
Tabla 3. Modelos ARIMA de la función auto.arima	63
Tabla 4. Medidas de Error del modelo ARIMA (0, 1, 3).....	63
Tabla 5. Matriz de Correlación.....	66
Tabla 6. Modelo Inicial (Todas las variables)	67
Tabla 7. Métricas de Regresión (Todas las variables).....	69
Tabla 8. Modelo # 2	70
Tabla 9. Métricas de Regresión Modelo # 2	71
Tabla 10. Modelo # 3.....	72
Tabla 11. Métricas de Regresión Modelo # 3.....	73
Tabla 12. Modelo # 4	73
Tabla 13. Métricas de Regresión Modelo # 4.....	74
Tabla 14. Modelo # 5.....	75
Tabla 15. Métricas de Regresión Modelo # 5.....	76
Tabla 16. Modelo # 6.....	77
Tabla 17. Métricas de Regresión Modelo # 6.....	78
Tabla 18. Métricas de Modelo # 1 Regresión con errores	79
Tabla 19. Modelos de Regresión con errores Arima (0,1,3)	80
Tabla 20. Métricas de Precisión de Errores y Pronósticos de Modelos Evaluados	86
Tabla 21. Resultados del Pronóstico del Modelo.....	92

CAPÍTULO 1

1. INTRODUCCIÓN

Actualmente, para las organizaciones, la planeación y el control de los inventarios son acciones de vital importancia, que permiten anticiparse ante las tendencias del comportamiento del mercado y generan una mejor toma de decisiones ante cualquier eventualidad, es por ello que, la predicción permite mejorar el control de inventarios reduciendo la incertidumbre. El uso de modelos de predicción acertados, permitirá contar con un pronóstico altamente eficiente, es decir, con una variación mínima, lo que conlleva a un alto cumplimiento de pedidos y obtención de mejores resultados.

La planificación de la demanda, juega un papel esencial debido a que si se mantiene un stock en exceso se incurren en gastos innecesarios; o, por el contrario, si se presentan inexistencias se traduce en la pérdida de clientes y como consecuencia un descenso en las ventas. Uno de los principales problemas que suelen presentarse en las empresas, es el manejo inadecuado de sus inventarios, los cuales se ven reflejados en aspectos como la rentabilidad, costos, satisfacción del cliente y disminución del inventario disponible, generando la falta de respuesta oportuna ante cambios propios del mercado tales como la demanda.

La previsión de la demanda, es un factor importante en las decisiones dentro de la organización. La demanda marca el inicio en la cadena de suministro y su integración en la gestión de inventarios, da lugar a políticas que minimizan los costos de almacenamiento, mediante el mantenimiento de cantidades de abastecimiento, ajustadas a las necesidades de la organización con niveles de inventarios seguros, evitando excesos y/o desabastecimientos sin afectar el nivel del servicio (Calderón Ardila & Supelano Aguirre, 2012).

En un mercado de constantes cambios y que cada vez se vuelve más competitivo, es de suma importancia para las empresas disponer de alguna técnica estadística que admita estimar un acontecimiento futuro a través del análisis de los datos históricos de las ventas, permitiendo de esta manera el tratamiento frente a la incertidumbre de la demanda, aumentar sus ganancias y reducir sus costos. Teniendo en cuenta, la premisa que el futuro depende de los eventos ocurridos en el pasado, es pertinente conocer lo acontecido en el mercado de la construcción para mantener una mejor percepción de objeto de estudio.

Ecuador a lo largo de los años ha mantenido un crecimiento económico sustentable, siendo así, que uno de los sectores que ha permitido fortalecer nuestra economía ha sido el sector de la construcción. Un factor principal que ha contribuido con esto son las construcciones de edificaciones públicas (centros educativos, unidades médicas, etc.) a través del Gobierno. Así como también, en el sector privado en donde la construcción se ha hecho presente en un sin número de proyectos urbanísticos. Y es gracias al apoyo de ambos sectores que han permitido que la construcción constituya un pilar económico fundamental.

La construcción en el Ecuador con el pasar del tiempo, ha tenido un gran avance e impacto en la economía, permitiendo su expansión y contribución dentro del PIB (Producto Interno Bruto) la misma que debido al crecimiento presentado se ha considerado de mayor importancia. Para Yandún (2020), la construcción representa el 10% del PIB, y si se suman las actividades ligadas a la misma (industrias de acero, cerámica, madera, ferretería, etc.) llega a constituir el 20%, además, la construcción es considerada como la principal actividad generadora de empleo.

El sector de la construcción posee aliados como lo son las ferreterías y/o locales de ventas de productos afines, ya que por medio de los mismos se pueden adquirir los productos necesarios e indispensables para la construcción de obras sean estas pequeñas, medianas o grandes, convirtiendo a las actividades del

sector ferretero en una de las más sobresalientes en la economía de nuestro país.

Para García Osorio & Tobar Cazares (2019) la construcción tiene una participación significativa no sólo por su ciclo productivo, sino también por el empleo de mano de obra, de tal forma, que según el Banco Central del Ecuador (BCE, 2018) contribuyó en promedio con el 8.4% de la producción total de la economía nacional.

Según Selberherr (2015) la construcción posee una contribución sustentable en la economía de un país, debido a su alto nivel de aportación a la economía nacional siendo el mayor generador económico que se encuentra presente en las etapas del desarrollo, coincidiendo criterios con Wilkinson, Yan Chang, & Sapeciay (2016) quienes indican que el sector de la construcción es una parte importante para la economía del país, la cual permite el desarrollo de otros sectores económicos.

La emergencia provocada por la pandemia del COVID-19, ha tenido un fuerte impacto en todos los sectores económicos, incluido el de la construcción y es que a pesar de constituir una de las actividades dinamizadoras debido a la crisis, la incertidumbre y demás causas, la reactivación de este sector es un poco compleja, sin embargo, la Organización Mundial del Trabajo estima que este sector sea la clave de la reactivación del resto de economías afectadas (Grupo Faro, 2020).

Según lo establecido en la SENPLADES (2017), el sector de la construcción y sus actividades conexas han experimentado cambios, con tendencias ascendentes y expansivas, lo que ha permitido un crecimiento constante debido a que la vivienda y el desarrollo en infraestructura integran uno de los principales objetivos del Plan Nacional del Buen Vivir, permitiendo promover el desarrollo sustentable.

En este contexto, el presente trabajo de investigación se enfoca en la aplicación de un modelo estadístico para el pronóstico de la demanda de productos de una empresa comercializadora de ítems de ferretería de la ciudad de Guayaquil.

1.1. Antecedentes

A lo largo de los años, ha existido cierto grado de interés por pronosticar o conocer eventos futuros por parte del ser humano; es por ello que, en la actualidad, para las empresas u organizaciones, el pronóstico de la demanda representa uno de los grandes retos a enfrentar, pues este beneficia la toma de decisiones con la menor incertidumbre posible. Estas predicciones no sólo se buscan por medio de las experiencias o conocimientos del personal (Modelos Cualitativos) sino a través del análisis de los datos históricos de las ventas basados en cálculos matemáticos (Modelos Cuantitativos), que permiten comprender e integrar en modelos estadísticos el impacto que pueden tener variables como promociones, eventos de mercadeo, leyes y demás factores propios del mercado.

Muchas de las técnicas que existen en la actualidad en relación a los pronósticos se desarrollaron en el siglo XIX; tales como, los análisis de regresión en el cual se destaca la metodología de Box-Jenkins.

Hanke & Reitsh (1996), afirman que “quienes toman decisiones lo harán mejor sí a partir de la comprensión de las técnicas de pronóstico, tanto cualitativas como cuantitativas, las utilizan de manera adecuada, en vez de que se vean forzados a planear el futuro sin el beneficio de esta valiosa información complementaria” (pág 2).

A nivel mundial, la planificación basada en pronósticos se ha convertido en un factor indispensable para las estimaciones estadísticas en el contorno empresarial, influyendo de manera trascendental en la administración. Estas estimaciones conllevan a realizar pronósticos oportunos siendo fundamental su evaluación para una toma de decisión confiable (Cadena Lozano, Ariza Garzón, & Palomo Zurdo, 2017).

En América Latina, cada vez es más frecuente la aplicación de los métodos estadísticos en áreas de todo tipo, al ser una opción para el planteamiento y solución de problemas que se presentan en las organizaciones, convirtiéndose así, en una base para la toma de decisiones, sin embargo, la aplicación errónea de los mismos, constituye la toma equívoca de decisiones derivándose en una baja rentabilidad. Los modelos que presentan una mayor cantidad de uso son: promedios móviles simples, suavización exponencial simple y doble, modelos ARIMA (Autoregressive Integrated Moving Average), entre otros.

Existen una serie de trabajos relacionados con los pronósticos de la demanda para mejorar la planeación de la demanda, los cuales se detallan a continuación:

Sánchez & Gavira Durón (2016) mantienen que para efectos de ejecutar una estimación a corto plazo se deben considerar todos los componentes ineludibles de la serie de tiempo que no han sido observados tales como, la tendencia, estacionalidad y las variaciones producidas de forma aleatoria, favoreciendo con ello al desarrollo de los pronósticos de las series.

Boada A. J. (2017) por su parte, efectúa el análisis de una serie de variables identificadas como causales para la proyección de la demanda, mediante el diseño y desarrollo de una herramienta estadística automatizada, que permite realizar estimaciones y predicciones.

Espino Timón (2017), realiza una recopilación de programas de libre acceso para el público en general, los cuales permitan realizar análisis para modelos predictivos de la demanda a través de diferentes herramientas estadísticas y así determinar el desempeño de los mismos mediante su evaluación. Entre las principales herramientas descritas que mantienen resultados más óptimos constan: la interfaz de RStudio y Weka.

Según datos del Servicio de Rentas Internas (SRI) en el año 2017 existió un incremento en las ventas de productos de ferretería a nivel nacional en relación al año anterior en el cual se produjo una caída en ventas del 16%.

Siendo así que el sector de la construcción ha tenido una contracción significativa a lo largo de estos años, generando retrocesos en los ingresos de las compañías de este sector y consecuentemente produciendo la reducción de participación en el Producto Interno Bruto (PIB) (Ekos, Mundo Constructor, 2020).

De acuerdo a un estudio realizado en el año 2018 por la revista Mundo Ferretero del Grupo Ekos, para determinar los productos más vendidos se realizaron entrevistas a propietarios de varias ferreterías (medianas y grandes), dando como resultado que los productos más vendidos corresponden a las categorías de Tuberías PVC (72.5%), grifería (60%), pintura (57.5%), materiales eléctricos (50%) entre los principales. (Ekos, Estudio de productos que toda ferretería debe tener, 2018).

Con lo anterior expuesto, podemos observar que existe gran contribución sobre este tema; no obstante, es evidente la falta de investigaciones al respecto dado que es un tema extenso en materia de predicción de la demanda, en este trabajo se aportará al área de comercialización de productos del sector ferretero pues se aplicarán modelos de predicción de la demanda.

1.2. Descripción del problema

La aleatoriedad de la demanda es frecuente debido a los cambios que se generan en el mercado por diferentes temas, tales como la crisis económica que afecta a nivel mundial, social, tecnológico, etc. En nuestra industria, el uso de métodos para la determinación de la demanda es muy escaso, normalmente, por el desconocimiento o falta de las herramientas necesarias que permitan efectuar estimaciones con mayor precisión, eliminando cierto grado de incertidumbre. Dado que una buena estimación conlleva a

mantener una planeación más idónea y una mejor toma de decisiones. Existe además un problema común basado en la existencia de excedentes y/o faltantes de inventarios, por ese motivo, toda empresa debe de mantener un control de los niveles de inventarios, los mismos que permitan optimizar sus utilidades.

En un análisis realizado a una empresa comercializadora de productos del sector ferretero en la ciudad de Guayaquil, se evidenció que no cuenta con herramientas cuantitativas para el diagnóstico de la demanda. Siendo así, que las estimaciones de ventas se realizan de una manera subjetiva por medio del conocimiento de la gerencia comercial y de compras entre las que se realiza una reunión para la toma de decisiones y determinación de la cantidad de estimaciones de ventas futuras, basándose en un análisis del historial de ventas, denominado también como Grupo de Consenso.

La gestión de aprovisionamiento se hace basándose en el sistema de control de inventarios EOQ (Economic Order Quantity), el cual mantiene el supuesto de que la demanda es conocida y constante, difiriendo de la realidad del mercado ya que la demanda no posee un comportamiento invariable. Además, no contemplan patrones en series de tiempo como la estacionalidad, tendencia, etc., afectando el nivel de stock de la empresa y provocando una gestión financiera deficiente.

De acuerdo a los comentarios emitidos por la gerencia comercial y de compras, en los meses de abril, junio y noviembre la organización realiza "Ferias" como estrategia comercial, lo cual les permite la captación de nuevos clientes y el aumento en la oferta y demanda de sus productos. En tanto, que en el mes de agosto al ser caracterizado como el mes de la construcción, se realiza una serie de descuentos y se otorgan facilidades de pago para los clientes que quieran realizar remodelaciones o construcciones.

La predicción de la demanda permite obtener una valiosa información para la toma de decisiones de la empresa al estimar las ventas de los productos, permitiendo mantener un mejor control del inventario.

Además, que esta información permite ser de gran utilidad para la planificación de la producción, obtención de los recursos de distribución, lanzamientos de promociones y acciones específicas cuando se prevé que los resultados sean más satisfactorios. Es por ello, que el presente trabajo de investigación propone la aplicación de un modelo estadístico para la predicción de la demanda en una empresa que ofrece productos de construcción.

1.3. Objetivos

1.3.1. Objetivo General

Evaluar modelos estadísticos para la predicción de la demanda de productos de una empresa comercializadora de ítems de ferretería mediante series de tiempo.

1.3.2. Objetivos Específicos

- ❖ Estimar modelos estadísticos para el pronóstico de la demanda.
- ❖ Realizar un análisis comparativo de los modelos estadísticos obtenidos que permita seleccionar el mejor modelo que se ajuste al comportamiento natural de los datos.
- ❖ Validar la efectividad de los pronósticos de la demanda con el modelo estadístico escogido.

1.4. Hipótesis

La aplicación de un modelo estadístico para el pronóstico de la demanda de productos de una empresa comercializadora de ítems de ferretería de la ciudad de Guayaquil permitirá determinar el comportamiento plausible de las ventas del producto seleccionado. Su correcto desempeño será evaluado con el pronóstico de 10 períodos de prueba para contrarrestarlos con los datos reales evaluando el error de estimación del modelo seleccionado.

1.5. Alcance

En la presente tesis se planteará la aplicación de un modelo estadístico predictivo para el SKU (Stock Keeping Unit / Unidad de Mantenimiento en Almacén) que presente una mayor demanda en el mercado del Ecuador de una empresa comercializadora de productos de ferretería. La base de datos que se analizará corresponde a las ventas mensuales de la compañía del período comprendido entre los años 2016 al 2020.

CAPÍTULO 2

2. MARCO TEÓRICO Y ESTADO DEL ARTE

En el presente capítulo se detalla el marco teórico y estado del arte del presente proyecto, en lo referente a los conceptos y definiciones de los métodos estadísticos a utilizarse, para proporcionar al lector una mayor comprensión de los temas tratados a lo largo de los siguientes capítulos.

2.1. Marco Teórico

2.1.1. Generalidades de los pronósticos

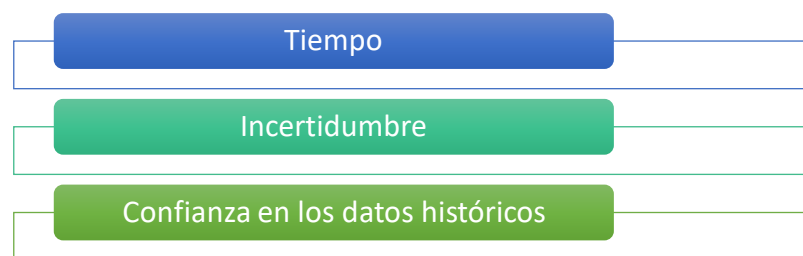
En la actualidad, debido al creciente incremento de competitividad en las industrias, es de suma importancia contar con las herramientas de análisis necesarias que permitan establecer e implementar estrategias eficaces, las cuales serán de ayuda para una planificación de actividades relacionadas al control de inventarios además de satisfacer de manera oportuna la demanda que se presente.

Para la predicción de valores futuros de una variable se emplean métodos de pronósticos, los cuales permiten reducir la incertidumbre de lo que puede suceder en un futuro, permitiendo información más real y confiable para la toma de decisiones. El pronóstico es de utilidad para la planificación y toma de acciones específicas cuando se estima resultados favorables. (Montemayor Gallegos, 2013).

Los pronósticos comúnmente se emplean para el establecimiento de objetivos a corto, mediano y/o a largo plazo, siendo así, una base para las organizaciones y para la consecución de los mismos. La aplicación de los pronósticos es bastante diversa y pueden ser usados en múltiples variables de las empresas tales como: producción, inventarios, finanzas, planeación estratégica, entre otras.

Dado que, no es posible eliminar el riesgo de los pronósticos es necesario que se considere de manera explícita la incertidumbre generada. Un factor de vital importancia que debe ser considerado es la disponibilidad de los datos ya que los datos históricos permiten establecer los procedimientos de pronósticos y las observaciones futuras. Es por ello que, al momento de realizar un pronóstico, los elementos que se mantienen de forma inherente son: el tiempo, la incertidumbre y la confianza de los datos históricos a utilizar (Montemayor Gallegos, 2013).

Figura 1. Elementos inherentes en pronósticos



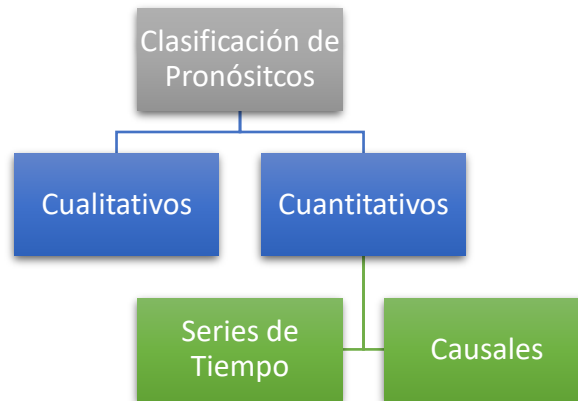
Fuente: Elaboración propia

2.1.2. Clasificación de los pronósticos

Los pronósticos pueden clasificarse de acuerdo a dos tipos: cuantitativos o cualitativos. Se debe de elegir aquel pronóstico que se adapte a las necesidades de cada organización. Los pronósticos de tipo cualitativos, son aquellos que no requieren de una estructura analítica de datos, utilizan el “juicio” de quien realiza el pronóstico por lo cual propende a ser subjetivo al efectuarse basado en la experiencia o intuición, podría estar sesgado y son usados cuando los datos que se poseen son escasos; mientras que, los de tipo cuantitativos se basan en datos o información histórica de la organización o el mercado lo que permite que sean objetivos.

Dentro de la clasificación de los pronósticos de tipo cuantitativos se encuentran las series de tiempo y los métodos causales.

Figura 2. Clasificación de los pronósticos



Fuente: Elaboración propia

Las series de tiempo son aquellos que utilizan la demanda histórica, se fundamentan en que la demanda pasada es un indicador confiable de la demanda futura.

Los métodos causales suponen una correlación entre el pronóstico y ciertos factores propios del entorno del mercado, existiendo una relación causa – efecto con las variables independientes.

2.1.3. Modelo Estadístico

2.1.3.1. Generalidades

Un modelo estadístico es una forma o modelo matemático que mantiene supuestos que guardan relación directa con la obtención de datos aproximados a la realidad, permitiendo la realización de predicciones a partir de dicha aproximación. El modelo estadístico general se define en la siguiente expresión matemática:

$$Y = f(X) + \epsilon$$

Donde $f(X)$ es la función matemática que manifiesta la relación existente entre las variables dependiente e independientes, y ϵ es la variable para el error.

La parte sistemática del modelo anteriormente detallado se define como $f(X_1, X_2 \dots X_p)$ el cual representa el valor esperado de la respuesta (\hat{Y}) para el conjunto de las variables predictoras ($X_1 = x_1, X_2 = x_2, \dots, X_p = x_p$) y \hat{f} es la estimación de f (Morales, 2018).

Se denomina modelo Predictivo a un método de análisis de datos y estadísticas para definir hipótesis o predecir resultados o sucesos futuros. El modelado genera predicciones con un grado de probabilidad según las variables analizadas y se busca estimar Y mediante la ecuación:

$$\hat{Y} = \hat{f}(X)$$

\hat{f} no es una estimación de suma precisión; por tanto, mantiene inherente un error.

La precisión del modelo está ligada a dos componentes: el error reducible y el error irreducible. El error reducible se encuentra relacionado con la precisión de los pronósticos para la estimación de $f(X)$; sin embargo, la incertidumbre siempre será un factor presente debido a la dependencia entre variables, el mismo que se lo conoce como error irreducible (Escalante, 2019).

Algunas de las técnicas más aplicadas del Modelado Predictivo son:

- Árboles de decisión
- Regresión lineal y logística
- Redes neuronales
- Análisis bayesiano
- Series temporales y datamining
- Máquinas de vectores de soporte
- K-vecinos más cercanos
- Ensemble models

- Gradient boosting
- Respuesta incremental
- Razonamiento con base en la memoria
- Regresión de mínimos cuadrados parciales

El modelo Predictivo permite descubrir oportunidades y prevenir situaciones desfavorables, es considerado capaz de generalizar el conocimiento aprendido de los datos históricos para predecir el futuro. Este tipo de modelo computarizan un resultado, al implementar una función de regresión y pueden ser usados en la implementación de una función de clasificación, en la que el resultado es una clase o categoría (Guazzelli, 2012).

2.1.3.2. Métodos de clasificación Paramétricos

Los métodos Paramétricos son un modelo estadístico, en donde provienen de una distribución normal y se estiman mediante los parámetros específicos como la μ (media) y σ (desviación estándar).

Estos modelos resultan ser eficaces cuando se sigue la distribución propuesta, aunque si bien es cierto, pueden ser muy sensibles cuando se utilizan muestras de un tamaño mínimo.

El caso más sencillo a considerar es siendo $f(x)$ la función lineal de todas las variables predictoras

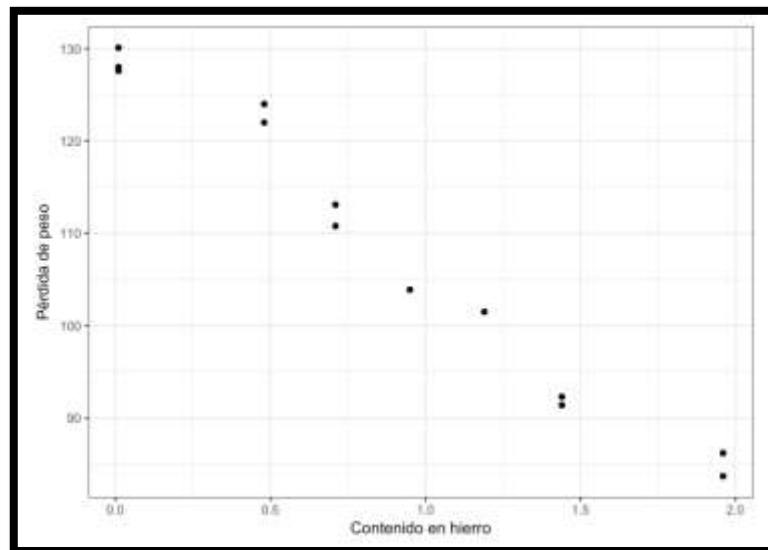
$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Donde $\beta_0 + \beta_1 + \beta_2 + \dots + \beta_p$ son los parámetros que corresponden al modelo y son estimados utilizando el método de los mínimos cuadrados, estos coeficientes estimados se califican como las pendientes de cada una de las variables predictoras. Mientras que, ϵ es una variable aleatoria conocida como el término del error del

modelo y representa la variabilidad de $f(X)$ que no puede ser explicada por la relación lineal entre $f(X)$ y las variables predictoras. En los casos en que únicamente se dispone de una sola variable predictora, el modelo matemático está representado de la siguiente manera:

$$f(X) = \beta_0 + \beta_1 X + \epsilon$$

Figura 3. Gráfico de Dispersión Tipo Lineal de la variable respuesta pérdida de peso y variable predictora contenido de Hierro



Fuente: (Morales, 2018)

2.1.4. Procesos estocásticos

Un proceso estocástico es un conjunto de variables aleatorias ordenadas que varían con el tiempo, las cuales evolucionan en función de otra variable. Estas variables aleatorias dependen de ciertos parámetros, cada una de ellas mantienen su distribución de probabilidad y existe la posibilidad de que estén correlacionadas

Existen dos tipos de procesos estocásticos, cuya diferencia en términos generales se relaciona con la predicción de una serie, estos son:

- **Procesos estocásticos estacionarios:** Mantiene a lo largo del tiempo sus características lo que hace que sea un proceso predecible.
- **Procesos estocásticos no estacionarios:** Es un proceso en el que su distribución tiene una variación no constante, lo que lo convierte en impredecible.

2.1.4.1. Definición

El proceso estocástico puede ser definido de dos formas matemáticas distintas:

- Como el conjunto de relaciones temporales cuyos índices mantienen la característica de evolucionar en el tiempo de manera aleatoria.
- Conjunto de variables aleatorias x_t , dado que $t \in T$.

2.1.4.2. Procesos estocásticos estacionarios

Los procesos estocásticos estacionarios se definen como aquellos cuya distribución de probabilidad es constante a lo largo de los períodos de tiempo. A través de esta información, se pueden elaborar modelos de predicción.

Una de las características que deben de mantener estos procesos es la estacionariedad. Un proceso estocástico es estacionario en sentido estricto cuando la distribución no sufre variación al desplazarse en k períodos de tiempo, se define como:

$$F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}] = F[Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_n+k}] \quad \forall (t_1, t_2, \dots, t_n)$$

Este proceso es estacionario en covarianza si cumple lo siguiente:

- La serie tiene la misma media:

$$E(Y_t) = \mu_t < \infty, \forall t$$

- La serie tiene la misma varianza:

$$V(Y_t) = E[Y_t - \mu]^2 = \sigma_Y^2 < \infty, \forall t$$

La autocovarianza depende del tiempo transcurrido entre dos períodos distintos de tiempo

$$Cov(Y_t, Y_s) = E[Y_t - \mu_t][Y_s - \mu_s] = Y_{|t-s|} = Y_k < \infty, \forall k$$

2.1.4.3. Primer y Segundo Momento de un Proceso Estocástico

Cuando es posible establecer las funciones de distribución para los conjuntos finitos de este proceso, se establece que el proceso estocástico está caracterizado. La determinación de los valores tiende a ser un procedimiento complejo. Los momentos de las variables aleatorias de un proceso estocástico son los valores esperados en cualquier instante del proceso. Debido a la complejidad para determinar la función de distribución se lo divide en momentos de diverso orden (Primer y Segundo momento).

El primer momento de un proceso estocástico se define por el conjunto de medias de todas las variables, dado por:

$$u_t = E(Y_t)$$

El segundo momento se define por el conjunto de varianzas de todas las variables; además, se consideran las covarianzas de los diferentes tiempos.

$$Y_{t,s} = Cov(Y_t, Y_s) = E[(Y_t - \mu_s)(Y_s - \mu_s)]$$

Cuando $t = s$ se tiene:

$$Var(Y_t) = E[(Y_t - \mu_t)^2]$$

Los coeficientes de autocorrelación, se encuentran determinados por:

$$\rho_{t,s} = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t) Cov(Y_s)}} \quad -1 < \rho_{t,s} < 1$$

Una vez determinados los dos momentos y si el proceso mantiene una distribución normal, se tendría definida su distribución.

2.1.5. Series de Tiempo

Diversidad de autores han establecido definiciones de las series de tiempo. Para Brockwell & Davis (2016), una serie de tiempo es un conjunto de observaciones X_t , cada una de las cuales se registra en un momento específico t .

Para el análisis de las series de tiempo se debe tener en consideración el orden y la dependencia de los datos.

Los análisis de las series de tiempo pueden ser de dos tipos: univariantes o multivariantes. Las series univariantes son aquellas que analizan una serie en función de su pasado y las series multivariantes son aquellas que analizan a la vez varias series de tiempo.

2.1.5.1. Objetivos del análisis de series de tiempo

Las series de tiempo consideran los siguientes objetivos principales:

- **Descripción:** al estudiar la serie se deben tener en cuenta las medidas descriptivas de la misma, teniendo en consideración su tendencia, la estacionalidad y la existencia de observaciones discordantes.

- **Predicción:** al analizar los valores se pretende entender el pasado, así como predecir los eventos futuros.

2.1.5.2. Componentes de una serie de tiempo

El estudio descriptivo de las series de tiempo se fundamenta en la descomposición de la serie a través de varios componentes, los cuales se detallan a continuación, según Marín Diazaraque (2014):

1. **Tendencia:** Se puede definir como un cambio a largo plazo que se produce en relación al nivel medio, o el cambio a largo plazo de la media.
2. **Efecto Estacional:** Las series presentan variaciones en ciertos períodos, cuyos tipos de efectos son fáciles de entender y se pueden medir explícitamente o incluso se pueden eliminar del conjunto de los datos, desestacionalizando la serie original.
3. **Componente Aleatorio:** Analizados los dos componentes antes mencionados, se pretende estudiar el comportamiento aleatorio.

2.1.6. Modelos de ajustes para Series de Tiempo

La estructura de un proceso estocástico se encuentra en función de las autocovarianzas y la autocorrelación, construyendo a partir de estas funciones un modelo que muestre el comportamiento de la serie y que pueda ser de utilidad para la predicción.

Los modelos de series univariantes descomponen a ésta en dos partes: la primera que es sistemática y la segunda que es aleatoria o llamada también innovación:

La parte sistemática es la parte predecible para la construcción del modelo de la serie Y_1, Y_2, \dots, Y_t

Dada una serie temporal cuya media es igual a cero, el valor resultante de Y para cada momento t depende de su pasado, el modelo que permite representar su comportamiento se describe como:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3} \dots) + a_t \quad t = 1, 2, 3 \dots$$

Para los casos en donde los procesos estacionarios mantienen una distribución normal con media cero, se puede representar Y_t como una combinación lineal de los valores pasados infinitos, descrita de la siguiente manera:

$$Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \dots + a_t \quad \forall t \quad t = 1, 2, 3 \dots$$

El proceso debe cumplir las siguientes condiciones generales:

No es un proceso anticipante; es decir, no está determinado por valores futuros.

El proceso es invertible; es decir, el presente depende de su pasado de forma convergente, lo cual se da si los parámetros cumplen la restricción detallada a continuación:

$$\sum_{i=1}^{\infty} \pi_i^2 < \infty$$

El modelo lineal mantiene tres representaciones (González Casimiro, 2009):

- Representación puramente autorregresiva AR
- Puramente medias móviles MA
- Representación finita denominado como Autorregresivo de medias móviles de orden (p, q)

2.1.6.1. Modelo Autoregresivos AR (p)

El modelo autorregresivo AR (p), es una aproximación al modelo lineal anteriormente descrito, y para su explicación se presenta la ecuación de los modelos de orden 1 y 2.

El orden uno, representado por su valor pasado, queda representado de la forma:

$$Y_t = \phi Y_{t-1} + a_t \quad t = 1, 2, \dots \quad a_t \sim RB(0, \sigma^2)$$

Donde a_t es ruido blanco (RB) con media cero y varianza σ^2 .

Es necesario comprobar si el modelo es estacionario para los valores del parámetro ϕ mediante las condiciones de estacionariedad:

- Estacionario en media (media constante y finita)

$$E(Y_t) = E(\phi Y_{t-1} + a_t) = \phi E(Y_{t-1})$$

Para que sea estacionario es necesario que $\phi \neq 1$

- Estacionario en covarianza (covarianza constante y finita)

$$\gamma_0 = \phi \gamma_0 + \sigma^2 \rightarrow (1 - \phi^2)\gamma_0 = \sigma^2 \rightarrow \gamma_0 = \frac{\sigma^2}{1 - \phi^2}$$

Para que sea estacionario es necesario que $|\phi| < 1$

El de orden dos, se encuentra representado de la forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t \quad a_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots$$

Este proceso mantiene las siguientes características:

- Media constante
- Función de autocovarianzas
- Coeficientes de autocorrelación definidos por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

$$k = 1, 2, 3, \dots$$

2.1.6.2. Modelo de Medias móviles MA (q)

Este modelo se representa con las medias móviles, es decir, que el polinomio autorregresivo es de orden cero y se define como:

$$Y_t = \phi_1 a_{t-1} - \phi_2 a_{t-2} - \dots - \phi_q a_{t-q} + a_t \quad a_t \sim RB(0, \sigma^2)$$

Dado los parámetros de naturaleza finita del modelo este es estacionario para cualquiera de sus valores y tiene como particularidad de no ser anticipante. Las principales características de este modelo son media constante y cero, varianza constante y finita. La función de autocorrelación está dada por:

$$\rho_k = \begin{cases} \rho_k \neq 0 & k = 1, 2, \dots, q \\ \rho_k = 0 & k > q \end{cases}$$

Este modelo se puede generalizar para modelos con media distinta de cero:

$$Y_t = \delta + a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \dots - \phi_q a_{t-q} \quad a_t \sim RB(0, \sigma^2)$$

2.1.6.3. Modelo Autorregresivo de Medias Móviles ARMA (p,q)

Los modelos autorregresivos de medias móviles determinan Y_t en función de su pasado hasta el retardo p y el retardo q: (Marín Diazaraque, 2014):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad a_t \sim RB(0, \sigma^2)$$

Este modelo se define de la siguiente forma en términos de retardos, como:

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) a_t$$

$$\phi_p(L) Y_t = \theta_q(L) a_t$$

donde $\phi_p(L)$ es el polinomio autorregresivo y $\theta_q(L)$ es el polinomio medias móviles.

Las condiciones de estacionariedad del modelo ARMA (p,q) vienen impuestas por la parte autorregresiva, dado que la parte de medias móviles finita siempre es estacionaria y las condiciones de invertibilidad vienen impuestas por la parte de medias móviles, dado que la parte autorregresiva finita siempre es invertible porque está directamente escrita en forma autorregresiva. Este modelo tiene media cero, varianza constante y finita y una función de autocovarianzas infinita (González Casimiro, 2009).

2.1.6.4. Modelo ARIMA

Los modelos de series de tiempo analizados hasta el momento se basan en la estacionariedad; es decir, media y varianzas constantes. Existen muchas series de tiempo que no son estacionarias dadas a las condiciones del tiempo. El modelo ARIMA (Autoregressive integrated moving average) es un modelo estadístico que utiliza las variaciones y regresiones para hacer predicciones del futuro mediante el estudio de patrones. Se lo denota como:

$$\begin{aligned}
 Y'_t &= \delta + \phi_1 Y'_{t-1} + \dots + \phi_p Y'_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} + a_t \\
 Y'_t - \phi_1 Y'_{t-1} - \dots - \phi_p Y'_{t-p} &= -\theta_1 a_{t-1} - \dots - \theta_q a_{t-q} + a_t \\
 (1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d Y'_t &= (1 - \theta_1 L - \dots - \theta_q L^q) a_t
 \end{aligned}$$

Donde:

Y'_t : es la serie diferenciada

p: orden autorregresivo

q: orden de promedio móvil

d: grado de primera diferenciación involucrada

L: función de retardo

a_t : ruido blanco

Este modelo también permite modelar datos estacionales, mediante los siguientes términos: ARIMA (p,d,q) (P,D,Q)_s, donde P es el orden autorregresivo del proceso estacional, Q el orden promedio móvil, D es el orden de diferenciación de la parte estacional y s representa la periodicidad de la serie.

Uno de los criterios que son de utilidad para determinar el orden del modelo estudiado es el Criterio de información Akaike (AIC).

En ciertos casos, es posible que ciertos valores de la serie de tiempo no mantienen una independencia entre sí; por el contrario, un valor depende de los anteriores, por lo cual es necesario medir o evaluar la dependencia que pudiera existir en los períodos de la serie, para lo cual existen varias funciones, que son:

- Función de autocorrelación (ACF)
- Función de Autocorrelación Parcial (PACF)
- Prueba de Ljung-Box

Para la construcción y análisis del modelo Autorregresivo Integrado de Media Móvil (ARIMA, del inglés AutoRegressive Integrated Moving Average) se deberán seguir los pasos que se detallan a continuación:

1. Una vez recolectados los datos (diarios, mensuales, anuales, etc.) se deberá realizar la representación gráfica para determinar que la serie sea estacionaria, requisito indispensable para la construcción del presente modelo. En el caso que los datos no presenten estacionariedad, se debe proceder a su transformación, ya que es necesaria la estabilización de la varianza, para esto se lo realiza mediante

la transformación logarítmica o la transformación de Box-Cox (soluciona problemas de normalidad y heterocedasticidad).

2. Si se determina que la serie no posee estacionariedad, se procede a calcular las primeras diferencias de los datos hasta que éstos sean estacionarios.
3. Se determina la función de autocorrelación (FAC) o ACF por sus siglas en inglés de Autocorrelation Function, de los órdenes de p (componente AR) y q (componente MA). Posterior a ello, se procede a ejecutar el método de máxima verosimilitud para estimar los parámetros desconocidos.
4. Mediante el Criterio de información Akaike Corregido (AICc) se procede a buscar el mejor modelo.
5. Una vez seleccionado el mejor modelo se procede a calcular la FAC de los residuos, se verifica que tenga comportamiento de ruido blanco.
6. Finalmente, se procede a realizar los pronósticos.

2.1.6.5. Modelo de Regresión Múltiple

La regresión lineal múltiple permite realizar modelos lineales en donde, más de una variable independiente conocida también como predictora son utilizadas para explicar la variable dependiente. Los modelos de regresión lineal consideran a los modelos de regresión múltiple como una ampliación de estos.

Los modelos de regresión múltiple son utilizados para describir la relación existente entre las variables predictoras y predecir el valor de la variable dependiente. Se los denota como:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon$$

Donde:

Y : es el valor de la predicción de la variable dependiente o también llamada endógena

$\beta_n X_n$: es el coeficiente de regresión de la variable independiente o exógena n

ε : es el error del modelo

Para efectuar estos modelos de regresión es necesario que cumplan los supuestos que se detallan a continuación:

- Linealidad
- Homocedasticidad
- Normalidad
- No colinealidad
- Independencia

Linealidad

Este supuesto indica que la relación establecida entre las variables dependientes e independientes debe ser lineal; es decir, que deben mantener una correlación entre ambas variables. Una forma posible de poder comprobar que existe linealidad es mediante la representación de diagramas de dispersión.

Homocedasticidad

La homocedasticidad se presenta cuando la varianza de los errores es constante; es decir, es igual para todas las variables independientes.

$$Var(\mu_t) = \sigma^2 \quad t = 1, 2, \dots, n$$

Normalidad

El supuesto de normalidad indica que todos los errores se encuentran distribuidos normalmente con media cero. La normalidad de estos errores, hace posible la estimación por intervalos de confianza.

$$\varepsilon \sim N(0, \sigma^2)$$

Una manera de comprobar el supuesto de normalidad es mediante un gráfico de histogramas o pruebas de normalidad.

No colinealidad

La no colinealidad comprende que no exista correlación entre las variables independientes o exógenas. Mientras que, la multicolinealidad se presenta cuando existe dependencia alguna entre las variables mencionadas.

Independencia

El supuesto de independencia establece que los residuos (diferencias entre los valores reales y los pronosticados) deben ser independientes y no estar correlacionados entre sí. Para comprobar este supuesto se puede recurrir a los diagramas de dispersión o pruebas como la de Durbin Watson.

$$E(\varepsilon_t|X_t) = E(\varepsilon_t) = 0 \quad t = 1, 2, \dots, n$$

2.1.6.6. Modelo de Regresión con errores ARIMA

En el apartado anterior se presentó el modelo ARIMA el cual utiliza los datos históricos (observaciones pasadas) de la serie y no permite la introducción de ninguna otra notable información; sin embargo, existe información adicional que permite efectuar mejores predicciones. Un ejemplo, puede ser el cambio de precio de un determinado producto que podría repercutir en los volúmenes de ventas en futuros períodos.

El modelo de Regresión con errores ARIMA permite la combinación de los datos históricos y de información considerable de variables predictoras. Su modelo tiene similitud a un modelo de regresión estándar, donde, las variables predictoras se encuentran en la parte derecha, su diferencia radica en el término del error.

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + a_t$$

Y_t es una función lineal de las k variables predictoras $(X_{1,t}, X_{2,t}, \dots, X_{k,t})$

En tanto que, en los modelos de regresión estándar el término del error es ruido blanco; es decir, que sus errores no se encuentran correlacionados, en el modelo de regresión con errores ARIMA permite que los errores contengan autocorrelación. Al reemplazar el término a_t por ε_t en la ecuación, la serie de regresión con errores n_t supone que el término del error proviene de un modelo ARIMA. Como ejemplo para un modelo ARIMA (1, 1, 1), la ecuación sería la siguiente:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t$$

$$(1 - \phi_1 B)(1 - B)\varepsilon_t = (1 + \theta_1 B)a_t$$

Se puede observar que el modelo contiene los dos términos del error: ε_t el error del modelo de regresión y a_t el error del modelo ARIMA, donde se supone que el error a_t es ruido blanco.

Hyndnan & Athanasopoulos (2018) manifiestan que un punto importante a tener en consideración al realizar la estimación del modelo de regresión con errores ARIMA es que las variables deben ser estacionarias para de esta manera evitar problemas que pudieran surgir, tales como:

- Al realizar alguna prueba estadística del modelo, ésta se encuentre incorrecta.
- Los valores correspondientes al AICc de los modelos ajustados no serán de utilidad para la elección del mejor modelo.
- No serán buenas estimaciones los coeficientes que se hayan obtenido.

2.1.7. Raíz Unitaria

Consiste en una forma estricta de validar si una serie es estacionaria, esta es una característica fundamental para los procesos estocásticos que de no cumplirse puede causar inconvenientes con los modelos de series de tiempo.

En la actualidad existen varios contrastes de hipótesis que nos permiten comprobar si las series de datos tienen raíz unitaria; tales como:

1. Prueba de Dickey–Fuller (DF)
2. Dickey–Fuller Aumentado (ADF)
3. Prueba de Phillips– Perron (PP)
4. Prueba Kwiatkowski, Phillips, Schmidt y Shin (KPSS), entre otras.

Una de las pruebas comúnmente utilizada es la prueba de Kwiatkowski, Phillips, Schmidt, Y Shin (KPSS), esta difiere de las otras pruebas de raíz unitaria, al partir con el supuesto que la serie es estacionaria bajo la hipótesis nula y la hipótesis alternativa es que los datos no son estacionarios.

Se consideran los residuos e_t ; $t = 1, 2, \dots, T$ al aplicar la regresión a la serie Y_t , donde:

$$Y_t = \alpha + \beta t + \mu_t$$

Luego se estima la varianza $\hat{\sigma}^2$, la suma parcial de los residuos se define:

$$SP_t = \sum_{t=1}^T e_t \quad t = 1, 2, \dots, T$$

El estadístico LM se define como: $LM = \sum_{t=1}^T \frac{SP_t^2}{\sigma^2}$

Finalmente, el estadístico de prueba KPSS se define:

$$KPSS = T^{-2} \sum_{t=1}^T \frac{SP_t^2}{\sigma_{\epsilon}^2}$$

Si por motivo alguno el proceso llegare a tener raíces unitarias múltiples, convendrá aplicar el operador de diferencia a la serie, el cual puede ser utilizado diversas veces.

2.1.8. Error del pronóstico

Una de las formas de conocer si el pronóstico está dando los resultados esperados es mediante la medición de los errores.

El error permite determinar el grado de exactitud del pronóstico frente de la realidad. Existen diferentes maneras de medir el error en los pronósticos y este se lo representa con la ecuación matemática:

$$E_t = Y_t - F_t$$

Donde E_t es el error del pronóstico en el período t , Y_t representa el valor real de ese período y F_t el valor pronosticado.

A continuación, se detallan las medidas de error de pronósticos más usadas:

- Error absoluto de la Media (MAE)
- Error absoluto porcentual de la media (MAPE)
- Desviación porcentual absoluta de la media (PMAD)
- Error cuadrático de la media (MSE)
- Raíz del error cuadrático de la media (RMSE)

Es de vital importancia estimar la exactitud de los pronósticos y ajustar los modelos de acuerdo a los resultados obtenidos. En la actualidad, existen diferentes formas para estimar los errores de los pronósticos los cuales tienen sus ventajas y desventajas, las

cuales se deben conocer para aminorar las desventajas. (Pavolini, 2019)

2.1.8.1. Error Absoluto de la Media (MAE)

Es el valor absoluto resultante de la diferencia entre la demanda real y el pronóstico, dividido para el número total de los períodos.

$$MAE = \frac{\sum_{t=1}^N |E_t|}{N}$$

2.1.8.2. Error Absoluto porcentual de la Media (MAPE)

Es el promedio del error absoluto o expresado también como la diferencia entre la demanda real y el pronóstico.

$$MAPE = \frac{\sum_{t=1}^N \left| \frac{E_t}{Y_t} \right|}{N}$$

2.1.8.3. Error cuadrático de la media (MSE)

Esta es una medida de dispersión, mide el promedio de errores al cuadrado.

$$MSE = \frac{\sum_{t=1}^N E_t^2}{N}$$

2.1.8.4. Raíz del Error cuadrático de la media (RMSE)

Esta es una medida que representa la raíz cuadrada del valor promedio de los errores elevados al cuadrado.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N E_t^2}{N}}$$

2.2. Estado del Arte

El hombre desde sus inicios se ha interesado por conocer el futuro, por este motivo, con el pasar de los años la predicción ha cobrado más importancia convirtiéndose en un tema de mucho interés, sobre todo para los sectores económicos e industriales. El mundo de los negocios cada vez se vuelve más competitivo y exigente, lo que lleva a las organizaciones a dirigir sus esfuerzos en disminuir los costos y/o aumentar sus ingresos, a través del desarrollo y del establecimiento de estudios predictivos, mediante herramientas y técnicas estadísticas que permitan procesar la información, con la finalidad de tener resultados confiables que conlleven a mejores tomas de decisiones y que permitan anticiparse a las necesidades actuales del mercado.

Si bien es cierto que la predicción de la demanda desempeña un papel fundamental en las organizaciones, resulta imposible obtener resultados que coincidan de forma exacta con los datos del futuro, por lo tanto, existe una atención particular en los errores, los cuales, se encuentran inherentes en el proceso y es de cierta manera inevitable su presencia, sin embargo, al realizar predicciones se intenta que estos sean los mínimos posibles. Para Cadena Lozano J. B. (2016) el pronóstico estratégico empresarial es un tema que no ha sido explorado a profundidad, por lo cual, su investigación se basó en conocer varios aspectos asociados a la planificación y al pronóstico como lo son: sus objetivos, las políticas de gestión, los métodos y las variables usadas, las áreas involucradas, así como también, las medidas de desempeño, con la finalidad de evaluar que el proceso de predicción se lleve a cabo junto con la planificación y les permita obtener resultados favorables a las organizaciones.

Dentro del proceso de planificación, los pronósticos tienen un papel ineludible, ya que forman parte del primer paso de este proceso y del que dependerán los resultados que se obtendrán. Los pronósticos, son

una herramienta que proporcionan un estimado cuantitativo de la probabilidad de eventos futuros.

Muchos autores exponen los diferentes tipos de técnicas de pronósticos que pueden emplearse y los clasifican conforme a su tendencia, estas pueden ser: cuantitativas (análisis de regresión y series de tiempo) y cualitativas (técnica subjetiva). Existen diversos trabajos de investigación desarrollados en la aplicación de estas técnicas, las cuales se detallan a continuación:

Lao León, Rivas Méndez, Pérez Pravia, & Marrero Delgado (2017), proponen un procedimiento para realizar el pronóstico de la demanda a través de Redes Neuronales Artificiales, aplicando el procedimiento propuesto en la Empresa Comercializadora y Distribuidora de Medicamentos Holguín (EMCOMED).

Un trabajo realizado en una empresa de consumo masivo dedicada a la venta de productos cosméticos por catálogo, permitió a Boada A. J. (2017) realizar un análisis detallado de las variables explicativas de la demanda de los productos, con el cual, diseñó una herramienta automatizada con la finalidad de realizar predicciones y un software para automatizar los procesos en relación con las estimaciones y predicciones de las ventas por catálogo, basados en datos matemáticos y estadísticos, mediante el estudio de variables como precio, fabricación, ventas, entre otras, con la finalidad de crear estrategias de planificación y predicción.

Los autores Asencio Cristóbal, González Ascencio, & Lozano Robles (2017) mencionaron que, debido a la competitividad existente en el mundo de los negocios, es de suma importancia mantener un control interno en la organización, que permita minimizar los costos y maximizar su utilidad. El trabajo efectuado, fue enfocado en la

distribución farmacéutica de la provincia del Guayas a través de un análisis detallado del manejo del inventario.

González (2020), en su artículo denominado “Un modelo de gestión de inventarios basado en estrategia competitiva”, utilizó la metodología denominada de cuatro pasos en una empresa líder en Chile de la sección de pernería y tornillería, utilizando un pronóstico de la demanda, donde se emplea el coeficiente de variación como medida de variabilidad y como técnicas de pronóstico la suavización exponencial (pág 133-142).

Mogrovejo Del Salto (2018) en su trabajo, mantiene como principal objetivo la mejora en el control de inventario de los productos mediante la implementación de políticas y la estimación de compras óptimas, a través de la simulación, considerando las predicciones de la demanda con la finalidad de evitar que exista inventario en exceso o por el contrario la falta de productos.

Escobar Sailema (2019), en su trabajo de investigación, evaluó un modelo de estimación estadística de predicción de la demanda en el mejoramiento de la rentabilidad de los emprendimientos, además los pronósticos son una herramienta que proporcionan un estimado cuantitativo de la probabilidad de eventos futuros, estableció los niveles de rentabilidad y las variables de incidencia asociadas.

Para Quinde Espinoza & Ramos Alvarado (2018), uno de los problemas más notables que enfrenan las organizaciones que tienen como principal fuente de ingreso las ventas de materiales, es la falta de una adecuada planificación de inventario, por ello su trabajo tuvo como finalidad el analizar la valuación y control de inventarios en la empresa Disensa Ramírez S.A y su efecto en la rentabilidad.

En un estudio realizado por Gallo Méndez (2020) a la empresa Roquette Laisa España, implementó una planificación de compra para las materias primas que son consideradas de suma importancia en la producción, debido a que su ausencia puede ocasionar una parada en la factoría. Para este propósito, utilizó un modelo para la obtención de pronósticos de la compra, basándose en una serie histórica de datos del 2015 al 2018. Con este modelo logró una predicción óptima para dar fundamento a la toma de decisiones de los directivos.

Gil-Vera (2017) realizó un trabajo de estudio de series de tiempo sobre el pronóstico de la demanda mensual de electricidad, tomando los datos de la demanda de la energía para efectuar el modelado y validación, concluyendo que las series de tiempo aplicadas permitieron obtener una predicción con alto nivel de precisión.

Contreras Juárez, Atziry Zuñiga, Martínez Flores, & Sánchez Partida (2016) presentan un proyecto de una empresa dedicada al almacenamiento de productos perecederos, en donde, emplean el uso de técnicas estadísticas para realizar la predicción de la demanda mediante series de tiempo. La metodología empleada fue la aplicación de un modelo estadístico de los datos históricos del volumen de ingreso y egreso, con la finalidad de establecer un pronóstico confiable para la determinación del volumen de productos a almacenar y prever los insumos y personal necesario para su cumplimiento.

Forero Gómez & Martínez Lozano (2020) obtuvieron un pronóstico de ventas para bolsas ecológicas (producto comercializado por la empresa Boleco S.A.) mediante el análisis de las variables explicativas relacionadas al producto, con lo cual se determinó la correlación existente entre las mismas, se elaboró un modelo de regresión múltiple y se efectuó su validación.

Alarcón Viracacha & Mancipe Romero (2016), presentaron un diagnóstico en el que se evidencia la gran importancia que tiene en la organización una buena metodología del pronóstico de demanda. En su artículo, presentan la metodología empleada para la gestión de la demanda en los diversos procesos de la cadena de suministro de una empresa líder en la fabricación de tuberías y polivinilos en Colombia. Su metodología empleada consistió en dos fases, la primera se fundamenta en la clasificación del producto y la segunda fase corresponde a la elaboración de un pronóstico según el orden de la clasificación realizada en la primera fase. La metodología usada para el pronóstico fue mediante la técnica de los promedios móviles. Con este estudio, se logró una reducción significativa en el inventario de la organización y una reducción del margen de error en la estimación de los pronósticos.

Por otra parte, Morales Tabares, Cabrera Campos, Vázquez Silva, & Caballero Mota (2016) en su artículo sobre “Modelo multivariado de predicción del stock de piezas de repuesto para equipos médicos”, indican la ausencia de aplicaciones de técnicas de predicción para productos de esta línea, motivo por el cual, proponen el modelado para la predicción del inventario de estos productos mediante la técnica estadística de Regresión Lineal Múltiple, la cual permitió obtener resultados confiables y a su vez la reducción de los errores de estimación de los pronósticos.

A pesar de existir una gran cantidad de trabajos, se evidencia que aún hay líneas de investigación por explorar.

Con una correcta estimación de la demanda, se proporciona a las empresas información valiosa sobre el mercado actual, permitiendo que los gerentes puedan tomar decisiones acertadas.

CAPÍTULO 3

3. METODOLOGÍA

En este capítulo se describe la base metodológica en la que se fundamenta el presente trabajo para la aplicación de un modelo estadístico que permita calcular pronósticos para la demanda de productos de una empresa comercializadora de ítems de ferretería. Para el desarrollo del presente trabajo de investigación se segmentará en las siguientes etapas:

3.1. Los datos

Se tuvieron datos de las ventas de los productos de ferretería como: tubos de presión roscable, tubos de desagüe, breakers, cintas temflex, interruptores, tee de desagüe, codos polipropileno, hojas de sierra, discos metal plano, etc. El producto seleccionado fue el tubo de presión roscable ½ (SKU 17050), el cual presentó el mayor volumen de ventas.

Se utilizó una serie de datos histórica con periodicidad mensual, que contiene las ventas totales del producto desde el período de enero 2016 hasta diciembre 2020, teniendo un total de $n=60$ datos. Los datos de las ventas son registrados de manera diaria a través de un ERP que mantiene la integración de procesos como compras e importaciones, facturación, ventas, entre otros. Además, permite la obtención de reportes de ventas en libros de Microsoft Excel, los cuales contienen los campos de SKU, descripción, marca, unidad de medida, cantidad vendida, costos, precio unitario y monto total, requeridos por el departamento del área Comercial para su control.

A continuación, se detallan las variables macroeconómicas que serán utilizadas para el análisis de la predicción de la demanda del producto seleccionado:

Producto Interno Bruto – Construcción

El Producto Interno Bruto (PIB) es considerado a nivel mundial uno de los indicadores macroeconómicos más importantes para medir la evolución y composición de la economía de un país. El PIB es la suma del valor monetario de todos los bienes y servicios que fueron producidos en el país durante un determinado período de tiempo, generalmente de un año. Este indicador es presentado de manera trimestral por el Banco Central del Ecuador.

Cuando se habla del PIB se debe tener en consideración que existen dos tipos de valoración: nominal y real.

El PIB nominal, es aquel cuyo cálculo se realiza en base a los precios del mercado de los productos y servicios producidos, teniendo en consideración la inflación o deflación, mientras que, el PIB real se basa al valor monetario de los bienes y servicios producidos en el período determinado, sin considerar las variaciones originadas producto de la inflación; es decir, el PIB nominal adiciona el efecto de la inflación y el PIB real no lo considera, calcula a precios constantes, lo que permite percibir con mayor claridad la evolución que presenta la economía de un país. Las diferencias de los precios que se originan son denominada como deflactor del PIB.

La forma de efectuar el cálculo del Producto Interno Bruto es compleja; por ello, existen tres metodologías o formas para efectuarlo:

- Enfoque de Ingresos
- Enfoque de Gastos
- Enfoque de la Producción

Uno de los pilares más importantes para el Ecuador es el sector de la construcción, el cual en año 2019 representó el 8.17% del PIB real, además de generar 6.1% del total de empleos y atraer inversión extranjera, es por esto que el aporte de este sector es altamente significativo a la economía de nuestro país (Lucero, 2020).

En el 2018, el sector de la construcción constituyó un 7% a nivel mundial en la generación de empleos.

Índice de Precios de Tuberías

El Instituto Nacional de Estadísticas y Censos (INEC) presenta de manera mensual los índices de precios de materiales, equipo y maquinaria de la construcción. Estos índices son calculados con los precios tanto de importador, como de productor e incluyen el Impuesto al Valor Agregado, salvo que exista exoneración en los productos. Los índices de precios se ejecutan en función a la base del 12 de abril del 2000 y tienen un valor de 100.

Los índices de precios son publicados con una periodicidad mensual por el INEC y tienen como propósito ser considerados para su aplicación en las fórmulas de reajuste de precios de las obras públicas. (Instituto Nacional de Estadísticas y Censos, 2020).

Desempleo Nacional

Otra de las variables macroeconómicas importantes para el país es la tasa de desempleo, en la cual expresa a nivel nacional, la desocupación de la Población Económicamente Activa (PEA).

El INEC es quien proporciona las estadísticas del mercado laboral con una periodicidad trimestral, obtenidas a través de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), la que tiene como principal objetivo proporcionar la información de las fuentes de ingreso y actividad económica de la población ecuatoriana. (Instituto Nacional de Estadística y Censos, 2021).

Gasto Público

El Gasto Público son todos aquellos gastos en bienes y servicios que realiza el Estado en todos sus niveles (locales, nacional, regional, etc.), es decir, comprende a la suma del consumo e inversión pública. El incremento del gasto público es considerado como una herramienta para el aumento de la

economía de un país, puesto que el PIB también tiende a aumentar. Este indicador es registrado por el Banco Central del Ecuador con una periodicidad mensual.

El primer trimestre del 2020 generó resultados preocupantes debido a la pandemia del COVID-19, motivo por el cual el Estado tuvo que realizar una redistribución del presupuesto para poder atender las necesidades que resultaron emergentes, aquellas relacionadas al sector de bienestar social, lo que conllevó a una disminución en la inversión pública.

Préstamos Hipotecarios del Banco del IESS

El Banco del Instituto Ecuatoriano de Seguridad Social (BIESS), presta servicios financieros (préstamos hipotecarios y quirografarios) para sus afiliados y jubilados, fue creado con la finalidad de apoyar proyectos de inversión que favorezcan a la economía ecuatoriana, además de la generación de empleos. El BIESS mantiene registros mensuales de las operaciones realizadas por los afiliados.

Desde su creación en el año 2009, mantuvo un crecimiento a lo largo de cuatro años en el otorgamiento de créditos hipotecarios lo que ha permitido que el sector de la construcción tenga una importante participación en el PIB. Estos préstamos pueden ser utilizados para la adquisición de viviendas, terrenos, locales comerciales, construcción de viviendas, remodelaciones, ampliaciones, etc.

Créditos otorgados en el Sector Privado

Los créditos son unas de las principales operaciones que realizan todas las entidades del sector financiero ecuatoriano, los cuales se encuentran clasificados de acuerdo a las necesidades de los diferentes tipos de personas (Natural y Jurídica).

Debido al sin número de necesidades que se presentan, sean estas por emprendimientos, estudios, construcción, viviendas, etc., el crédito representa uno de los pilares fundamentales para el sector financiero,

favoreciendo además el desarrollo económico. Estas operaciones realizadas en instituciones financieras del sector privada son registradas por el BCE y mantiene registros mensuales.

Las instituciones del Sistema Financiero Privado destinan créditos para viviendas conforme lo establecido en la Ley Orgánica para la regulación de los Créditos para Viviendas y Vehículos sea para compra o construcción de viviendas, remodelaciones o adecuaciones. Desde la creación del BIESS las instituciones de este sector han disminuido su participación de las operaciones crediticias en el segmento de vivienda.

3.2. Métodos estadísticos utilizados

Para el desarrollo de este estudio, se realizó la segmentación del conjunto de datos en dos grupos:

1. Datos de entrenamiento (Training Set) y
2. Datos de prueba (Test Set).

Los datos de entrenamiento, están conformados por las primeras 50 observaciones de la serie de datos y corresponden al período desde enero del 2016 a febrero del 2020, fueron tomados con la finalidad de entrenar los modelos. Mientras que, los datos de prueba, corresponden a las últimas 10 observaciones de la serie y pertenecen al período de marzo a diciembre del 2020, estos datos se utilizaron para la evaluación del rendimiento de los modelos.

El desarrollo analítico del presente trabajo de investigación, se compone en tres etapas que se describen a continuación:

1. Análisis de los datos: Se analizarán los datos reales de las ventas mensuales de la empresa, luego se procederá a seleccionar el producto más representativo en relación al volumen de ventas.

2. Obtención del modelo estadístico: Se calcularán varios modelos estadísticos de pronóstico cuantitativo de la demanda, para este propósito se utilizan los métodos descritos en el Capítulo 2 que son: ARIMA, modelo de regresión múltiple y de regresión con errores. Para determinar su comportamiento y ajuste, se seleccionará aquel que mejor describa el comportamiento plausible de las ventas del producto seleccionado.
3. Evaluación del desempeño del modelo: Se realizará la validación del modelo para demostrar su correcto desempeño, realizando el pronóstico de 10 períodos de prueba (test set) para compararlos con los datos reales y evaluar el error de estimación del modelo seleccionado aplicando las métricas de los errores mencionadas en el Capítulo 2.

Por los procesos de análisis antes mencionados, se determina que esta investigación está desarrollada como un estudio del tipo observacional, descriptivo, longitudinal.

Se utilizó el software RStudio versión 4.0.4 (2021-02-15), con los paquetes: forecast, astsa, tseries, lmtest, ggplot2, car, TSstudio, tidyverse, urca, modelr, corrplot y psych, los cuales sirvieron de ayuda para efectuar un análisis estadístico a través de las transformaciones, ajustes, creación y evaluación de los tres modelos empleados en el presente proyecto de investigación.

Mediante el uso del software RStudio se pueden establecer de manera manual los componentes del modelo ARIMA; sin embargo, para obtener un mayor provecho del software se usará la función auto.arima, la cual proporciona el mejor modelo teniendo en consideración la estacionariedad, diferencias, etc., basado en el Criterio de Información Akaike (AIC por sus siglas en inglés de Akaike Information Criteria), el AIC corregido (AICc) y el Criterio de Información Bayesiano (BIC).

Esta función realiza el ARIMA de manera automática, en donde efectúa iteraciones en búsqueda del mejor modelo que se ajuste a los datos.

3.3. Análisis de datos

Esta etapa, se fundamenta en efectuar el análisis del comportamiento de los datos que son objeto de estudio, mediante la representación gráfica que nos permitirá obtener una mejor visión sobre los cambios presentados en la demanda durante los años desde el 2016 al 2020.

3.4. Obtención del Modelo estadístico

Luego de analizar los datos macroeconómicos (Producto Interno Bruto – Construcción, Índice de Precios de Tuberías, Desempleo Nacional, Gasto Público, Préstamos Hipotecarios del Banco del IESS y Créditos otorgados en el Sector Privado) en el apartado 3.2.2 y los datos históricos de las ventas totales del SKU 17050 cuyas ventas totales ascienden a US\$ 145.346.790 (del 2016 al 2020), se procedió a obtener tres modelos de series de tiempo para posteriormente determinar el que mejor se ajuste a los datos estudiados.

CAPÍTULO 4

4. RESULTADOS

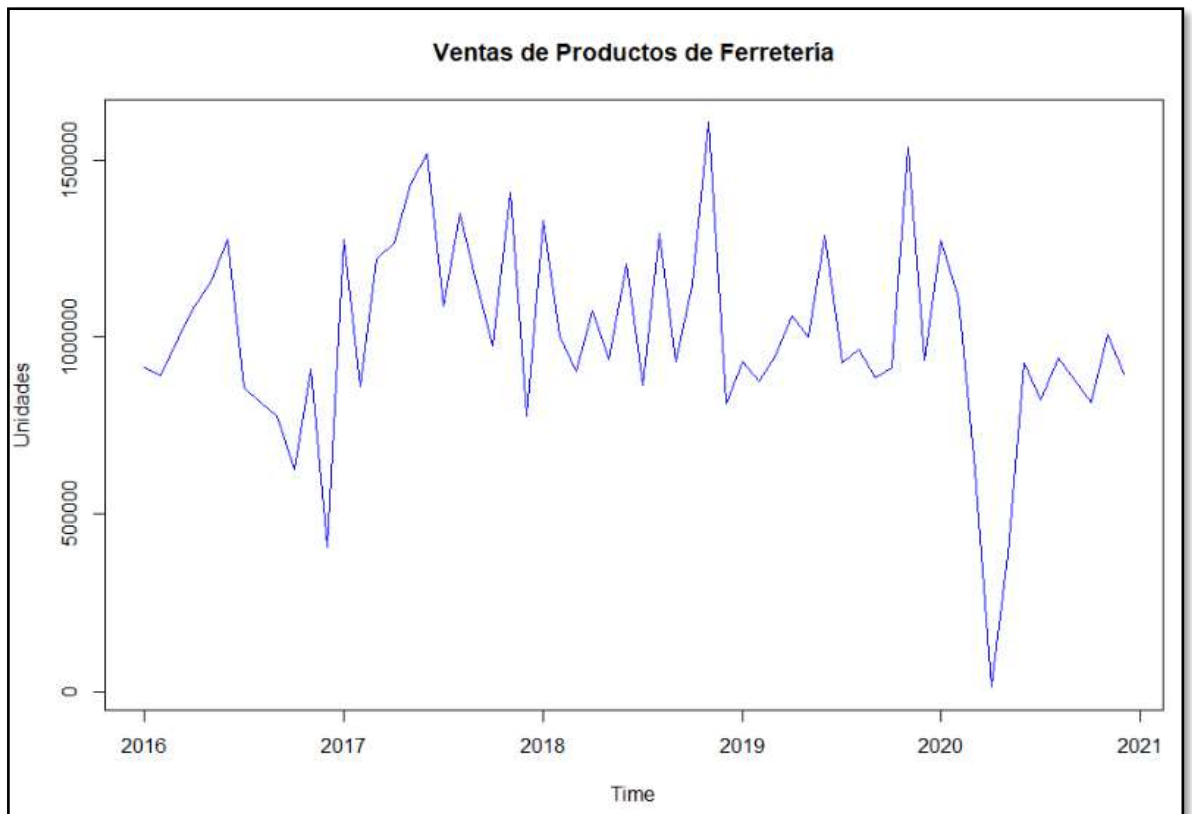
En esta sección se presentan los resultados obtenidos del análisis realizado a la serie de datos de la demanda, así como la aplicación de los modelos descritos con anterioridad.

4.1. Comportamiento de los datos

4.1.1. Análisis de la Variable Ventas Totales

En la Figura 4 se presentan las ventas totales de los productos de la línea de ferretería comercializados por la compañía.

Figura 4. Ventas de los Productos de Ferretería

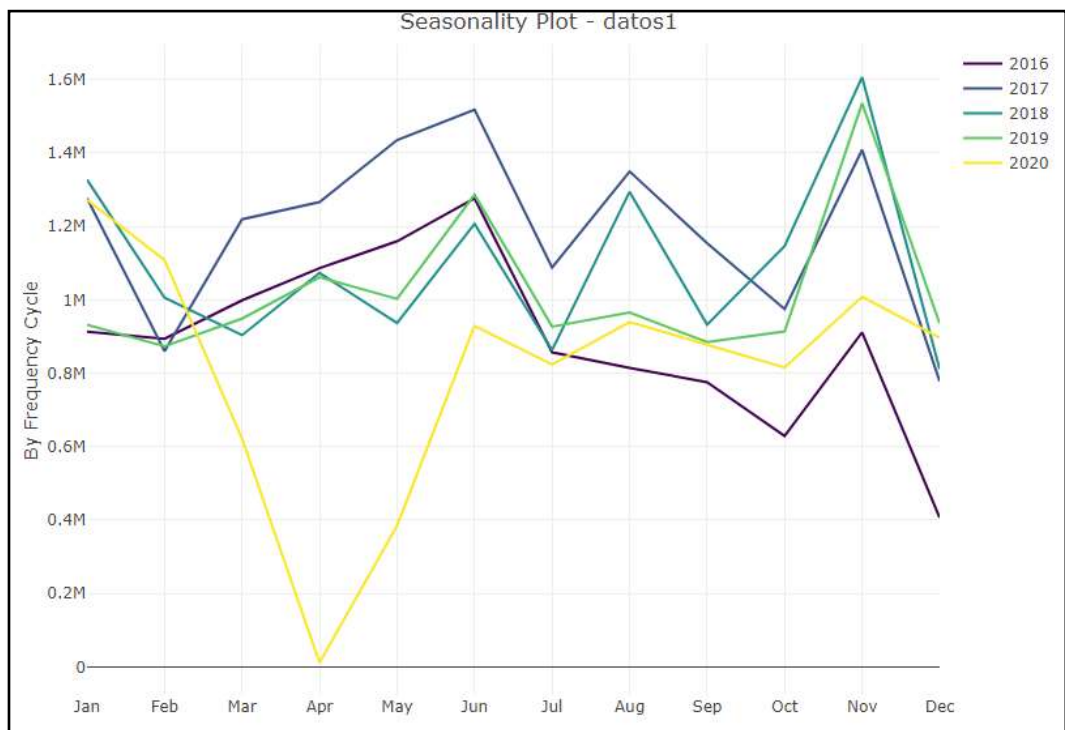


Fuente: Elaboración propia

Como se observa en la Figura 4, los picos altos en las fluctuaciones se evidencian en el mes de agosto que corresponde al mes de la construcción y es cuando se realizan promociones y descuentos y en los meses de abril, junio y noviembre en donde la empresa de forma anual realiza ferias otorgando atractivos descuentos a sus clientes por volúmenes de compras.

En el período 2020 se presentan ciertas caídas drásticas desde el mes de marzo, debido a la emergencia sanitaria producida por la pandemia del COVID-19 siendo los meses de marzo, abril y mayo los que tuvieron menor demanda en todos los períodos.

Figura 5. Ventas estacionales de los Productos de Ferretería

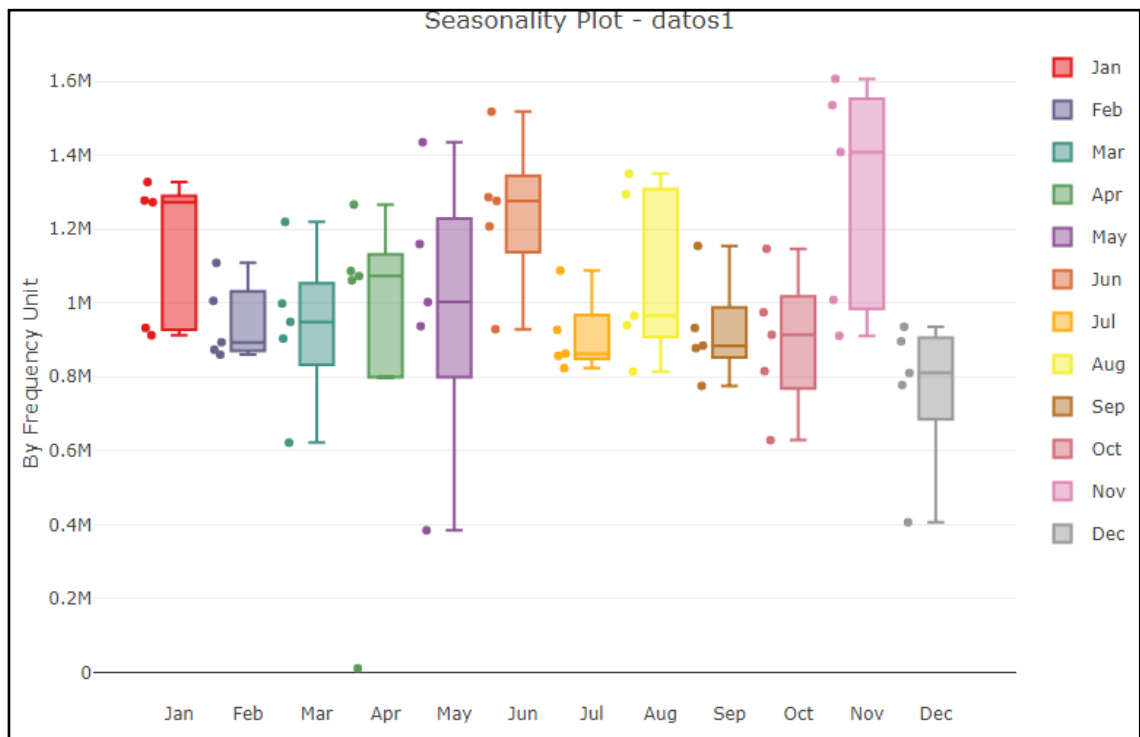


Fuente: Elaboración propia

Como se puede apreciar en las ventas por meses de los productos de ferretería (Figura 5 y 6), existen incrementos de ventas en los meses de abril, junio, agosto y noviembre en los años del 2016 al

2019, mientras que en el año 2020 se registra una clara disminución producto del impacto ocasionado por la emergencia de salud debido a la pandemia del coronavirus (COVID-19).

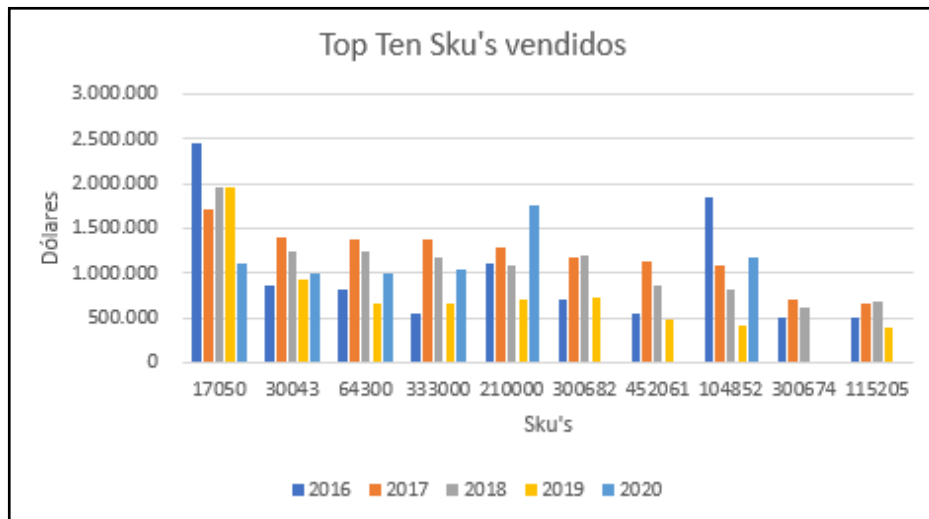
Figura 6. Ventas estacionales de Productos de Ferretería – Diagrama de Cajas



Fuente: Elaboración propia

Para determinar el SKU que presenta una mayor demanda, se procedió a representar de manera gráfica los diez productos que presentaron mayores ingresos para la organización.

Figura 7. Top Ten de Sku's más vendidos



Fuente: Elaboración propia

Como se puede evidenciar en la figura 7 el ítem que ha tenido mayor demanda es el SKU 17050, en el período comprendido del 2016 al 2109 obtuvo ventas anuales por encima del resto de productos comercializados por la organización, mientras que en el año 2020 se posicionó en el tercer lugar de los productos con mayor demanda. El SKU 17050 pertenece al grupo de Tubería Roscable PVC y representa el 6.31% del total de los ingresos obtenidos durante los períodos de enero del 2016 a diciembre del 2020.

Conforme a lo establecido en el alcance de este trabajo, el análisis y estudio se realiza al SKU 17050 por los motivos anteriormente expuestos.

4.1.2. Análisis de Variables Macroeconómicas

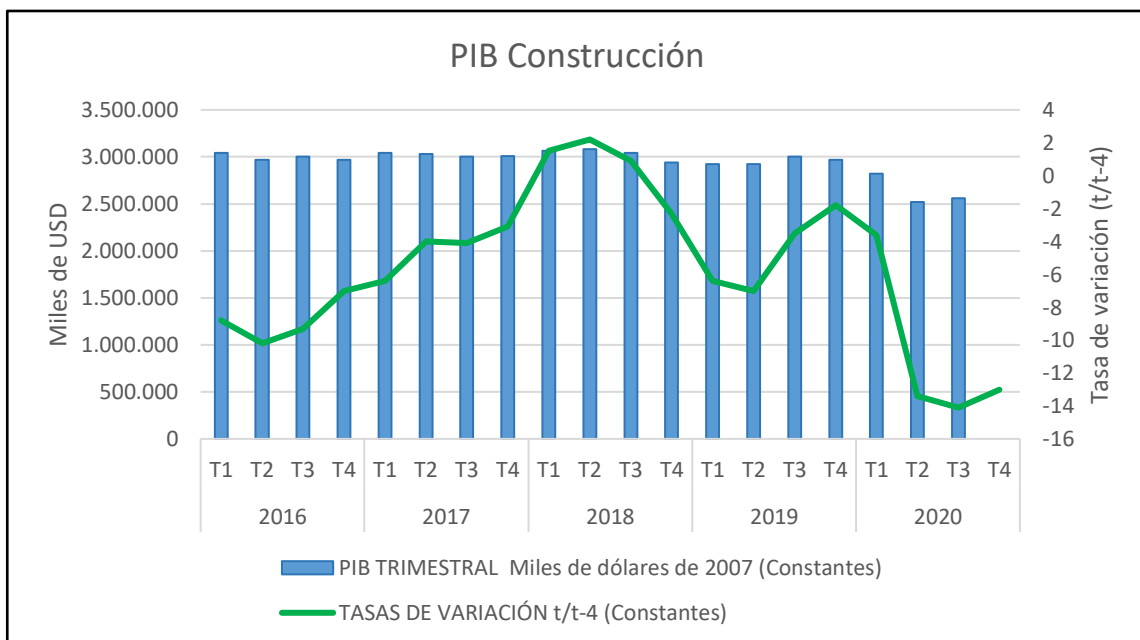
En el siguiente apartado se realizará el análisis del comportamiento de las variables macroeconómicas que se utilizarán en el presente trabajo de investigación, para la aplicación de un modelo que permita calcular el pronóstico de la demanda del producto seleccionado, estas variables son:

- Producto Interno Bruto – Construcción
- Índice de Precios de Tuberías
- Desempleo Nacional
- Gasto Público
- Préstamos Hipotecarios del Banco del IESS
- Créditos otorgados en el Sector Privado

Producto Interno Bruto - Construcción

A continuación, en la figura 8 se presenta el indicador trimestral del PIB sectorial de Construcción y la tasa de variación del período 2016 al 2020 en miles de dólares (Constantes) como base el 2007=100, utilizado por el Banco Central del Ecuador.

Figura 8. Producto Interno Bruto Construcción



Fuente: Elaboración propia, datos tomados del Banco Central del Ecuador (BCE)

Analizando el gráfico se puede observar que durante el primer trimestre del año 2016 al segundo trimestre del año 2018 existe una tendencia creciente, mientras que a partir del tercer trimestre del año 2018 al segundo trimestre del año 2019 se registró una

tendencia decreciente, la misma que mejoró en los dos trimestres siguientes.

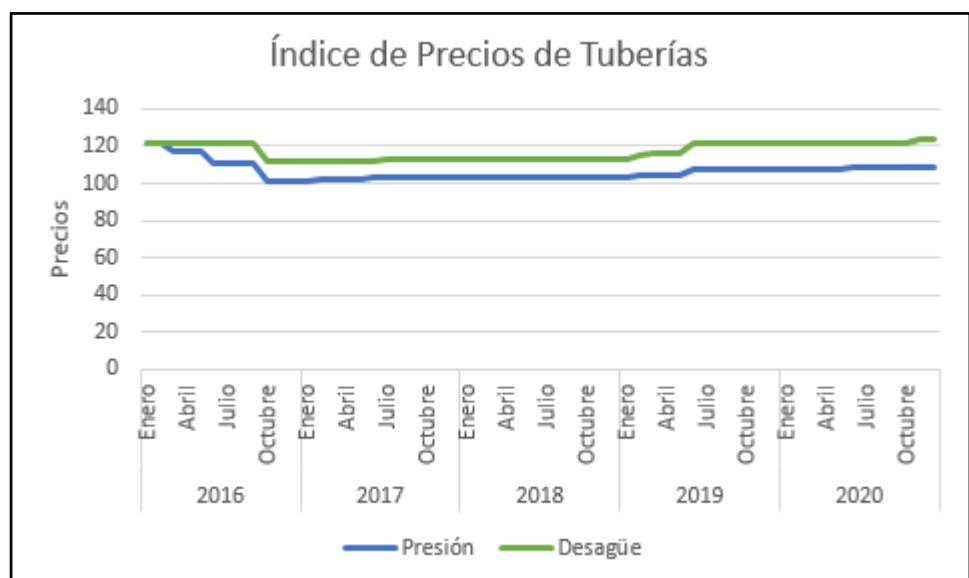
El año 2020 no fue lo esperado, debido a que desde el primer trimestre se registraron bajas, siendo así que en el segundo trimestre se produjo una baja altamente significativa producto de la emergencia sanitaria a nivel mundial originada por la pandemia del COVID-19.

En el cuarto trimestre del año 2020, el PIB en el sector de la construcción presentó un incremento en relación al tercer trimestre del mismo año, existiendo una leve recuperación económica para el país a consecuencia de la emergencia sanitaria.

Índice de Precios de Tuberías

Los datos de los índices de precios de tuberías se muestran a continuación:

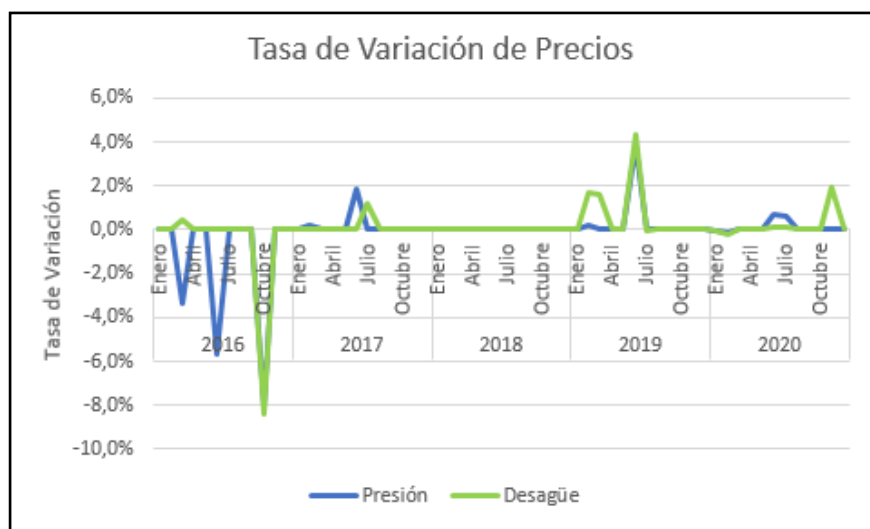
Figura 9. Índice de Precios de Tuberías



Fuente: Elaboración propia, datos tomados del INEC

En la figura 9 se observa que durante los trimestres del año 2016 se registró el decrecimiento de los precios de tuberías de presión como de desagüe y del primer trimestre del año 2017 al segundo trimestre del 2019 hubo un leve crecimiento. Finalmente, en los dos últimos trimestres del año 2020 se ha mantenido constante los precios de las tuberías de presión y un similar comportamiento se mantuvo con las tuberías de desagüe.

Figura 10. Tasa de Variación de Precios de Tuberías



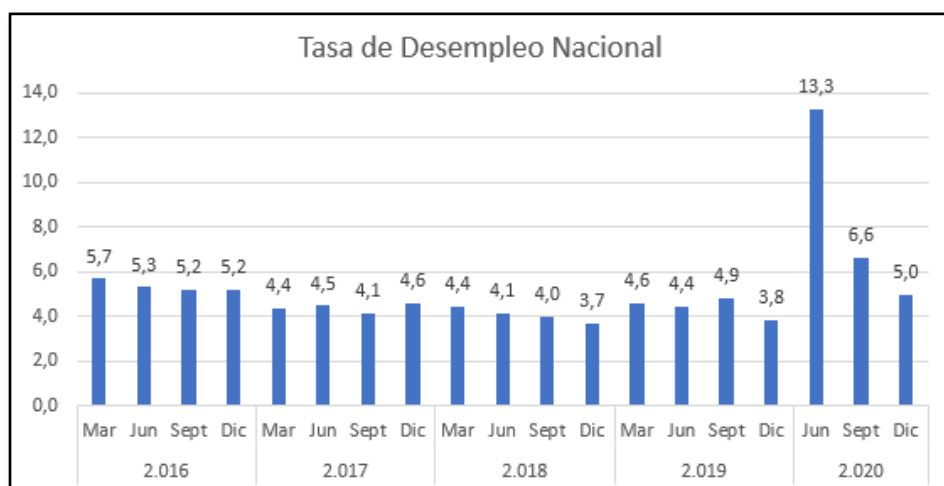
Fuente: Elaboración propia

En el tercer trimestre del 2016 se registró una tasa decreciente del 8.1% en comparación con los otros trimestres de los otros años donde no hubo variación alguna en los precios de tubería de presión y de igual manera en los precios de tubería de desagüe en donde se registró una tasa decreciente del 8.4% siendo las tasas de variación más altas registradas durante los 5 años (2016 al 2020).

Desempleo Nacional

En la Figura 11 que se muestra a continuación, se presentan las tasas de desempleo a partir del 2016 al 2020:

Figura 11. Tasa de Desempleo Nacional



Fuente: Elaboración propia, datos tomados del INEC

Las tasas de desempleo más bajas en el período del 2016 al 2020 se dieron en el cuarto trimestre de los años 2018 y 2019 registrando porcentajes del 3.7% y 3.8% respectivamente.

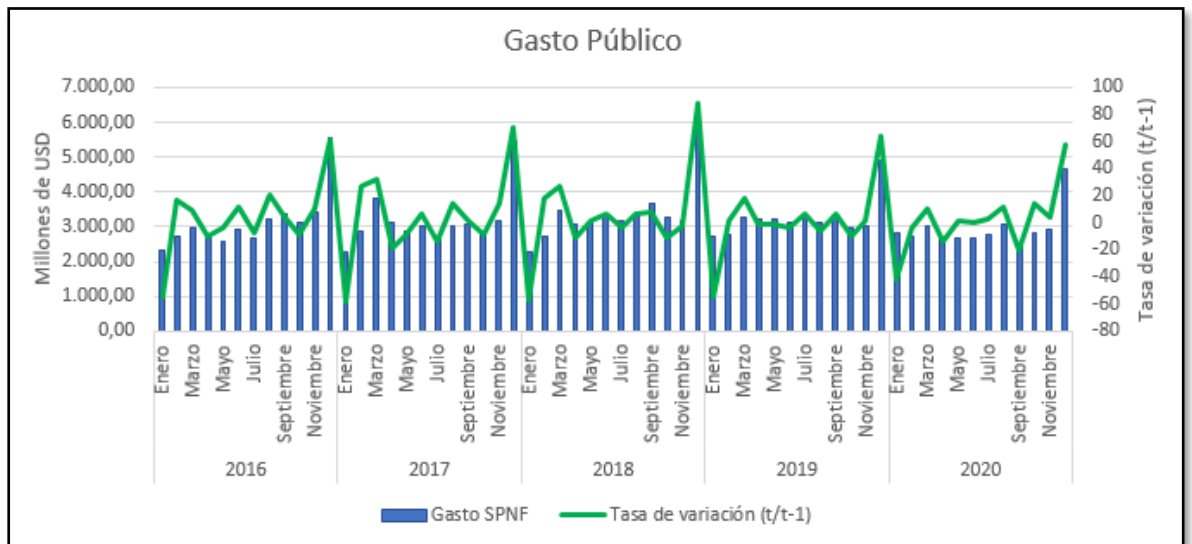
El desempleo nacional en el año 2020 alcanzó el 13.3% siendo de las peores tasas registradas en la historia del Ecuador, esto como consecuencia de la emergencia sanitaria presentada a nivel mundial, la misma que desencadenó una grave crisis económica nacional en la que se produjeron reducciones salariales, de jornadas laborales y un sin número de despidos de personal tanto en el sector público como privado.

En el tercer y cuarto trimestre del 2020 se registró una importante reducción de la tasa de desempleo llegando al 6.6% y 5.0% respectivamente, gracias a la reactivación económica de ciertas actividades comerciales.

Gasto Público

En relación al Gasto Público se muestra a continuación, en la Figura 12 su evolución:

Figura 12. Gasto Público del Estado



Fuente: Elaboración propia, datos tomados del BCE

El gasto público experimentó una baja en el primer y segundo trimestre del 2020 en relación al año 2019. En el período 2019 presentó una baja de 1.9 puntos en relación al período 2018, en donde el gasto representó el 38.5% del Producto Interno Bruto. Ecuador ha empeorado su situación en relación al gasto lo que ha generado que exista un descenso de puesto en la escala a nivel mundial según el gasto público per cápita.

Préstamos Hipotecarios del Banco del IESS

En el gráfico siguiente (Figura 13), se detallan las operaciones de créditos hipotecarios otorgados por el Banco del Instituto Ecuatoriano de Seguridad Social (BIESS) en el período comprendido entre el 2016 al 2020:

Figura 13. Operaciones de Créditos Hipotecarios en el Biess



Fuente: Elaboración propia, datos tomados de Subgerencia de Operaciones Biess

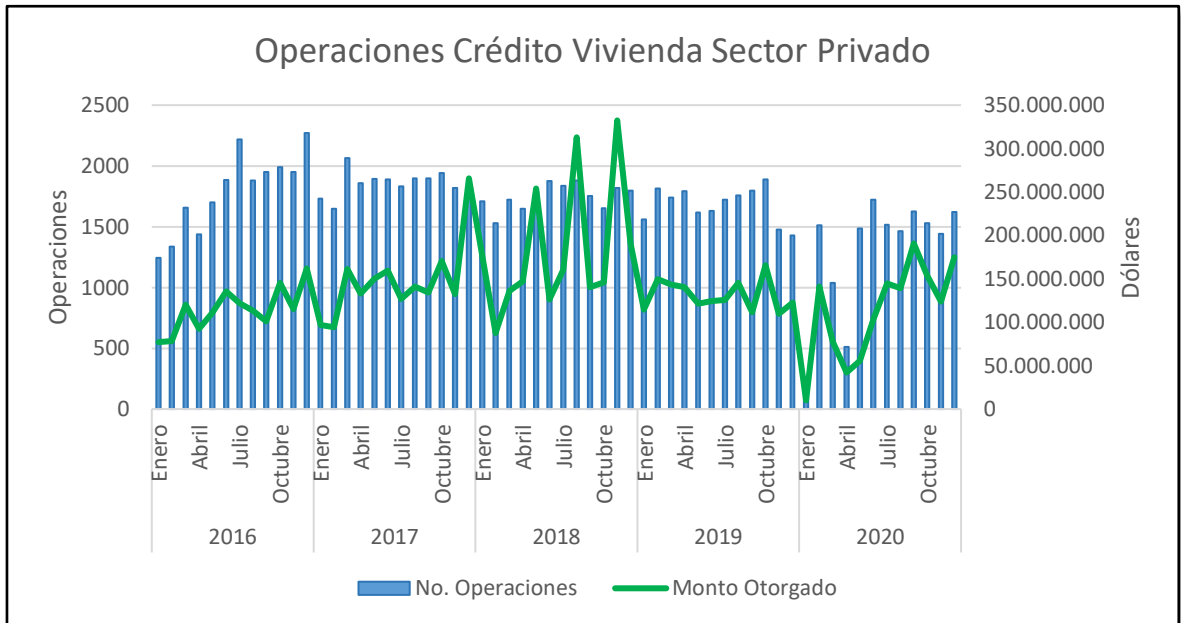
En la figura 13 se puede apreciar el número de operaciones netas generadas en los períodos del 2016 al 2020, las mismas que mantienen una tendencia decreciente conforme la situación económica del país. En el año 2018 se realizaron un total 18.039 operaciones por un monto de USD 829 millones de dólares, en el año 2019 se registró un monto de USD 564.7 millones en 12.045 operaciones (disminución del 31%) mientras que en el año 2020 se registraron un total de 6.795 operaciones por el monto de USD 350 millones de dólares.

La baja significativa generada en el período 2020 se debió a la crisis económica que desencadenó la pandemia por el COVID-19, al tener un impacto negativo en la adquisición de préstamos debido a la reducción de sueldos y remuneraciones, así como el aumento del desempleo a nivel nacional.

Créditos otorgados en el Sector Privado

En la Figura 14 que se presenta a continuación, se muestra el número de operaciones de créditos para viviendas otorgados por el Sistema Financiero Privado:

Figura 14. Operaciones de Crédito Vivienda en Sector Privado



Fuente: Elaboración propia, datos tomados de la SBE

De acuerdo a los datos tomados de la Superintendencia de Bancos del Ecuador (SBE) se observa en la figura 14 el número de operaciones crediticias y montos asignados del segmento de viviendas generadas en los períodos del 2016 al 2020, este último mantiene una tendencia creciente hasta el año 2018 y a partir del 2019 se observa un deceso en el mismo. Con relación al número de operaciones realizadas, en el año 2016 y 2017 se generó un crecimiento en sus operaciones mientras que en el período del 2017 al 2020 existe una tendencia decreciente registrando 21.011, 20.211 y 15.593 operaciones respectivamente.

4.2. Construcción del Modelo

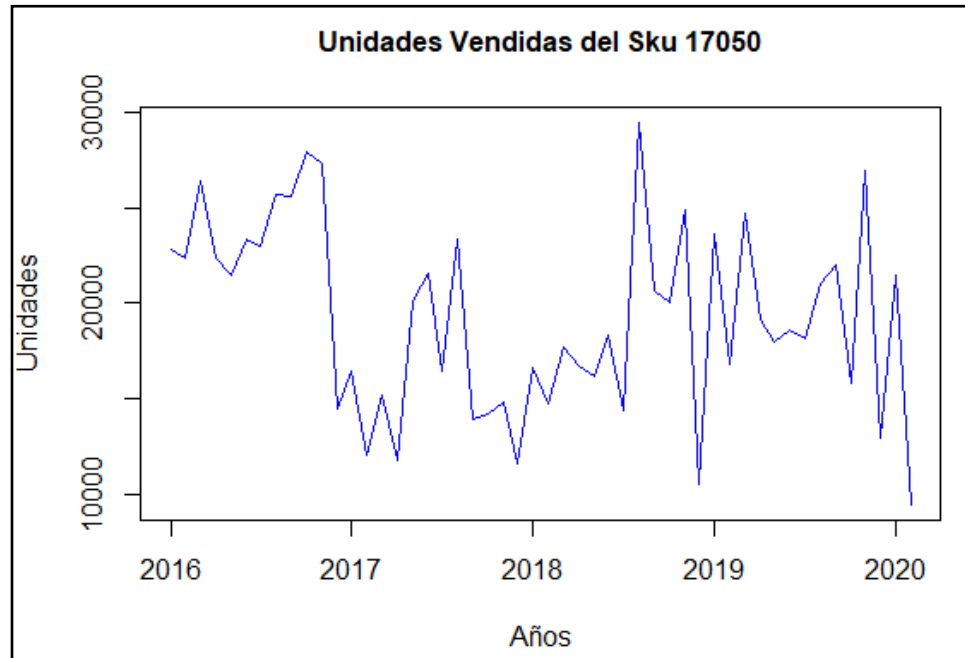
En el siguiente apartado se realizará la construcción del modelo para cada uno de los modelos estadísticos que a continuación, se detallan:

- Modelo ARIMA no estacionario
- Modelo de Regresión Múltiple
- Modelo de Regresión con errores ARIMA

4.2.1. Modelo ARIMA no estacionario

Se inicia la construcción del modelo de la demanda del SKU 17050 mediante la representación gráfica de la serie de tiempo.

Figura 15. Demanda del SKU 17050



Fuente: Elaboración propia

Se puede observar en la figura 15 que la serie de tiempo no presenta estacionariedad dadas ciertas características como las tendencias y niveles variantes, sin embargo, para comprobar lo mencionado se aplica el test de estacionariedad de Kwiatkowski, Phillips, Schmidt y Shin (KPSS) el cual permite contrarrestar la estacionariedad en nivel (media) con significancia $\alpha = 0.05$ en donde establece la siguiente hipótesis:

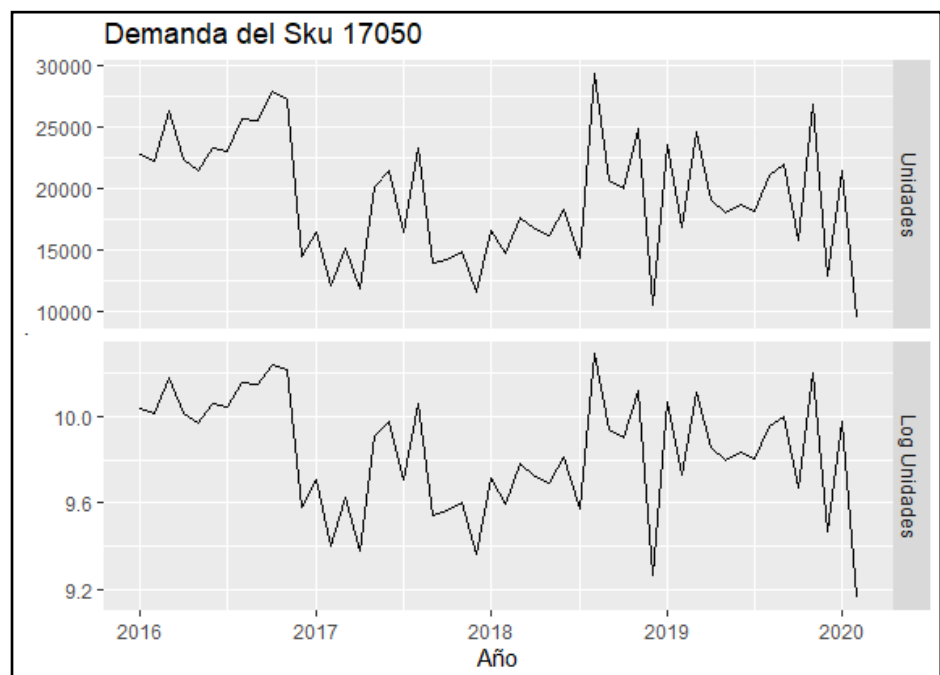
H_0 : La serie tiene nivel estacionario y,

H_1 : La serie no tiene nivel estacionario.

Como resultado del test de KPSS se obtuvo un p-value de 0.02576; dado que el valor p obtenido es menor al nivel de significancia entonces se rechaza la hipótesis nula con un 95% de confianza.

Como se determinó que la serie no es estacionaria, se procede a realizar la transformación logarítmica de los datos de la serie; para lo cual se realiza este paso mediante la transformación logarítmica.

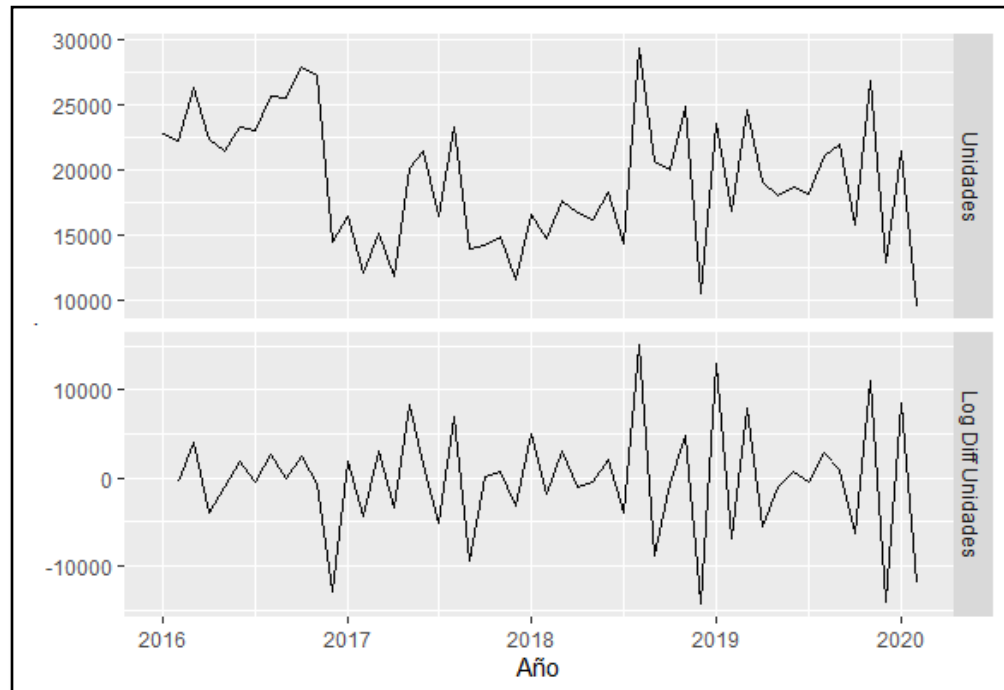
Figura 16. Transformación logarítmica del SKU 17050



Fuente: Elaboración propia

De acuerdo a la figura 16 se puede observar que luego de la transformación no se logra obtener una estabilización en la media, para afirmar esto efectuaremos nuevamente el test de KPSS, dando como resultado un p-value de 0.02736; por lo tanto, se puede corroborar que la serie no es estacionaria. Al no poder lograr estacionariedad en los datos es necesario efectuar la primera diferencia, misma que se muestra a continuación:

Figura 17. Primera Diferencia de los datos



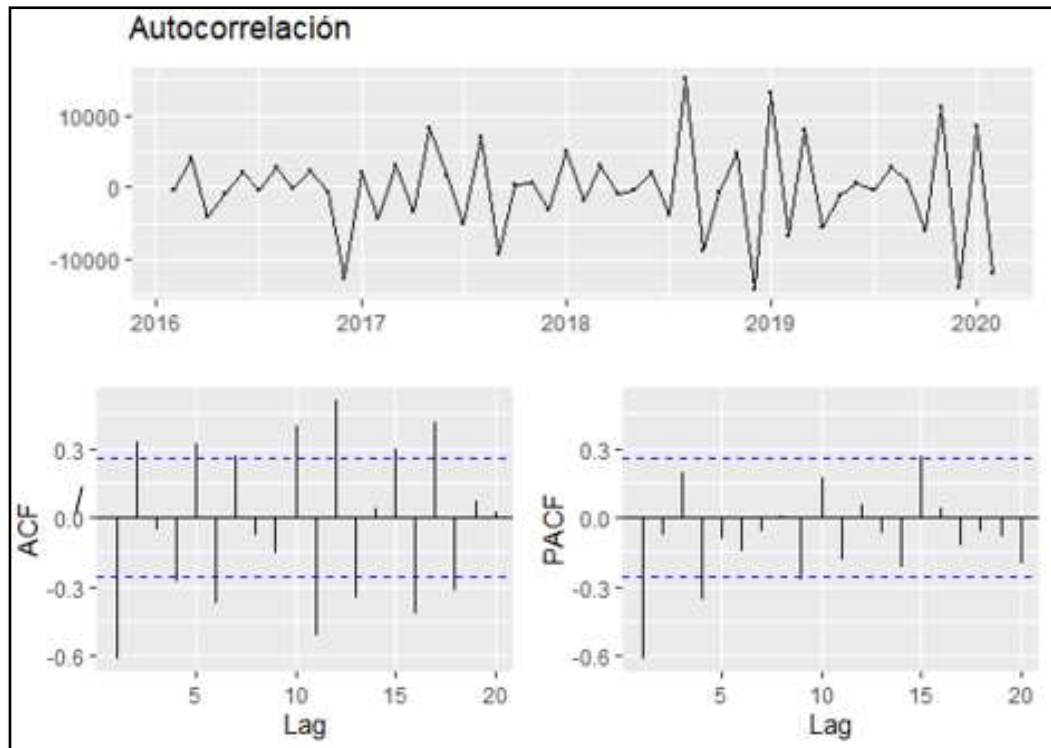
Fuente: Elaboración propia

Como se puede apreciar en la figura 17 con la primera diferenciación generada (Log Diff Unidades) se logra la estabilización en la media de la serie de datos del SKU 17050. Se efectúa nuevamente el test de KPSS para determinar si existe estacionariedad en la cual se obtiene el p-value de 0.1, por tanto, se establece que con un 95% de confianza, los datos analizados no presentan evidencia estadística para rechazar la hipótesis nula (H_0 : La serie tiene nivel estacionario). Lo anteriormente indicado, se puede corroborar con la función ndiffs, la cual indica el número de diferencias requeridas para lograr estacionariedad en los datos. Para este caso, mostró como resultado que era necesario realizar una diferenciación.

Realizados los pasos 1 y 2, se procede a determinar los valores de los órdenes de p (componente AR) y q (componente MA) a través

la función de autocorrelación (ACF) y autocorrelaciones parciales (PACF) del acrónimo en inglés Partial Autocorrelation Function.

Figura 18. Funciones de Autocorrelación ACF y PACF



Fuente: Elaboración propia

Analizando los correlogramas se puede establecer de manera manual los componentes p y q de los datos que fueron transformados. De acuerdo a los patrones de cada una de los gráficos ACF y PACF se podrían establecer varios modelos de los cuales se determina como un posible modelo el ARIMA (1, 1, 3), debido a que al observar el correlograma de PACF la primera barra correspondiente al primer retardo es significativa lo que sugiere un AR (1) y en la gráfica de ACF se muestran rezagos significativos que pueden sugerir un MA (3). Una vez que se ha determinado el orden del modelo, mediante el uso del software RStudio se hallan los valores de los parámetros que permiten la obtención de los datos

estudiados al maximizar las probabilidades, para ello utiliza la estimación de máxima verosimilitud (MLE, por sus siglas en inglés).

Al ejecutar el modelo ARIMA, este nos proporciona los coeficientes, la capacidad predictiva y los criterios en los cuales se basa el modelo. Las medidas de error / ajuste que utiliza son dependientes de la escala y están basadas en porcentajes:

Medidas dependientes de la escala

- ME (Error Medio)
- RMSE (Raíz del Error Cuadrático Medio)
- MAE (Error Absoluto Medio)

Medidas basadas en porcentajes

- MPE (Error Porcentual Medio)
- MAPE (Error Porcentual Absoluto Medio)

Como resultado nos arroja el coeficiente de 0.3173 de la parte autorregresiva (p) y los coeficientes -1.0332, 0.8153 y -0.6447 correspondientes de la parte de media móvil (q), como se detallan en la tabla a continuación:

Tabla 1. Coeficientes del modelo ARIMA (1, 1, 3)

Coefficients:	ar1	ma1	ma2	ma3
	0.3173	-1.0332	0.8153	-0.6447
s.e.	0.1999	0.1642	0.1735	0.1326

Fuente: Elaboración propia

Además, se presentan las medidas de error del modelo ARIMA (1, 1, 3) (Ver tabla 2), las cuales permitirán establecer la elección del modelo que mejor se ajuste a los datos y muestra un Criterio de Información de Akaike (AIC) de 1165.15 y un Criterio de Información Akaike Corregido (AICc) de 1166.28.

Tabla 2. Medidas de Error del modelo ARIMA (1, 1, 3)

Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
Training set	-448.8089	4781.5	3869.315	-9.9297	22.3632

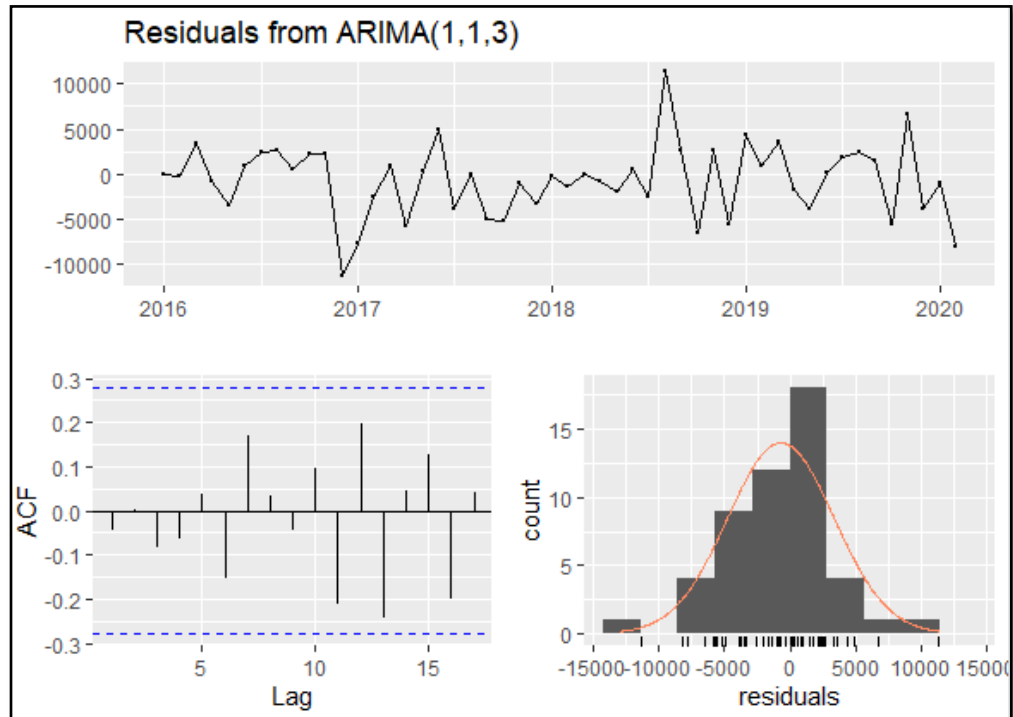
Fuente: Elaboración propia

Este modelo puede ser descrito de la forma: $Y'_t = 0.3173Y_{t-1} + 1.0332a_{t-1} - 0.8153a_{t-2} + 0.6447a_{t-3} + a_t$

Luego, se procede con la verificación de la autocorrelación ACF de los residuos estandarizados del modelo (Figura 19), en la cual se observa que los residuos si tienen un comportamiento de ruido blanco, esto se comprueba mediante el estadístico de Ljung-Box test donde se obtuvo un p- value = 0.7546, al ser mayor a 0.05 nos indica que los errores son independientes; es decir, cumplen con el supuesto de independencia (Figura 19). Se verifica, que los residuos cumplen con el supuesto de normalidad mediante la prueba de Shapiro-Wilk al obtener un p-value = 0.5548.

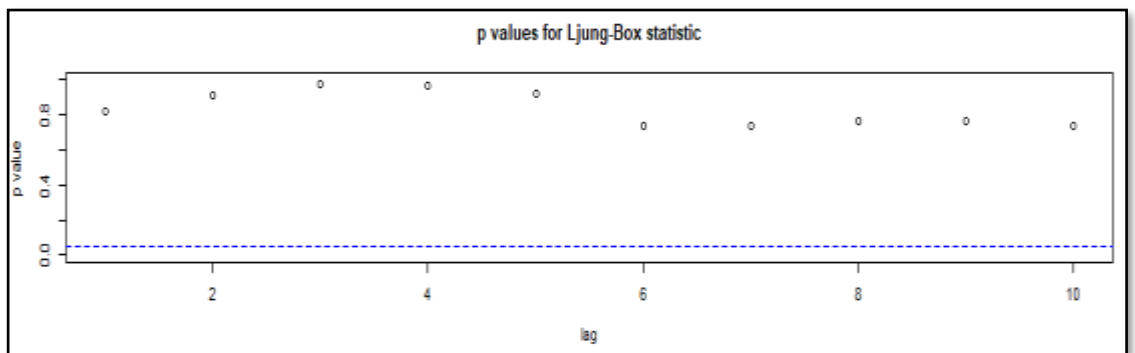
Otra forma de validar que exista el comportamiento de ruido blanco es mediante el diagnóstico del modelo (Figura 20) en donde nos indica que en los primeros 10 rezagos alcanzaron valores p muy altos y que los valores p son mayores a 0.05 (línea azul), lo que indica que los errores cumplen con la independencia.

Figura 19. Autocorrelación ACF de los Residuos



Fuente: Elaboración propia

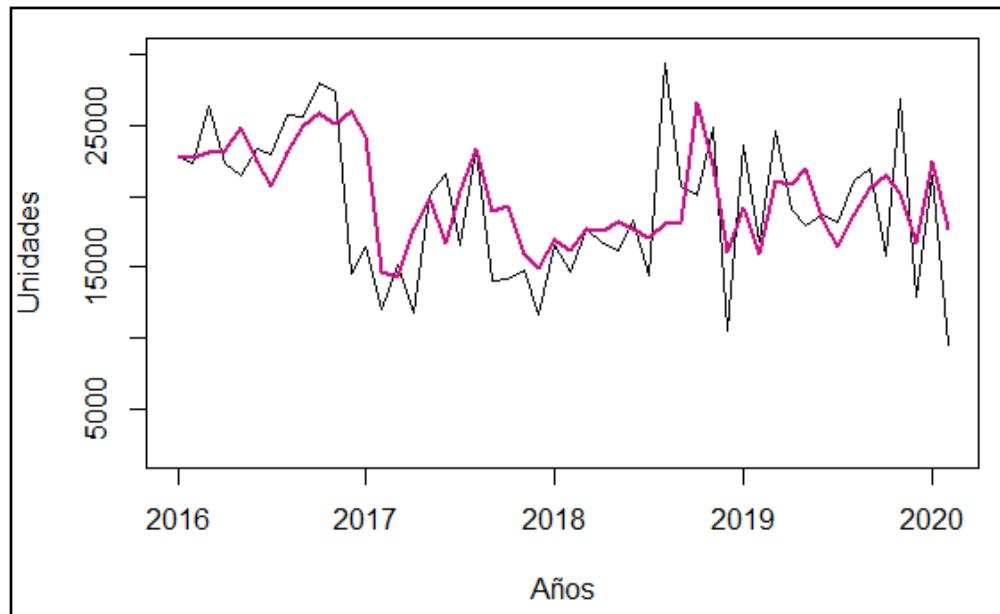
Figura 20. Valores p del Estadístico Ljung-Box



Fuente: Elaboración propia

A continuación, en la figura 21 se muestra la curva ajustada de los datos con el modelo ARIMA (1, 1, 3):

Figura 21. Curva ajustada del Modelo ARIMA (1,1,3)



Fuente: Elaboración propia

Al finalizar la corrida de la función `auto.arima`, presenta una lista de los modelos analizados (Ver tabla 3) y el mejor modelo seleccionado basándose en el AICc.

Como resultados obtenidos de esta función, se muestra como mejor modelo el ARIMA (0, 1, 3) siendo el que tuvo un menor AICc entre todos los modelos analizados, obtuvo un AIC de 1165.07 y un AICc de 1165.81, al comparar el modelo anterior (manual) con el modelo en mención (automático), se puede apreciar que el valor del AICc del modelo ARIMA (0, 1, 3) es mejor al modelo planteado previamente.

El modelo puede ser descrito de la forma: $Y'_t = 1.0315a_{t-1} - 0.8271a_{t-2} + 0.6360a_{t-3} + a_t$

Tabla 3. Modelos ARIMA de la función auto.arima

Modelos Arima		AICc
ARIMA(2,1,2)	with drift	1.170.013
ARIMA(0,1,0)	with drift	1.198.707
ARIMA(1,1,0)	with drift	1.172.748
ARIMA(0,1,1)	with drift	1.178.147
ARIMA(0,1,0)	with drift	1.196.609
ARIMA(1,1,2)	with drift	1.174.875
ARIMA(2,1,1)	with drift	1.173.800
ARIMA(3,1,2)	with drift	1.176.322
ARIMA(2,1,3)	with drift	Inf
ARIMA(1,1,1)	with drift	1.174.890
ARIMA(1,1,3)	with drift	Inf
ARIMA(3,1,1)	with drift	1.173.176
ARIMA(3,1,3)	with drift	Inf
ARIMA(2,1,2)		1.167.588
ARIMA(1,1,2)		1.172.517
ARIMA(2,1,1)		1.171.544
ARIMA(3,1,2)		1.173.811
ARIMA(2,1,3)		1.168.705
ARIMA(1,1,1)		1.172.706
ARIMA(1,1,3)		1.166.284
ARIMA(0,1,3)		1.165.808
ARIMA(0,1,2)		1.174.789
ARIMA(0,1,4)		1.166.165
ARIMA(1,1,4)		1.168.249

Fuente: Elaboración propia

A continuación, en la tabla 4 se detallan las medidas de Error del mejor modelo obtenido determinado por auto.arima, las cuales permiten medir la precisión del pronóstico.

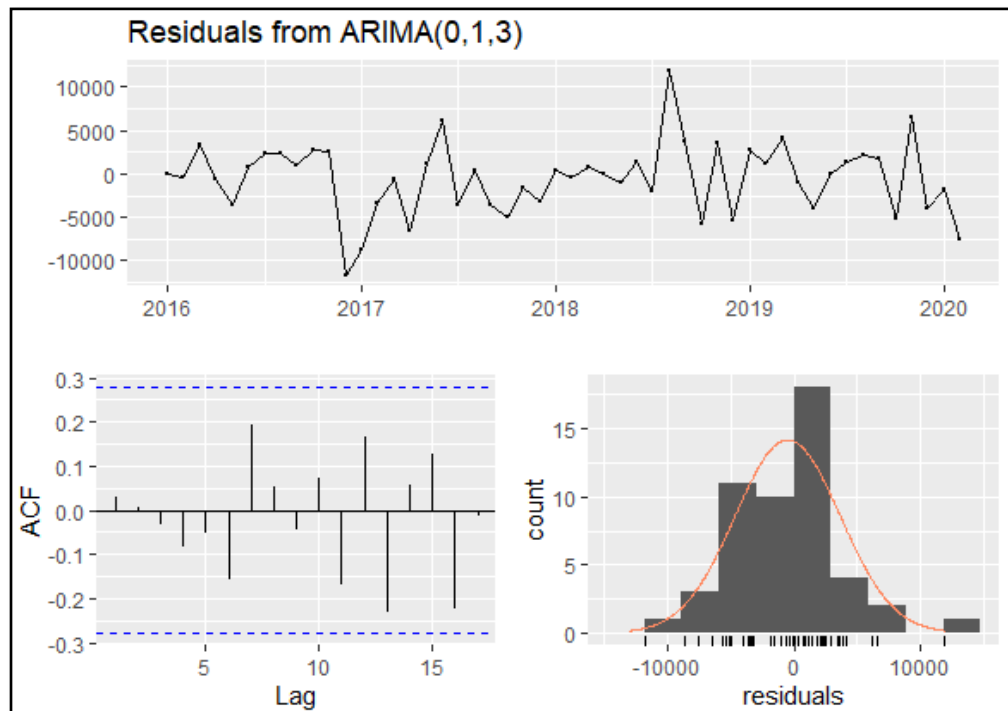
Tabla 4. Medidas de Error del modelo ARIMA (0, 1, 3)

Training set error measures:					
	ME	RMSE	MAE	MPE	MAPE
Training set	-654.5487	4713.746	3797.547	-11.0261	22.0102

Fuente: Elaboración propia

Se procede a realizar una verificación de los residuos del modelo obtenido (Ver figura 22) en donde se visualiza que el modelo se ajusta bien, mantiene una distribución normal, las autocorrelaciones de la función ACF se encuentra dentro de los límites, el error tiene una media igual a cero y varianza constante, existiendo ruido blanco.

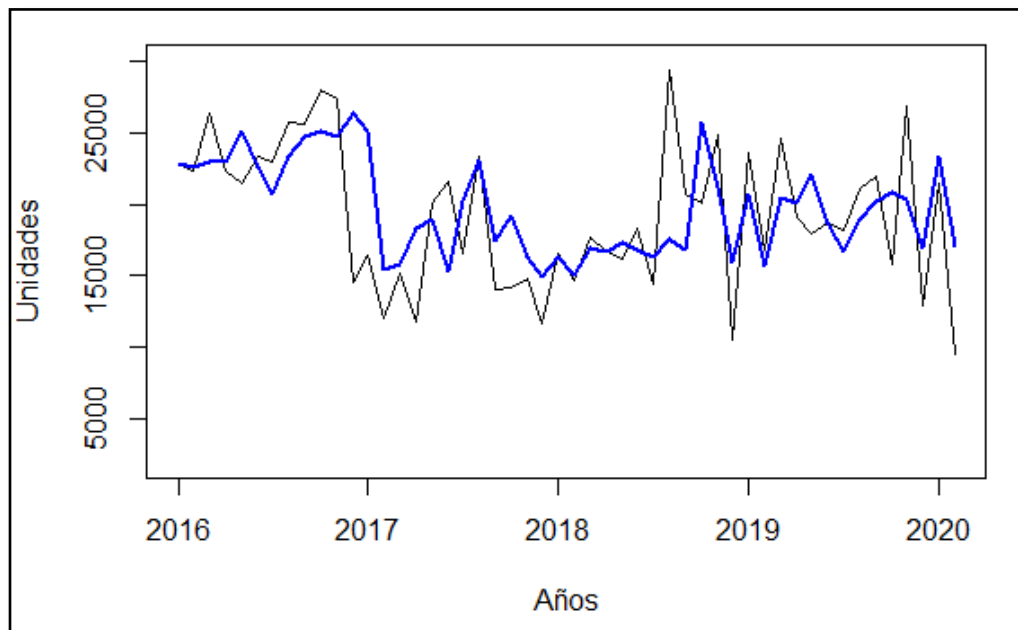
Figura 22. Autocorrelación ACF de los Residuos



Fuente: Elaboración propia

Se procedió a graficar la curva ajustada del modelo automático ARIMA (0, 1, 3), en donde se puede observar que el ajuste obtenido se aproxima a los datos originales de la serie objeto de estudio en el presente trabajo de investigación. A continuación, se muestra el gráfico:

Figura 23. Curva ajustada del Modelo ARIMA (0,1,3)



Fuente: Elaboración propia

4.2.2. Modelo de Regresión Múltiple

El siguiente modelo a obtener es mediante la aplicación de la técnica de regresión múltiple, el cual establece la relación funcional existente entre la variable de interés o independiente y la o las variables explicativas o regresoras.

Las variables independientes consideradas en el presente trabajo son variables macroeconómicas del Ecuador, tales como, el Producto Interno Bruto (PIB), Índice de Precios de Tuberías, Tasa de Desempleo Nacional, Gasto Público, Préstamos Hipotecarios del Banco del IESS y Créditos otorgados en el sector privado.

Como primer paso para la modelación se procedió con la elaboración de la matriz de correlación de Pearson para medir el grado de relación existente entre la variable de interés o variable endógena y las variables macroeconómicas antes descritas; es decir, las variables independientes o exógenas.

Tabla 5. Matriz de Correlación

	Ventas	PIB Sector Construcción	Índice de Precios de Tuberías	Tasa de Desempleo Nacional	Gasto Público	Préstamos Hipotecarios Biess	Créditos Viviendas SFP
Ventas	1	0.22	0.15	-0.24	-0.13	0.51	0.25
PIB Sector Construcción	0.22	1	-0.16	-0.14	-0.05	0.42	0.19
Índice de Precios de Tuberías	0.15	-0.16	1	0.19	-0.26	-0.09	-0.35
Tasa de Desempleo Nacional	-0.24	-0.14	0.19	1	-0.16	-0.24	-0.04
Gasto Público	-0.13	-0.05	-0.26	-0.16	1	0.04	0.24
Préstamos Hipotecarios Biess	0.51	0.42	-0.09	-0.24	0.04	1	0.56
Créditos Viviendas SFP	0.25	0.19	-0.35	-0.04	0.24	0.56	1

Fuente: Elaboración propia

Como se aprecia en la tabla 5, existen cuatro variables que mantienen una relación positiva con la variable de interés y son el PIB sector construcción, el índice de precios de tuberías, las operaciones de créditos de viviendas en el sector privado (bancos, cooperativas de ahorro y crédito, mutualistas, etc.) y las operaciones de préstamos hipotecarios del Biess, siendo esta última variable la que presenta una mayor correlación.

Por otra parte, tenemos las variables de desempleo nacional y gasto público las cuales mantienen una relación negativa con la variable dependiente (demanda) con una correlación de -0.24 y -0.13 respectivamente. Lo que indica que el aumento en la tasa de desempleo genera una disminución en las ventas totales del producto SKU 17050, debido a que no hay demanda para la utilización de este producto en el sector de la producción.

Una vez analizada la matriz de correlación la cual generó una idea más clara de la relación existente entre las variables exógenas y la variable endógena, se procede con la generación de los respectivos modelos. Al disponer de más de una variable y para una mejor selección de variables que expliquen la variable de interés, se usará uno de los métodos más comunes y recomendados en la literatura que es el método de eliminación hacia atrás (Backward Stepwise Regression).

Este método consiste en realizar el modelo inicial incluyendo todas las variables exógenas en la ecuación, para su posterior eliminación una a una. En cada una de las etapas del presente método se elimina aquella variable cuyo p-value sea el mayor para la prueba t, hasta lograr obtener el mejor modelo.

A continuación, se presentan los resultados obtenidos en el modelo inicial que incluyen todas las variables macroeconómicas antes descritas.

Tabla 6. Modelo Inicial (Todas las variables)

VENTAS ~ PIBC+ PPT + DN + GP + BIESS + CSP			
Coefficients:			
		p-value	
	PIBC	0.86317	
	PPT	0.05861	
	DN	0.08864	
	GP	0.21839	
	BIESS	0.00968	
	CSP	0.39031	
VENTAS = 0.0001908 PIBC + 306.80 PPT - 528.80 DN - 1.186 GP + 4.546 BIESS + 2.270 CSP - 18890			
Multiple R-squared:	0.3544	Adjusted R-squared:	0.2813

Fuente: Elaboración propia

Donde:

PIBC = Producto Interno Bruto - Construcción

PPT = Índice de Precio de Tuberías

DN = Desempleo Nacional

GP = Gasto Público

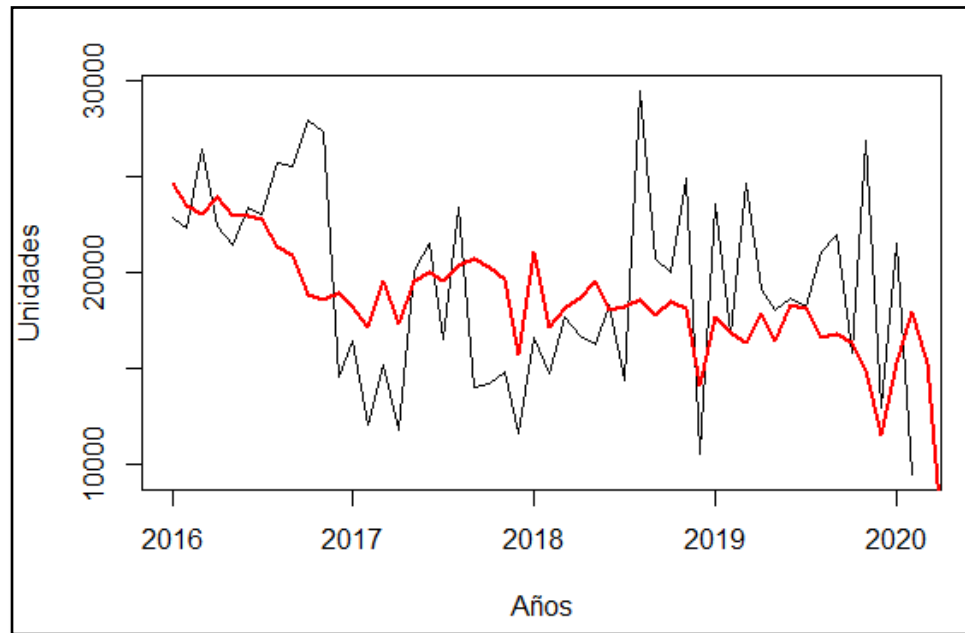
BIESS = Préstamos Hipotecarios del Banco del IESS

CSP = Créditos otorgados en el sector privado

En la tabla 6, se puede observar que el modelo inicial obtenido nos indica que con el aumento de las operaciones por solicitudes de préstamos con el BIESS y de créditos en el sector privado existe un aumento en la demanda y mientras que el aumento del índice de desempleo implicaría la disminución de la demanda en más de 500 unidades.

Este modelo presenta como resultado un R^2 ajustado de 0.2813 lo que indica que no se ajusta completamente a los datos. Se observa que la variable que posee un mayor valor p en la prueba t es el Producto Interno Bruto en Construcción (PIBC) que arrojó un resultado de 0.86317 (Ver Tabla 6). Una vez realizado el ajuste del modelo, se muestra a continuación la representación gráfica en la figura 24:

Figura 24. Ajuste de Modelo Inicial (Todas las Variables)



Fuente: Elaboración propia

De igual manera, se muestran los resultados obtenidos de los ajustes para los datos respectivos de entrenamiento y de prueba:

Tabla 7. Métricas de Regresión (Todas las variables)

VENTAS ~ PIBC + PPT + DN + GP + BIESS + CSP					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	4688.805	3688.419	-9.505	25.132
Test set	-2474.98	4501.513	3432.191	31.104	38.411

Fuente: Elaboración propia

Como se indicó con anterioridad, la variable eliminada fue el Producto Interno Bruto en Construcción (PIBC); por lo cual, se procedió a efectuar el modelo sin considerar esta variable, obteniendo los siguientes resultados:

Tabla 8. Modelo # 2

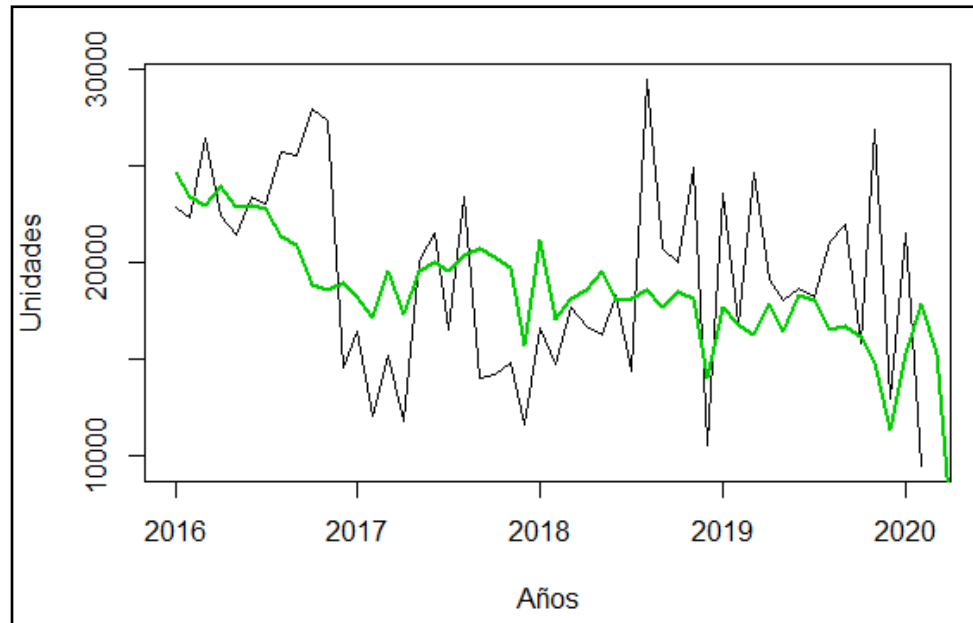
VENTAS ~ PPT + DN + GP + BIESS + CSP			
Coefficients:			
			p-value
PPT			0.05640
DN			0.08539
GP			0.20655
BIESS			0.00409
CSP			0.39262
VENTAS = 302.0 PPT - 529.30 DN - 1.202 GP + 4.658 BIESS + 2.229 CSP -17870			
Multiple R-squared:	0.354	Adjusted R-squared:	0.2942

Fuente: Elaboración propia

En la tabla 8, muestra que los coeficientes del modelo mantienen una tendencia similar al primer modelo. Analizando la variable BIESS se observa que por cada operación adicional de préstamos hipotecarios con el BIESS se puede esperar que la demanda aumente en un promedio de 4.658 unidades y en el caso de la variable DN por cada aumento en el índice de desempleo se espera una disminución en la demanda.

Este modelo presenta como resultado un R^2 ajustado de 0.2942 el cual presenta un ajuste ligeramente mejor que el modelo anterior. Se observa que la variable que posee un mayor valor p en la prueba t es la de créditos en el sector privado (CSP) que arrojó un resultado de 0.3926 (Ver Tabla 8). Realizado el ajuste al modelo, se efectúa la gráfica correspondiente:

Figura 25. Ajuste de Modelo # 2



Fuente: Elaboración propia

A continuación, se muestran los errores del ajuste para los datos de entrenamiento y de prueba:

Tabla 9. Métricas de Regresión Modelo # 2

VENTAS ~ PPT + DN + GP + BIESS + CSP					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	4690.131	3683.262	-9.473	25.059
Test set	-2579.532	4425.979	3429.123	31.550	37.874

Fuente: Elaboración propia

Al realizar la comparación del modelo inicial con el modelo # 2 (Variable PIBC excluida) se muestra que los valores de las métricas de medición del presente modelo presentan mejores resultados que el modelo que incluye todas las variables exógenas con excepción del RMSE. Continuamos con el proceso descrito por lo cual se procedió a efectuar el modelo sin considerar la variable eliminada, obteniendo los siguientes resultados:

Tabla 10. Modelo # 3

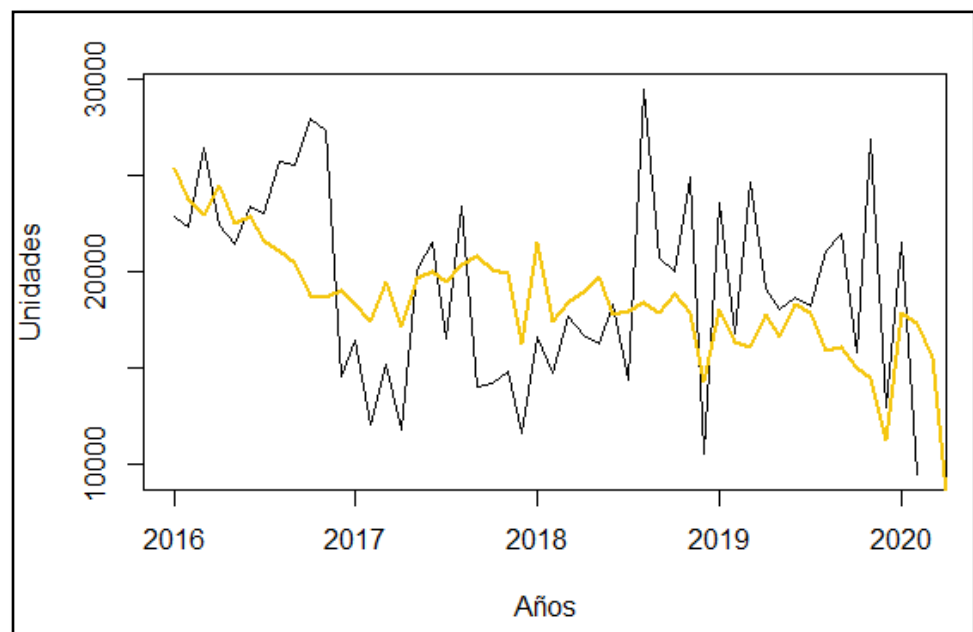
VENTAS ~ PPT + DN + GP + BIESS			
Coefficients:			
		p-value	
PPT		0.0834	
DN		0.1140	
GP		0.2671	
BIESS		5.41e-05	
VENTAS = 255.1 PPT - 472.2 DN - 1.206 GP + 5.454 BIESS - 11020			
Multiple R-squared:		0.3451	Adjusted R-squared: 0.2975

Fuente: Elaboración propia

En la tabla 10, se muestra el modelo obtenido, cuyos coeficientes mantienen una tendencia similar a los modelos anteriores.

El modelo presenta un R^2 ajustado de 0.2975 siendo ligeramente superior al modelo # 2 y al modelo inicial (todas las variables) y muestra que la variable Gasto Público (GP) es la siguiente en ser eliminada al tener un mayor valor p. En la figura 26 se muestra el ajuste del modelo.

Figura 26. Ajuste de Modelo # 3



Fuente: Elaboración propia

En la tabla 11, se muestran los valores de las métricas de ajuste para los datos de entrenamiento y de prueba:

Tabla 11. Métricas de Regresión Modelo # 3

VENTAS ~ PPT + DN + GP + BIESS					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	4722.272	3762.384	-9.745	25.523
Test set	-2592.851	4409.929	3429.746	33.221	38.289

Fuente: Elaboración propia

En la tabla 12 se presentan los resultados obtenidos del modelo # 4 sin considerar la variable GP.

Tabla 12. Modelo # 4

VENTAS ~ PPT + DN + BIESS			
Coefficients:			
		p-value	
PPT		0.0424	
DN		0.1445	
BIESS		5.18e-05	
VENTAS = 292.88 PPT - 432.98 DN + 5.473 BIESS - 18526.03			
Multiple R-squared:		0.3302	Adjusted R-squared: 0.2943

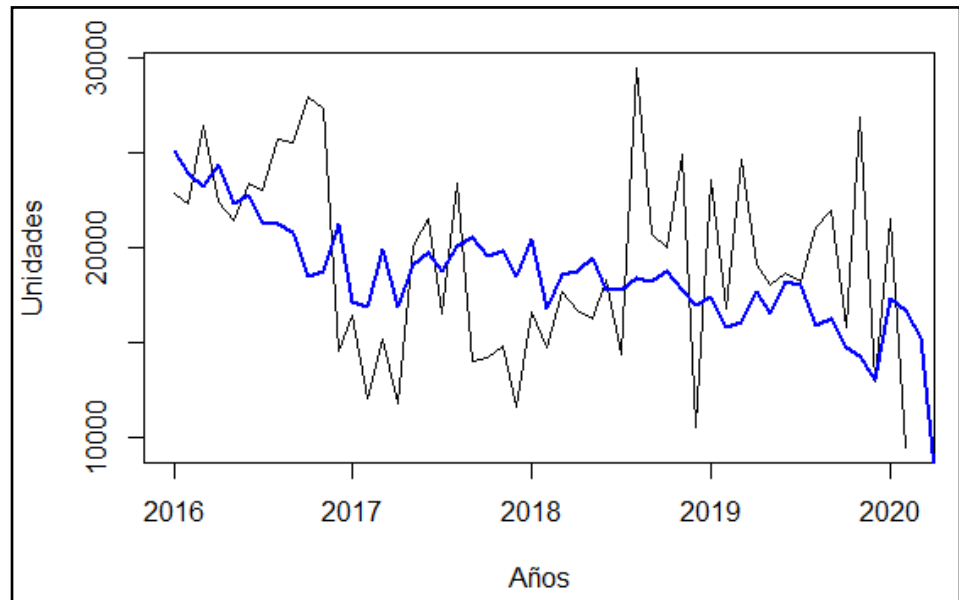
Fuente: Elaboración propia

El presente modelo muestra como variables significativas el índice de precios de tuberías y las operaciones de préstamos hipotecarios con el BIESS. Por cada operación adicional de préstamos hipotecarios con el BIESS se puede esperar que la demanda aumente en un promedio de 5.473 unidades.

El modelo # 4 tiene como resultado un R^2 ajustado de 0.2943, siendo similar al del modelo # 2 el cual presenta un R^2 de 0.2942. Se muestra que la variable con un valor p mayor es DN (Desempleo

Nacional) con un valor de 0.1445. Realizado el ajuste correspondiente, se muestra en la figura 27:

Figura 27. Ajuste de Modelo # 4



Fuente: Elaboración propia

En la tabla 13 se presentan los valores de los errores obtenidos del modelo para los datos de entrenamiento y de prueba

Tabla 13. Métricas de Regresión Modelo # 4

VENTAS ~ PPT + DN + BIESS					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	4775.925	3779.745	-9.869	25.546
Test set	-2141.565	4271.821	3353.256	31.926	36.438

Fuente: Elaboración propia

Realizando la comparación de este modelo con los anteriores, se puede observar que es superado ampliamente por el modelo que incluye todas las variables y los modelos # 2 y # 3, los cuales presentan mejores métricas en relación al RMSE con valores de 4688, 4690 y 4722 respectivamente (Ver Tablas 7, 9 y 11) y en

relación al MAE con valores de 3688, 3683 y 3762 respectivamente (Ver Tablas 7, 9 y 11), siendo inferiores al presente modelo.

Conforme el método de selección empleado, se realizó el siguiente modelo excluyendo la variable menos influyente en el anterior modelo, teniendo los siguientes resultados:

Tabla 14. Modelo # 5

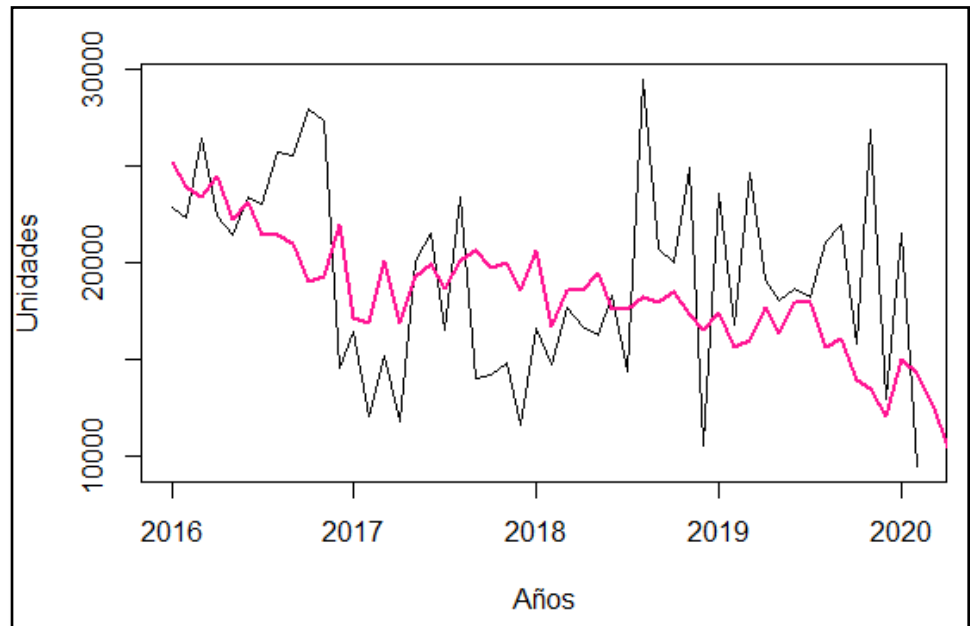
VENTAS ~ PPT + BIESS			
Coefficients:			
		p-value	
	PPT	0.0736	
	BIESS	1.23e-05	
VENTAS = 255.38 PPT + 5.887 BIESS - 17198.8			
Multiple R-squared:	0.304	Adjusted R-squared:	0.2796

Fuente: Elaboración propia

Como se muestra en la tabla 14, este modelo presenta un R^2 ajustado de 0.2796, siendo el menor porcentaje obtenido en relación a los anteriores modelos detallados; es decir, se logra una menor explicación del modelo con las variables exógenas.

La siguiente y última variable a ser eliminada corresponde al índice de precios de tuberías (PPT). En la figura 28 se muestra el ajuste del modelo.

Figura 28. Ajuste de Modelo # 5



Fuente: Elaboración propia

En la tabla 15, se tienen las métricas de errores del modelo con dos variables, para los datos de entrenamiento y de prueba. Las métricas de errores del presente modelo son superadas por los modelos anteriores.

Tabla 15. Métricas de Regresión Modelo # 5

VENTAS ~ PPT + BIESS					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	4868.400	3944.222	-10.702	27.253
Test set	3250.967	4551.638	4032.189	32.171	46.190

Fuente: Elaboración propia

Finalmente, se realiza el modelo con la variable exógena restante representado en la tabla 16 como sigue:

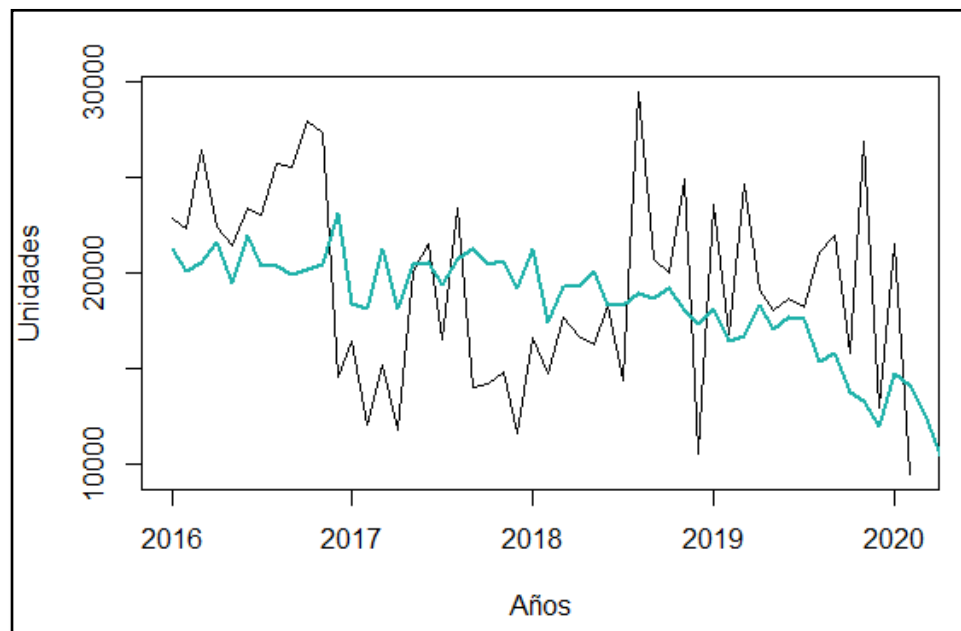
Tabla 16. Modelo # 6

VENTAS ~ BIESS			
VENTAS = 5.687 BIESS - 10284.26			
Multiple R-squared:	0.2634	Adjusted R-squared:	0.2507

Fuente: Elaboración propia

El modelo con una variable presenta un R^2 ajustado de 0.2507, siendo el modelo que menor porcentaje posee entre los seis modelos realizados; lo cual indica ser el que mejor se ajusta para explicar la variable de interés. A continuación, se muestra la figura 29 en donde se puede visualizar el ajuste correspondiente:

Figura 29. Ajuste de Modelo # 6



Fuente: Elaboración propia

El modelo con una variable presenta un R^2 ajustado de 0.2507, siendo el modelo que menor porcentaje posee entre los seis modelos realizados.

Tabla 17. Métricas de Regresión Modelo # 6

VENTAS ~ BIESS					
	ME	RMSE	MAE	MPE	MAPE
Training set	0	5008.221	4095.591	-11.0650	28.229
Test set	-2881.061	4391.781	3772.276	35.859	43.443

Fuente: Elaboración propia

El modelo con una variable exógena muestra valores de métricas de ajuste, los cuales son superados por todos los modelos presentados al tener los valores más altos del ME, RMSE, MAE, MPE, MAPE.

4.2.3. Modelo de Regresión con errores ARIMA

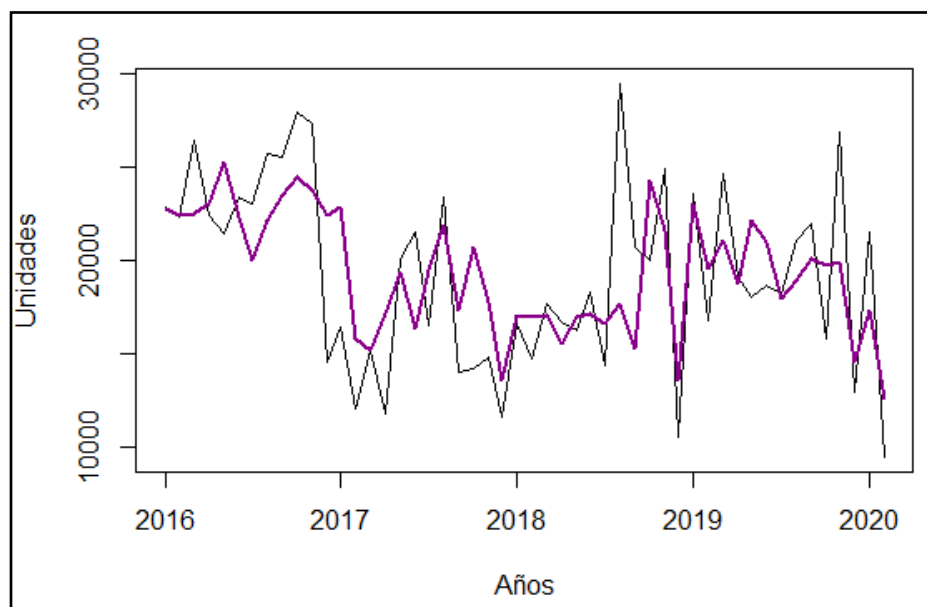
El presente modelo de Regresión con errores ARIMA combina los dos modelos estadísticos presentados con anterioridad, permitiendo mantener dos términos de error (Error del modelo ARIMA y Error del modelo de Regresión).

El modelo ARIMA con el cual se efectuará la regresión con errores será el modelo que fue seleccionado en el presente capítulo; es decir, tomando el modelo ARIMA (0, 1, 3).

A continuación, se presentan los resultados obtenidos de aplicar el modelo de Regresión con errores ARIMA, utilizando el modelo ARIMA (0,1,3):

$$\begin{aligned}
 Ventas = & -0.00019 PIBC - 277.90 PPT - 93.12 DN - 0.85 GP \\
 & + 2.89 BIESS - 0.96 CSP + 0.625a_{t-1} - 0.496a_{t-2} \\
 & + 0.437a_{t-3}
 \end{aligned}$$

Figura 30. Modelo # 1 Regresión con errores ARIMA (0, 1, 3)



Fuente: Elaboración propia

Tabla 18. Métricas de Modelo # 1 Regresión con errores ARIMA

Con todas variables regresoras					
	ME	RMSE	MAE	MPE	MAPE
Training set	491.236	4734.281	3768.741	-6.508	8.229
Test set	2915.013	4988.532	3638.004	40.019	25.433

Fuente: Elaboración propia

Al efectuar una comparación de las métricas de ajustes presentadas en la tabla 18 del modelo de regresión con errores ARIMA con todas las variables regresoras frente al modelo de ARIMA (0, 1, 3) y al modelo de regresión múltiple con todas las variables exógenas se evidencia que las métricas de los dos modelos mencionados muestran mejores resultados tales como en el RMSE donde se registran valores de 4713.74 y 4688.80 respectivamente.

Adicionalmente, se procedió a probar varios modelos con errores ARIMA (0,1, 3) con variaciones de las variables predictoras y así

observar sus valores de AIC. En la tabla que se presenta a continuación, se muestran los resultados obtenidos:

Tabla 19. Modelos de Regresión con errores ARIMA (0,1,3)

VARIABLES	AIC	Training Set	Test Set
		MAE	MAE
PIBC + PPT + DN + GP + BIESS + CSP	1174.44	3768.74	3638.00
PIBC + PPT + DN + GP + BIESS	1171.85	3965.99	3506.98
PIBC + PPT + DN + GP + CSP	1173.13	3785.47	4696.72
PIBC + PPT + DN + BIESS + CSP	1174.48	3841.54	3554.24
PIBC + PPT + GP + BIESS + CSP	1172.71	3916.26	4571.96
PIBC + DN + GP + BIESS + CSP	1172.03	3998.61	4553.23
PPT + DN + GP + BIESS + CSP	1171.12	3821.62	3771.23
PIBC + PPT + DN + GP	1170.18	3964.01	4513.82
PPT + DN + GP + BIESS	1169.98	3833.33	3963.75
DN + GP + BIESS + CSP	1170.34	3736.36	3460.14
PIBC + DN + GP + BIESS	1170.11	3841.91	4483.19
PIBC + PPT + BIESS + CSP	1169.97	3963.74	3940.35
PIBC + PPT + DN + CSP	1170.14	3856.37	3859.10
PIBC + GP + BIESS + CSP	1168.99	3951.20	3478.16
PPT + DN + GP + CSP	1170.47	3674.34	4519.18
PIBC + DN + GP + CSP	1169.22	3961.57	4310.27
PIBC + PPT + DN + BIESS	1169.75	3771.70	3832.17
PIBC + PPT + DN	1165.95	3767.32	3417.00
PPT + DN + GP	1168.17	3818.36	3761.01
DN + GP + BIESS	1167.75	3883.46	3757.96
GP + BIESS + CSP	1166.00	3675.16	4243.37
PIBC + DN + GP	1165.99	3957.10	4602.01
PIBC + GP + BIESS	1168.13	3818.56	3987.42
PIBC + BIESS + CSP	1168.47	3693.39	4336.04
PPT + GP + BIESS	1167.11	3755.23	4115.77
PPT + BIESS + CSP	1167.36	3941.11	4198.50
DN + BIESS + CSP	1166.87	3809.99	4660.91
PIBC + DN + CSP	1166.68	3894.56	4644.32
PIBC + PPT + GP	1168.03	3861.18	4538.85
PIBC + PPT + BIESS	1167.23	3723.24	3517.11
PIBC + PPT + CSP	1167.48	3757.33	4244.04
PPT + DN + CSP	1168.26	3823.10	4018.51
DN + GP + CSP	1167.10	3891.47	4365.49
PIBC + PPT	1166.54	3786.31	3430.24
PIBC + DN	1165.99	3925.24	4490.12
PIBC + GP	1167.42	3691.10	3505.01

PIBC + BIESS	1165.83	3748.49	4673.06
PIBC + CSP	1165.91	3766.39	4410.31
PPT + DN	1165.14	3772.12	3926.88
PPT + GP	1167.09	3670.96	4269.67
PPT + BIESS	1168.22	3770.01	4274.39
PPT + CSP	1167.17	3876.99	4389.26
DN + GP	1165.96	3697.55	3714.10
DN + BIESS	1168.54	3983.28	3532.55
DN + CSP	1165.92	3750.11	4340.44
GP + BIESS	1167.09	3808.44	4550.09
GP + CSP	1166.13	3718.95	4514.13
BIESS + CSP	1166.00	3767.19	4181.77
PIBC	1166.82	3995.76	3729.50
PPT	1166.74	3736.68	4381.62
DN	1167.43	3874.30	4330.87
GP	1166.24	3957.88	4253.93
BIESS	1166.36	4017.35	4993.73
CSP	1167.96	3945.10	4058.17

Fuente: Elaboración propia

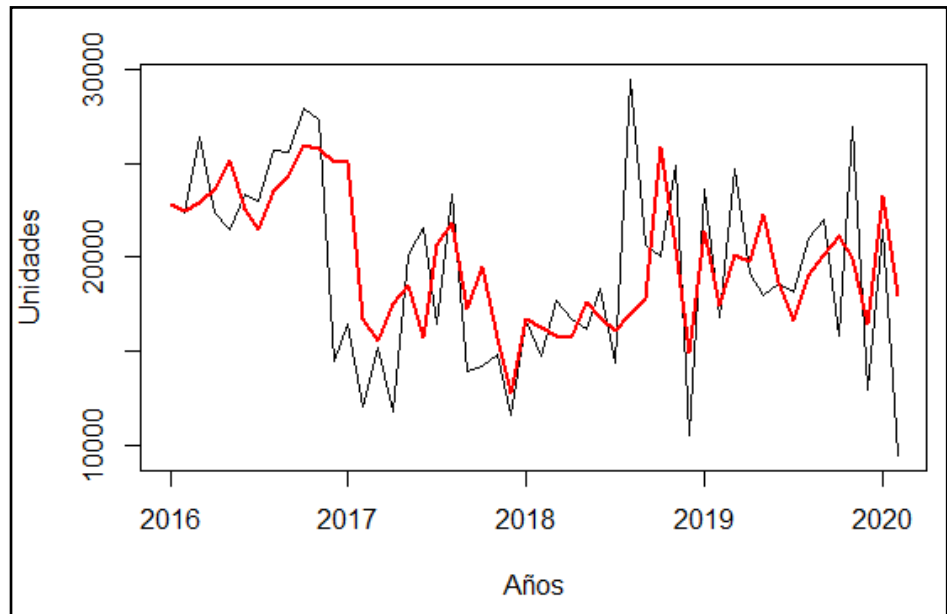
Realizando un análisis de los resultados de los modelos evaluados se puede observar que el modelo que tiene el mejor AIC es el modelo que cuenta con las variables predictoras PPT y DN (1165.14). Otro modelo que presenta un AIC similar al indicado es el modelo con las variables predictoras GP, BIESS y CSP (1166.00) siendo las que mostraron las correlaciones más altas con la variable objeto de estudio.

Por otra parte, en relación a los modelos que presentan los mejores valores del error absoluto medio (MAE) son:

- En los datos de entrenamiento el mejor modelo es el que cuenta con las variables predictoras PPT y GP con 3670.96.
- En los datos de prueba el mejor modelo es el que tiene las variables PIBC, PPT y DN con 3417.00.

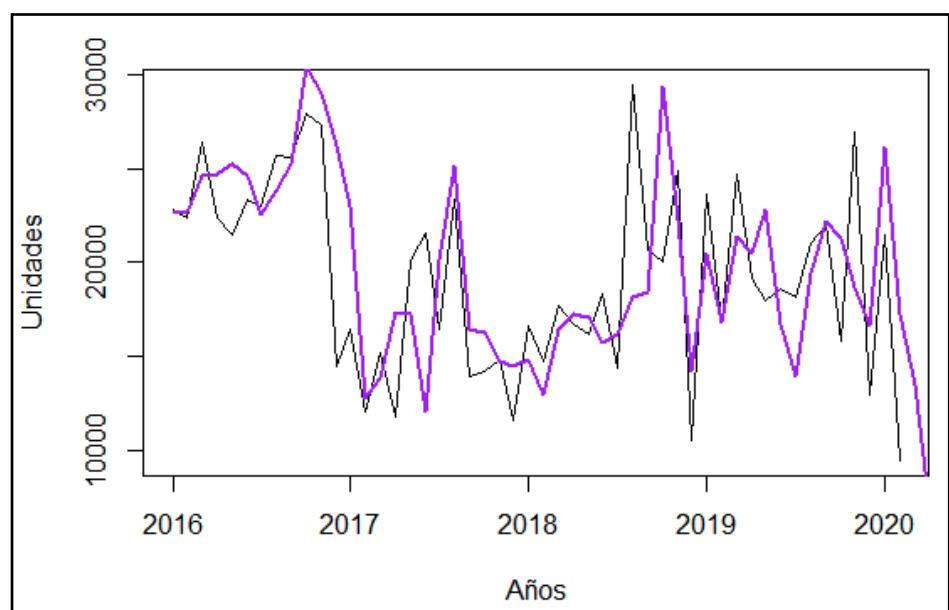
Para los modelos antes descritos se muestra el ajuste de la curva de los mismos en las figuras 31 y 32 respectivamente:

Figura 31. Modelo Regresión con errores ARIMA (PPT, GP)



Fuente: Elaboración propia

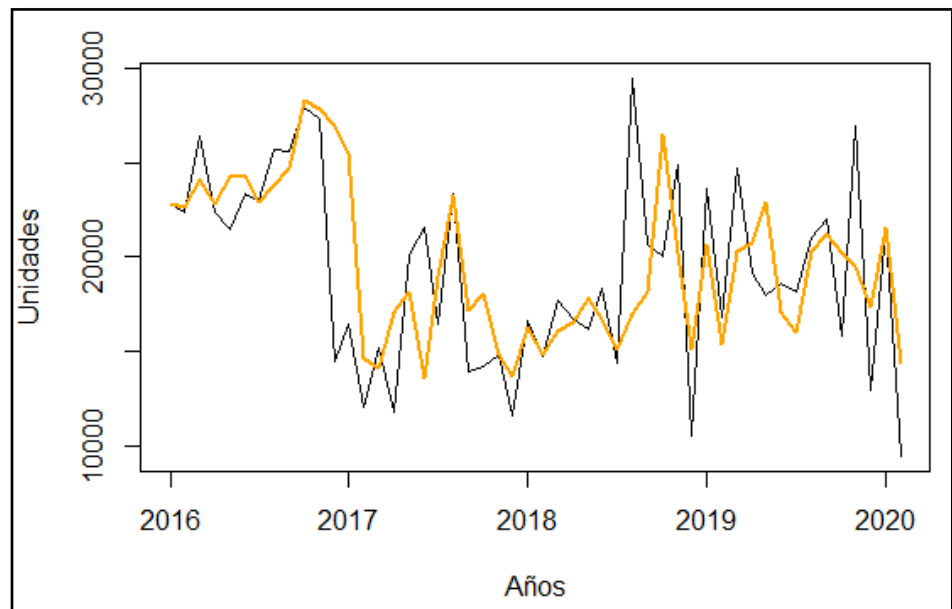
Figura 32. Modelo Regresión con errores ARIMA (PIBC, PPT y DN)



Fuente: Elaboración propia

Uno de los modelos que mantiene un MAE más cercano al modelo ARIMA (0, 1, 3) es el que contiene 2 variables predictoras (PIBC y PPT) con el valor de 3786.31, a continuación, en la figura 33 se muestra el ajuste de la curva del modelo en mención:

Figura 33. Modelo Regresión con errores ARIMA (PIBC y PPT)



Fuente: Elaboración propia

4.3. Selección del Modelo

Una vez que han sido presentados los modelos propuestos en los apartados anteriores para determinar el modelo que mejor se ajuste a la demanda del producto objeto de estudio del presente trabajo de investigación (SKU 17050), se procede con la selección del mejor modelo.

Con la finalidad de realizar esta selección en la tabla 20 se muestra un resumen general de cada uno de los modelos que fueron evaluados con sus respectivas métricas de precisión de errores y los pronósticos correspondientes. Cabe indicar que las métricas bajo las cuales se realizará la selección del modelo son:

- RMSE (Raíz del Error Cuadrático Medio)
- MAE (Error Absoluto Medio)

La tabla muestra las variables exógenas utilizadas en cada uno de dichos modelos de los tres métodos de análisis (ARIMA, Regresión Múltiple y Regresión con Errores ARIMA), donde 1 indica que la variable fue incluida en el modelo y 0 lo contrario.

Empezando por los modelos del método de análisis ARIMA se puede observar que el modelo que muestra mejores resultados en relación al Error Absoluto Medio es el modelo ARIMA (0, 1, 3), el cual fue determinado mediante el auto.arima; sin embargo, en los datos de prueba no fueron los mejores resultados obtenidos.

En el caso de los modelos efectuados con el método de Regresión Múltiple, se tiene que el mejor modelo en relación a la RMSE es aquel que contiene todas las variables exógenas (PIBC, PPT, DN, GP, BIESS y CSP) presentando un valor de 4688.81 y el mejor modelo en relación al MAE es aquel que contiene las variables de índice de precio de tuberías (PPT), desempleo nacional (DN), gasto público (GP), préstamos hipotecarios del Banco del IESS (BIESS) y créditos otorgados en el sector privado (CSP) con el valor mínimo de 3683.26.

Finalmente, con los modelos de Regresión con errores ARIMA, se obtuvo como resultado que el modelo con un mejor RMSE es el modelo que contiene la variable exógena PIBC con un valor de 4511.16, mientras que el mejor modelo considerando el MAE es el modelo que contiene las variables de índice de precio de tuberías (PPT) y gasto público (GP) con un valor de 3670.96; sin embargo, en cuanto a la precisión con los datos de prueba no son aquellos que presenten los mejores resultados. En cuanto a los modelos que mostraron mejores resultados con los datos de prueba en comparación con este método de análisis son los modelos de regresión múltiple, en donde al considerar el MAE son aquellos que generaron los mínimos valores.

En base a lo anteriormente expuesto y revisados los ajustes del comportamiento de los datos, se procederá a seleccionar el modelo de Regresión con errores ARIMA que contiene las variables: producto interno bruto en construcción (PIBC), índice de precio de tuberías (PPT) y el desempleo nacional (DN) cuya capacidad predictiva se ajusta de mejor manera a las características de los datos reales de la demanda del producto 17050, el mismo que presenta un MAE de 3417.00 en los datos de prueba.

Tabla 20. Métricas de Precisión de Errores ARIMA y Pronósticos de Modelos Evaluados

Método de Análisis	Variables Exógenas						Training Set		Test Set	
	PIBC	PPT	DN	GP	BIESS	CSP	RMSE	MAE	RMSE	MAE
ARIMA (1,1,3)	0	0	0	0	0	0	4781.50	3869.32	4419.38	3435.11
ARIMA (0,1,3)	0	0	0	0	0	0	4713.75	3797.55	4430.37	3398.71
Regresión Múltiple	1	1	1	1	1	1	4688.81	3688.42	4501.51	3432.19
	0	1	1	1	1	1	4690.13	3683.26	4425.98	3429.12
	0	1	1	1	1	0	4722.27	3762.38	4409.93	3429.75
	0	1	1	0	1	0	4775.93	3779.75	4271.82	3353.26
	0	1	0	0	1	0	4868.40	3944.22	4551.64	4032.19
	0	0	0	0	1	0	5008.22	4095.59	4391.78	3772.28
Regresión con Errores ARIMA	1	1	1	1	1	1	4734.28	3768.74	4988.53	3638.00
	1	1	1	1	1	0	4743.19	3965.99	4836.48	3506.98
	1	1	1	1	0	1	4728.23	3785.47	4984.34	4696.72
	1	1	1	0	1	1	4834.10	3841.54	4766.22	3554.24
	1	1	0	1	1	1	4705.87	3916.26	4656.56	4571.96
	1	0	1	1	1	1	4714.37	3998.61	4771.46	4553.23
	0	1	1	1	1	1	4780.03	3821.62	4879.00	3771.23
	1	1	1	1	0	0	4743.07	3964.01	4841.82	4513.82
	0	1	1	1	1	0	4912.12	3833.33	4971.26	3963.75
	0	0	1	1	1	1	4699.65	3736.36	4671.13	3460.14
	1	0	1	1	1	0	4895.13	3841.91	4841.84	4483.19
	1	1	0	0	1	1	4704.41	3963.74	4856.34	3940.35
	1	1	1	0	0	1	4996.45	3856.37	4341.23	3859.10
	1	0	0	1	1	1	4687.77	3951.20	4693.89	3478.16

0	1	1	1	0	1	4729.74	3674.34	4508.40	4519.18
1	0	1	1	0	1	4676.45	3961.57	4518.63	4310.27
1	1	1	0	1	0	4719.65	3771.70	4366.20	3832.17
1	1	1	0	0	0	4853.56	3767.32	4482.72	3417.00
0	1	1	1	0	0	4792.00	3818.36	4315.18	3761.01
0	0	1	1	1	0	4927.23	3883.46	4475.28	3757.96
0	0	0	1	1	1	4822.65	3675.16	4396.45	4243.37
1	0	1	1	0	0	4791.27	3957.10	4945.23	4602.01
1	0	0	1	1	0	4943.19	3818.56	4719.73	3987.42
1	0	0	0	1	1	4964.46	3693.39	4807.90	4336.04
0	1	0	1	1	0	4801.20	3755.23	4927.00	4115.77
0	1	0	0	1	1	4697.49	3941.11	4552.15	4198.50
0	0	1	0	1	1	4670.96	3809.99	4621.34	4660.91
1	0	1	0	0	1	4892.75	3894.56	4799.21	4644.32
1	1	0	1	0	0	4933.32	3861.18	4423.15	4538.85
1	1	0	0	1	0	4988.09	3723.24	4908.05	3517.11
1	1	0	0	0	1	4837.35	3757.33	4642.91	4244.04
0	1	1	0	0	1	4959.66	3823.10	4493.44	4018.51
0	0	1	1	0	1	4755.04	3891.47	4494.18	4365.49
1	1	0	0	0	0	4762.45	3786.31	4305.93	3430.24
1	0	1	0	0	0	4682.45	3925.24	4585.49	4490.12
1	0	0	1	0	0	4673.83	3691.10	4531.98	3505.01
1	0	0	0	1	0	4784.37	3748.49	4686.23	4673.06
1	0	0	0	0	1	4821.23	3766.39	4582.50	4410.31
0	1	1	0	0	0	4971.19	3772.12	4424.44	3926.88
0	1	0	1	0	0	4996.18	3670.96	4395.17	4269.67
0	1	0	0	1	0	4801.95	3770.01	4556.21	4274.39

0	1	0	0	0	1	4678.04	3876.99	4655.31	4389.26
0	0	1	1	0	0	4811.87	3697.55	4473.36	3714.10
0	0	1	0	1	0	4769.31	3983.28	4831.03	3532.55
0	0	1	0	0	1	4734.01	3750.11	4812.37	4340.44
0	0	0	1	1	0	4907.39	3808.44	4396.10	4550.09
0	0	0	1	0	1	4685.88	3718.95	4840.65	4514.13
0	0	0	0	1	1	4889.16	3767.19	4865.30	4181.77
1	0	0	0	0	0	4511.16	3995.76	4435.37	3729.50
0	1	0	0	0	0	4755.26	3736.68	4468.88	4381.62
0	0	1	0	0	0	4832.34	3874.30	4526.54	4330.87
0	0	0	1	0	0	4780.85	3957.88	4807.45	4253.93
0	0	0	0	1	0	4950.46	4017.35	4586.35	4993.73
0	0	0	0	0	1	4845.04	3945.10	4779.78	4058.17

Fuente: Elaboración propia

4.4. Validación del Modelo

Según Hyndnan & Athanasopoulos (2018) luego de la selección de variables y ajuste del modelo es importante efectuar la validación del modelo mediante la verificación del cumplimiento de diferentes características; por lo cual, a continuación, se procede con el análisis de las mismas del modelo de Regresión con errores ARIMA con las variables producto interno bruto en construcción (PIBC), índice de precio de tuberías (PPT) y el desempleo nacional (DN) seleccionado en el capítulo anterior.

$$Y_t = -0.00022 PIBC - 0.7039 PPT - 0.5271 DN + 0.6652 a_{t-1} \\ - 0.6450 a_{t-2} + 0.4698 a_{t-3}$$

Las características a validar son:

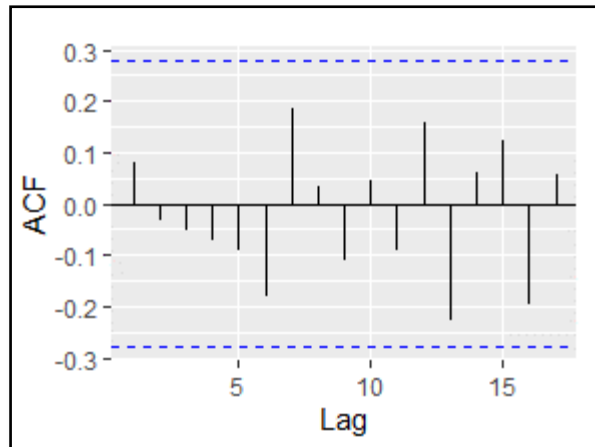
- Autocorrelación de los residuos
- Distribución normal de los residuos
- Gráfico de residuos vs ajustes

4.4.1. Autocorrelación de los residuos

Al ajustar un modelo de regresión es frecuente la existencia de autocorrelación de los residuos, es por ello que una prueba útil es efectuar el gráfico de autocorrelación (ACF).

Como se puede visualizar en la figura 34, las autocorrelaciones se encuentran dentro de los límites, lo que indican que existe ruido blanco y por ende mantiene una media igual a cero y una varianza constante.

Figura 34. Gráfica de ACF de los residuos

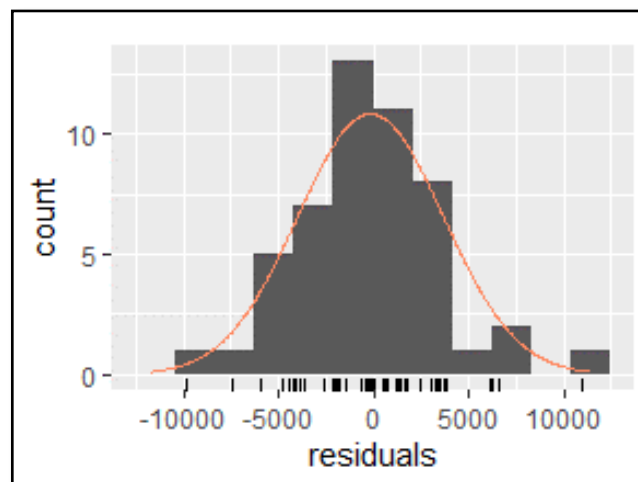


Fuente: Elaboración propia

4.4.2. Distribución normal de los residuos

Se debe comprobar que los residuos mantienen una distribución normal, lo cual se puede observar mediante el histograma de los mismos.

Figura 35. Histograma de los residuos



Fuente: Elaboración propia

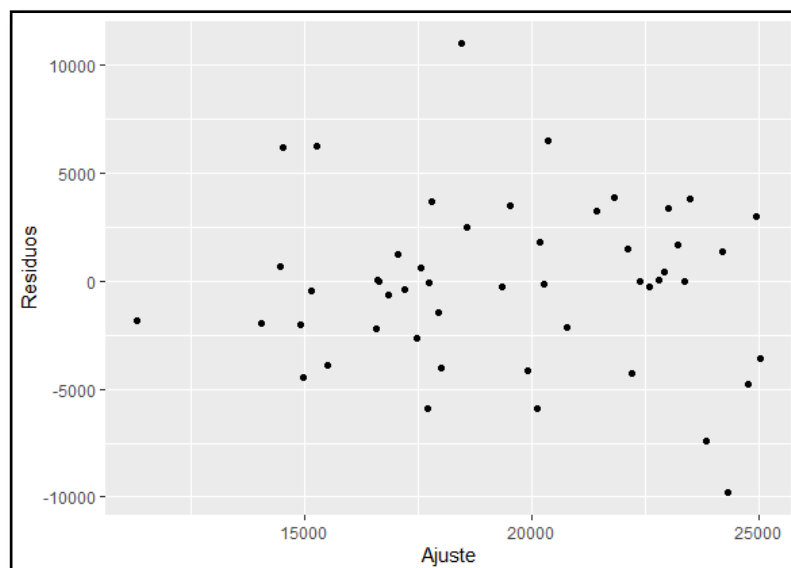
Como se aprecia en la figura 35 los residuos mantienen una distribución normal, no muestran un aparente sesgo, por tanto, no se ven afectados los intervalos de predicción.

Para efectos de verificar lo indicado, se realizó la prueba de normalidad mediante el test de Shapiro-Wilk en el cual se obtuvo un p-value de 0.6905; por tanto, no se rechaza la hipótesis nula, concluyendo que los residuos tienen una distribución normal.

4.4.3. Gráfico de residuos vs ajustes

Otra característica importante a considerar es que los residuos mantengan una distribución aleatoria; es decir, que no presenten patrón alguno y para ello se lo realiza mediante la representación del diagrama de dispersión de los residuos vs los valores del ajuste, lo que permitirá además determinar si existe heterocedasticidad; es decir, que la varianza de los residuos no es constante.

Figura 36. Gráfico de dispersión Residuos vs Ajuste



Fuente: Elaboración propia

Se puede apreciar en la figura 36 que no existe ningún patrón que permita pensar la existencia de heterocedasticidad.

4.4.4. Predicción del Modelo seleccionado

Una vez que ha sido efectuada la validación del modelo para el cumplimiento de los supuestos y poder realizar la predicción, se procede con el análisis de la capacidad de predecir del modelo, en la cual arrojaron los resultados mostrados en la tabla 21, correspondientes a los pronósticos de los meses de marzo a diciembre del 2020 (datos de prueba).

Tabla 21. Resultados del Pronóstico del Modelo

Mes	Demanda Real	Pronóstico	Límite Inferior	Límite Superior
Marzo	5060	8576	6985	17608
Abril	6530	9428	7725	18780
Mayo	4360	8171	6576	17134
Junio	11015	10983	8846	18947
Julio	10593	11550	9781	20239
Agosto	10540	12120	10015	21181
Septiembre	11070	10986	8971	19345
Octubre	13000	14196	12059	22568
Noviembre	19438	19134	17435	23906
Diciembre	12830	12497	10236	21822

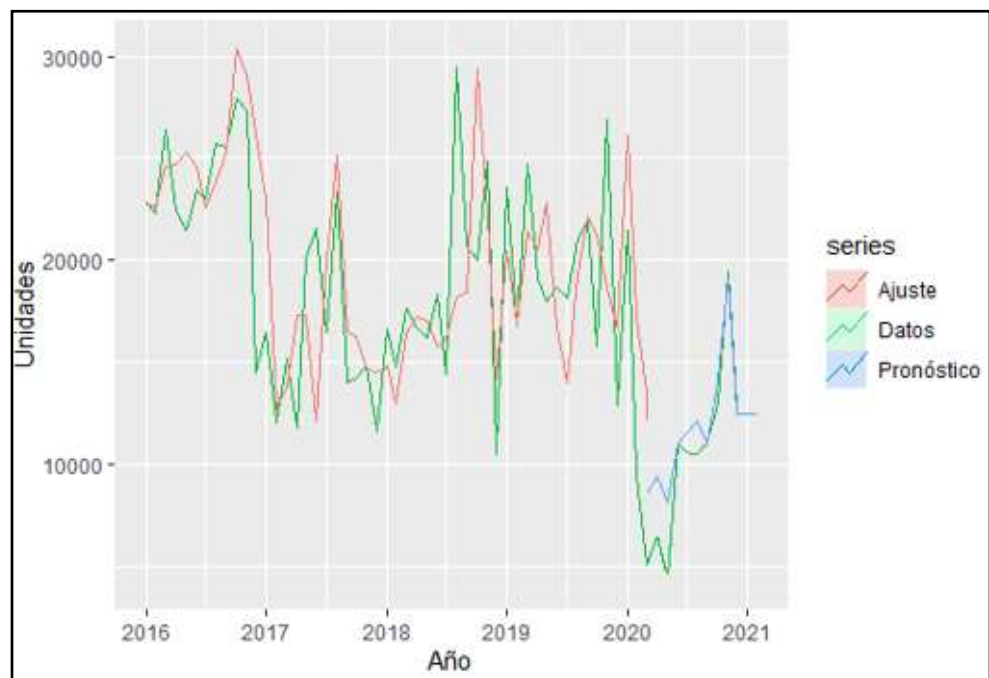
Fuente: Elaboración propia

Se puede apreciar que la demanda real del producto en los primeros meses (marzo, abril y mayo) no se encuentran dentro de los límites inferiores de la predicción del modelo cuyo intervalo es al 95% de confianza, esto debido al deterioro que sufrió la economía en el Ecuador a causa de la pandemia por el COVID-19 lo que llevó a meses de confinamiento ocasionando la suspensión temporal de las actividades comerciales, registrando en dichos meses la caída drástica de muchos indicadores macroeconómicos, cifras que fueron reportadas por el Banco Central del Ecuador.

Del total de períodos utilizados como prueba, se observa que el 40% de los pronósticos fueron subestimados por el modelo; mientras que, el 60% restante de los pronósticos se encuentran sobreestimados. Ninguno de los meses sobrepasa de los límites superiores pronosticados al 95% de confianza.

En los meses de junio a diciembre se registró el aumento de la demanda real del producto gracias a la reactivación de las distintas actividades comerciales, presentando una leve mejora en la economía del país, con lo cual el modelo seleccionado logra recoger el comportamiento de los datos reales, lo cual se aprecia en la figura 37.

Figura 37. Pronóstico de la demanda del modelo seleccionado



Fuente: Elaboración propia

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- La previsión de la demanda es una estrategia de suma importancia para las organizaciones dado que contribuyen a la eliminación de la incertidumbre y permite una mejor toma de decisiones. Siendo el objetivo principal del presente trabajo de investigación la evaluación de modelos estadísticos para la predicción de la demanda de productos de una empresa comercializadora de ítems de ferretería mediante series de tiempo, se puede concluir que dicho objetivo fue cumplido en su totalidad, en donde mediante la ayuda del software RStudio se modelaron y evaluaron tres modelos de series de tiempo (Modelo ARIMA, Modelo de Regresión Múltiple y Modelo de Regresión con errores ARIMA), utilizando datos de entrenamiento para entrenar los modelos, determinar el que mejor se ajustara a los datos históricos y que presentara la mejor predicción (Demanda real vs Pronóstico).
- En el modelo ARIMA (Uno de los enfoques más usados para los pronósticos de series de tiempo) se obtuvieron dos modelos: manual y automático (auto.arima); siendo este último el que presentó el mejor modelo con este enfoque. En el caso de los modelos de Regresión Múltiple se ajustaron varios modelos lineales entre la variable de interés y las seis variables exógenas (PIBC, PPT, DN, GP, BIESS, CSP) usando el método de eliminación hacia atrás. Con los modelos antes descritos se logró cumplir el objetivo específico de la estimación de modelos estadísticos para el pronóstico de la demanda; y es que actualmente existe diversidad de modelos para explicar y predecir el comportamiento de los datos, por lo cual no se puede aseverar con

libertad que determinado modelo o enfoque sea el apropiado para cada caso; sin antes realizar el análisis correspondiente y validar que el modelo seleccionado proporcione los mejores resultados.

- Se realizó un análisis comparativo de los modelos estadísticos obtenidos para seleccionar el mejor modelo que se ajuste al comportamiento natural de los datos y con mayor precisión en la predicción, cumpliendo de esa manera con otro de los objetivos específicos del presente trabajo. Las medidas de error utilizadas para la comparación fueron: la raíz del error cuadrático de la media (RMSE) y el error cuadrático de la media (MAE). Al analizar las medidas del error en los datos de entrenamiento se obtuvo que el modelo de Regresión con errores ARIMA utilizando tres variables (DN, BIESS, CSP) presentó el mejor resultado del RMSE y en relación al MAE el modelo de Regresión con errores ARIMA con dos variables (PPT, GP) superó a los otros modelos. En los datos de prueba el modelo de Regresión Múltiple con tres variables (PPT, DN, BIESS) obtuvo el mejor RMSE, y respecto al MAE el modelo de Regresión con errores ARIMA con tres variables (PIBC, PPT, DN) registró el mejor MAE. El modelo seleccionado fue el de Regresión con Errores ARIMA con tres variables (PIBC, PPT, DN), el cual presentó el mejor ajuste en los datos de prueba y cuyo pronóstico tuvo mayor efectividad; a pesar de no ser el que obtuvo el menor de los errores.
- Se logró validar la efectividad de los pronósticos de la demanda del producto estudiado con el modelo estadístico escogido cumpliendo el último de los objetivos específicos. Los resultados obtenidos con el pronóstico del modelo evidencian una proyección próxima a la demanda real; presentando hasta un 9% de diferencias con los datos reales que corresponden a los meses entre junio a diciembre, exceptuando el mes de agosto en que la diferencia es mayor. Los pronósticos permitirán a la organización facilitar el entendimiento del comportamiento de las ventas para de esta manera poder generar

alternativas que ayuden a la organización a maximizar sus ingresos y una mejor toma de decisiones.

5.2. Recomendaciones

- Mantener un seguimiento en el tiempo de la evolución de la demanda del producto proporcionará un mayor control de los registros en el aumento o disminución de la misma y permitirá a la organización obtener estimaciones más idóneas.
- Es necesario que todo el personal de la organización que interviene en la cadena del proceso, tenga el pleno conocimiento de la demanda, así como la estrategia a llevar a cabo para el logro de sus objetivos.
- Implementar el modelado del pronóstico de la demanda a los demás productos de ferretería que son comercializados por la compañía, lo cual permitirá una mejor planificación en las compras, importaciones, niveles de inventarios, etc.
- Utilizar el modelo propuesto como un instrumento agregado a los utilizados por la organización para que los pronósticos obtenidos logren ser calculados con mayor precisión.
- Es recomendable efectuar evaluaciones periódicas y mantener una actualización del modelo mediante nuevas herramientas o aplicaciones estadísticas y así evitar trabajar con metodologías obsoletas; además, de la constante actualización de la información de la demanda que permitirá contar con una data completa de los datos históricos para las futuras predicciones.
- En la literatura existe un sin número de técnicas que permiten obtener diversidad de modelos para la predicción de la demanda, tales como, el utilizado en el presente trabajo de investigación, el cual corresponde a un modelo clásico, sin embargo, hay otras alternativas para la predicción y que no son ampliamente utilizadas como son las

Redes Neuronales las cuales se fundamentan en técnicas de optimización para la predicción de la demanda y mantener así niveles óptimos de inventarios.

6. Bibliografía

- Alarcón Viracacha, J., & Mancipe Romero, A. (2016). *Diseño del Método de pronóstico de la demanda para la empresa de importaciones y exportaciones Mundo Industrial Ltda.* Colombia.
- Arias Vargas, M. (2017). Impacto en el inventario de seguridad por la utilización de la desviación estándar de los correos de pronósticos. *Tecnología en Marcha*, 49-54.
- Asencio Cristóbal, L., González Ascencio, E., & Lozano Robles, M. (2017). El inventario como determinante en la rentabilidad de las distribuidoras farmacéuticas. *RETOS - Revista de Ciencias de la Administración y Economía*.
- Banco Central del Ecuador (BCE). (2018). *Información Estadística Mensual Marzo-Abril 2018*. Obtenido de <https://contenido.bce.fin.ec/home1/estadisticas/bolmensual/IEMensual.jsp>
- Boada, A. J. (2017). Modelo lineal dinámico bayesiano como procedimiento de actualización automática para modelos estadísticos predictivos. *Dimensión Empresarial*, 33-52.
- Boada, A. J. (2017). Sistema de proyección de la demanda. Caso práctico de predicción automatizada en empresas de venta por catálogo. *Revista Perspectiva Empresarial*, 23-41.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Switzerland: Springer International.
- Cadena Lozano, J. B. (2016). *Gestión del Pronóstico Estratégico*. Bogotá: CESA - Colegio de Estudios Superiores de Administración.
- Cadena Lozano, J. B., Ariza Garzón, M. J., & Palomo Zurdo, R. J. (2017). La gestión de pronósticos en las decisiones empresariales: un análisis empírico. *Espacios*, 39(13), 1-16.
- Calderón Ardila, F., & Supelano Aguirre, L. (2012). *Diseño del Protocolo de Pronóstico de demanda basado en modelos teóricos para el sistema de información en empresa Comertex S.A.* Bucaramanga.
- Contreras Juárez, A., Atziry Zuñiga, C., Martínez Flores, J. L., & Sánchez Partida, D. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *ScienceDirect*, 387-396.
- Ekos. (2018). Estudio de productos que toda ferretería debe tener. *Mundo Constructor*. Obtenido de <https://www.mundoconstructor.com.ec/estudio-de-productos-que-toda-ferreteria-debe-tener/>
- Ekos. (Noviembre de 2020). *Mundo Constructor*. Obtenido de <https://www.mundoconstructor.com.ec/ranking-de-la-construccion-el-sector-busca-su-reactivacion/>
- Escalante, G. (Agosto de 2019). *Instituto Tecnológico Itmd*. Obtenido de <https://www.itmasterd.es/blog/informatica/big-data-aprendizaje-estadistico>
- Escobar Sailema, J. C. (2019). Modelo de estimación estadística «Programa Inclusión Productiva» MIPRO-Ecuador. *RETOS - Revista de Ciencias de la Administración y Economía*.
- Espino Timón, C. (2017). *Análisis predictivo: técnicas y modelos utilizados* y. España.
- Forero Gómez, G., & Martínez Lozano, J. (2020). *Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa Boleco S.A., en la ciudad de Bogotá DC*. Bogotá DC.

- Gallo Méndez, G. (2020). *RiuNet: Repositorio Institucional UVP*. Obtenido de <https://riunet.upv.es/bitstream/handle/10251/139782/Gallo%20-%20An%c3%a1lisis%20de%20la%20previsi%c3%b3n%20de%20la%20demanda%20en%20la%20empresa%20Roquette%20Laisa%20Espa%c3%b1a%20S.A..pdf?sequence=1&isAllowed=y>
- García Osorio, N. E., & Tobar Cazares, X. (2019). La construcción en el Producto Interno Bruto del Ecuador. *Podium*, 57-68. doi:<http://dx.doi.org/10.31095/podium.2019.35>.
- Gil-Vera, V. (2017). Pronóstico de la demanda mensual de electricidad con series de tiempo. *Revista EIA*, 111–120. doi:<https://doi.org/10.24050/reia.v13i26.749>
- González Casimiro, M. P. (2009). *Universidad del País Vasco*. Obtenido de Archivo digital docencia investigación: <https://addi.ehu.es/bitstream/handle/10810/12492/04-09gon.pdf;jsessionid=4220E9FC94F07A7AC00DDCFF8138C6F9?sequence=1>
- Gonzalez, A. (2013). *Modelación y pronósticos de demanda en un simulador de negocios*. Obtenido de Modelación y pronósticos de demanda en un simulador de negocios.: <http://tesis.pucp.edu.pe/repositorio/handle/123456789/4816>
- González, A. (2020). Un modelo de gestión de inventarios basado en estrategia competitiva. *Ingeniare - Revista Chilena de Ingeniería*.
- Grupo Faro. (4 de Septiembre de 2020). *Grupo Faro, ideas y acción colectiva*. Obtenido de <https://grupofaro.org/sector-de-la-construccion-en-el-ecuador-en-epoca-de-covid-19/>
- Guazzelli, A. (17 de 12 de 2012). Obtenido de <https://developer.ibm.com/es/technologies/predictive-analytics/articles/ba-predictive-analytics2/>
- Hanke, J. E., & Reitsh, A. G. (1996). *Pronósticos en los negocios*. Mexico: Maryland Composition.
- Hyndnan, R. J., & Athanasopoulos, G. (2018). *Forecasting Pinciples and Practice* (Second ed.). Melbourne, Australia: OTexts. Obtenido de <https://otexts.com/fpp2>
- Instituto Nacional de Estadística y Censos. (Enero de 2021). *Ecuador en cifras - INEC*. Obtenido de <https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2021/Enero-2021/Boletin%20tecnico%20de%20empleo%20ene21.pdf>
- Instituto Nacional de Estadísticas y Censos. (2020). *Índices de Precios de Materiales, Equipo y Maquinaria de la Construcción (IPCO)*. Obtenido de <https://www.ecuadorencifras.gob.ec/indice-de-precios-de-la-construccion-ipco-2/>
- Lao León, Y., Rivas Méndez, A., Pérez Pravia, M., & Marrero Delgado, F. (2017). Procedimiento para el pronóstico de la demanda mediante redes neuronales artificiales. *Ciencias Holguín*, 1-18.
- Lucero, K. (2020). La construcción, un pilar de la economía debilitado por la pandemia. *Gestión Digital*.
- Marín Diazaraque, J. M. (2014). *Portal Universidad Carlos III de Madrid*. Obtenido de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/EDescrip/tema7.pdf>
- Mogrovejo Del Salto, A. (Junio de 2018). *UPCommons. Portal de acceso abierto al conocimiento de la UPC*. Obtenido de <https://upcommons.upc.edu/handle/2117/118369>
- Montemayor Gallegos, J. E. (2013). *Métodos de pronósticos para negocios*. Mexico: Editorial Digital Tecnológico de Monterrey.
- Montgomery Douglas C. (2007). *Introduction to statistical quality control*. New York: 6th ed., vol. 7 John Wiley & Sons.

- Morales Tabares, Z. E., Cabrera Campos, A., Vázquez Silva, E., & Caballero Mota, Y. (2016). MPREDSTOCK : Modelo multivariado de predicción del stock de piezas de repuesto para equipos médicos. *Revista Cubana de Ciencias Informáticas*, 88-104.
- Morales, J. (2018). *Modelos Estadísticos. Grado Biotecnología*. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/354914_503ddf0d5ca5490e9a2f3bccaa8d9637.html
- Pavolini, G. (24 de Julio de 2019). *Pavolini.net*. Obtenido de <https://pavolini.net/2019/07/24/estimando-errores-de-pron%C3%B3stico/>
- Quinde Espinoza, C., & Ramos Alvarado, T. (2018). Valuación y control del inventario y su efecto en la rentabilidad. *Revista Observatorio de la Economía Latinoamericana*.
- Sánchez Córdova, O. (2018). *Construcción de un modelo estadístico para el pronóstico de la demanda de alimento balanceado para camarones en una empresa comercializadora de la ciudad de Guayaquil*. Guayaquil.
- Sánchez, R., & Gavira Durón, N. (2016). Pronóstico de la demanda de visitantes internacionales:. *Journal of Research in Accounting and Management Science*, 29-54.
- Selberherr, J. (2015). Sustainable life cycle offers through cooperation. *Smart and Sustainable Built Environment*. 4-24.
- SENPLADES. (2017). *Secretaría Nacional de Planificación y Desarrollo*. Obtenido de Plan Nacional para el Buen Vivir: <https://www.gobiernoelectronico.gob.ec/wp-content/uploads/downloads/2017/09/Plan-Nacional-para-el-Buen-Vivir-2017-2021.pdf>
- Wilkinson, S., Yan Chang, R. A., & Sapeciay, Z. (2016). Improving construction sector resilience. *International journal of Disaster Resilience in the Built Environment. Emerald Insight*, 173-185.
- Yandún, H. (11 de Agosto de 2020). *El Telégrafo*. Obtenido de <https://www.eltelegrafo.com.ec/noticias/sociedad/6/construimos-ecuador-casa-para-todos#:~:text=La%20construcci%C3%B3n%20representa%20el%2010,actividades%20conexas%20constituye%20el%2020%25.&text=Seg%C3%BAn%20el%20INEC%2C%20en%20septiembre,mano%20de%20obra>

7. Apéndices y anexos

Códigos del Software RStudio

```
# Número de diferencias para serie estacionaria
ndiffs(datossttrain)

# Número de diferencias estacionales
nsdiffs(datossttrain)

# Cálculo de la primera diferencia para datos estacionarios
datosdiff <- diff(datossttrain)
cbind("Unidades"=datossttrain, "Log Diff Unidades"=diff(datossttrain))
%>% autoplot(facets=TRUE) + xlab("Año")

# Prueba de KPSS serie con primera diferencia datos estacionarios
ur.kpss(datosdiff)
summary(ur.kpss(datosdiff))
```

```
# Número de diferencias para serie estacionaria
ndiffs(datossttrain)

# Número de diferencias estacionales
nsdiffs(datossttrain)

# Cálculo de la primera diferencia para datos estacionarios
datosdiff <- diff(datossttrain)
cbind("Unidades"=datossttrain, "Log Diff Unidades"=diff(datossttrain))
%>% autoplot(facets=TRUE) + xlab("Año")

# Prueba de KPSS serie con primera diferencia datos estacionarios
ur.kpss(datosdiff)
summary(ur.kpss(datosdiff))
```

```
# Función de Autocorrelación ACF
acf(diff(datossttrain))
ggAcf(diff(datossttrain))

# Función de Autocorrelación Parcial PACF
pacf(diff(datossttrain))
ggPacf(diff(datossttrain))

# Funciones ACF y PACF
ggtsdisplay(diff(datosst), main="Autocorrelación")
```

```

# Transformación logarítmica
logdatostra <- log(datossttrain)

# Transformación logarítmica vs datos originales
cbind("Unidades"=datossttrain, "Log Unidades"=log(datossttrain))
%>% autoplot(facets=TRUE) + xlab("Año") + ggtitle("Demanda del sku 17050")

# Prueba de KPSS serie transformación logarítmica
ur.kpss(logdatostra)
summary(ur.kpss(logdatostra))
kpss.test(logdatostra)

```

```

# Modelo ARIMA

# Modelo #1 Arima manual
modeloar1 <- Arima(datostrain, order = c(1,1,3))
modeloar1
tsdiag(modeloar1)
summary(modeloar1)

# Ljung-Box
Box.test(residuals(modeloar1), type = "Ljung-Box")
error <- residuals(modeloar1)
plot(error)
checkresiduals(modeloar1)

```

```

# Ajuste de curva modelo ARIMA (1,1,3)
forecast1 <- fitted(modeloar1)
plot(datostrain, xlab="Años",ylab="Unidades",type="l",ylim=c(2000,30000))
lines(forecast1,lwd=2,col="mediumvioletred")

# Modelo #2 Arima automático
autoarima <- auto.arima(datossttrain, trace = TRUE, seasonal= FALSE)
summary(autoarima)

```

```

modeloar2 <- Arima(datossttrain, order = c(0,1,3))
modeloar2
summary(modeloar2)
checkresiduals(modeloar2)

# Gráficos residuales del modelo Arima elegido
checkresiduals(modeloar2)

# Ajuste de curva modelo ARIMA (0,1,3)
forecast2 <- fitted(modeloar2)
plot(datossttrain, xlab="Años",ylab="Unidades",type="l",ylim=c(2000,30000))
lines(forecast2,lwd=2,col="blue")

```

```

# Modelo de Regresión Múltiple

variables_Macro <- read_excel("E:/Maestría Estadística/Unidad de
                             Titulación/Tesis Maestría/Tesis Proyección
                             de Demanda/Datos/Variables Macro.xlsx")

view(variables_Macro)
names(variables_Macro)
class(variables_Macro)

datosrmst <- ts(variables_Macro,start = c(2016,1),frequency = 12)

datarmtrain <- ts(variables_Macro[1:50,],start = c(2016,1),
                  end = c(2020,2),frequency = 12)
view(datarmtrain)

datarmtest <- ts(variables_Macro[51:60,],start = c(2020,3),
                  end = c(2020,12),frequency = 12)
view(datarmtest)

```

```

# Matriz de Correlación
varmacro_corr <- cor(variables_Macro, method = "pearson")
round(varmacro_corr, digits = 2)
names(data_num)

# Modelo 1
modelormst1 <- tslm(VENTAS~PIBC+PPT+DN+GP+BIESS+CSP, data = datarmtrain)
summary(modelormst1)

# Métricas de ajuste
data.frame(R2=rsquare(modelormst1,data=datarmtrain),
           RMSE=rmse(modelormst1,data=datarmtrain),
           MAE=mae(modelormst1,data=datarmtrain))

# Ajuste
ajustem1 <- fitted(modelormst1)
plot(datarmtrain[, 'VENTAS'],xlab="Años",ylab="Unidades",type="l")
lines(ajustem1,lwd=2,col="red")

```

```

# Pronóstico
pronom1 <- forecast(modelormst1,data.frame(datarmtest[,2:7],level=c(95)))
autoplot(pronom1)

round(accuracy(pronom1,data.frame(datarmtest),digits = 2))
accuracy(pronom1,data.frame(datarmtest))

# Gráfico de Datos, Ajuste y Pronóstico
autoplot(datosrmst[, 'VENTAS'],series="Datos")+
  autolayer(ajustem1,series="Ajuste")+
  autolayer(pronom1,series="Pronóstico")+
  xlab("Año")+ylab("Unidades")+ggtitle("Datos, Ajuste y Pronóstico")

```

```

# Modelo 2
modelormst2 <- tslm(VENTAS~PPT+DN+GP+BIESS+CSP, data = datarmtrain)
summary(modelormst2)

# Metricas de ajuste
data.frame(R2=rsquare(modelormst2,data=datarmtrain),
           RMSE=rmse(modelormst2,data=datarmtrain),
           MAE=mae(modelormst2,data=datarmtrain))

# Ajuste
ajustem2 <- fitted(modelormst2)
plot(datarmtrain[, 'VENTAS'],xlab="Años",ylab="Unidades",type="l")
lines(ajustem2,lwd=2,col="green3")

```

```

# Modelo 3
modelormst3 <- tslm(VENTAS~PPT+DN+GP+BIESS, data = datarmtrain)
summary(modelormst3)

# Ajuste
ajustem3 <- fitted(modelormst3)
plot(datarmtrain[, 'VENTAS'],xlab="Años",ylab="Unidades",type="l")
lines(ajustem3,lwd=2,col="#F5C710")

# Pronostico
pronom3 <- forecast(modelormst3,data.frame(datarmtest[,2:7],level=c(95)))
autoplot(pronom3)

round(accuracy(pronom3,data.frame(datarmtest),digits = 2))
accuracy(pronom3,data.frame(datarmtest))

```

```

# Modelo 4
modelormst4 <- tslm(VENTAS~PPT+DN+BIESS, data = datarmtrain)
summary(modelormst4)

# Ajuste
ajustem4 <- fitted(modelormst4)
plot(datarmtrain[, 'VENTAS'],xlab="Años",ylab="Unidades",type="l")
lines(ajustem4,lwd=2,col="blue")

# Pronostico
pronom4 <- forecast(modelormst4,data.frame(datarmtest[,2:7],level=c(95)))
autoplot(pronom4)

round(accuracy(pronom4,data.frame(datarmtest),digits = 2))
accuracy(pronom4,data.frame(datarmtest))

```

```

# Modelo 5
modelormst5 <- tslm(VENTAS~PPT+BIESS, data = datarmtrain)
summary(modelormst5)

# Ajuste
ajustem5 <- fitted(modelormst5)
plot(datarmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(ajustem5, lwd=2, col="deeppink1")

# Pronostico
pronom5 <- forecast(modelormst5, data.frame(datarmtest[, 2:7], level=c(95)))
autoplot(pronom5)

round(accuracy(pronom5, data.frame(datarmtest), digits = 2))
accuracy(pronom5, data.frame(datarmtest))

```

```

# Modelo 6
modelormst6 <- tslm(VENTAS~BIESS, data = datarmtrain)
summary(modelormst6)

# Metricas de ajuste
data.frame(R2=rsquare(modelormst6, data=datarmtrain),
           RMSE=rmse(modelormst6, data=datarmtrain),
           MAE=mae(modelormst6, data=datarmtrain))

# Ajuste
ajustem6 <- fitted(modelormst6)
plot(datarmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(ajustem6, lwd=2, col="lightseagreen")

# Pronostico
pronom6 <- forecast(modelormst6, data.frame(datarmtest[, 2:7], level=c(95)))
autoplot(pronom6)

```

```

# Modelo de Regresión con errores Arima
# Regresión con errores ARIMA (0,1,3)
fit1 <- Arima(datarmtrain[, 'VENTAS'], xreg =datarmtrain[, 2:7],
             order=c(0,1,3))
summary(fit1)
checkresiduals(fit1)

# Modelo 1 todas las variables
modelord1 <- fitted(fit1)
plot(datarmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord1, lwd=2, col="darkmagenta")

pronomrd1 <- forecast(fit1, xreg =datarmtest[, 2:7], h=10, level=c(95))

autoplot(pronomrd1)

accuracy(pronomrd1, datossttest)
round(accuracy(pronomrd1, h=10, datossttest), digits = 2)

```

```

# Modelo 2
fit2 <- Arima(datartrain[, 'VENTAS'], xreg = datarmtrain[, 2:6],
              order=c(0,1,3))
summary(fit2)
checkresiduals(fit2)

modelord2 <- fitted(fit2)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord2, lwd=2, col="red")

pronomrd2 <- forecast(fit2, xreg = datarmtest[, 2:6], h=10, level=c(95))

autoplot(pronomrd2)

accuracy(pronomrd2, datossttest)
round(accuracy(pronomrd2, h=10, datossttest), digits = 2)

```

```

# Modelo 3
fit3 <- Arima(datartrain[, 'VENTAS'], xreg = datarmtrain[, 2:5],
              order=c(0,1,3))
summary(fit3)
checkresiduals(fit3)

modelord3 <- fitted(fit3)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord3, lwd=2, col="green")

pronomrd3 <- forecast(fit3, xreg = datarmtest[, 2:5], h=10, level=c(95))

autoplot(pronomrd3)

accuracy(pronomrd3, datossttest)
round(accuracy(pronomrd3, h=10, datossttest), digits = 2)

```

```

# Modelo 4
fit4 <- Arima(datartrain[, 'VENTAS'], xreg = datarmtrain[, 2:4],
              order=c(0,1,3))
summary(fit4)
checkresiduals(fit4)

modelord4 <- fitted(fit4)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord4, lwd=2, col="purple")

pronomrd4 <- forecast(fit4, xreg = datarmtest[, 2:4], h=10, level=c(95))

autoplot(pronomrd4)

accuracy(pronomrd4, datossttest)
round(accuracy(pronomrd4, h=10, datossttest), digits = 2)

```

```
# Residuos vs Ajustes
cbind(Ajuste=fitted(fit4),Residuos=residuals(fit4)) %>%
  as.data.frame() %>% ggplot(aes(x=Ajuste, y=Residuos))+geom_point()

# Test de Normalidad
shapiro.test(fit4$residuals)
```

```
# Modelo 5
fit5 <- Arima(datararmtrain[, 'VENTAS'], xreg = datarmtrain[, 2:3],
             order=c(0,1,3))
summary(fit5)
checkresiduals(fit5)

modelord5 <- fitted(fit5)
plot(datararmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord5, lwd=2, col="orange")

pronomrd5 <- forecast(fit5, xreg = datarmtest[, 2:3], h=10, level=c(95))
accuracy(pronomrd5, datossttest)
round(accuracy(pronomrd5, h=10, datossttest), digits = 2)

# Gráfico de Datos, Ajuste y Pronóstico
autoplot(datosrmst[, 'VENTAS'], series="Datos")+
  autolayer(modelord5, series="Ajuste")+
  autolayer(pronomrd5, series="Pronóstico")+
  xlab("Año")+ylab("Unidades")+ggtitle("Datos, Ajuste y Pronóstico")
```

```
# Modelo 6
fit6 <- Arima(datararmtrain[, 'VENTAS'], xreg = datarmtrain[, 2], order=c(0,1,3))
summary(fit6)
checkresiduals(fit6)

modelord6 <- fitted(fit6)
plot(datararmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord6, lwd=2, col="brown")

pronomrd6 <- forecast(fit6, xreg = datarmtest[, 2], h=10, level=c(95))
autoplot(pronomrd6)

accuracy(pronomrd6, datossttest)
round(accuracy(pronomrd6, h=10, datossttest), digits = 2)
```

```
# Modelo 7
fit7 <- Arima(datararmtrain[, 'VENTAS'], xreg = datarmtrain[, 3], order=c(0,1,3))
summary(fit7)
checkresiduals(fit7)

modelord7 <- fitted(fit7)
plot(datararmtrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord7, lwd=2, col="yellow")

pronomrd7 <- forecast(fit7, xreg = datarmtest[, 3], h=10, level=c(95))
accuracy(pronomrd7, datossttest)
round(accuracy(pronomrd7, h=10, datossttest), digits = 2)
```



```

# Modelo 8
fit8 <- Arima(datartrain[, 'VENTAS'], xreg = datartrain[, 4], order=c(0,1,3))
summary(fit8)
checkresiduals(fit8)

modelord8 <- fitted(fit8)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord8, lwd=2, col="pink")

pronomrd8 <- forecast(fit8, xreg = datarmtest[, 4], h=10, level=c(95))
autoplot(pronomrd8)

accuracy(pronomrd8, datossttest)
round(accuracy(pronomrd8, h=10, datossttest), digits = 2)

```

```

# Modelo 9
fit9 <- Arima(datartrain[, 'VENTAS'], xreg = datartrain[, 5], order=c(0,1,3))
summary(fit9)
checkresiduals(fit9)

modelord9 <- fitted(fit9)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord9, lwd=2, col="blue")

pronomrd9 <- forecast(fit9, xreg = datarmtest[, 5], h=10, level=c(95))
autoplot(pronomrd9)

accuracy(pronomrd9, datossttest)
round(accuracy(pronomrd9, h=10, datossttest), digits = 2)

```

```

# Modelo 10
fit10 <- Arima(datartrain[, 'VENTAS'], xreg = datartrain[, 6],
              order=c(0,1,3))
summary(fit10)
checkresiduals(fit10)

modelord10 <- fitted(fit10)
plot(datartrain[, 'VENTAS'], xlab="Años", ylab="Unidades", type="l")
lines(modelord10, lwd=2, col="#F5C710")

pronomrd10 <- forecast(fit10, xreg = datarmtest[, 6], h=10, level=c(95))
autoplot(pronomrd10)

accuracy(pronomrd10, datossttest)
round(accuracy(pronomrd10, h=10, datossttest), digits = 2)

```