UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

FACULTAD DE CIENCIAS FISICAS Y MATEMÁTICAS

ESCUELA PROFESIONAL DE ESTADÍSTICA

Serie de tiempo para pronosticar la demanda de la línea de producto de bebidas gasificadas negras en la empresa Mercantil Inca S.A. - Chiclayo 2022.

Para optar el Título Profesional de:

Licenciado en Estadística

Presentado por:

Bach. Toro Vallejos Jhon Kenedy

Bach. Wong Altamirano Lear Scoth

Asesor:

Lic. Albañil Santisteban Víctor Teófilo

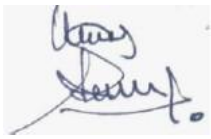
Lambayeque - Perú, 2026

Serie de tiempo para pronosticar la demanda de la línea de producto de bebidas gasificadas negras en la empresa Mercantil Inca S.A. - Chiclayo 2022.

Presentado por:

BACH. TORO VALLEJOS JHON KENEDY

AUTOR



BACH. WONG ALTAMIRANO LEAR SCOTH

AUTOR

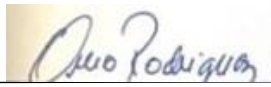
LIC. ALBAÑIL SANTISTEBAN VÍCTOR TEÓFILO

ASESOR

Lambayeque – Perú 2026

Serie de tiempo para pronosticar la demanda de la línea de producto de bebidas gasificadas negras en la empresa Mercantil Inca S.A. - Chiclayo 2022.

Aprobado por:



Dr. Lic. Est. Wilver Omero Rodriguez López

Presidente

Dr. Lic. Est. Wilder Alejandro Alvarado Castillo

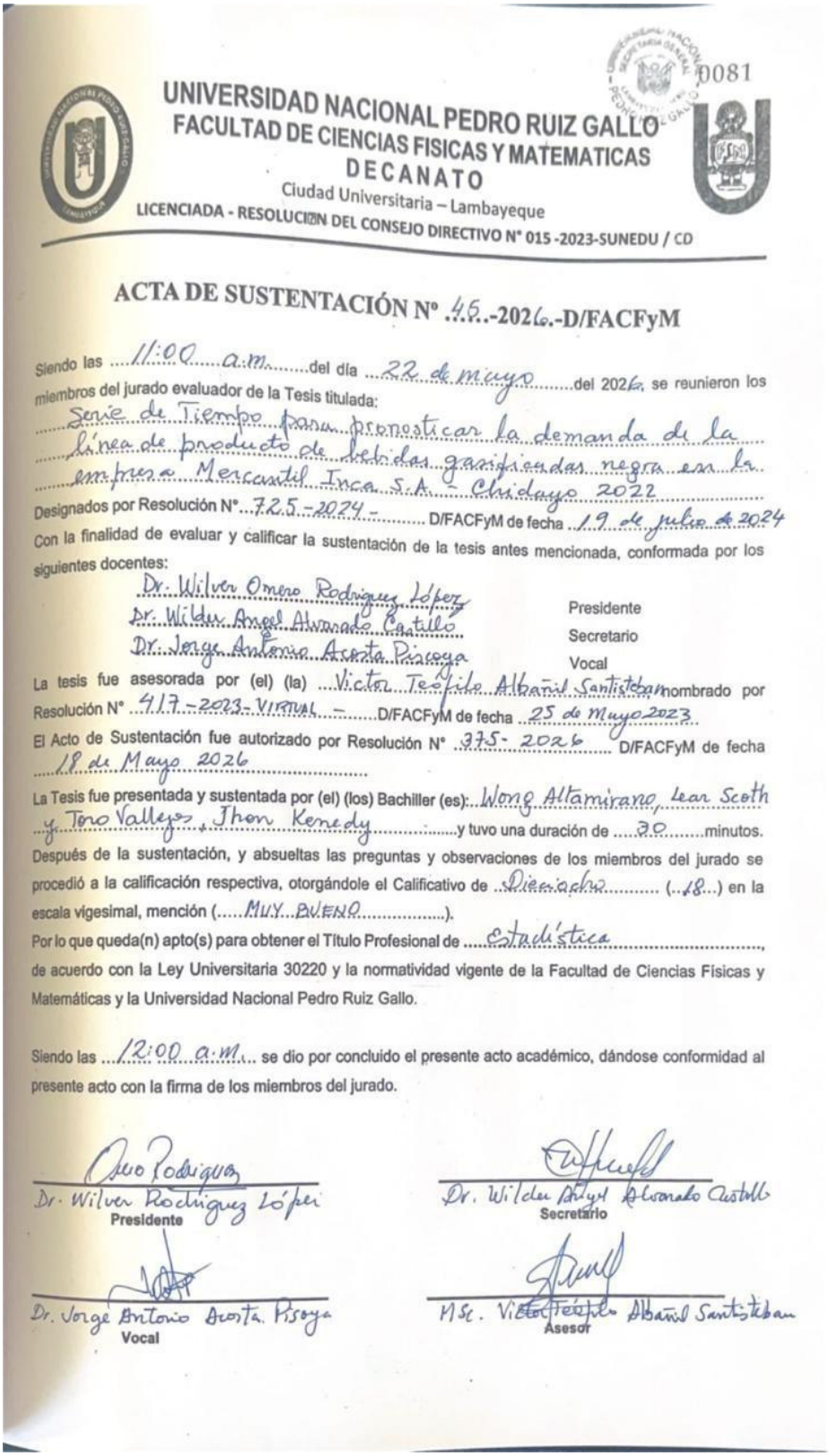
Secretario

Dr. Lic. Est. Jorge Antonio Acosta Piscoya

Vocal

Lambayeque – Perú 2026

ACTA DE SUSTENTACIÓN



0081

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
DECANATO
Ciudad Universitaria – Lambayeque
LICENCIADA - RESOLUCIÓN DEL CONSEJO DIRECTIVO N° 015 -2023-SUNEDU / CD

ACTA DE SUSTENTACIÓN N° .46..-2026.-D/FACFyM

Siendo las ...11:00 a.m.....del día ...22 de mayo.....del 2026, se reunieron los miembros del jurado evaluador de la Tesis titulada:
.....Serie de Tiempo para pronosticar la demanda de la línea de producto de bebidas gasificadas negra en la empresa Mercantil Inca S.A - Chiclayo 2022.....
Designados por Resolución N°...725-2024-..... D/FACFyM de fecha ..19 de julio de 2024
Con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

Dr. Wilver Omero Rodriguez López — Presidente
Dr. Wilder Angel Alvarado Castillo — Secretario
Dr. Jorge Antonio Acosta Pisoya — Vocal

La tesis fue asesorada por (el) (la) ...Victor Teófilo Albañil Santisteban.....nombrado por Resolución N° ...417-2023-VIRTUAL-.....D/FACFyM de fecha ..25 de Mayo 2023.....
El Acto de Sustentación fue autorizado por Resolución N° ..375-2026..... D/FACFyM de fecha ..18 de Mayo 2026.....
La Tesis fue presentada y sustentada por (el) (los) Bachiller (es):..Wong Altamirano, Sear Scoth y Toro Valleyos, Jhon Kenedy.....y tuvo una duración de .....20.....minutos.
Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de ..Dieciocho..... (..18...) en la escala vigesimal, mención (.....MUY BUENO.....).
Por lo que queda(n) apto(s) para obtener el Título Profesional de .....estadística.....,
de acuerdo con la Ley Universitaria 30220 y la normatividad vigente de la Facultad de Ciencias Físicas y Matemáticas y la Universidad Nacional Pedro Ruiz Gallo.

Siendo las ...12:00 a.m.... se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

Dr. Wilver Rodriguez López
Presidente

Dr. Wilder Angel Alvarado Castillo
Secretario

Dr. Jorge Antonio Acosta Pisoya
Vocal

Msc. Victor Teófilo Albañil Santisteban
Asesor

CONSTANCIA DE VERIFICACIÓN DE ORIGINALIDAD DE ORIGINALIDAD

Yo, Lic. Albañil Santisteban Víctor Teófilo, usuario revisor del documento titulado, "Serie de tiempo para pronosticar la venta de la línea de producto de bebidas gasificadas negras en la empresa mercantil inca s.a. - Chiclayo 2022". Cuyos autores son: Wong Altamirano Lear Scoth , identificado con Documento de identidad N.° 43735039 y Toro Vallejos Jhon Kenedy, identificado con Documento de Identidad N.° 46011469, declaro que la evaluación realizada por el Programa Informático, ha arrojado un porcentaje de similitud de 17 %, verificable en el Resumen de Reporte automatizado de similitudes que se acompaña.

El suscrito analizó dicho reporte y concluyó que cada una de las coincidencias detectadas dentro del porcentaje de similitud permitido no constituyen plagio y que el documento cumple con la integridad científica y con las normas para el uso de citas y referencias establecidas en los protocolos respectivos.

Lambayeque 25 de Mayo, 2026

FIRMA

Victor Teófilo Albañil Santiesteban…

DNI: 17633734
ASESOR

RESULTADO DE INFORME DE SIMILITUD

SERIE DE TIEMPO PARA PRONOSTICAR LA DEMANDA DE LA LÍNEA DE PRODUCTO DE BEBIDAS GASIFICADAS NEGRAS EN LA EMPRESA MERCANTIL INCA S.A. - CHICLAYO 2019

INFORME DE ORIGINALIDAD

17%	17%	3%	7%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	9%
2	Submitted to Universidad Nacional de Frontera Trabajo del estudiante	1%
3	repositorio.unprg.edu.pe:8080 Fuente de Internet	1%
4	Submitted to Universidad Nacional Pedro Ruiz Gallo Trabajo del estudiante	1%
5	www.scielo.org.mx Fuente de Internet	1%
6	alicia.concytec.gob.pe Fuente de Internet	<1%
7	revistas.unitru.edu.pe Fuente de Internet	<1%
8	repositorio.unasam.edu.pe Fuente de Internet	<1%
9	cipres.sanmateo.edu.co Fuente de Internet	<1%
10	repositorio.uns.edu.pe Fuente de Internet	<1%
11	risti.xyz Fuente de Internet	<1%

12	repositorio.unprg.edu.pe Fuente de Internet	<1 %
13	Submitted to Universidad Técnica Nacional de Costa Rica Trabajo del estudiante	<1 %
14	dokumen.pub Fuente de Internet	<1 %
15	Submitted to Unviersidad de Granada Trabajo del estudiante	<1 %
16	repositorio.uncp.edu.pe Fuente de Internet	<1 %
17	www.coursehero.com Fuente de Internet	<1 %
18	www.dspace.espol.edu.ec Fuente de Internet	<1 %
19	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	<1 %
20	colposdigital.colpos.mx:8080 Fuente de Internet	<1 %
21	repositorio.unap.edu.pe Fuente de Internet	<1 %
22	vdocuments.es Fuente de Internet	<1 %
23	bellman.ciencias.uniovi.es Fuente de Internet	<1 %
24	cybertesis.uni.edu.pe Fuente de Internet	<1 %
25	Submitted to University of Hong Kong Trabajo del estudiante	<1 %
26	mat.izt.uam.mx Fuente de Internet	<1 %

27	www.buenastareas.com Fuente de Internet	<1 %
28	www.info.gov.hk Fuente de Internet	<1 %
29	www.researchgate.net Fuente de Internet	<1 %

Excluir citas	Activo	Excluir coincidencias	< 15 words
Excluir bibliografía	Activo		

Asesor
VICTOR TEOFILO ALBAÑIL SANTISTEBAN
DNI 17633734

turnitin

Recibo digital

Este recibo confirma quesu trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	jhon Kenedy Toro Vallejos
Título del ejercicio:	Quick Submit
Título de la entrega:	SERIE DE TIEMPO PARA PRONOSTICAR LA DEMANDA DE LA LÍ...
Nombre del archivo:	TESIS_TORO-WONG-2025.docx
Tamaño del archivo:	2.04M
Total páginas:	85
Total de palabras:	11,697
Total de caracteres:	66,132
Fecha de entrega:	10-jul-2025 12:31p. m. (UTC-0500)
Identificador de la entrega:	2712939754

ASESOR:VICTOR TEOFILO ALBAÑIL SANTISTEBAN
DNI: 17633734

DEDICATORIA

Reconocemos con gratitud la mano providencial de Dios, quien nos fortaleció en cada etapa de este trabajo y dispuso las circunstancias y colaboraciones que hicieron posible su realización.

Y por supuesto, a nuestros padres. Gracias por darnos las herramientas para llegar hasta aquí. Nos enseñaron el valor de la perseverancia, la importancia de la paciencia, y la motivación para competir y dar lo mejor de nosotros mismos. Más importante aún, nos rodearon de amor y buenas vibras, creando el ambiente perfecto para que pudiéramos florecer y alcanzar este sueño. ¡Los queremos!

Dedicamos esta investigación con profunda gratitud a los profesores que compartieron su conocimiento, a los compañeros que brindaron su aliento, y a las familias que ofrecieron su paciencia y comprensión. Su apoyo incondicional y valiosas perspectivas fueron fundamentales para superar los desafíos que encontramos en el camino, haciendo posible la culminación de esta tesis y la realización de nuestro sueño. Con la esperanza de que este trabajo inspire y sirva de guía a futuras generaciones de estudiantes de estadística, quienes encontrarán en estas páginas un punto de partida para sus propias exploraciones y descubrimientos.

Los autores

AGRADECIMIENTO

Expresamos nuestra profunda gratitud a una fuerza superior por brindarnos la fortaleza necesaria para finalizar esta tesis, por sostenernos en los momentos más desafiantes y por hacer posible la realización de este proyecto académico

Extendemos nuestro agradecimiento a nuestras familias por su apoyo incondicional en la culminación de este proceso, por la inversión realizada en nuestra formación académica y por el aliento moral que nos brindaron cuando más lo necesitamos

Reconocemos la valiosa guía de nuestro asesor, Victor Albañil, por sus orientaciones expertas, su paciencia durante nuestro aprendizaje y su constante disponibilidad para facilitar el progreso de esta tesis

Agradecemos también la colaboración de los docentes que contribuyeron a nuestra comprensión de conceptos clave, los cuales fueron esenciales para la construcción y el desarrollo de esta investigación

El trabajo se a realizado gracias a sus enseñanzas y a nuestra disposición.

Los autores

INDICE GENERAL

Contenido

INDICE DE TABLAS

INDICE DE FIGURAS

RESUMEN

Hacemos la investigación tuvo como objetivo el uso de la metodología Box Jenkins para encontrar el modelo que mejor se ajustaba a la serie de la demanda de cierre mensual de las ventas de bebidas gasificadas negras de la empresa Mercantil Inca S.A. Chiclayo, desde diciembre de 2017 hasta noviembre de 2022. El estudio se basó en un diseño de investigación observacional, predictivo, retrospectivo y longitudinal. La muestra del estudio comprendió 54 observaciones de la variable principal. Además, se incluyeron 6 observaciones adicionales para contrastar el pronóstico realizado. En total, el estudio abarcó 60 observaciones. Tras llevar a cabo el análisis con la metodología establecida, se realizó un pronóstico para los seis meses siguientes. Posteriormente, se compararon los resultados pronosticados con los datos reales observados, con el fin de determinar cuál de los pronósticos generados demostró ser el más preciso. Se observó que la metodología de Box Jenkins con el menor error, para ello se utilizó el indicado IAC y los errores promedio como MAE, MAPE, RMSE . Por consecuente se considera a la metodología de Box Jenkins con el modelo Sarima (1,1,0) (1,1,0) para el procedimiento de pronóstico de la serie de tiempo en estudio.

Palabras claves: Metodología Box Jenkins, Ventas de bebidas gasificadas negras.

ABTRACT

The research aimed to use the Box-Jenkins methodology to find the model that best fit the series of monthly closing demand for black carbonated beverage sales of the company Mercantil Inca S.A. Chiclayo, from December 2017 to November 2022. The study was based on an observational, predictive, retrospective, and longitudinal research design. The study sample comprised 54 observations of the main variable. In addition, 6 additional observations were included to contrast the forecast made. In total, the study covered 60 observations. After carrying out the analysis with the established methodology, a forecast was made for the following six months. Subsequently, the forecasted results were compared with the real observed data, in order to determine which of the generated forecasts proved to be the most accurate.It was observed that the Box Jenkins methodology had the least error, for which the indicated IAC and the average errors such as MAE, MAPE, RMSE were used. Consequently, the Box Jenkins methodology is considered with the Sarima model (1,1,0) (1,1,0) for the forecast procedure of the time series under study.

Keywords: Box Jenkins Methodology, sales of black carbonated beverages.

INTRODUCCION

En la presente investigación, se abordó el tema del pronóstico de ventas del cierre mensual de las bebidas gasificadas negras en la empresa Mercantil Inca S.A.; esta empresa distribuidora cuenta con más de 10 agencias en el país, es una de las distribuidoras con gran presencia en el norte del país.

En su investigación llamada "Método de modelos ARIMA y Box Jenkins", Fournies P. mencionó que el método Box Jenkins comenzó en el '70. Se trataba de facilitar las estadísticas creando un modelo para una serie a corto plazo. Este modelo trató de descubrir cómo se reunió la serie y adivina lo que podría hacer a continuación. El autor describió una serie temporal como un montón de puntos de datos sobre algo que cambia con el tiempo, recopilado a intervalos de tiempo regulares. Fournies P. habló sobre esta caja de Jenkins, que es una forma de modificar estos modelos ARIMA para que se ajusten a un montón de datos.

Creemos que el método Box Jenkins se ajusta para analizar series temporales, y es por eso que lo vamos a utilizar; Además, es un nuevo enfoque en el campo.

Como consecuencia se planteó en la investigación el siguiente **problema:** ¿Cuál sería el modelo de la serie de tiempo para el pronóstico de la demanda de la línea de producto de bebidas gasificadas negras en la empresa mercantil inca S.A.?

Ante la pregunta, formulamos el siguiente **objetivo General:** Determinar el modelo de serie de tiempo para pronosticar la demanda en la línea de producto de bebidas gasificadas negras de la empresa Mercantil Inca S.A.

La hipótesis planteada fue que el modelo de pronóstico para las ventas de cierre mensual en las bebidas gasificadas negras de la empresa Mercantil Inca S.A. en Chiclayo, desde diciembre

de 2017 hasta noviembre de 2022, se ajustó y proporcionó pronósticos con menor margen de error que un modelo de serie de tiempo basado en la metodología de Box Jenkins7

La investigación se desarrolló a través de capítulos, los cuales se detallan a continuación.

En el Capítulo I, se llevó a cabo el diseño teórico.

El Capítulo II se dedicó al desarrollo de los conceptos relacionados con los métodos que se utilizarían a lo largo de la ejecución del proyecto.

En el Capítulo III, se detallaron las metodologías.

El Capítulo IV presentó el análisis e interpretación de los.

Finalmente, en el Capítulo V, se propuso una solución al problema, lo cual se expresó como una recomendación concreta.

I. DISEÑO TEÓRICO

1.1 Antecedentes de la investigación

1.1.1 A nivel Internacional

Amézquita (2021) investigó un poco sobre la predicción de la demanda de viajes aéreos en Colombia para el futuro cercano utilizando modelos de series de tiempo. Compartieron los resultados de verificar qué tan bien diferentes modelos pueden predecir las tendencias futuras en los datos de la serie temporal que verificaron los números mensuales de transporte aéreo de pasajeros, observando la escena local y global local, de 1994 a 2019. Observaron un grupo de modelos, como la regresión armónica, uno, los resultados de Holt-Winters Exponenciales, el modelo de autor móvil móvil, y el sarimas, y los resultados de los resultados de los Mobile Mobile, y el Sarima, y los resultados de los resultados de los Mobile Mobile, y el Sarima, y los resultados de los Resultados, y los resultados de los resultados de los resultados móviles, y el Sarima, y los resultados, y los resultados de los resultados de los Mobile Mobile, y los resultados de los resultados de Sarima, y los resultados de los resultados de los Mobile. Los modelos Sarima y Sarimax son de primera categoría, con sus predicciones súper precisas, alcanzando ese punto óptimo de menos del 3% de error en promedio. El modelo multivariado de Sarimax supera a los univariados, y está claro que el PIB aumenta el rendimiento del modelo, especialmente en la industria del aire. Incluso después de considerar otras cosas como los precios del petróleo y los eventos inesperados de todo el mundo, todavía marcó una gran diferencia. El modelo de regresión armónica hizo un gran trabajo al adivinar el número exacto de personas que viajan internacionalmente por aire, con una muy buena precisión del otro lado, el modelo Holt-Winters

lo clavó con predicciones súper precisas para los datos de tráfico, tanto en una escala global como en general.

Tobón (2019), en su investigación: Desarrollo de un modelo de pronóstico en la metodología de box-jenkins para la demanda de estudiantes de pregrado de la fundación universitaria tecnológico COMFENALCO de la ciudad de Cartagena de indias desde 2002 a 2018, Los resultados mostraron una identificación que incluía el historial de inscripciones semestrales desde el primer semestre de 2001 hasta el segundo semestre de 2018, comprendiendo un total de 36 registros. El análisis gráfico reveló que la serie temporal exhibió una tendencia estacionaria en ciertos segmentos de su evolución. Se observó que la variable había aumentado en las últimas dos décadas representadas, aunque en los últimos tres años se registró un decremento. Luego se analizaron los modelos ARIMA(2,1,2), ARIMA(3,1,1), ARIMA (1,1,1), ARIMA (6,0,2) con fixed=c(0,NA,0,NA,0,NA,0,0,NA), ARIMA (6,0,1) con fixed=c(NA,0,NA,0,NA,0,0,NA), Tras el ajuste efectuado, se determinó que el modelo lineal presentado era el que mejor se ajustaba al comportamiento observado yt = 0,5568yt−1 + 0,1914yt−3 − 0,1885yt−5 + 1398,58 + Ɛ, Utilizando el modelo ARIMA, se observó que la serie temporal demostraba una marcada dependencia de los datos más recientes, dependencia que decrecía al analizar períodos más alejados en el tiempo. El modelo, en su aplicación, no incluyó un componente MA ni requirió la diferenciación de los datos, dado que la serie temporal original presentaba estacionariedad. Aunque la predicción resultante del modelo ARIMA no se alineó perfectamente con la serie temporal y requería información adicional para mejorar su solidez, sí ofreció una aproximación al comportamiento de la serie. Dicha predicción sugirió una estabilidad en la demanda, caracterizada por una mayor actividad en el primer semestre de cada año y una disminución en el segundo, reflejando el patrón observado en años anteriores. Se hizo hincapié en la importancia de refinar la

estimación del modelo a través de la incorporación de más datos históricos y la consideración de variables exógenas representativas de la economía local y nacional. Este proceso permitiría obtener una estimación más fiable con el transcurso del tiempo.

Fierro et al. (2022), En su investigación: Análisis comparativo de los modelos tradicionales y modernos para el pronóstico de la demanda: enfoques y características presentadas los siguientes resultados: El propósito de esta investigación documental era establecer la clasificación de los principales tipos de pronósticos y proponer modelos representativos para la implementación en pequeñas y medianas empresas. Encontraron tres formas clave de configurar modelos de pronóstico a corto plazo. La primera parte fue sobre el uso de estadísticas o modelos de la vieja escuela, como lo de Holt-Winters para suavizar los datos y el modelo ARIMA para los promedios. Eran las opciones de referencia. El segundo método habló sobre cosas de aprendizaje automático, como la cosa del bosque aleatorio y el modelo LSTM para patrones a corto plazo por último, observaron un tipo de modelo de mezcla y combinación, hablando sobre el combo ARIMA-LSTM y el Profeta de Facebook. Descubrieron que nadie tiene una bola de cristal para adivinar cuántas cosas que necesitaremos en el futuro resulta, cuán preciso es un modelo realmente depende de cuánto y cuán buenos son los datos, además de cómo se comporta la serie temporal cuando se mira cuando miran los datos, no se dieron cuenta de que tenía un patrón constante, lo que hace que fuera más fácil predecir las cosas futuras con menos conjeturas. Cuando la serie temporal no era estable, utilizaron un truco matemático llamado diferenciación para que sea más fácil trabajar con los modelos de pronóstico a corto plazo de la vieja escuela solo dependen de las estadísticas, pero hoy en día, tenemos dos grandes tendencias: aprendizaje automático que aprende por sí mismo y una combinación de diferentes métodos.

Roque (2021), en su investigación: Modelo de Box-Jenkins para el pronóstico de la recaudación de ingreso de la genérica 1.3 de una entidad pública, 2017 – 2021, El estudio aplicó la metodología de Box-Jenkins en el período comprendido entre julio de 2017 y junio de 2021 con el objetivo de pronosticar los ingresos mensuales de la entidad, específicamente los correspondientes a la genérica de ingresos 1.3. La metodología facilitó el análisis y la predicción de los ingresos mensuales de la entidad en esta categoría, que abarca la venta de servicios, derechos administrativos y bienes. Los datos estadísticos revelaron características importantes que permitieron identificar patrones específicos en los ingresos mensuales de la entidad.

El pronóstico para el período de julio a diciembre de 2021, basado en el modelo ARIMA (1,1,2), indicó un aumento en comparación con el año 2020. Este incremento se atribuyó a las medidas implementadas por el gobierno peruano en relación con el protocolo sanitario y a la gradual reapertura de diversos sectores y servicios a nivel nacional. Estas acciones permitieron una mayor afluencia de público y una mayor conciencia sobre la salud en la población peruana.

Izquierdo et al. (2022), en su investigación modelo de pronóstico de la exportación minera en el perú, 2020, El estudio tuvo como objetivo establecer un modelo de pronóstico de la exportación minera en el Perú para el período de enero de 2016 a diciembre de 2020, buscando comprender los cambios producidos por la coyuntura mundial, específicamente la pandemia de COVID-19, Los resultados arrojaron que el modelo ARIMA (1,0,0) se ajustaba adecuadamente a la exportación minera en el Perú durante el periodo analizado, correspondiendo a un modelo autorregresivo de orden uno sin diferencia y sin medias móviles: yt = 2270.176 + 0.749596yt−1 + ϵ

Montenegro (2019), en su investigación: Evaluación de métodos de pronóstico para predecir las ventas de los productos de la línea cardiológica del boticario de San Carlos - Chiclayo, presente los siguientes resultados: el análisis se centró en cuatro medicamentos principales ya comercializados: 100 mg de aspirina, bisoprolol de 5 mg, clip -clipidogrel de 75 mg e isosorburo de 10 mg. Se sumergieron profundamente en cómo estos productos relacionados con el corazón se vendieron con el tiempo. Eligieron el método de drogas que tenía la menor posibilidad de equivocarse después de revisar un montón de opciones, descubrieron que usar el promedio de doble móvil era la mejor manera de manejar la aspirina de 100 mg. Para la dosis de bisoprolol de 5 mg, el modelo ARIMA (1,0,2) salió a la cima para predecir el mejor resultado. Para el clopidogrel de 75 mg, eligieron el modelo ARIMA (2,0,2), y para el isosorburo de 10 mg, fueron con el modelo ARIMA (1,0,1).

Arévalo y Santisteban (2020), en su modelo de investigación de investigación sobre la cantidad de producción de electricidad con la metodología de la caja de Jenkins de la región de Lambayeque en el período 2005-2017, obtuvieron los siguientes resultados: la serie original que representa la producción de electricidad en la región de Lambayeque durante el período 2005-2017 no mostró una tendencia apreciable, sin estacionamiento cada 12 meses. Pero una vez que separaron la serie con el tiempo, podrían establecer estacionamiento con una probabilidad súper confiable del 95%, basada en la versión más dura de la prueba de Dickey Fuller. Se identificaron varios modelos que demostraron ser satisfactorios para este contexto específico: SARIMA (1,1,5) (1,1,0); SARIMA (2,1,2) (1,0,0); SARIMA (2,1,2) (1,1,0).

El modelo óptimo hallado fue el SARIMA (2,1,2) (1,1,0), cuyo modelo es:$(1 - \propto 1\ L - \propto 2\ L2)(1 - \beta 1L12)((Yt - Yt{-}1) - (Yt{-}12 - Yt{-}13)) = (1 - 1L - \emptyset 2L2)et$, y cuya ecuación es:

Yt = 1.2858 Yt−1 + 0.7365 Yt−2 − 0.4507 Yt−3 +0.5022 Yt−12 − 0.07412876 Yt−13 − 0.81858722 Yt−14 + 0.22634154 Yt−15 +0.4978 Yt−24 − 0.35552876 Yt−25 − 0.08208722 Yt−26 + 0.22435846 Yt−27 + et +0.5496 et−1 + 0.4345 et−2, por cumplir con todos los criterios de modelo satisfactorios y

por ser parsimonioso con el menor BIC.

1.2 Base Teórica

1.2.1 Series de tiempo

Oliva (2019) La primera forma de abordar el análisis de series de tiempo es a través de la visualización de los datos en una gráfica, donde uno de los ejes representa el tiempo. Sin embargo, existen dos enfoques para este tipo de análisis. Por un lado, está el enfoque de la importancia del tiempo, que consiste en reconocer cómo lo que sucede hoy es afectado por lo que sucedió ayer. Existía también el enfoque del análisis de frecuencia, el cual buscaba reconocer la importancia que los ciclos tenían para los investigadores.

López (2020) mencionó que una serie de tiempo era una secuencia de mediciones que describían la evolución de un fenómeno o variable a lo largo del tiempo. Cada medición se asociaba con una marca de tiempo, lo que permitía que los datos se presentaran de forma estructurada. En otras palabras, una serie de tiempo era una secuencia de observaciones sobre intervalos de tiempo regulares.

1.2.2 Modelo de serie de tiempo de un ARIMA y SARIMA

ARIMA

Ayala y Bucio (2020) encontraron que los modelos Arima, que son como una mezcla de auto-representantes y calcetines móviles, son básicamente una extensión de los modelos de armas tradicionales. La letra "I" se coloca para mostrar las cosas que no se producen en algunas series de tiempo, este problema de no estelar se puede solucionar simplemente eliminando partes de la serie original, deshaciéndose de una posible tendencia polinómica del orden D que está en la serie que estamos viendo.

En el modelo ARIMA, la parte "integrada" mostró que YT provenía de tomar la relación (1) y usar el operador ∇D, lo que básicamente significa agregar todos los términos XT al infinito. Los modelos ARIMA (P, D, Q) tenían piezas 'P' que podían recuperarse, 'D' se trataba de cuán diferentes eran las cosas, y 'Q+1' era la cantidad de bits desordenados por cosas aleatorias. Mostró su idea con esta ecuación:

$$\nabla dYt = \varphi 1 \nabla dYt{-}1 + \varphi 2 \nabla dYt{-}2 + \cdots + \varphi p \nabla dYt{-}p + \theta 1 \varepsilon t{-}1 + \theta 2 \varepsilon t{-}2 + \cdots + \theta q \varepsilon t{-}q + \varepsilon t$$

Peréz (2022) Es posible que la serie temporal que estamos examinando no se origine a partir de un proceso estocástico estacionario y, por lo tanto, no cumpla con las suposiciones de los modelos mencionados. En el caso de que la serie temporal exhiba estacionariedad, se procede a aplicar un modelo ARMA. En situaciones contrarias, se busca una serie transformada para lograr estacionariedad, comúnmente mediante la diferenciación de la serie original. Esta necesidad da lugar a la definición de la clase de modelos ARIMA, que no es más que una ampliación de los modelos ARMA diseñada para abordar series no estacionarias.

Si tomamos un número entero no negativo representado por d, se afirma que {Yt} es un proceso ARIMA(p, d, q) si la serie Xt = (1 − B)^d Yt es un proceso ARMA(p, q) que exhibe estacionariedad.

Esta definición implica que la serie {Yt} satisface la relación

$$\phi(B)(1 - B)^d Y_t = c + \theta(B)\varepsilon_t$$

En la expresión, ϕ(z) y θ(z) representan los polinomios autorregresivo y de media móvil de grado p y q, respectivamente, sin compartir raíces comunes, y {εt} sigue una distribución normal N(0, σ^2). Es importante destacar que el proceso es estacionario únicamente cuando d = 0, lo que resultaría en un modelo ARMA(p, q).

Modelos Sarima

Los modelos SARIMA tienen la capacidad de representar diversos patrones estacionales. En su configuración (p, d, q) (P, D, Q)m, incorporan elementos de datos estacionales (P, D, Q)m y los modelos ARIMA (p, d, q), donde m representa el número de observaciones en un año.

Osorio (2019), El modelo SARIMA puede ser conceptualizado como una ampliación del modelo ARIMA, incorporando un componente estacional conocido como “Seasonal Autoregressive Integrated Moving Average (Media Móvil Integrada Autorregresiva Estacional)”. Este nuevo componente agrega elementos (P, D, Q) a la forma tradicional (p, d, q) del modelo ARIMA. La expresión resultante es:

$$\phi(B)+\Phi(Bs)(1-B)d(1-Bs)Dzt=\theta(B)\Theta(Bs)$$

Donde (1−B)^d representa la cantidad de diferenciaciones no estacionales en el modelo, y (1−Bs)^D se refiere a la diferenciación estacional.

Metodología Box-Jenkins

Roque (2021), En 1976, Box & Jenkins introdujeron este modelo mediante su obra "Time Series Analysis: Forecasting and Control", marcando el comienzo de las herramientas de pronóstico. La metodología de Box-Jenkins, también llamada metodología ARIMA, implica la modelación del proceso generador de un conjunto de datos. El enfoque se desglosó en un procedimiento iterativo de cuatro fases. Primero, se identificó el modelo. Luego, se estimaron los parámetros necesarios. Después, se realizó la validación del modelo. Finalmente, se generó el pronóstico basado en los resultados obtenidos.

Identificación del modelo

Marroqui (2019), antes de encontrar el modelo correcto para que coincida con los datos, verificaron cómo funcionaban las cosas estacionales. Durante esa parte, descubrieron la autocorrelación total y parcial para la serie de datos. Hicieron esto pensando que todo era estable y predecible con las cosas aleatorias que suceden.

Asimismo, se comparan los correlogramas de manera adecuada con los de los modelos "AR(p), MA(q) o ARMA(p, q)".

Prueba de raíz unitaria Dickey-Fuller

En el pasado, este contraste se empleó para evaluar la no estacionariedad. La hipótesis nula ($H0$) planteaba la existencia de una raíz unitaria en el proceso generador de los valores de la serie analizada. La formulación básica fue la siguiente:

H1: yt=a1yt−1+εt

H0: yt=yt−1+εt

Para el parámetro estimado "$a1$", su distribución de probabilidad presenta una discontinuidad cuando $a1=1$a1=1, indicando la no estacionariedad en la varianza de yt. El modelo utilizado se expresa como:

$$\Delta Yt=a0+(a1-1)yt-1+\varepsilon t$$

$$\Delta Yt=a0+\gamma yt-1+\varepsilon t$$

Donde H0:γ=0 en comparación con H1:γ<0En el estudio, se señaló que los datos base utilizados para contrastar el parámetro "g" dependían de la selección del proceso generador de valores. En este sentido, se presentaron tres modelos distintos.:

Modelo 1 (simple) (τμ):

$\Delta Yt=\gamma yt-1+\varepsilon t$

Modelo 2 (con constante) (τμ):

$\Delta Yt=a0+\gamma yt-1+\varepsilon t$

Modelo 3 (con constante y tendencia determinística) (ττ):

$\Delta Yt=a0+\gamma yt-1+a2t+\varepsilon t$

¿Que es raiz unitaria?

Arevalo (2020), En el texto se afirmaba que la prueba en cuestión confirmaba si las variables eran estacionarias, evaluando posibles desigualdades en comparación con un estado estacionario. Se indicaba, además, que esta prueba podía prevenir la regresión espuria.

Estimación del modelo

En esa fase, se determinaron los valores óptimos (significativos) de los parámetros autorregresivos y de media móvil que se habían incorporado en el modelo.

Etapa de Verificación del modelo

En esa fase, se determinó la idoneidad del modelo mediante la prueba de ruido blanco intermedio aplicada a los residuales.

Etapa de Pronóstico del modelo

Después de que terminaron los últimos pasos de su plan, obtuvieron un modelo que es realmente bueno para hacer predicciones que van más allá del corto plazo, superando a los de la vieja escuela.

Criterios de información de Akaike y Bayesiano

El criterio de información de Akaike (AIC) se presenta como una estimación muestral de la esperanza de la log-verosimilitud y se define mediante la fórmula:

$$AIC(k) = -2\ lnL[\Theta\hat{}(k)] + 2k.$$

En esta expresión, $L[\Theta\hat{}(k)]$ representa la función de máxima verosimilitud de las observaciones, $\Theta\hat{}(k)$ En el modelo, se utilizó la estimación de máxima verosimilitud para determinar el vector de parámetros Θ. La variable k representó el número de parámetros independientes que se estimaron.

El primer término de AIC se interpreta como una medida de la bondad de ajuste, mientras que el segundo término actúa como una penalización que aumenta con el número de parámetros, siguiendo el Principio de Parsimonia. El objetivo es seleccionar el modelo con el menor valor de

AIC, ya que este criterio no busca identificar el modelo verdadero, sino el más adecuado entre los candidatos.

El AIC puede volverse menos fiable con muestras pequeñas, lo cual motiva la introducción del concepto de AIC corregido (AICc), especialmente útil en tales situaciones. Para muestras grandes, el comportamiento de ambos es similar.

Por lo cual queda definido de la siguiente manera

$$〖AIC〗_c=AIC+ \frac{2k(k+1)}{(n-k-1)}$$

Criterio de Información Bayesiano (BIC) Por otro lado, el criterio de información Bayesiano (BIC), propuesto por Schwarz en 1978, modifica el AIC penalizando el número de parámetros con ln n en lugar de 2. La fórmula de:

$$BIC(k, n) = -2 \ln L[\hat{\Theta}(k)] + k \ln n$$

En retrospectiva, el análisis se centró en varios componentes clave. Se examinó la función de verosimilitud máxima de las observaciones, denotada como $L[\hat{\Theta}(k)]$, que cuantificaba la probabilidad de los datos observados dado el modelo. Además, se determinó la estimación de máxima verosimilitud del vector de parámetros, representada como $\hat{\Theta}(k)$, que proporcionaba los valores de los parámetros que mejor se ajustaban a los datos. El número de parámetros independientes estimados dentro del modelo, simbolizado por k, se tuvo en cuenta, al igual que el tamaño de la muestra (n) utilizada en el análisis.

Este criterio comparte numerosas características con el AIC, destacando que su objetivo principal es estimar la pérdida de información al aproximar un modelo real con uno estimado. En otras palabras, a medida que el valor del BIC disminuye, se obtiene un modelo de mejor calidad. Al aumentar el tamaño de la muestra, la estimación se vuelve más confiable, y el BIC funciona como una medida global de la bondad del ajuste del modelo.

Sin embargo, a diferencia del AIC, este criterio impone restricciones más estrictas a los parámetros, penalizando con el valor de ln(n)k. Por lo tanto, al introducir más parámetros en el modelo, se mejora el ajuste y se reduce la desviación medida por -2 lnL[Θˆ(k)]. Esto representa una compensación entre la desviación y la penalización en el contexto del BIC.

Diferencias entre el criterio AIC y el BIC.

Aunque AIC y BIC comparten características, como su enfoque en estimar la pérdida de información al aproximar un modelo real, la principal diferencia radica en cómo penalizan los modelos con más parámetros. BIC tiende a seleccionar modelos de orden inferior en comparación con AIC, ya que la penalización es más fuerte. Además, los modelos seleccionados por AIC son más apropiados para tamaños de muestra grandes. La elección entre AIC y BIC depende de las condiciones específicas de la estimación del modelo.

II. DISEÑO METODOLÓGICO

Definición y Operacionalización de variables:

Tipo de variable	Nombre de la variable	Dimensión	Indicadores	Escala
Independiente	Serie de tiempo	Base de datos a través del tiempo	Registro de cierre de ventas mensuales	Cuantitativo continuo

2.1 Tipo de Investigación

La presente investigación es de tipo aplicada, ya que tiene como finalidad generar un modelo de pronóstico que contribuya a la toma de decisiones a partir de datos históricos. Asimismo, es de nivel descriptivo y predictivo, debido a que analiza el comportamiento de la serie de datos y proyecta valores futuros mediante técnicas estadísticas.

2.2 Método de Investigación

El método de investigación utilizado es cuantitativo, ya que se basa en el análisis de datos numéricos. Se empleó el enfoque de series de tiempo, específicamente la metodología Box-Jenkins (modelos ARIMA), para la construcción del modelo de pronóstico, utilizando el software estadístico R para el procesamiento y modelamiento de los datos

2.3 Diseño de contrastación

El diseño de la investigación es descriptiva, retrospectiva y predictiva

2.4 Población, Muestra y Muestreo

La población está conformada por los registros de cierre de ventas de la empresa Mercantil Inca S.A. en la sede Chiclayo

Está conformada por los registros de ventas de producto de bebidas gasificadas negras en la empresa mercantil inca s.a. durante el periodo 2017 -2022. no amerita utilizar muestreo.

2.5 Técnicas, Instrumentos, Equipos y Materiales de Recolección de Datos

No amerita utilizar técnicas e instrumentos de recolección de datos debido a que se está utilizando una población especifica.

2.5.1 Técnicas de recolección:

La técnica utilizada será por observación de acuerdo a la base de datos que se registran mensualmente al cierre de ventas.

2.5.2 Técnicas estadísticas:

Se emplearon las técnicas de modelado de series de tiempo correspondientes a la metodología de Box-Jenkins.

2.5.3 Procedimiento Estadístico

1. Metodología Box Jenkins

1.1 Identificación del Modelo

La identificación del modelo consistió en determinar las componentes del modelo ARIMA (p,d,q) o SARIMA (p,d,q)(P,D,Q), donde:

p: Representó el número de componentes autorregresivos AR(p) no estacionales.

d: Indicó el número de diferencias no estacionales.

q: Correspondió al número de componentes de promedio móvil MA(q) no estacionales.

P: Denotó el número de componentes autorregresivos AR(P) estacionales.

D: Se refirió al número de diferencias estacionales.

Q: Representó el número de componentes de promedio móvil MA(Q) estacionales. La identificación del modelo también incluirá:

Hicieron una serie rápida para ver cómo el estacionamiento, las tendencias y las temporadas se desarrollaron, siguiendo el método Box y Jenkins de (1973) Después de eso, hicieron algunos gráficos interesantes que muestran cómo los puntos de datos originales se relacionan entre sí con el tiempo, tanto directa como indirectamente. Todo se trataba de descubrir si había pedazos y piezas autoprendiendo, y si hubo algún patrón promedio que apareció, si estaban atados a las estaciones o no (Box y Jenkins, 1973). Si la serie no se quedó quieta, verificaron la serie diferenciada de primer orden gráficamente (d = 1) para ver si podían detectar algún patrón (Box y Jenkins, 1973). Por lo tanto, crujieron los números para descubrir los bits no estacionales y estacionales de los datos, utilizando las funciones de autocorrelación y autocorrelación parcial, para ver si había algún patrón que se repitiera con el tiempo (Box y Jenkins el modelo que realmente se ajustan a los datos, eso es lo que llamaron el buen modelo (Box & Jenkins, 1973)..

Estimación del modelo

Estimación del modelo

Utilizaron Rstudio para ejecutar el método ARIMA Box y Jenkins para la estimación del modelo, basado en el trabajo original de Box y Jenkins de 1973, descubrieron los números y cuánto podrían variar, verificando si son sólidos con un nivel de confianza del 95% utilizando la prueba "T" de (Box & Jenkins, 1973).

Adecuación del modelo

En el estudio, se analizó la adecuación del modelo propuesto. Para esta evaluación, se tomaron en cuenta diversos indicadores, siguiendo la metodología establecida por Box y Jenkins en 1973.

1. Cuantificación del error de estimación del modelo.
2. Análisis de los residuales, que comprende:

2.1 En el estudio, se implementó la estadística de Ljung-Box, una modificación de la prueba de Box-Pierce, para evaluar los rezagos 12, 24, 36 y 48. La fórmula utilizada para este propósito fue la presentada por Box y Jenkins en 1973.

$$Q_{LB} = T(T+2)\sum_{J=1}^{K} \frac{r_j^2}{(T-j)}$$

T: "Longitud de la serie de tiempo".

K: 2Número de rezagos que se analizan (k = 12, 24, 36, 48)".

rj : "Coeficientes de autocorrelación".

2.2 Evaluación de la Función de Autocorrelación (FAC) y de la Función de Autocorrelación Parcial (FACP) de los residuales. (Box & Jenkins, 1973).

2.3 En su análisis, se evaluó la normalidad de los residuales. Esto se llevó a cabo mediante la prueba de Kolmogorov-Smirnov y el uso de métodos gráficos. Entre estos últimos, se incluyeron el papel probabilístico, el histograma y la representación gráfica de los valores pronosticados en relación con los valores distribuidos en el tiempo, siguiendo la metodología propuesta por Box y Jenkins en 1973.

Pronóstico del modelo

Según el modelo estimado, se llevaron a cabo seis pronósticos de la variable en estudio. El análisis incluyó estimaciones interválicas con un nivel de confianza del 95%, las cuales fueron representadas gráficamente. (Box & Jenkins, 1973)

Validez del pronóstico

Revisaron las seis predicciones hechas por el modelo contra los datos reales del estudio que estos números, que coinciden con la predicción de seis meses, provienen de la base de datos.

En la investigación, verificaron cómo las predicciones estaban utilizando diferentes medidas, por lo que cuando usaron el enfoque Box y Jenkins, descubrieron los coeficientes de información AKIKE (AIC), porque esos son los que funcionan con ese método.

2.6 Procesamiento y Análisis de Datos

Para el procesamiento de los datos se utilizó el software R, con una base de datos compuesta por 60 observaciones. De estas, 54 datos fueron utilizados para la estimación y ajuste del modelo, mientras que los 6 datos restantes se destinaron a la validación y comparación del pronóstico.

Se aplicó la metodología Box-Jenkins, que incluye las etapas de identificación, estimación y validación del modelo ARIMA. Posteriormente, se evaluó la precisión del modelo comparando los valores pronosticados con los datos reales, lo que permitió determinar el nivel de ajuste y la confiabilidad del pronóstico obtenido.

“

III. RESULTADOS

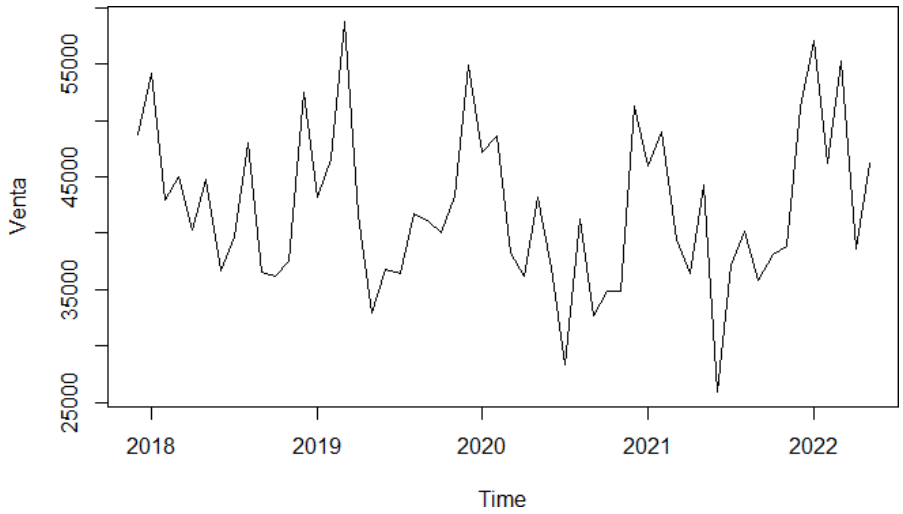
3.1 Metodología Box Jenckis

3.1.1 Paso 1: Modelo a identificar.

- Gráfico de la serie para observar su estacionamiento, su tendencia y estacionalidad.

Figura 1

La empresa Mercantil Inca S.A. en Chiclayo cerró sus ventas en diciembre de 2017 y mayo de 2022.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

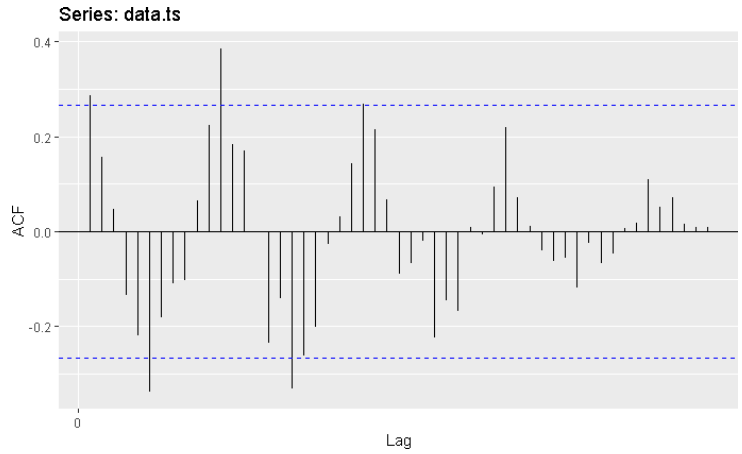
En el período comprendido entre diciembre de 2017 y mayo de 2019, la serie exhibió una tendencia ascendente, con un aumento de 48 684.76 cajas físicas a 46 206.28 cajas físicas. El análisis gráfico reveló que la serie no era estacionaria ni en su media ni en su varianza.

A continuación, se procedió a examinar la posible presencia de componentes autorregresivos y de media móvil en la serie.

- Gráfico de función de autocorrelación simple

Figura 2

Comportamiento de repetición de los valores de la serie original a lo largo del tiempo.

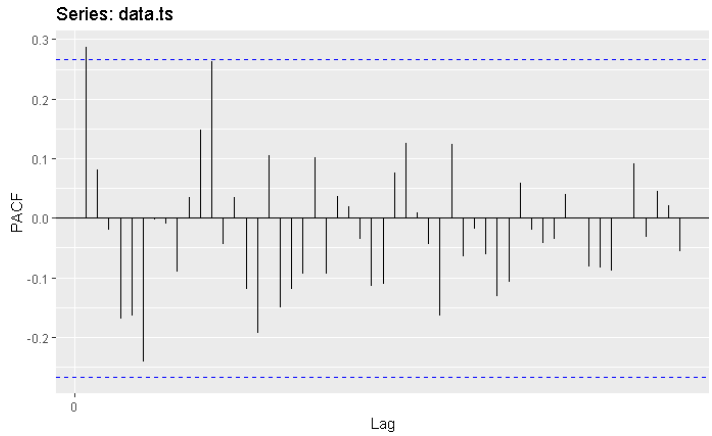


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

Dado que el decaimiento era lento, se presumió que la serie no era estacionaria y que el modelo incluía un componente de media móvil.

Figura 3

Relación directa entre valores actuales y pasados sin interferencia de otros.



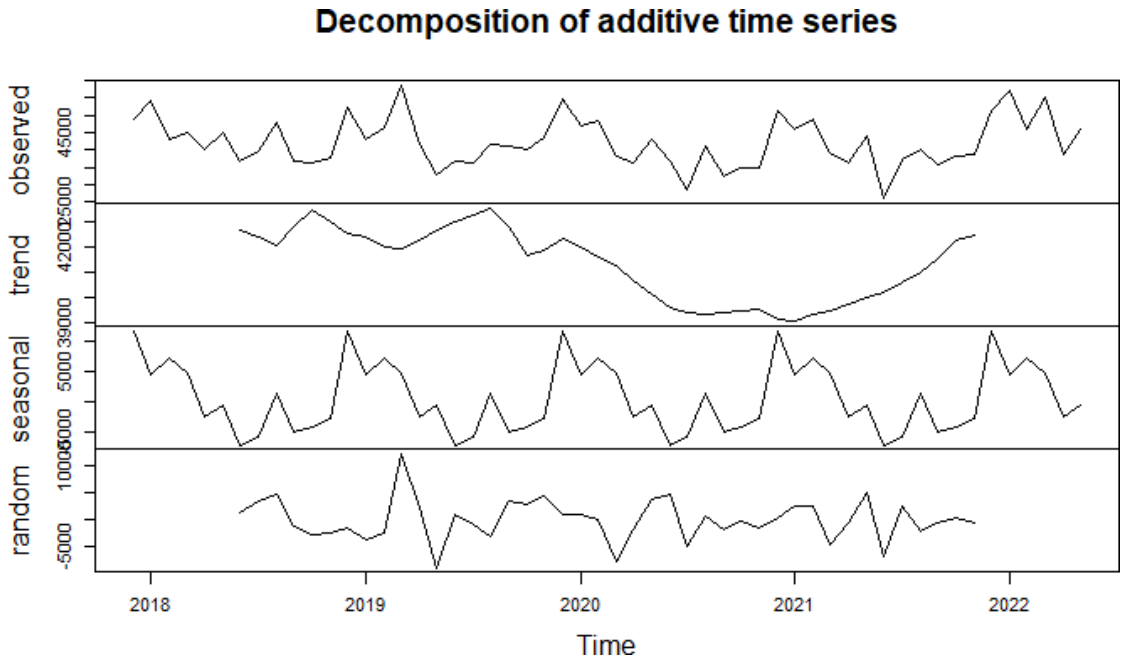
Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

El gráfico de la función de autocorrelación parcial en realidad no nos dio una señal clara si el modelo tuviera piezas autorregresivas o no, parece que el modelo ARIMA estacionario podría ser una combinación de algunas tendencias a largo plazo y movimientos a mediano plazo.

La serie se dividió en partes, cada una diseñada para ayudarnos a profundizar y comprender mejor.

Figura 4

Análisis de la serie de datos y sus partes: comportamiento constante, patrones que se repiten y dirección del cambio.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

El código R a continuación desglosa una serie a corto plazo en su estacionamiento, piezas de estacionamiento y tendencias Este método hace que sea más fácil cavar y obtener una imagen más clara de los patrones en la serie de datos a corto plazo

Plot (Descompos (data.ts))

La Figura 4 muestra que debemos deshacernos de esa tendencia ascendente para que los datos también establezcan, insinúa que el programa podría necesitar partes que cambien con las estaciones.

3.1.2 Paso 2: Determinamos la estacionariedad o estacionalidad

Usaron un modelo SARIMA para esto, la secuencia basada en el tiempo se trata de hacer que el estacionamiento funcione mejor para descubrir cuántas veces necesitamos diferenciar, estamos revisando las funciones NDIFFS y NSDIFFS en R. Estas funciones descubren cuántos cambios regulares y estacionales se necesitan para hacer que la serie temporal estable.

CÓDIGOS:

data.ts: base de datos original

ndiffs (data.ts): Función en R para calcular diferenciaciones regulares.

nsdiffs (data.ts): Función en R para calcular diferenciaciones estacionales.

Número de diferenciaciones regulares.

Ndiffs = 0

El análisis arroja que la serie en cuestión alcanza la estacionariedad sin requerir una diferenciación regular repetida. En otras palabras, no es necesario diferenciar la serie más de una vez para lograr que sea estacionaria.

- Número de diferenciaciones estacionales.

  Nsdiffs = 1

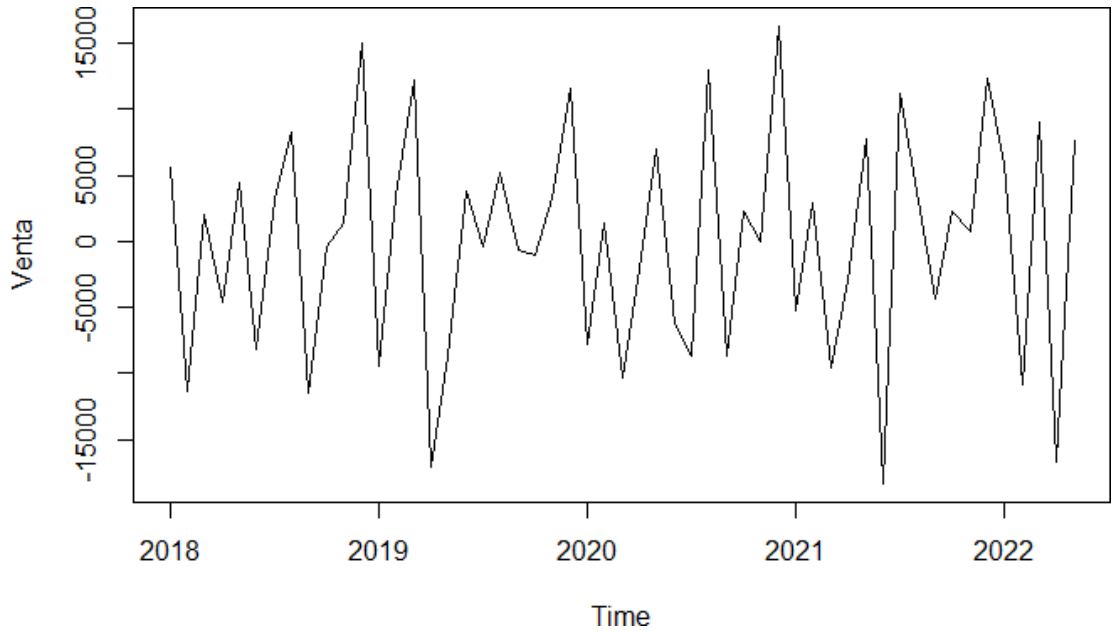
  Según los resultados de la función, la serie requiere una diferenciación estacional, además de la diferenciación regular.

- Gráfico de la serie diferenciada.

En el primer trimestre de 2023, las cifras de ventas han mostrado un aumento constante, con un pico notable en marzo. Desde noviembre de 2017 hasta mayo de 2022, notaron patrones que cambiaron con las estaciones, por lo que tuvieron que modificar un poco los datos para hacerlo más estable y consistente..

Figura 5

Serie original diferenciada.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Después de aplicar la primera diferenciación, la serie muestra un comportamiento estacional más estable en relación con su media. Ahora se observa que la serie ha eliminado su componente de estacionalidad. Una vez que esta componente se ha eliminado, se aplican las pruebas de estacionariedad pertinentes para verificar la hipótesis.

- **Prueba de estacionariedad.**

  Serie original: tsd

  ✓ **Dickey Fuller aumentada**

  Hipótesis:

  H0: La serie en primera diferencia sigue siendo inestable.

  Ha: La serie en primera diferencia ya es estable.

  Significancia:

  $\infty = 0.05$

  Solución:

```
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -11.17    0.01
[2,]   1  -7.32    0.01
[3,]   2  -5.21    0.01
[4,]   3  -4.48    0.01
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -11.07    0.01
[2,]   1  -7.25    0.01
[3,]   2  -5.16    0.01
[4,]   3  -4.43    0.01
Type 3: with drift and trend
     lag    ADF p.value
[1,]   0 -10.99    0.01
[2,]   1  -7.19    0.01
[3,]   2  -5.12    0.01
[4,]   3  -4.37    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
Phillips-Perron Unit Root Test
alternative: stationary
```

  Decisión:

  Rechazar Ho.

Conclusión:

El análisis revela que la serie diferenciada una vez exhibe estacionariedad. Esto se deduce del p-valor de 0.01, el cual lleva al rechazo de la hipótesis nula que postula la no estacionariedad.

- ✓ **Con la Prueba de Philip Perron:**

Hipótesis:

H0: La serie en primera diferencia no es estacionaria.

Ha: La serie en primera diferencia es estacionaria.

Significancia: $\infty = 0.05$

Solución:

```
Note: in fact, p.value = 0.01 means p.value <= 0.01
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
 lag Z_rho p.value
   3   -68    0.01
-----
 Type 2: with drift no trend
 lag Z_rho p.value
   3 -68.1    0.01
-----
 Type 3: with drift and trend
 lag Z_rho p.value
   3 -67.9    0.01
---------------
Note: p-value = 0.01 means p.value <= 0.01
```

Decisión: Rechazar H0.

Conclusión: La serie es estacionaria; los dos test confirmaron idea gráfica.

3.1.3 Paso 3: Gráficos de las funciones de Autocorrelación (ACF) y Autocorrelación parcial (PACF).

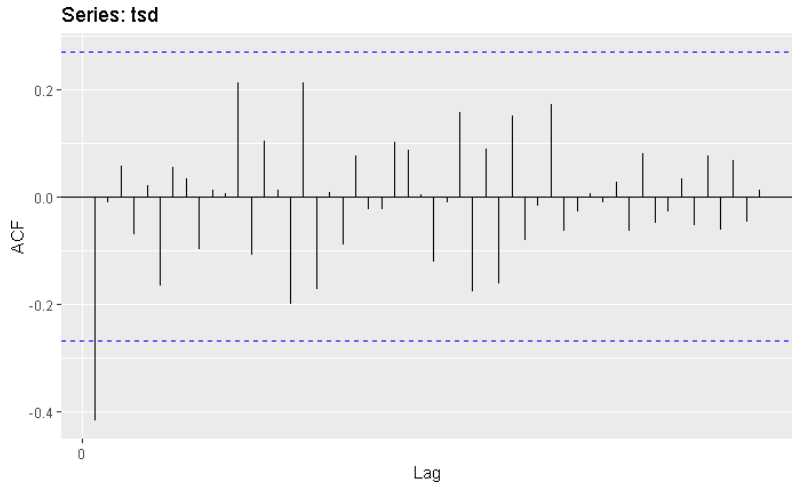
- **Gráfico de función de autocorrelación simple.**

En el período de enero de 2019 a diciembre de 2020, se obtuvieron los siguientes resultados: después de tomar la serie de datos de las ventas mensuales de bebidas gasificadas negras en Mercantil Inca S. de noviembre de 2014 a mayo de 2019, descubrieron los modelos adecuados para usar para hacer predicciones. Cuando se trata de este, están mirando la primera diferencia. Para averiguar el modelo, debe consultar las funciones de autocorrelación simples y parciales.

- **Gráfico de función de autocorrelación.**

Figura 6

Se evaluó si la serie diferenciada aún mantiene relación con sus datos anteriores.



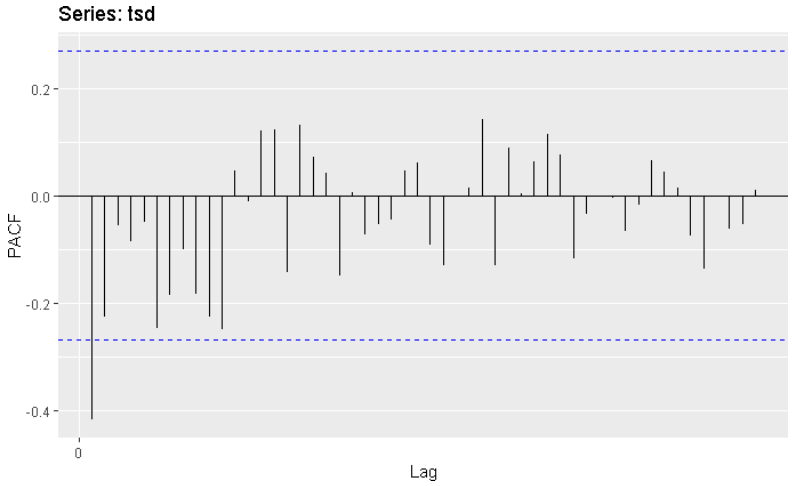
Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

Cuando estás revisando el gráfico de autocorrelación, es como si estuvieras apostando a que el modelo tiene bits del móvil promedio en él, puedes preparar algunos borradores con una pieza del rompecabezas.

- Gráfico de función de autocorrelación parcial.

Figura 7

Relación entre valores actuales y pasados tras quitar tendencias.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

Basándose en la evaluación del autocorrelograma parcial, el modelo bajo consideración presenta la posibilidad de contener componentes autorregresivos AR. Adicionalmente, se observa una sutil manifestación de comportamiento estacional en el modelo.

A partir del análisis de los correlogramas, se proponen los siguientes modelos de pronóstico para ajustar a la serie.

3.1.4 Paso 4: Determinación del modelo SARIMA(p,d,q)(P;D;Q)

Estimación de modelos SARIMA's

Tabla 1

Cálculo de los valores que definen el comportamiento de los modelos propuestos para la serie de datos.

Modelo 1: SARIMA (1,1,0)(0,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-0.45321	0.14069	-3.2214	0.001275
Modelo 2: SARIMA (1,1,0)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-0.54417	0.13412	-4.0573	4.96E-05
sar1	-0.52589	0.1423	-3.6956	0.0002194
Modelo 3: SARIMA (1,1,1)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ma1	-0.88394	0.1018	-8.6835	< 2.2e-16
sar1	-0.48514	0.15251	-3.181	0.001468
Modelo 4: SARIMA (0,1,1)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ma1	-0.88398	0.1018	-8.6838	< 2.2e-16
sar1	-0.48512	0.15259	-3.1793	0.001476
Modelo 5: SARIMA (0,1,0)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
sar1	-0.41982	0.16839	-2.4931	0.01266
Modelo 6: SARIMA (0,0,0)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
sar1	-0.48121	0.15009	-3.2061	0.001345
Modelo 7: SARIMA (1,0,1)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
sar1	-0.48121	0.15002	-3.2077	0.001338
Modelo 8: SARIMA (1,1,0)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-0.44658	0.12324	-3.6236	0.0002905
sar1	0.33564	0.14251	2.3552	0.0185124
Modelo 9: SARIMA (1,1,1)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ma1	-1	0.056785	-17.6101	< 2.2e-16
sar1	0.526126	0.125975	4.1764	2.96E-05
Modelo 10: SARIMA (1,1,1)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ma1	-1	0.056786	-17.6101	< 2.2e-16
sar1	0.526123	0.125909	4.1786	2.93E-05

Modelo 11: SARIMA (0,1,0)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
sar1	0.29626	0.15154	1.9549	0.05059
Modelo 12: SARIMA (0,0,0)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
sar1	5.06E-01	1.25E-01	4.0337	5.49E-05
intercept	4.23E+04	1.39E+03	30.3683	< 2.2e-16
Modelo 13: SARIMA (2,1,0)(1,0,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-5.87E-01	1.32E-01	-4.4412	8.95E-06
ar2	-3.26E-01	1.41E-01	-2.3203	0.020324
sar1	4.26E-01	1.41E-01	3.0272	0.002468
Modelo 14: SARIMA (2,1,0)(0,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-0.63959	0.1416	-4.5169	6.28E-06
ar2	-0.44325	0.14546	-3.05E+00	0.00231
Modelo 15: SARIMA (2,1,0)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ar1	-0.79908	0.14203	-5.626	1.85E-08
ar2	-0.48616	0.14806	-3.2836	0.001025
sar1	-0.57226	0.13629	-4.1989	2.68E-05
Modelo 16: SARIMA (2,1,1)(1,1,0)				
	Estimate	Std. Error	z value	Pr(>|z|)
ma1	-0.88394	0.1018	-8.6835	< 2.2e-16
sar1	-0.48514	0.15251	-3.181	0.001468

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

Una vez que se obtienen todos los modelos posibles que cumplen con el criterio de significancia, con el fin de seleccionar el mejor modelo se utilizará el criterio AKAIKE (AIC) en cada modelo estimado, el modelo con mejor modelo SARIMA para la serie IPC y también observar los errores para luego validarlos en la serie D. El análisis descarta ciertos modelos, ya que estos no presentan parámetros significativos, tal como se detalla en el (Anexo 2).

Tabla 2

Modelos SARIMA.

Modelos	ME	RMSE	MAE	MPE	MAPE	AIC
Modelo 1: SARIMA (1,1,0)(0,1,0)	-22.08281	7700.934	4864.749	-1.332635	12.11602	865.700468561923
Modelo 2: SARIMA (1,1,0)(1,1,0)	14.23845	6550.031	4350.57	-1.249435	10.75209	858.432286131180
Modelo 3: SARIMA (1,1,1)(1,1,0)	263.8783	5538.194	3607.925	-0.6629666	8.83698	845.386722640465
Modelo 4: SARIMA (0,1,1)(1,1,0)	263.8872	5538.181	3607.934	-0.6629579	8.837016	845.386722741976
Modelo 5: SARIMA (0,1,0)(1,1,0)	-56.78498	7895.793	5562.985	-1.382498	13.90564	869.846196708100
Modelo 6: SARIMA (0,0,0)(1,1,0)	-315.8568	5433.28	3690.64	-2.198958	9.131034	859.331341078776
Modelo 7: SARIMA (1,0,1)(1,1,0)	-315.8569	5433.28	3690.64	-2.198958	9.131034	859.331341078819
Modelo 8: SARIMA (1,1,0)(1,0,0)	-69.98622	7139.801	5568.834	-2.125574	13.58714	1099.639507577460
Modelo 9: SARIMA (1,1,1)(1,0,0)	-733.8536	6083.943	4687.88	-3.880776	11.64788	1087.804790323020
Modelo 10: SARIMA (1,1,1)(1,0,0)	-733.8553	6083.945	4687.883	-3.880783	11.64789	1087.804790322460
Modelo 11: SARIMA (0,1,0)(1,0,0)	-23.98595	8008.083	6490.696	-1.915758	15.84592	1109.250377997040
Modelo 12: SARIMA (0,0,0)(1,0,0)	-31.62971	6101.702	4681.023	-2.342503	11.48425	1104.156089705000
Modelo 13: SARIMA (2,1,0)(1,0,0)	-21.72377	6733.245	5244.385	-2.001856	12.82939	1096.615030707360
Modelo 14: SARIMA (2,1,0)(0,1,0)	132.9544	6936.059	4531.675	-0.9035839	11.3076	859.548774534310
Modelo 15: SARIMA (2,1,0)(1,1,0)	208.2225	5737.197	3901.082	-0.6143145	9.54616	850.957689387551

3.1.5 Paso 5: Validación de los posibles modelos

En esta parte, están verificando si los desechos de los modelos que mostraron los números importantes son normales, independientes y siguen un cierto patrón.

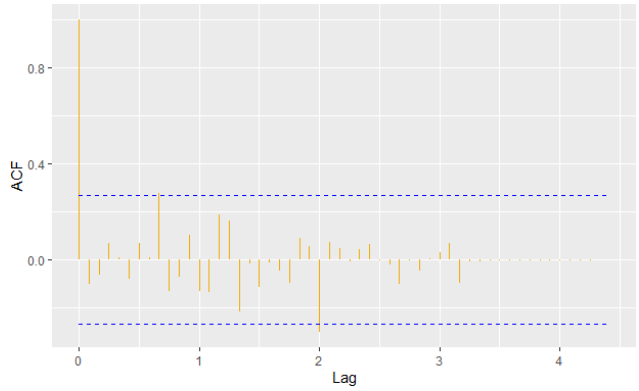
A. MODELO SARIMA

(1) MODELO (1): SARIMA (1,1,1)(1,1,0)

Análisis de los residuales

Figura 8

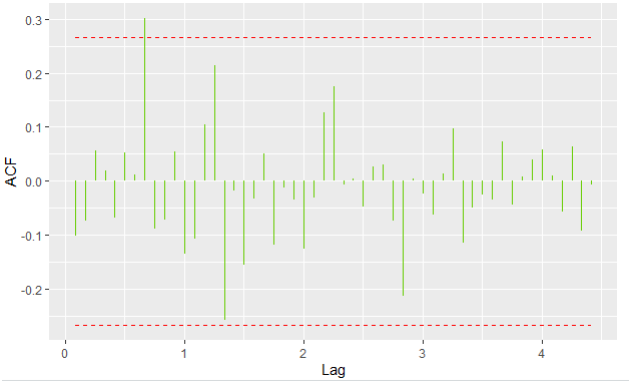
Función de autocorrelación simple de los residuos del modelo SARIMA (1,1,1) (1,1,0).



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Figura 9

La autocorrelación parcial de los errores sobrantes del modelo Sarima (1,1,1) (1,1,0).

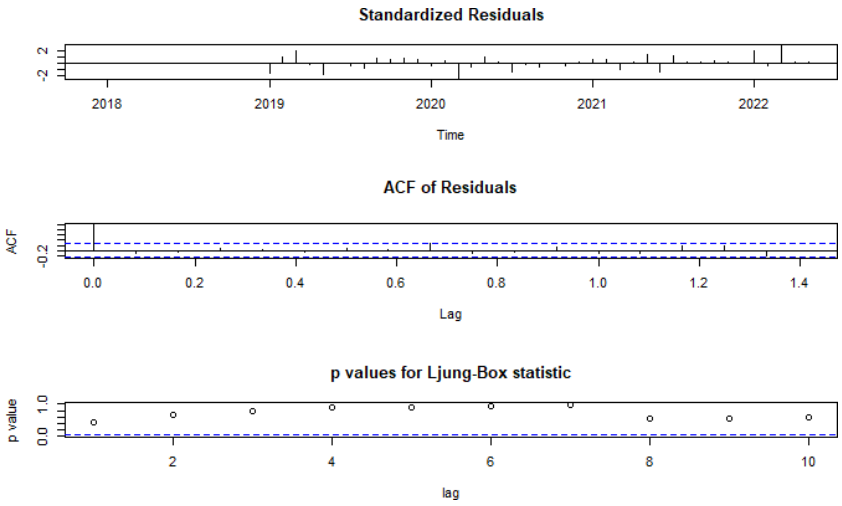


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En la propuesta del modelo, la mayoría de los coeficientes de autocorrelación de los residuos se sitúan dentro de las bandas de confianza. Existe un único valor atípico que no sigue un patrón discernible. Por lo tanto, se considera que los residuos obtenidos presentan un comportamiento similar al de un ruido blanco.

Figura 10

Evaluación de los errores del modelo SARIMA (1,1,1)(1,1,0) para comprobar su ajuste.

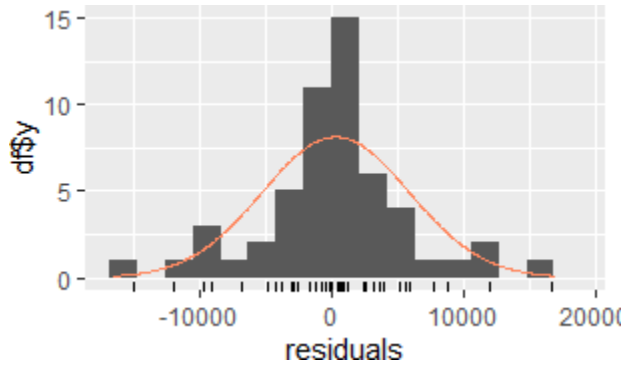


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En la Figura No. 10, los residuos estandarizados se extienden uniformemente alrededor de la media. Los coeficientes de autocorrelación de los residuos se mantienen dentro del rango de confianza del 95%, pero tampoco muestran ningún patrón de repetición, las estadísticas de la caja de Ljung-Box muestran que las cosas sobrantes no están realmente haciendo una diferencia, parece que las sobras están por todas partes y no siguen ningún patrón, lo que significa que solo están al azar..

Figura 11

Histograma de la distribución de los errores del modelo SARIMA (1,1,1)(1,1,0).



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa

La forma en que se extienden los desechos generalmente sigue un patrón bastante normal, esto está claro de cómo se alinean los puntos en el siguiente Figura 12.

Figura 12

Comprobación de si los errores del modelo ARIMA (2,1,2) siguen un patrón lineal.

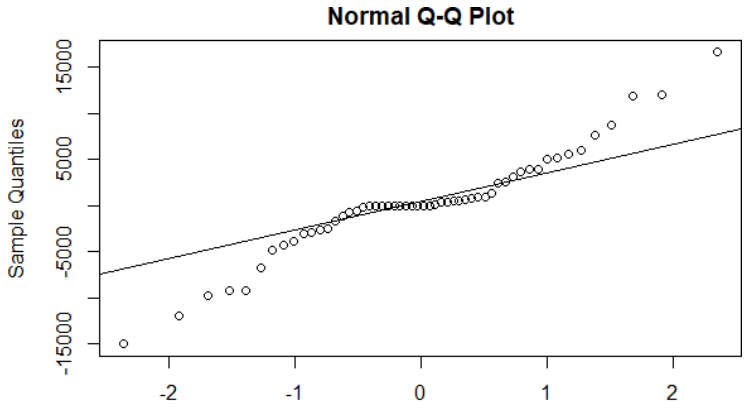


figura: Grafico Q para evaluar normalidad

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

El gráfico se genera a partir de los valores del modelo, los cuales se obtienen mediante la distribución normal estándar. Se asume un comportamiento normal basándose en la observación de un comportamiento lineal.

PRUEBAS DE HIPOTESIS DE LOS TEST's

Test de Ljung-Box

H0: Los residuales se distribuyen de forma independiente (El texto describe una situación donde no existen correlaciones en la población de la que se extrae una muestra. Por lo tanto, cualquier correlación que se observe en los datos de la muestra es atribuible únicamente al azar inherente al proceso de muestreo)

Ha: Los residuales no se distribuyen de forma independiente

Test de kolmogorov smirnov.

H0: Los residuales proceden de una distribución normal.

H1: Los residuales no proceden de una distribución normal.

Tabla 3

Resultados de las pruebas para verificar si los errores del modelo SARIMA (1,1,1)(1,1,0) son aleatorios, no están relacionados entre sí y tienen distribución normal.

SARIMA (1,1,1,)(1,1,0)	NE	P-Valor	Decisión	Conclusión
Test Ljung-Box	9.1106	0.3331	No rechazamos H0.	Los residuales se distribuyen de forma independiente.
Test Kolmogorov Smirnov	0.15707	0.1249	Rechazamos H0.	Los residuos proceden de una distribución normal.

Nota: Obtuvimos algunos resultados interesantes utilizando el programa R y la base de datos original de la compañía.

El modelo que se les ocurrió hace algunos números sobrantes que deben verificarse con esas pruebas de normalidad, Ljung-Box y Kolmogorov-Smirnov. Los resultados muestran que los

desechos generalmente se extienden de manera normal, con un 95% que cae dentro de un cierto rango. Además, los bits sobrantes se extienden al azar, no se afectan entre sí después de mirar todo, el modelo está mejor de lo que esperábamos en este punto.

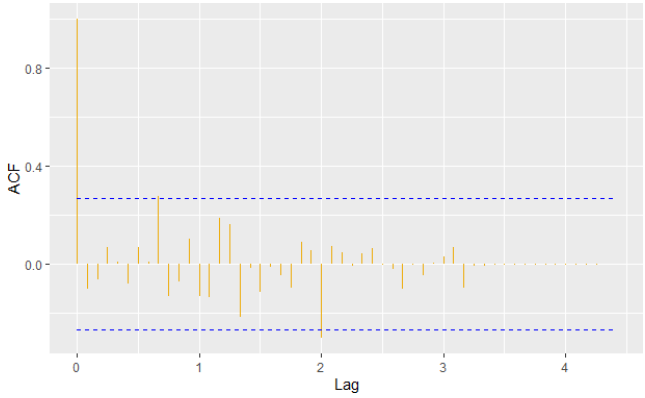
B. MODELOS SARIMA'S

(2) MODELO (4): SARIMA (0,1,1) (1,1,0)

Análisis de los residuales

Figura 13

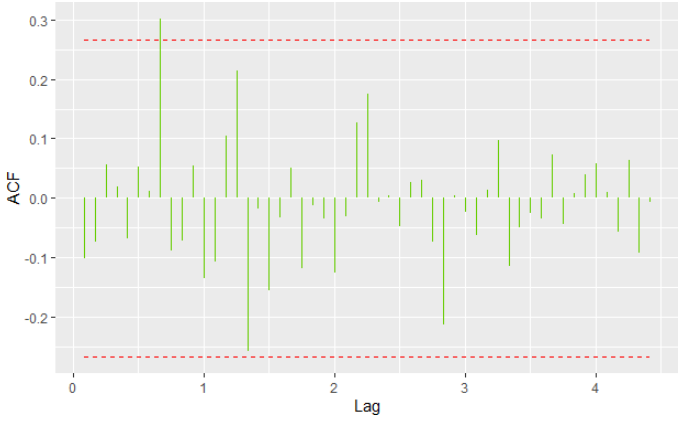
Análisis de la relación entre los errores del modelo SARIMA (0,1,1)(1,1,0) a distintos periodos de tiempo.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Figura 14

Estudio de la relación directa entre los errores del modelo SARIMA (0,1,1)(1,1,0) con sus propios valores pasados, excluyendo influencias intermedias.

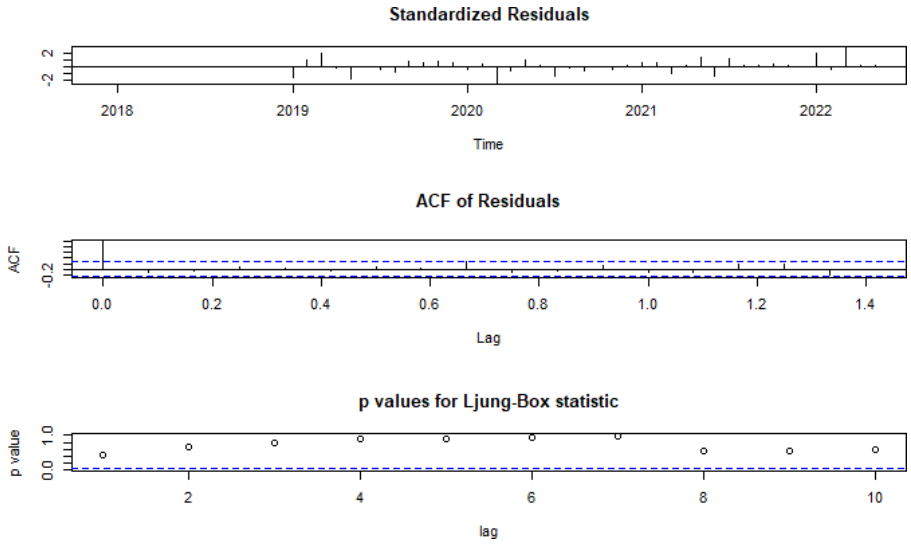


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En la propuesta del modelo, la mayoría de los coeficientes de autocorrelación de los residuales se encuentran dentro de las bandas de confiabilidad. Existe una excepción con un coeficiente que está fuera de estas bandas, pero la distribución general se considera aceptable. Por lo tanto, se asume que los residuos obtenidos presentan un comportamiento similar al de un ruido blanco.

Figura 15

Evaluación de los errores del modelo SARIMA (0,1,1)(1,1,0) para comprobar su buen ajuste.

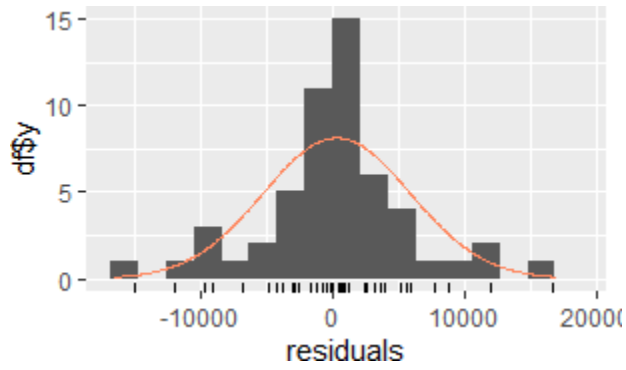


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En la Figura N°. 15, los residuos se extienden uniformemente alrededor de la media. A pesar de que algunos de los números de autocorrelación de residuos no están en la zona de confianza del 95%, eso no significa que debamos descartar el modelo. Pero cuando verificaron los residuos para cualquier patrón, la mayoría de las veces no encontraron nada significativo. Básicamente, a las cosas sobrantes no les importa lo que vino antes y simplemente actúa al azar.

Figura 16

Histograma que muestra cómo se distribuyen los errores del modelo SARIMA (0,1,1)(1,1,0).



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

La forma en que se extienden los desechos generalmente sigue un patrón normal. Los puntos en la figura adjunta se extienden en línea recta, mostrando que es seguro.

Figura 17

Verificación de si los errores del modelo SARIMA (0,1,1)(1,1,0) siguen un patrón recto o consistente.

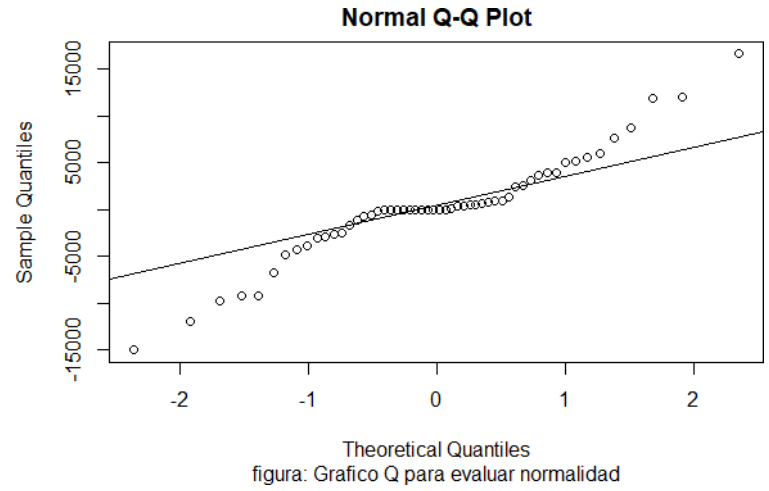


figura: Grafico Q para evaluar normalidad

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

La gráfica se genera a partir de los valores del modelo, los cuales se obtienen mediante la distribución normal estándar. Se asume un comportamiento normal, sustentado en la observación de un comportamiento lineal.

PRUEBAS DE HIPOTESIS DE LOS TEST's

Test de Ljung-Box

H0: Los residuales se distribuyen de forma independiente

Ha: Los residuales no se distribuyen de forma independiente

Test de kolmogorov smirnov.

H0: Los residuales proceden de una distribución normal.

H1: Los residuales no proceden de una distribución normal.

Tabla 4

Resultados de las pruebas que determinan si los errores del modelo SARIMA (0,1,1)(1,1,0) son aleatorios, no están relacionados entre sí y tienen una distribución normal.

SARIMA (0,1,0)(1,0,1)	NE	P-Valor	Decisión	Conclusión
Test Ljung-Box	9.1097	0.4272	No rechazamos H0.	Los residuos se distribuyen de forma independiente.
Test Kolmogorov Smirnov	0.15707	0.1249	No rechazamos H0.	Los residuos proceden de una distribución normal.

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En el modelo en el que están pensando, usan las pruebas Ljung-Box y Kolmogorov-Smirnov para verificar si los desechos siguen un patrón normal. Los resultados muestran que la basura generalmente se extiende de una manera que podemos estar 95% seguros de que es normal, también, en este modelo, los bits sobrantes se extienden al azar, como si no estén conectados de

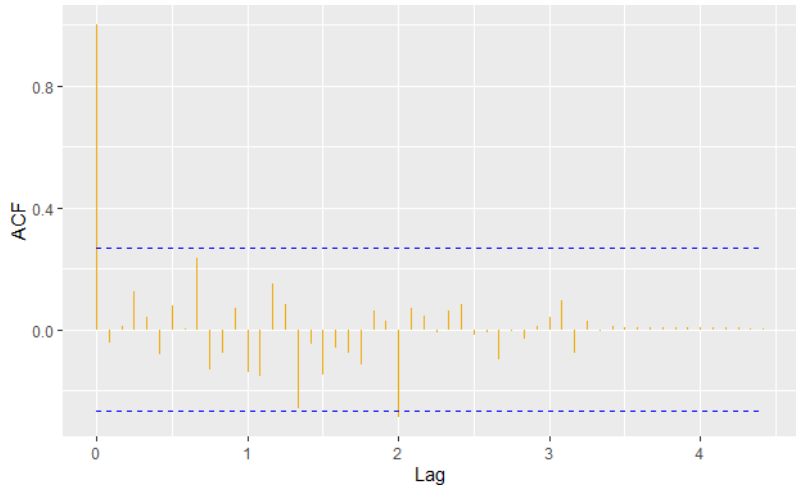
ninguna manera. Entonces, después de mirar todo, el modelo está mejor de lo que pensábamos en este momento.

(3) MODELO (7): SARIMA (1,0,1) (1,1,0)

Análisis de los residuales

Figura 18

La función de autocorrelación simple de los residuos del modelo Sarima (0.1.0) (2,0,2).



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Figura 19

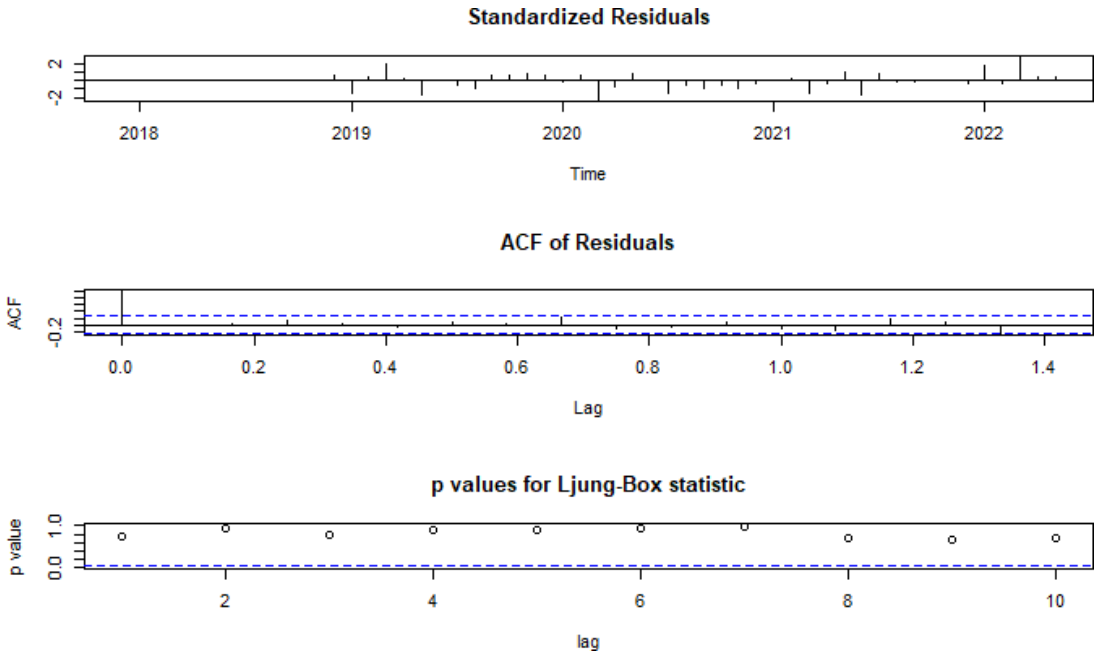
Análisis de la relación directa entre los errores actuales y pasados del modelo SARIMA (0,1,0)(2,0,2), eliminando influencias intermedias.

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En el modelo propuesto, la mayoría de los coeficientes de autocorrelación de los residuos se sitúan dentro de las bandas de confianza. Aunque hay un coeficiente fuera de las bandas en la FAC y otro en la PFAC, la distribución general se considera aceptable. Por lo tanto, se puede concluir que los residuos obtenidos se comportan como ruido blanco.

Figura 20

Evaluación de los errores del modelo SARIMA (0,1,0)(2,0,2) para comprobar su comportamiento y validez.

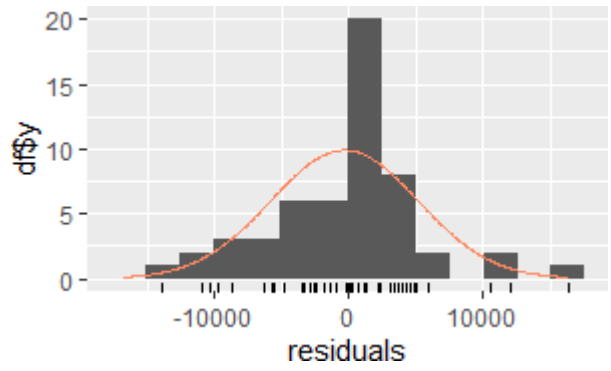


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

En la Figura N°. 20, los residuos se extienden uniformemente alrededor de la media. Pero bueno, no cada vez que verificamos las estadísticas de desechos, caen dentro del rango habitual que esperamos. Aunque ese es el caso, la prueba de Ljung-Box muestra que la mayoría de las cosas de autocorrelación sobrantes no son realmente significativas. Entonces, resulta que las cosas sobrantes (residuos) no siguen ningún patrón y solo actúan por sí solo.

Figura 21

Histograma que representa cómo se distribuyen los errores del modelo SARIMA (0,1,0)(2,0,2).



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Se observa que la distribución de los residuos no tiende a seguir un comportamiento normal. Esto se constata con la distribución lineal de los puntos en la figura siguiente.

Figura 22

Comprobación de si los errores del modelo SARIMA (0,1,0)(2,0,2) siguen un comportamiento recto y coherente.

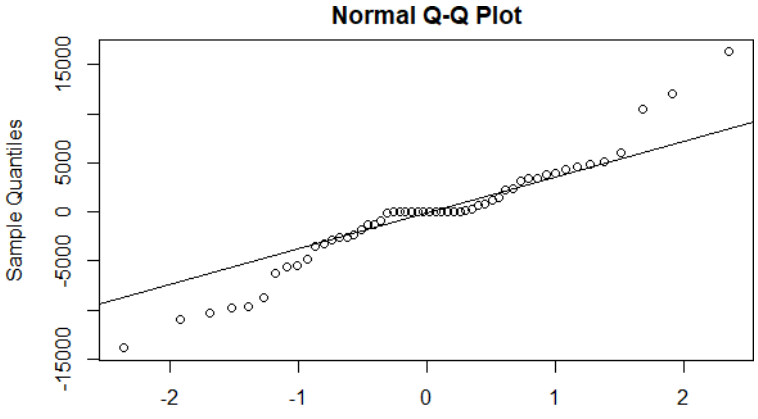


figura 09: Grafico Q para evaluar normalidad

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

La gráfica se genera a partir de los valores del modelo, los cuales se obtienen mediante la distribución normal estándar. Se asume un comportamiento normal debido a la observación de un comportamiento lineal.

Tabla 5
Resultados del test de ruido blanco, independencia y Normalidad de los residuales del modelo propuesto SARIMA(1,0,1)(1,1,0).

SARIMA (0,1,0)(2,1,0)	NE	P-Valor	Decisión	Conclusión
Test Ljung-Box	7.4249	0.4916	No rechazamos H0.	Los residuos se distribuyen de forma independiente.
Test Kolmogorov Smirnov	0.10952	0.7407	Rechazamos H0.	Los residuos proceden de una distribución normal al 5%

Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

Verificaron si los desechos del modelo que se les ocurrió siguieron una distribución normal usando la caja de Ljung y Kolmogorov-Smirnov, los resultados muestran que los desechos generalmente se extienden de una manera que la mayor parte cae dentro de un cierto rango, como el 95% del tiempo. Además, también cumplen con la condición de ser distribuido de forma independiente. Entonces, estamos pasando a la siguiente parte del proceso.

Luego muestran las predicciones de los modelos que pasaron el análisis de la prueba de normalidad.

”

3.1.6 Paso 6: Pronóstico del modelo.

Pronósticos de modelos SARIMA's.

En esta etapa, solo los modelos que satisfacen todos los requisitos de la fase de adecuación avanzan. Únicamente se consideran los modelos con estructura SARIMA, excluyendo los modelos ARIMA, dado que los modelos de prueba de esta última estructura no cumplieron con los criterios establecidos.

De los modelos SARIMA's propuestos los modelos: 3 – SARIMA (1,1,1) (1,1,0), 4 – SARIMA (0,1,1) (1,1,0) y 5 - SARIMA (1,0,1) (1,1,0), Se cumplen los requisitos establecidos. Los residuales se distribuyen como un ruido blanco, son independientes y siguen una distribución normal. Ahora, se procede a realizar los pronósticos respectivos para los próximos seis meses (Mayo 2019 – Noviembre de 2022).

Tabla 6

Pronóstico de las ventas de colas negras de una empresa distribuidora de bebidas en Chiclayo 2022.

	VENTA REAL	SARIMA (3) (1,1,1) (1,1,0)	SARIMA (4) (0,1,1) (1,1,0)	SARIMA (7) (1,0,1) (1,1,0)
Jun-22	32,143	33,091	33,090	31,255
Jul-22	36,069	34,632	34,632	32,874
Ago-22	39,828	42,507	42,507	40,710
Set-22	35,661	36,071	36,070	34,291
Oct-22	41,657	38,324	38,324	36,544
Nov-22	39,255	38,708	38,707	36,930

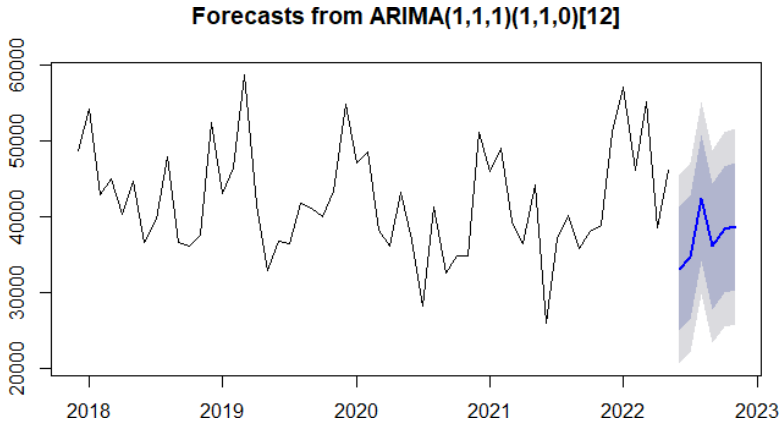
Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

De la Tabla N° 6 Se puede afirmar que las predicciones generadas por el modelo SARIMA (4) exhiben una mayor proximidad a los valores reales en comparación con el otro modelo evaluado. Por lo tanto, se considera, hasta el momento, como un modelo de pronóstico tentativo adecuado. El color verde en la representación visual denota el modelo de pronóstico que, en cada fila, demuestra la mayor cercanía al valor real u original de la serie temporal.

(1) SARIMA 3: (0,1,1) (1,1,0)

Figura 23

Valores pronosticados del modelo SARIMA (1,1,1)(1,1,0) de la venta mensual de colas negras en una distribuidora entre los meses Diciembre 2017 – Noviembre 2022.

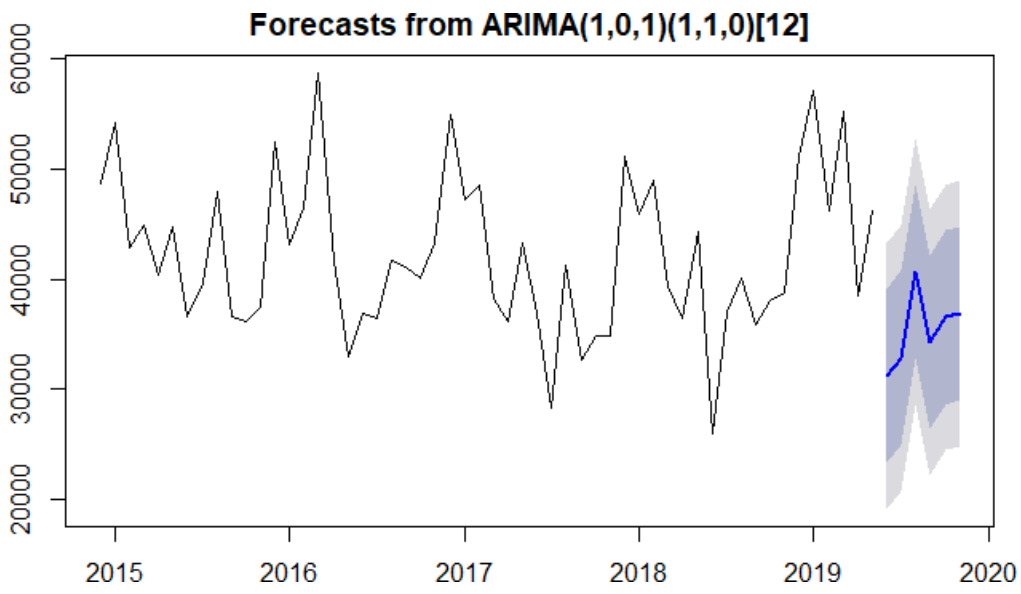


Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

(3) SARIMA 7: (1,0,1) (1,1,0)

Figura 24

Valores pronosticados del modelo SARIMA (1,0,1)(1,1,0) de la venta mensual de colas negras en una distribuidora entre los meses Diciembre 2017 – Noviembre 2022.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

“

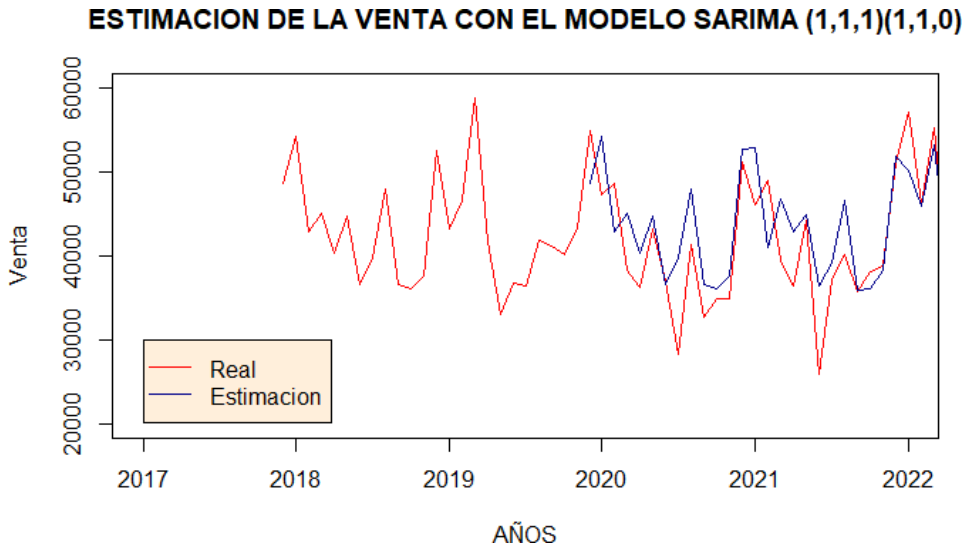
3.2 Modelo SARIMA ganador de la metodología Box Jenkins.

(2) MODELO 3: SARIMA (1,1,1) (1,1,0) vs Serie original

$$IPC_t = -0.88398 IPC_z(t-1) - 0.48512 IPC_z(t-12) + z_t$$

Figura 25

Comparación de los valores estimados y pronosticados del modelo N° 3: SARIMA (1,1,1) (1,1,0) de la venta mensual de colas negras en una distribuidora entre los meses diciembre 2017 – noviembre 2022.



Nota: Los resultados se obtuvieron utilizando el programa R y la base de datos original proporcionada por la empresa.

”

IV. DISCUSIÓN DE RESULTADOS.

Se realizo una serie de tiempo de la venta de colas negras de una empresa distribuidora; durante el periodo diciembre 2017 a noviembre 2022

- Al analizar la serie de datos de las ventas de colas negras de la empresa distribuidora se pudo comprobar que uno de los pasos más importantes en la selección del modelo para pronosticar uno adecuado a la realidad, con los datos de la serie de tiempo es considerar los diferentes modelos de datos, pudiendo determinar los 4 tipos generales: horizontal, tendencias, estacionales y cíclicos (Peralta Panduro, 2021)

- Para el pronóstico univariado, se utilizó la metodología de Box – Jenkins; con la serie de la venta se estimó el modelo SARIMA [(1,1,1) (1,1,0)], previamente la serie tuvo que realizar una diferencial regular y estacional para lograr la estacionariedad; la serie presento todos los coeficientes de regresión significativos ($p<0.01$), residuos estacionarios y normales, en cuanto a los errores de pronóstico presenta error absoluta media (MAE = 3607.934 cajas físicas), error cuadrático medio equivalente (RMSE = 5538.181 CF), porcentaje de error absoluto (MAPE = 8.84%), sin embargo en la evaluación de métodos para estimar la demanda de la línea de productos Ajinomoto en la ciudad de Chiclayo, en donde se compararon 3 tipos de herramientas para pronosticar; método de la empresa, Winter y descomposición estacional, donde error absoluta media (MAE = 0.73 gramos), error cuadrático medio equivalente (RMSE = 1.19 gramos), porcentaje de error absoluto (MAPE = 11.68%) un porcentaje muy por encima del encontrado por nuestro valor estimado.

- Al realizar el análisis para tomar la decisión del modelo que utilizaremos para realizar el pronostico no solo tenemos en cuenta los indicadores MAE, MAPE, RMSE; si no también el indicador AKAINE (AIC), que como en el resto de indicadores también se define porque cual de todos los modelos tenga el menor número, no establecemos que las investigaciones de los antecedentes no cumplan con este ultimo indicador, pero con lleva a que la investigación sea mas robusta.

V. CONCLUSIONES

- Se analizaron los datos históricos de ventas de bebidas gasificadas negras correspondientes al periodo diciembre 2017 – noviembre 2022, identificándose tendencia y comportamiento estacional en la serie temporal.
- Mediante la metodología Box-Jenkins y el análisis de autocorrelación (ACF y PACF), se identificó el modelo SARIMA (1,1,1)(1,1,0) como el más adecuado para representar el comportamiento de la demanda.
- El modelo fue validado mediante pruebas de estacionariedad, significancia estadística y Akaike, obteniendo resultados confiables y un AIC de 845.386722640465 considerado aceptable para pronósticos.

El modelo SARIMA permitió generar pronósticos ajustados de la demanda, contribuyendo a mejorar la planificación y toma de decisiones en la empresa Mercantil Inca S.A. con el modelo:

$$IPC_t = -0.88398 IPC_z(t-1) - 0.48512 IPC_z(t-12) + z_t$$

RECOMENDACIONES

- A la empresa distribuidora de bebidas colas negras se recomienda utilizar el modelo SARIMA (1,1,1)(1,1,0)[12] para pronosticar las ventas mensuales, retroalimentando la serie y evaluando periódicamente el modelo.
- Para futuras investigaciones con la METODOLOGIA BOX JENKINS se recomienda evaluar series con datos que presentan diferentes patrones en su comportamiento (Estacionalidad, Estacionariedad, Ciclos, etc.) y verificar la influencia de utilizar más datos en el entrenamiento y predicción para una mejor toma de decisiones.

Referencias Bibliográficas

Antunez Irgoin, C. H. (Febrero de 2011). *eumed.net*. Obtenido de http://www.eumed.net/ce/2011a/chai.htm

Ballesteros, A. (s.f.). *Neural Networks Framework*. Obtenido de http://www.redes-neuronales.com.es/tutorial-redes-neuronales/tutorial-redes.htm

Box, G., & Jenkins, G. (1973). *Metodología Box Jenkins.*

BVL Bolsa de Valores de Lima. (26 de abril de 2018). *Bolsa de Valores de Lima*. Obtenido de http://www.bvl.com.pe/eduayuda_glosario.html

CICA. (s.f.). *Centro Informático Científico de Andalucía.* Obtenido de https://thales.cica.es/rd/Recursos/rd98/TecInfo/07/capitulo3.html

Coghlan, A. (2014). *A Little Book of R For Time Series.* Cambridge.

Colina, E., & Rivas, F. (1998). *Itroduccion a la inteligencia artificial Mérida.* Venezuela.

Cryer, J. D., & Sick Chang, K. (2008). *Time Series Analysis.* New York: Springer Science.

De la Cruz, A., & Làzaro, A. (2015). *Análisis comparativo de Redes Neuronales Artificiales y la metodología de BOX-JENKINS en el pronotico de ídices de precios al consumirdor en la ciudad de Huaraz, periodo 2000 - 2014.* Huaraz.

debitoor. (2017). *DEBITOOR*. Obtenido de https://www.weforum.org/search?query=problemas+de+series+financieras

Econometría. (30 de Abril de 2015). *Econometría*. Obtenido de http://laeconometria.blogspot.com/2015/04/parsimonia.html?m=1

Estrategias de traiding. (26 de octubre de 2016). *Estrategias de traiding*. Obtenido de https://estrategiastrading.com/sobre-este-blog/

Faraway, J., & Chatfield, C. (1998). *Metodología de Redes Nuronales.* UK.

FINANCIALRED. (10 de Agosto de 2010). *FINANCIAL RED*. Obtenido de http://aprenderbolsa.com/cotizaciones-de-apertura-maxima-minima-y-de-cierre/

Fournies P., A. (2015). *Modelos ARIMA y Método Box y Jenkins.*

García, C. (2014). *Predicción del comportamiento de la rentabilidad de la acción de Isagen, a patir de la técnica de Redes Neuronales Artificiales.* Bogotá.

Gedesco. (29 de Agosto de 2012). *GEDESCO*. Obtenido de https://www.gedesco.es/blog/accion-bursatil-como-se-fija-y-como-varia-su-precio/

Gekko. (24 de Junio de 2017). *NEGOCIOS*. Obtenido de http://blablanegocios.com/sarima-que-es/

Gómez, W. (2010). *Pronostico de una serie temporal usando Redes Neuronales Artificiales.* Oaxaca.

Granados Flores, S. G. (2010). *Pronóstico de demanda de efectivo en cajeros automaticos aplicando Redes Neuronales Artificiales.* Lima.

Gvaladze, S. (2015). *Evaluación de métodos para pronóstico de series de tiempo; aplicado al consumo de energía predicciones para Hvaler.* Halden, Norway.

Haykin, S. (2009). *Neural networks and learning machines.* Nueva York: Prentice Hall/Pearson.

INGENIO EMPRESA. (7 de Marzo de 2016). *INGENIO EMPRESA*. Obtenido de https://ingenioempresa.com/medicion-error-pronostico/

Lewis, N. D. (2017). *Neuronal Networks for Time Series Forecasting with R.*

Melo Mayta, E. (2016). *Modelo de predicción mensual de mortalidad general intrahospitalaria en el hospital regional Manuel Nuñez Butrón - Puno, 2008 - 2016-I.* Puno.

Moncayo, L., Naranjo, J., & García, I. (2016). *Uso de Redes Neuronales Artificiales en la valoración económica de ruido de tráfico.* Quito: EuroRegio.

Nicole Roldan, P. (2015). *ECONOMIPEDIA*. Obtenido de http://economipedia.com/definiciones/bolsa-de-valores.html

Peralta Panduro, R. (2017). *Evaluación de métodos de pronóstico de serie de tiempo para estimar la demanda de la línea de producto ajinomoto en la ciudad de Chiclayo – periodo 2014 – 2016.* Chiclayo.

Pérez Ma, L., & Quintin, M. (2003). *aplicaciones de las redes neuronales artificiales a la estadística.* Madrid, España: La Muralla.

Pfaff, B. (2011). *Analysis of Integrated and Cointegrated time series with R.* Washington: Boltimore.

Ráez Guevara , L. R. (2012). *Metodología para la medición de la atención en una central telefónica usando Box-Jenkins.* Lima.

Rios, G. (2008). *Series de tiempo.* Chile.

Rodríguez, M., & Valdivia, H. (2012). *Modelo para pronóstico de precios de acciones en la bolsa de valores de Lima basado en Redes Neuronales Artificiales.* Lima.

Rubalcaba, J. (21 de Febrero de 2016). *Modelización y estadística en Biología*. Obtenido de https://jgrubalcaba.wordpress.com/2016/02/21/cosas-que-conviene-saber-al-usar-aic-dic-y-otros-criterios-de-informacion/

Saavedra Sánchez, F. (2018). *Bolsa de Valores de Lima y objetivos.* Lima.

Shumway, R. H., & Stoffer, D. S. (2015). *Time Series Analysis and Its Applications.* USA.

Sim, N. (1999). *Planeación de la información y control de inventarios.* North-Holland.

SOPORTE DE MINITAB 18. (Enero de 2019). *SOPORTE DE MINITAB 18*. Obtenido de https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/basics/forecasting/

Universidad Iberoamericana. (29 de Enero de 2018). *Composi.info*. Obtenido de http://composi.info/universidad-iberoamericana.html?page=16#5.5.2_Error_cuadr%C3%A1tico_medio_(MSD)

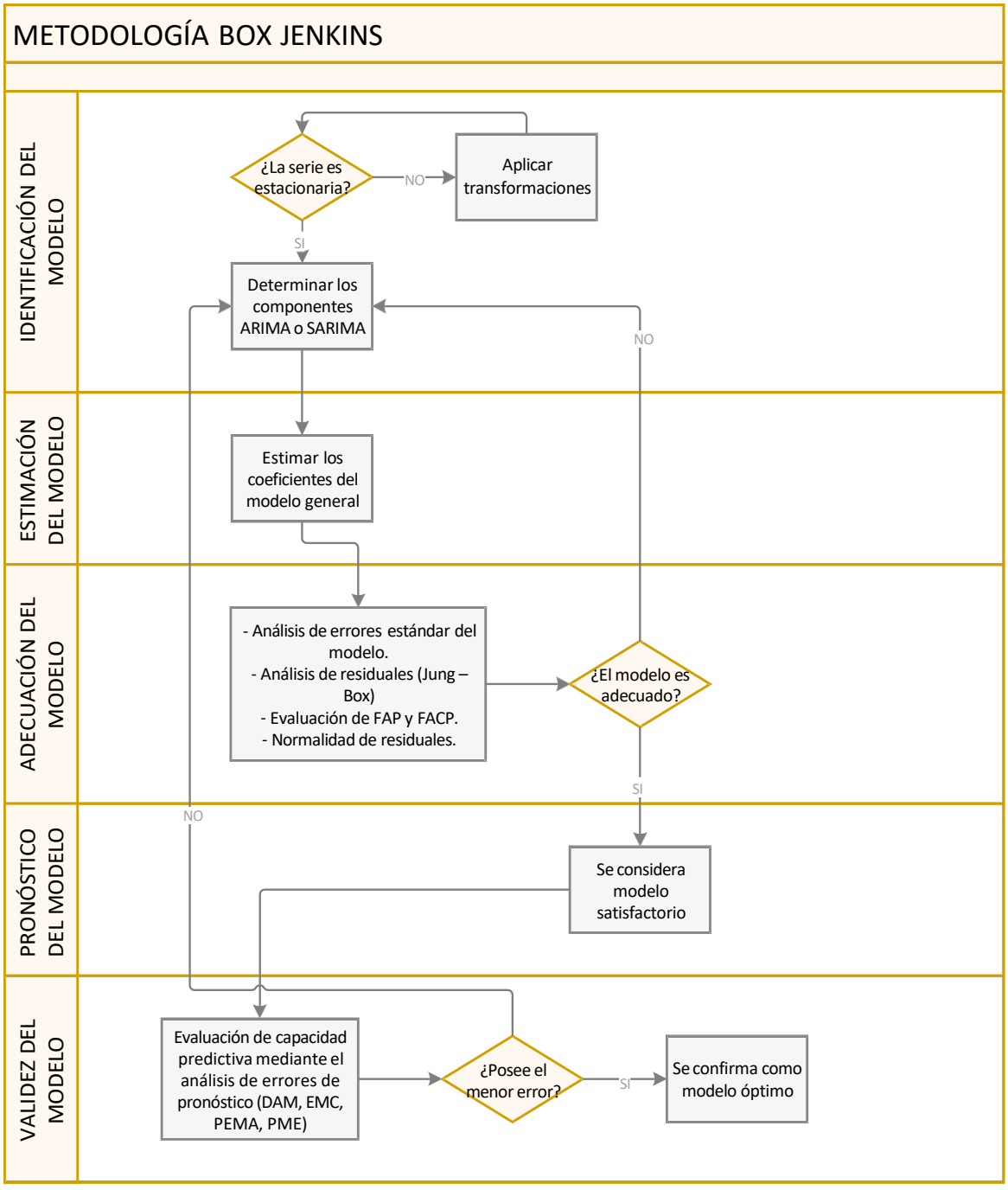
Villalobos, P. (2014). *Pronóstico de ingresos tributarios recaudados de impuestos generales a las ventas a nivel nacional, por la SUNAT, mediante la metodología de Box y Jenkins y Redes Neuronales Artificiales para el año 2014.* Lambayeque.

Villavicensio, J. (s.f.). *Introducción a las series de tiempo.*

Walker, B., Hollinger, C., Carpenter, S., & Kinzing, A. (2004). *Resilience, Adaptability and Transformability in Social-ecological Systems.*

ANEXOS

Anexo 1: Metodología de Box Jenkins



Anexo 2: Modelos descartados de la metodología Box Jenkins debido a que no cuentan con coeficientes significativos.

SARIMA (1,0,0)(0,1,0)	SARIMA (0,0,0)(1,0,)	SARIMA (0,0,1)(0,,)
SARIMA (0,0,1)(0,1,0)	SARIMA (0,1,0)(1,0,)	SARIMA (1,0,1)(0,,)
SARIMA (1,0,1)(0,1,0)	SARIMA (0,1,0)(1,0,)	SARIMA (1,0,1)(0,,)
SARIMA (1,0,1)(1,0,0)	SARIMA (0,1,1)(0,0,)	SARIMA (1,1,0)(0,,)
SARIMA (1,0,0)(1,0,0)	SARIMA (0,0,1)(0,0,)	SARIMA (0,1,0)(0,,)
SARIMA (1,1,0)(1,0,0)	SARIMA (1,0,1)(0,0,)	SARIMA (0,1,0)(0,,)
SARIMA (0,0,1)(0,0,0)	SARIMA (0,1,0)(0,0,)	SARIMA (1,0,0)(0,,)
SARIMA (1,0,1)(0,0,0)	SARIMA (0,1,0)(0,0,)	SARIMA (1,0,0)(0,,)
SARIMA (1,0,1)(1,0,0)	SARIMA (0,1,1)(0,0,)	SARIMA (1,1,0)(0,,)

Anexo 15: Código R usado para el procesamiento estadístico de ambas metodologías

```
#######METODOLOGIA BOX JENKINS######

data<-read.delim("clipboard")

data

##Generar vector de fechas, clase "Date"

fechas<- seq(as.Date("1/12/2014"), length.out = length(data), by = "months")

##convertir a serie de tiempo

data.ts<-ts(data, start = c(2017,12), frequency = 12)

data.ts

#grafica con flechas

ts.plot(data.ts, main = "Comportamiento")

#Grafica de tendencia, estacionalidad y ruido

plot(decompose(data.ts))

##Cargarmos paquetes

install.packages("FinTS")

install.packages("tseries")

install.packages("lmtest")

install.packages("forecast")

install.packages("nortest")

install.packages("ggplot2")

install.packages("ggfortify")

install.packages("CADFtest"
```

```
library(FinTS)
library(tseries)# para pruebas adf.test y pp.test
library(lmtest)
library(forecast)


library(nortest)
library(ggfortify)
library(CADFtest)
# GRAFICO DE LA SERIE
library(ggfortify)
library(CADFtest)
library(lmtest)
# PASO 1: IDENTIFICAR EL MODELO#
## observamos su grafico de autocorrelacion parcial y autocorrelacion total
autoplot(acf(data.ts,lag.max = 54, plot = FALSE),main = "Serie: data.ts")
#Determinamos los ordenes q del modelo
# q = (0,1) ademas  Q (0)_12      3,4,8,9     1,2,3,4
autoplot(pacf(data.ts,lag.max = 54, plot = FALSE),main = "Serie: data.ts")
#Determinamos los ordenes q del modelo
# p = (0,1,2) ademas P (1)_12    3,4  1,2
## Paso 2: Determinamos la estacionariedad o estacionalidad
ndiffs(data.ts) #Indica el numero de diferencias ordinales que se deberian realizar (d)
nsdiffs(data.ts)# Indica el numero de diferencias estacionales que se deberian realizar (D)
```

```
## Prueba de raiz unitaria Dickey - fuller (ads.test) y phillips - Perron (pp.test)
adf<- adf.test(data.ts)
app<- pp.test(data.ts)
## suavisamos la serie de tiempo
tsd<- diff(data.ts)
plot(data.ts)
plot(tsd)
ndiffs(tsd)
nsdiffs(tsd)
## Prueba de raiz unitaria Dickey - fuller (ads.test) y phillips - Perron (pp.test)
adf<- adf.test(tsd)
app<- pp.test(tsd)
##Paso 3: Graficos de las funciones de Autocorrelacion (ACF) y Autocorrelacion parcial
(PACF)
autoplot(acf(tsd,lag.max = 54, plot = FALSE))
#Determinamos los ordenes q del modelo
# q = (0,1) ademas  Q (0)_12

autoplot(pacf(tsd,lag.max = 54, plot = FALSE))
#Determinamos los ordenes q del modelo
# p = (0,1,2) ademas P (0,1)_12

## Encontramos el posible modelo arima
```

```
Posiblesarima<-auto.arima(y = data.ts)
Posiblesarima2<-auto.arima(data.ts,max.p = 2, max.q = 1, max.P = 1, max.Q = 1, max.order = 12, max.d = 1, max.D = 1)
summary (Posiblesarima)
summary (Posiblesarima2)

##estimacion de los coeficientes de los parametros de la serie de los modelos tentativos
instalamos el lmtest, para conocerlos.
coeftest(arima1)
coeftest(arima2)
coeftest(arima3)
coeftest(arima4)
coeftest(arima5)
coeftest(arima6)
coeftest(arima7)
coeftest(arima8)
coeftest(arima9)
coeftest(arima10)
coeftest(arima11)
coeftest(arima12)
coeftest(arima13)
coeftest(arima14)
coeftest(arima15)
```

```
##Seleccion del modelo

aic1 <- arima1 $ aic

aic2 <- arima2 $ aic

aic3 <- arima3 $ aic

aic4 <- arima4 $ aic

aic5 <- arima5 $ aic

aic6 <- arima6 $ aic

aic7 <- arima7 $ aic

aic8 <- arima8 $ aic

aic9 <- arima9 $ aic

aic10 <- arima10 $ aic

aic11 <- arima11 $ aic

aic12 <- arima12 $ aic

aic13 <- arima13 $ aic

aic14 <- arima14 $ aic

aic15 <- arima15 $ aic

aic1

aic2

aic3

aic4

aic5

aic6

aic7
```

aic8

aic9

aic10

aic11

aic12

aic13

aic14

aic15

summary (arima1)

summary (arima2)

summary (arima3)

summary (arima4)

summary (arima5)

summary (arima6)

summary (arima7)

summary (arima8)

summary (arima9)

summary (arima10)

summary (arima11)

summary (arima12)

summary (arima13)

summary (arima14)

summary (arima15)

```
#Paso5: validacion del modelo
residuales3 <- arima3$residuals #crearemos una variable con las residuales del modelo seleccionado
plot(residuales3)
```

Verificamos los residuos del modelo Arima3

```
autoplot(acf(arima3$residuals, plot = FALSE,lag.max = 60), colour = "darkgoldenrod2",conf.int.colour = "blue")
autoplot(pacf(arima3$residuals, plot = FALSE, lag.max = 60),colour = "chartreuse3",conf.int.colour = "red")
```

```
checkresiduals(arima3) #Genera la prueba lung-box, ademas del grafico de ACF y el histograma para los residuales.
Box.test(residuales3) #genera la prueba de box-pierse
tsdiag(arima3) #Genera los residuales estandarizados, ACF y el grafico de la prueba de lung - box para los residuales
ks.test(residuales3,"pnorm",mean=mean(residuales3), sd=sd(residuales3)) #prueba de kolmogorov smirnof para la normalidad de los residuales
qqnorm(residuales3, sub= "figura: Grafico Q para evaluar normalidad"); qqline(residuales3)
```

Verificamos los residuos del modelo Arima4

```
residuales4 <- arima4$residuals #crearemos una variable con las residuales del modelo seleccionado
```

```
plot(residuales4)

autoplot(acf(arima4$residuals, plot = FALSE,lag.max = 60), colour = "darkgoldenrod2",conf.int.colour = "blue")
autoplot(pacf(arima4$residuals, plot = FALSE, lag.max = 60),colour = "chartreuse3",conf.int.colour = "red")


checkresiduals(arima4) #Genera la prueba lung-box, ademas del grafico de ACF y el histograma para los residuales.
Box.test(residuales4) #genera la prueba de box-pierse
tsdiag(arima4) #Genera los residuales estandarizados, ACF y el grafico de la prueba de lung - box para los residuales
ks.test(residuales4,"pnorm",mean=mean(residuales4), sd=sd(residuales4)) #prueba de kolmogorov smirnof para la normalidad de los residuales
qqnorm(residuales4, sub= "figura: Grafico Q para evaluar normalidad"); qqline(residuales4)
Verificamos los residuos del modelo Arima7
residuales7 <- arima7$residuals #crearemos una variable con las residuales del modelo seleccionado
plot(residuales7)
autoplot(acf(arima7$residuals, plot = FALSE,lag.max = 60), colour = "darkgoldenrod2",conf.int.colour = "blue")
```

```
autoplot(pacf(arima7$residuals, plot = FALSE, lag.max = 60),colour = "chartreuse3",conf.int.colour = "red")


checkresiduals(arima7) #Genera la prueba lung-box, ademas del grafico de ACF y el histograma para los residuales.
Box.test(residuales7) #genera la prueba de box-pierse
tsdiag(arima7) #Genera los residuales estandarizados, ACF y el grafico de la prueba de lung - box para los residuales
ks.test(residuales7,"pnorm",mean=mean(residuales7), sd=sd(residuales7)) #prueba de kolmogorov smirnof para la normalidad de los residuales
qqnorm(residuales7, sub= "figura 09: Grafico Q para evaluar normalidad"); qqline(residuales7)
##Paso 6: Pronostico
##PRonostico del modelo 3
prediccion3 <- forecast(arima3,6)
prediccion3
plot(prediccion3)
##PRonostico del modelo 4
prediccion4 <- forecast(arima4,6)
prediccion4
plot(prediccion4)
##PRonostico del modelo 7
```

```
prediccion7 <- forecast(arima7,6)

prediccion7

plot(prediccion7)

Pronosticos arima3

fit3s <- ts(matrix(prediccion3$fitted),end = c(2022,11), frequency = 12)

mean3s<-ts(matrix(prediccion3$mean),end = c(2022,11), frequency = 12)

estimacion3s<-rbind(fit3s,mean3s)

pronosticos4s<-ts(estimacion3s,start=c(2017,12),end=c(2022,11),frequency = 12)
```