

Análisis de series temporales para predicción de demanda de productos

Cristian Camilo Ortega Betancur

Samuel Fernando Sánchez Cuervo

Asesor

Edgar Andres Villabon Aldana

Universidad Nacional Abierta y a Distancia – UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Dedicatoria

A nosotros mismos por formar un equipo que trabajo en el objetivo de hacer realidad el ser científicos de datos

Agradecimientos

Agradecemos al equipo de trabajo de docentes de la UNAD que nos brindaron sus experiencias y conocimientos que nos ayudaron a sacar el mayor provecho a la especialización de ciencia de datos y analítica; al doctor Edgar Andrés Villabon que dirigió la asesoría de este proyecto y finalmente a mi compañero de estudio Cristian Ortega Betancur y a Samuel Fernando Sánchez Cuervo por el esfuerzo y dedicación al desarrollo de este trabajo de grado.

Resumen

Este trabajo se organiza en cuatro secciones principales que abordan de manera integral el análisis de series temporales aplicado a la predicción de demanda. La primera sección presenta el origen de los datos y realiza un análisis descriptivo de las ventas históricas, proporcionando una visión general inicial. La segunda sección explora los modelos de estrategia utilizados en el proyecto, como Power BI, ARIMA, SARIMAX y Prophet, herramientas clave en el análisis de datos temporales. La tercera sección se enfoca en un informe sobre la predicción de demanda técnica basada en el modelo SARIMAX, mientras que la cuarta sección detalla el análisis de series temporales del producto en estudio y concluye con recomendaciones y observaciones generales. Los métodos de análisis de series temporales se dividen en dos categorías: cualitativos y cuantitativos. Los métodos cualitativos, basados en el juicio humano y la experiencia, permiten identificar patrones y tendencias de forma subjetiva, mientras que los métodos cuantitativos emplean modelos matemáticos para detectar patrones en los datos de manera objetiva. Entre los métodos cuantitativos más comunes se encuentran la regresión lineal, el análisis de componentes principales y el modelo ARIMA. La estructura metodológica planteada en este trabajo no solo destaca la importancia de seleccionar la técnica adecuada, sino también de adaptar el modelo a los objetivos específicos de predicción y análisis. Esto proporciona una base sólida para que el análisis de series temporales no solo mejore las proyecciones de demanda, sino también optimice la gestión de recursos en función de las necesidades reales del mercado.

Palabras claves: Métodos cualitativos, Métodos cuantitativos, ARIMA, SARIMAX, Power BI.

Abstract

This paper is organized into four main sections that comprehensively address time series analysis applied to demand forecasting. The first section presents the origin of the data and performs a descriptive analysis of historical sales, providing an initial overview. The second section explores the strategy models used in the project, such as Power BI, ARIMA, SARIMAX and Prophet, key tools in time-series data analysis. The third section focuses on a report on technical demand forecasting based on the SARIMAX model, while the fourth section details the time series analysis of the product under study and concludes with general recommendations and observations. Time series analysis methods fall into two categories: qualitative and quantitative. Qualitative methods, based on human judgment and experience, allow patterns and trends to be identified subjectively, while quantitative methods use mathematical models to detect patterns in the data objectively. Among the most common quantitative methods are linear regression, principal component analysis and the ARIMA model. The choice of the most appropriate method depends on factors such as the nature and availability of the data and the level of precision required for the analysis. The methodological structure put forward in this paper not only highlights the importance of selecting the appropriate technique, but also of tailoring the model to the specific prediction and analysis objectives. This provides a solid basis for time series analysis to not only improve demand projections, but also to optimize resource management based on actual market needs.

Keywords: Qualitative methods, quantitative methods, ARIMA, SARIMAX, Power BI.

Tabla de contenido

Introducción	10
Justificación	11
Objetivos.....	13
Objetivo General.....	13
Objetivos Específicos.....	13
Planteamiento del Problema	14
Marco de Referencia.....	15
Estado del Arte.....	15
Marco Conceptual.....	18
Técnicas de Análisis de Series Temporales	18
Marco Teórico.....	20
Gráfico ACF (Autocorrelation Function)	21
Gráfico PACF (Partial Autocorrelation Function).....	21
Estacionariedad y Varianza en Series Temporales	22
Metodología	23
Tipo de Investigación.....	23
Diseño de la Investigación	23
Técnicas y Herramientas.....	23
Población y Muestra	24
Procedimientos de Recolección de Datos	25
Resultados.....	27
Análisis de Datos	27

Análisis Exploratorio	31
Modelos de Predicción.....	34
Gráfico PACF (Partial Autocorrelation Function).....	37
Creación del Modelo de Predicción SARIMAX	41
Creación del Modelo de Predicción PROPHET	42
Criterios de Selección del Mejor Modelo	44
Selección del Modelo de Predicción.....	45
Predicción de Demanda de Ventas Basada en el Modelo SARIMAX	45
Sistema de Análisis de Series Temporales	46
Conclusiones	48
Recomendaciones	50
Referencias.....	51

Lista de Tablas

Tabla 1 <i>Datos Proporcionados por el Establecimiento</i>	27
Tabla 2 <i>Formato Transformado de los Datos con Estructura Fecha-Valor</i>	28
Tabla 3 <i>Estructura de los Datos por Día</i>	29
Tabla 4 <i>Estructura de los Datos por Meses</i>	30
Tabla 5 <i>Resumen Estadístico de los Datos</i>	32
Tabla 6 <i>Datos con Index por Fecha</i>	36
Tabla 7 <i>Resultado Prueba ADF</i>	38
Tabla 8. <i>Resultado de la Nueva Prueba</i>	38
Tabla 9 <i>Resultados del Código del p, q</i>	40
Tabla 10 <i>Reestructuración del Dataframe para PROPHET</i>	43
Tabla 11 <i>Resultados Criterio 1</i>	44
Tabla 12 <i>Resultados Criterio 2</i>	45

Lista de Figuras

Figura 1 <i>Datos Configurables en el Eje x Para Periodos de Tiempo</i>	29
Figura 2 <i>Gráfico de los Datos por Meses</i>	31
Figura 3 <i>Gráfico de Varianza</i>	33
Figura 4 <i>Previsión de Enero de 2021</i>	34
Figura 5 <i>Previsión de Enero de 2022</i>	35
Figura 6 <i>Previsión de Enero de 2023</i>	35
Figura 7 <i>Gráficos de ACF y PACF</i>	36
Figura 8 <i>Modelo ARIMA</i>	39
Figura 9 <i>Modelo ARIMA</i>	40
Figura 10 <i>Modelo SARIMAX</i>	42
Figura 11 <i>Modelo PROPHET</i>	43

Introducción

En el campo de la estadística, el análisis de series temporales se ha convertido en una herramienta fundamental, especialmente para predecir la demanda de productos. Este análisis se enfoca en estudiar la evolución de una variable a lo largo del tiempo, lo que permite a las empresas identificar patrones y tendencias en sus datos históricos de ventas. Al comprender estos patrones, las empresas pueden anticiparse a las fluctuaciones del mercado y tomar decisiones informadas que optimicen la gestión de sus recursos.

El objetivo principal del análisis de series temporales es facilitar la estimación de la demanda futura, al hacerlo, no solo se mejora la precisión en la planificación de la producción y el inventario, sino que también se contribuye a una gestión más eficiente de los recursos disponibles, esto es crucial en un entorno empresarial donde la capacidad de responder rápida y precisamente a las demandas del mercado puede marcar la diferencia. Para lograr estos objetivos, es esencial desarrollar un sistema de análisis de series temporales que permita anticiparse a la demanda de productos en un horizonte de tiempo determinado. Este sistema debe ser capaz de aplicar estrategias tecnológicas avanzadas, técnicas de programación sofisticadas y desarrollos propios que se adapten de manera óptima a las necesidades específicas de la empresa.

Las fluctuaciones impredecibles en la demanda pueden llevar a problemas como exceso de inventario o escasez de productos, lo que afecta negativamente la rentabilidad y la satisfacción del cliente. Ante este desafío, surge la necesidad de implementar un sistema de análisis de series temporales que permita entender cómo se comercializa la cerveza Águila Light en un restaurante-bar ubicado en la ciudad de Medellín, y de esta forma, el dueño del local tenga claridad en el manejo de este producto en su negocio.

Justificación

Las series temporales es una herramienta estadística que permite predecir la evolución de una variable en el tiempo a partir de datos históricos. Esta herramienta es ampliamente utilizada en la predicción de la demanda de productos, ya que permite a las empresas tomar decisiones más informadas sobre la planificación de la producción, el inventario y las estrategias de marketing y es así como se pretende aplicar a una base de datos propia para toma de decisiones eficientes en cuanto a inventario y distribución que servirá como modelo para aplicarlos a más de un producto.

Algunos de los beneficios del análisis de series temporales para la predicción de la demanda de productos son los siguientes:

Mejorar la precisión de los pronósticos: El análisis de series temporales permite identificar patrones y tendencias en los datos históricos, lo que puede ayudar a mejorar la precisión de los pronósticos de demanda.

Reducir el riesgo de falta de existencias: Al contar con pronósticos de demanda más precisos, las empresas pueden reducir el riesgo de quedarse sin existencias, lo que puede afectar negativamente a las ventas y la satisfacción del cliente.

Tomar decisiones informadas sobre marketing y promociones: Los pronósticos de demanda pueden ayudar a las empresas a identificar oportunidades de marketing y promoción, lo que puede aumentar las ventas y la participación de mercado.

En cuanto a algunas investigaciones recientes sobre análisis de series temporales para la predicción de la demanda de productos, en los últimos años se han realizado una serie de investigaciones que demuestran el impacto positivo del análisis de series temporales en la predicción de la demanda de productos. Por ejemplo, el análisis de series temporales puede

ayudar a mejorar la precisión de los pronósticos de demanda en un 20% (Sepúlveda & Benavides, 2023).

Un estudio realizado por Molina (2020) encontró que el análisis de series temporales puede ayudar a reducir el exceso de inventario en un 15%. Los estudios realizados por parte de Angulo & Ruiz (2023) aplicaron métodos de series temporales para predecir la demanda de energía eléctrica en Colombia, los resultados mostraron que los métodos de series temporales fueron capaces de predecir la demanda de energía eléctrica con un alto grado de precisión.

Moreno (2022) utilizó métodos de series temporales para predecir la producción de leche en Panamá. Los resultados mostraron que los métodos de series temporales fueron capaces de predecir la producción de leche con un alto grado de precisión, otros investigadores como Cruz & Correa (2017) aplicaron métodos de series temporales y redes neuronales artificiales para predecir la demanda intermitente. Los resultados mostraron que los métodos de series temporales fueron más precisos que las redes neuronales artificiales. Boada (2017), desarrolló un sistema de proyección de la demanda para empresas de venta por catálogo.

Objetivos

Objetivo General

Desarrollar un sistema de análisis de series temporales que permita anticiparse a la demanda de productos en un horizonte de tiempo determinado, aplicando estrategias ya sean de tipo tecnológico, de programación o desarrollos propios que mejor se adapten a la propuesta planteada.

Objetivos Específicos

Analizar los datos determinando patrones, tendencias estacionales y cualquier factor relevante en la demanda.

Aplicar técnicas de modelado de series temporales mediante modelos de aprendizaje automático, para creación de modelos de predicción.

Generar un informe general que ayude a la toma de decisiones estratégicas basadas en las predicciones de demanda.

Planteamiento del Problema

La gestión eficiente de la demanda de productos es crucial tanto para pequeños como grandes negocios en un entorno comercial altamente competitivo y en constante cambio. Las fluctuaciones impredecibles en la demanda pueden llevar a problemas como exceso de inventario o escasez de productos, afectando negativamente la rentabilidad y la satisfacción del cliente. Ante este desafío, surge la necesidad de implementar un sistema de análisis de series temporales que permita entender cómo se comercializa la cerveza Águila Light en un restaurante-bar ubicado en la ciudad de Medellín, proporcionando al dueño del local claridad en el manejo de este producto. Un sistema de análisis de series temporales utiliza datos históricos para identificar patrones y tendencias en la demanda, permitiendo a las empresas tomar decisiones más informadas sobre la planificación de la producción, la gestión del inventario y la fijación de precios.

Estudios recientes han demostrado la eficacia de estos sistemas para predecir la demanda de productos. Por ejemplo, Contreras et al., (2016), encontraron que el análisis de series temporales predijo con precisión la demanda de almacenamiento de productos perecederos, mientras que Fernández López et al., (2019) destacaron su eficacia en la predicción de la demanda turística. Además, Llanes & Benítez (2020), revisaron la literatura sobre modelos para predecir las producciones de leche y concluyeron que los modelos de series temporales son los más utilizados. Estos hallazgos subrayan la importancia de los sistemas de análisis de series temporales en la gestión eficiente de recursos empresariales y en la mejora de la competitividad en el mercado.

Marco de Referencia

Estado del Arte

El análisis de series temporales es una técnica crucial en la predicción de demanda, empleada ampliamente en campos como la economía, la energía, la industria y el turismo. A lo largo del tiempo, diversos autores han propuesto y perfeccionado modelos matemáticos para identificar patrones y tendencias en datos históricos y predecir comportamientos futuros. Uno de los textos clásicos en este campo es el de Box et al., (2015) desarrollaron los modelos ARIMA e hicieron una base metodológica para el análisis de series temporales univariadas. Este modelo sigue siendo relevante debido a su capacidad de descomponer series en componentes como tendencia, estacionalidad y ruido, lo cual facilita la creación de predicciones precisas y su implementación en múltiples aplicaciones prácticas.

Hyndman & Athanasopoulos (2021) amplían este marco metodológico con su enfoque en la práctica de la previsión, ofreciendo una visión moderna y práctica del uso de series temporales. Su obra es reconocida por combinar fundamentos teóricos con aplicaciones en software estadístico como R.

Makridak et al., (2020) complementan estas ideas al abordar una variedad de métodos de prevención para distintas aplicaciones. Su enfoque va más allá de las series temporales, explorando métodos cualitativos y cuantitativos y su implementación en contextos empresariales y financieros. Este enfoque multidimensional subraya la importancia de seleccionar el método de pronóstico adecuado según la naturaleza de los datos y la precisión requerida. En la misma línea, Montgomery & Johnson (2021) presentan una visión técnica que integra modelos econométricos y modelos de series temporales, enfocándose en los análisis predictivos en entornos

empresariales, enfatizando la importancia de combinar modelos estadísticos para mejorar la exactitud y confiabilidad en la predicción de la demanda.

Diversos estudios empíricos complementan la teoría con aplicaciones prácticas en la predicción de demanda de productos específicos. Contreras et al., (2016) emplean series temporales para pronosticar la demanda de almacenamiento de productos perecederos, un área donde la precisión en el manejo de inventarios es crucial. Este estudio demuestra cómo el análisis de series temporales puede optimizar los procesos logísticos y reducir el desperdicio. En el sector turístico, Fernández et al., (2019) revisan modelos de predicción de demanda, destacando la necesidad de modelos específicos debido a la variabilidad estacional y las características propias del turismo.

Otros estudios recientes, como el de Llanes & Benítez (2020), aplican modelos de predicción para estimar la producción de leche, evidenciando la utilidad de los modelos en la gestión agrícola. Asimismo, Sepúlveda & Benavides (2023) exploran el análisis predictivo de la demanda de servicios, demostrando cómo los modelos ARIMA y SARIMAX pueden aplicarse en el sector de servicios, donde la variabilidad es alta. Este último estudio resalta cómo los modelos más avanzados pueden adaptarse para incluir variables exógenas; que son variables externas al modelo que puede influir en la serie temporal que se está modelando, pero no es afectada por las variables internas del modelo. Estas variables se utilizan para mejorar la precisión de las predicciones al incorporar información adicional relevante, mejorando la precisión del pronóstico en entornos complejos.

El modelo SARIMAX se ha destacado como una de las herramientas más versátiles en análisis de series temporales, como lo evidencian Avendaño & Vega (2020), quienes lo utilizan para mejorar la gestión de inventarios en sectores farmacéuticos y clínicos, respectivamente. La

capacidad de este modelo para incorporar factores externos como la estacionalidad y eventos especiales lo convierte en un recurso valioso para diversos sectores. Además, aplican modelos de series temporales para predecir rendimientos agrícolas en México, comparando su efectividad con métodos clásicos, mientras que Cruz & Correa (2017) introdujeron redes neuronales para complementar las predicciones en demandas intermitentes, lo que representa una integración de técnicas de máquina.

Estudios como el de Rivero (2022) sobre predicción de demanda en farmacias, y el de Moreno (2022) sobre la producción de leche en Panamá, confirman la aplicabilidad de los modelos de series temporales en contextos de negocio. Angulo & Ruiz (2023) aplican esta metodología en la predicción de demanda de energía eléctrica en Colombia, lo cual destaca la adaptabilidad de los modelos de series temporales a distintos entornos, desde el comercio minorista.

Finalmente, el uso de plataformas tecnológicas como Power BI ha facilitado la implementación de estos modelos en el ámbito empresarial. Power BI permite la integración de modelos de aprendizaje automático para análisis predictivo, simplificando la visualización y el pronóstico de datos complejos (Microsoft, 2024).

En conclusión, el análisis de series temporales es fundamental en la predicción de la demanda, con aplicaciones que abarcan desde la gestión de inventarios hasta la planificación de servicios públicos. La combinación de modelos estadísticos avanzados con herramientas tecnológicas permite una implementación más efectiva y accesible, facilitando una toma de decisiones basada en datos en sectores cada vez más diversos y competitivos.

Marco Conceptual

El análisis de series temporales para la predicción de demanda de productos se fundamenta en la premisa de que los datos históricos de ventas contienen información valiosa que puede utilizarse para anticipar la demanda futura. Esta información se manifiesta a través de patrones y tendencias en los datos. Los patrones son secuencias de valores que se repiten a intervalos regulares, mientras que las tendencias representan cambios graduales en los valores de una serie temporal.

Los modelos de predicción de series temporales se emplean para identificar patrones y tendencias en los datos históricos, permitiendo así predecir la demanda futura. Por ejemplo, en estudios como el de Velasteguí & Argoti (2023) se aplicaron métodos de series temporales para predecir las precipitaciones pluviales en el cantón Huaca, Ecuador, demostrando una alta precisión en sus predicciones. De manera similar Saldarriaga et al., (2023), analizaron la reducción del tiempo del plan de demanda en Pintuco Medellín, encontrando que dicha reducción puede mejorar la eficiencia de la empresa. Asimismo, Rivero et al., (2022) desarrollaron un modelo para predecir la demanda en farmacias utilizando métodos de series temporales, lo que subraya la eficacia de estos modelos en la predicción de la demanda futura y la optimización de la gestión empresarial.

Técnicas de Análisis de Series Temporales

Existen una amplia variedad de técnicas de análisis de series temporales que pueden utilizarse para predicción de demanda de productos. Estas técnicas se pueden clasificar en dos categorías principales:

Métodos clásicos: Estos métodos se basan en el análisis de los datos históricos sin utilizar ningún modelo estadístico. Los métodos clásicos más comunes incluyen:

Método de promedios móviles: Este método calcula un promedio de los valores de la serie temporal en un período de tiempo determinado.

Método de suavización exponencial: Este método utiliza un peso decreciente para los valores más antiguos de la serie temporal.

Método de descomposición de series temporales: Este método descompone la serie temporal en sus componentes de tendencia, estacionalidad y ruido.

Métodos estadísticos: Estos métodos utilizan modelos estadísticos para identificar patrones y tendencias en los datos históricos. Los métodos estadísticos más comunes incluyen:

Modelos autorregresivos: Estos modelos predicen el valor futuro de una variable en función de sus valores pasados.

Modelos de media móvil: Estos modelos predicen el valor futuro de una variable en función de la media móvil de sus valores pasados.

Modelos de ARIMA: Estos modelos combinan modelos autoregresivos, modelos de media móvil e ideas de descomposición de series temporales.

Modelos de SARIMAX: Extienden los modelos ARIMA para incluir componentes estacionales y variables exógenas.

Modelos de PROPHET: Desarrollados por Facebook, son modelos aditivos que descomponen la serie temporal en componentes de tendencia, estacionalidad y efectos de días festivos.

Marco Teórico

Una serie de tiempo es una secuencia ordenada de datos recolectados a intervalos regulares para medir cómo evoluciona una variable a lo largo del tiempo, como el precio de una acción o la temperatura diaria. Estas series presentan patrones clave como tendencias (cambios a largo plazo), estacionalidad (repeticiones periódicas) y ciclos (fluctuaciones prolongadas), además de interrelaciones entre sus valores, lo que permite realizar pronósticos y analizar el impacto de distintos factores. El análisis de series temporales, además de facilitar la comprensión de fenómenos pasados, es una herramienta estadística fundamental para anticipar comportamientos futuros, especialmente en el ámbito empresarial, donde identificar patrones y prever la demanda permite optimizar procesos como la producción, distribución y gestión de recursos, mejorando la toma de decisiones estratégicas.

En este contexto, el proyecto analiza diversas herramientas que destacan en el ámbito del análisis de series temporales, como Power BI, ARIMA, SARIMAX y Prophet. Power BI, además de su capacidad de visualización de datos, permite la integración de modelos de aprendizaje automático para el pronóstico, generando paneles interactivos y accesibles. ARIMA, un modelo estadístico clásico, permite realizar predicciones a corto y mediano plazo al descomponer la serie temporal en componentes como tendencia y estacionalidad. SARIMAX, por su parte, incorpora variables exógenas, mejorando la precisión de los pronósticos al considerar factores externos, mientras que Prophet, desarrollado por Meta, es capaz de modelar patrones complejos y facilitar la visualización de datos, siendo de gran utilidad en aplicaciones diversas.

Tras evaluar las herramientas, se eligió SARIMAX como la más adecuada para este proyecto debido a su precisión y capacidad de análisis predictivo. Estudios previos refuerzan esta elección; por ejemplo, investigaciones como la de Avendaño & Vega (2020) evidencian la

eficacia de modelos de series temporales para prever la demanda de bienes y servicios, reduciendo el desperdicio y mejorando la satisfacción de los clientes. También destacan otros estudios, como los de Cruz & Correa (2017) subrayan la superioridad de los modelos de series temporales en comparación con otros métodos. Así, el uso de SARIMAX en este proyecto se justifica no solo por su aplicabilidad directa en los datos, sino también por su éxito demostrado en otros contextos empresariales.

Gráfico ACF (Autocorrelation Function)

Muestra la correlación entre una serie temporal y sus valores rezagados en distintos intervalos de tiempo. Este gráfico es útil para identificar la presencia de correlaciones a diferentes retrasos. En el contexto del modelo ARIMA, el gráfico ACF es clave para determinar el número de términos de media móvil (q) que deben incluirse en el modelo. Si el gráfico ACF muestra una caída rápida (es decir, la correlación se vuelve insignificante después de unos pocos rezagos), esto sugiere que se debe incluir un número limitado de términos MA (media móvil) en el modelo.

Gráfico PACF (Partial Autocorrelation Function)

Muestra la correlación entre una serie temporal y sus valores rezagados, pero eliminando el efecto de los rezagos intermedios. En otras palabras, mide la relación entre una observación y su rezago, controlando la influencia de otras observaciones intermedias. En el contexto del modelo ARIMA, el gráfico PACF se utiliza para determinar el número de términos autorregresivos (p) que deben incluirse en el modelo. Si el gráfico PACF muestra una caída rápida (es decir, la correlación se vuelve insignificante después de unos pocos rezagos), esto indica que solo se necesita un número limitado de términos AR (autorregresivos).

Estacionariedad y Varianza en Series Temporales

Una serie temporal se considera estacionaria si sus propiedades estadísticas, como la media y la varianza, permanecen constantes a lo largo del tiempo. Sin embargo, muchas series de tiempo presentan tendencias o patrones que cambian con el tiempo, lo que puede hacer que la serie no sea estacionaria. En estos casos, se deben realizar diferencias para estabilizar estos patrones antes de aplicar el modelo.

La diferenciación de una serie temporal implica calcular la diferencia entre cada valor y el valor anterior en la serie. Este proceso puede repetirse varias veces, y el número de diferencias necesarias se representa como d . Para determinar si una serie temporal es estacionaria, se realiza la prueba ADF (Augmented Dickey-Fuller). Si el valor del estadístico ADF es menor que el valor crítico (o si el valor p es menor que el nivel de significancia, típicamente 0.05), se rechaza la hipótesis nula, lo que indica que la serie es estacionaria.

La varianza mide qué tanto se dispersan los valores de una serie respecto a su media. En modelos de predicción, una alta varianza indica volatilidad, dificultando la detección de patrones. Para estabilizarla, se usan transformaciones como la diferenciación o logaritmos. También se aplican métodos de suavizado, como promedios móviles o modelos ARIMA, y en algunos casos, se descompone la serie en tendencia, estacionalidad y residuo. Estas técnicas mejoran la calidad de los datos y optimizan la precisión de los modelos predictivos.

Metodología

Tipo de Investigación

El presente proyecto se enmarca en un estudio cuantitativo de tipo predictivo. Este tipo de investigación busca analizar datos históricos para identificar patrones y prever tendencias futuras, en este caso, la demanda de un producto específico basado en series temporales. Al utilizar modelos estadísticos y de aprendizaje automático, se pretende hacer proyecciones precisas de las ventas a partir de los datos históricos obtenidos.

El enfoque cuantitativo es adecuado ya que permite utilizar técnicas matemáticas y estadísticas para modelar y predecir comportamientos futuros basándose en la relación que existe entre variables históricas y patrones de comportamiento observados.

Diseño de la Investigación

El diseño de la investigación es de tipo descriptivo y correlacional. En primer lugar, se realizará un análisis descriptivo para explorar las características de los datos históricos, identificando las tendencias estacionales y la relación entre variables. Posteriormente, se utilizarán modelos predictivos (ARIMA, SARIMAX y Prophet) para evaluar la relación entre las variables en el tiempo y hacer proyecciones de demanda futura.

Descriptivo: Se describirán las principales características de los datos, tales como las tendencias de ventas, las fluctuaciones estacionales, las variaciones de la demanda por días de la semana, meses, y otros posibles factores.

Correlacional: El análisis exploratorio también ayudará a identificar la relación entre diferentes factores y cómo estas variables afectan la demanda de productos a lo largo del tiempo.

Técnicas y Herramientas

Para la ejecución de este proyecto, se utilizarán las siguientes técnicas y herramientas:

Python: Se utilizará para implementar los modelos ARIMA, SARIMAX y Prophet. Las bibliotecas específicas como statsmodels (para ARIMA y SARIMAX) y Prophet serán fundamentales para entrenar los modelos y realizar las predicciones.

Power BI: Se utilizará para la visualización interactiva de los datos, permitiendo que el propietario del negocio pueda explorar los datos y realizar análisis visuales.

Análisis de Estacionariedad: Se realizarán pruebas como el test de Dickey-Fuller para verificar la estacionariedad de las series temporales y, en caso de que no lo sean, se aplicarán transformaciones para lograrla (como diferenciación).

Análisis de autocorrelación (ACF y PACF): Estos gráficos se usarán para identificar la relación entre las observaciones pasadas y ayudar a definir los parámetros del modelo ARIMA.

ARIMA (AutoRegressive Integrated Moving Average): Este modelo estadístico se utilizará para analizar las ventas históricas y realizar predicciones basadas en los componentes de autorregresión, diferenciación e integración.

SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Regressors): Esta técnica permitirá incorporar variables exógenas que puedan influir en la demanda, como días festivos o eventos especiales, que afecten las ventas del producto.

Prophet: Un modelo de predicción desarrollado por Meta, que será útil para gestionar patrones complejos no lineales y estacionales en la serie temporal.

Población y Muestra

La población del estudio está constituida por los datos históricos de ventas de un producto específico de un establecimiento en el área metropolitana del Valle de Aburrá. La muestra está compuesta por las ventas de este producto desde enero de 2021 hasta septiembre de 2023, con información detallada sobre las unidades vendidas por día, y la fecha correspondiente.

Aunque los datos no son sensibles, se manejan como información privada y se asegurará que el acceso a los resultados sea exclusivo para el propietario del negocio.

Los datos utilizados están organizados por fecha, y cada observación corresponde a una venta diaria. Para el análisis de series temporales, los datos serán agrupados mensualmente para reducir la variabilidad de las ventas diarias y facilitar la identificación de patrones estacionales.

Procedimientos de Recolección de Datos

La recolección de datos se basará en la información proporcionada por el establecimiento, la cual ya está disponible en un archivo cronológico con los datos de ventas y fechas. Dado que los datos han sido organizados en función de las transacciones diarias, no será necesario un proceso de recolección adicional.

El análisis de los datos se realizará en las siguientes etapas:

Se realizará un análisis descriptivo para identificar tendencias, patrones estacionales y posibles anomalías. Se visualizarán las ventas en función del tiempo utilizando herramientas de gráficos interactivos en Power BI.

Los datos serán agrupados por meses para suavizar las fluctuaciones diarias y hacer los datos más eficientes para los modelos predictivos. También se verificará la estacionariedad de la serie temporal utilizando pruebas estadísticas y, si es necesario, se realizarán transformaciones para asegurar la adecuación de los modelos ARIMA y SARIMAX.

Se entrenarán los modelos ARIMA, SARIMAX y Prophet. Para ARIMAX y SARIMAX, se ajustarán los parámetros de autorregresión, diferenciación e integración. Para Prophet, se configurará el modelo para capturar tendencias no lineales y estacionales.

Se calculará el error cuadrático medio (RMSE) de las predicciones para cada modelo. Esta métrica cuantitativa permitirá medir el rendimiento y la precisión de las predicciones.

Además de calcular el RMSE, se realizará una evaluación subjetiva visual de los resultados. Se compararán gráficamente las predicciones con los datos reales, observando cómo cada modelo captura las tendencias y fluctuaciones estacionales. Esta evaluación cualitativa ayudará a determinar cuál modelo refleja mejor los patrones observados.

El modelo con el menor RMSE se seleccionará como el más preciso. No obstante, la evaluación visual subjetiva también jugará un papel importante, considerando cómo los modelos ajustan las tendencias y variaciones estacionales en los datos.

Se presentará un informe detallado con los resultados obtenidos, que incluirá las predicciones generadas, las comparaciones entre los modelos utilizados, las recomendaciones estratégicas para el negocio y las conclusiones del análisis.

La metodología propuesta se basa en el uso de técnicas estadísticas robustas y modelos avanzados de aprendizaje automático, que son capaces de capturar tanto patrones estacionales como tendencias a largo plazo en series temporales. El enfoque cuantitativo se justifica porque la predicción precisa de la demanda requiere un análisis objetivo y basado en datos históricos.

El uso de modelos como ARIMA, SARIMAX y Prophet es apropiado, ya que estos modelos están especialmente diseñados para manejar series temporales, y han sido validados en estudios previos relacionados con la predicción de demanda. La evaluación del RMSE asegura una medida objetiva de precisión, mientras que la evaluación visual subjetiva garantiza que el modelo seleccionado sea el más adecuado para reflejar las tendencias y variaciones de los datos. Finalmente, el uso de Python y Power BI facilitará tanto la implementación de los modelos como la interpretación y visualización de los resultados.

Resultados

Análisis de Datos

Tabla 1

Datos Proporcionados por el Establecimiento

Mes	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Enero							
Semana 1	35	40	34	28	60	100	115
Semana 2	90	35	40	30	70	110	101
Semana 3	30	36	37	40	86	121	88
Semana 4	25	27	33	38	75	98	80
Febrero							
Semana 5	33	28	36	41	90	140	98
Semana 6	21	31	41	45	100	114	90
Semana 7	20	37	40	25	78	88	70
Semana 8	33	39	42	45	99	123	99
Marzo							
Semana 9	40	34	38	37	85	112	93
Semana 10	23	30	35	32	79	98	90
Semana 11	20	40	45	44	100	113	150
Semana 12	90	32	30	50	200	130	100
Abril							
Semana 13	30	47	80	90	100	140	67
Semana 14	33	42	40	50	89	133	100
Semana 15	19	46	38	40	98	100	101
Semana 16	22	35	33	41	114	135	88

Nota. Esta tabla muestra la cantidad de cerveza águila Light vendida diariamente.

Aunque los datos originales están organizados por meses, días de la semana y cantidades vendidas, esta estructura no es la más adecuada para aplicar técnicas de análisis de series

temporales. Por ello, se realizó una transformación del formato inicial como se muestra en la Tabla 1, que cuenta con columnas separadas para cada día de la semana, a un formato más lineal, compuesto por dos columnas: una para la fecha y otra para el valor correspondiente al producto estrella. Esta reorganización, sin modificar ninguna variable, permite un tratamiento más eficiente en herramientas de análisis y programación orientada a series temporales.

Tabla 2

Formato Transformado de los Datos con Estructura Fecha-Valor

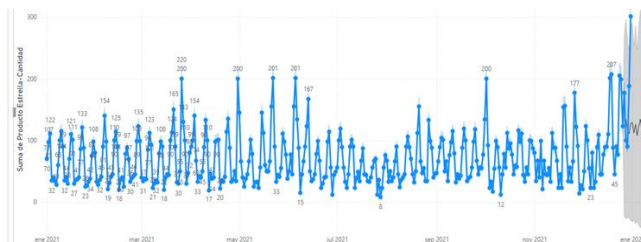
Fecha	Producto Estrella-Cantidad
01/01/2021	70
02/01/2021	97
03/01/2021	111
04/01/2021	35
05/01/2021	40
06/01/2021	34
07/01/2021	28
08/01/2021	60
09/01/2021	100
10/01/2021	115
11/01/2021	90
12/01/2021	35
13/01/2021	40
14/01/2021	30
15/01/2021	70
16/01/2021	110
17/01/2021	101

Nota. Esta tabla muestra los datos organizados de forma más eficaz

Una de las herramientas iniciales para previsión de la demanda del producto trabajado en este proyecto fue de inteligencia de negocios como lo es Power BI; en donde se pueden analizar los datos a través de una tabla calendario distribuida por año, trimestre, mes y día como se muestra en la figura 1.

Figura 1

Datos Configurables en el Eje x Para Periodos de Tiempo



Al trabajar los datos por mes es posible visualizar hacia atrás y hacia adelante los datos reales vs la previsión que hace Power BI mediante una combinación de algoritmos de interpolación y técnicas de modelado de series temporales para realizar previsiones. Generalmente, emplea modelos como el método de Holt-Winters (también conocido como suavizado exponencial triple) para capturar tendencias y estacionalidades en los datos.

Tabla 3

Estructura de los Datos por Día

	Date	Sales
0	2021-01-01	70
1	2021-01-02	97
978	2023-09-06	94
979	2023-09-07	109

Nota. Esta tabla muestra los datos estructurados por día.

En la tabla 4 se puede ver el resultado de los datos por organizada por meses

Tabla 4

Estructura de los Datos por Meses

	Month_Year	Sales
0	2021-01-01	1980
1	2021-02-01	1746
2	2021-03-01	2127
3	2021-04-01	2147
4	2021-05-01	2199
5	2021-06-01	2173
6	2021-07-01	1701
7	2021-08-01	1961
8	2021-09-01	2075
9	2021-10-01	2309
10	2021-11-01	2051
11	2021-12-01	3253
12	2022-01-01	2102
13	2022-02-01	1470
14	2022-03-01	1764
15	2022-04-01	1960
16	2022-05-01	2353
17	2022-06-01	3068
18	2022-07-01	3753
19	2022-08-01	3955

Nota. Esta tabla muestra los datos estructurados por mes.

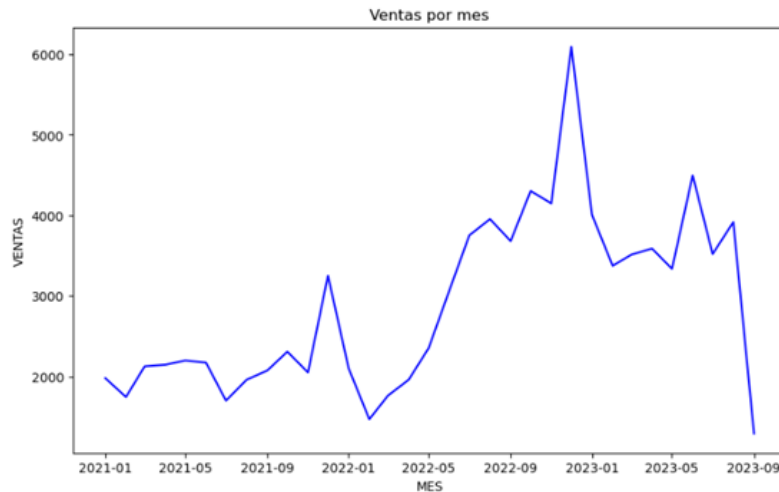
En este momento tenemos los datos estructurados de una manera más eficiente de procesar.

Análisis Exploratorio

Ahora se grafican los datos respectivamente para hacer un análisis exploratorio y como resultado se obtiene el gráfico que relaciona cantidad de ventas por mes.

Figura 2

Gráfico de los Datos por Meses



Luego de analizar visualmente este gráfico podemos concluir varios hallazgos:

Se observa una tendencia general al alza en las ventas a lo largo del período analizado, con algunas fluctuaciones estacionales.

Existe una considerable variabilidad en las ventas de un mes a otro, lo que indica la presencia de ruido o aleatoriedad en los datos.

No se observan valores atípicos evidentes en el gráfico, lo que simplifica el análisis.

En base de esto se genera un resumen estadístico (análisis descriptivo) de la variable ventas.

Tabla 5*Resumen Estadístico de los Datos*

Dato	Valor
Count	33.00000
Mean	2891.69697
Std	1103.68914
Min	1292.00000
25%	2051.00000
50%	2353.00000
75%	3682.00000
Max	6095.00000

Nota. Esta tabla muestra el resumen estadístico que incluye media de los datos, el máximo, mínimo, la desviación estándar y los cuartiles.

Durante el período analizado, el promedio de ventas mensuales fue de aproximadamente 2892 unidades, lo que nos da una buena referencia del rendimiento general. Sin embargo, al considerar la mediana, que se sitúa en 2353 unidades, observamos que la distribución de las ventas no es perfectamente simétrica. Esto significa que la mitad de los meses tuvieron ventas por debajo de este valor y la otra mitad por encima, lo que sugiere cierta variabilidad en el comportamiento de las ventas.

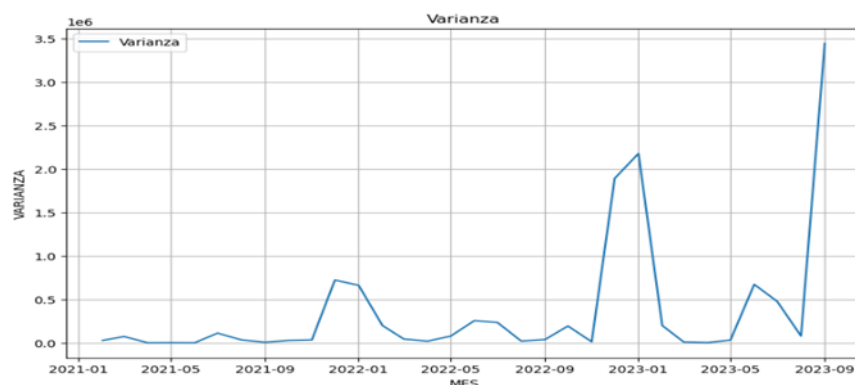
En cuanto a la dispersión de los datos, la desviación estándar es de 1103.69, lo que indica que las ventas pueden variar considerablemente de un mes a otro. Este valor relativamente alto refleja fluctuaciones significativas en las ventas mensuales. Además, el rango de las ventas va desde un mínimo de 1292 unidades hasta un máximo de 6095 unidades, lo que nos muestra la amplitud de los datos y la diferencia entre los meses con mayores y menores ventas.

Por último, los cuartiles proporcionan una visión más detallada de cómo se distribuyen las ventas. El primer cuartil, con un valor de 2051 unidades, indica que el 25% de los meses tuvieron ventas por debajo de este umbral. Mientras tanto, el tercer cuartil, que se sitúa en 3682 unidades, señala que el 75% de los meses tuvieron ventas menores a este valor.

Ahora se grafica la varianza de los datos por cada 2 meses y como resultado se obtiene el gráfico de la figura 3.

Figura 3

Gráfico de Varianza



Luego de analizar visualmente este gráfico podemos concluir varios hallazgos.

Diciembre - enero 2022/2023: En Los picos, en estos meses son típicos en muchos mercados debido a la temporada de fin de año y las festividades navideñas, cuando la demanda de productos suele aumentar drásticamente.

Septiembre 2023: este es un mes muy particular, una posibilidad es que el pico sea consecuencia de cambios en las tendencias de consumo.

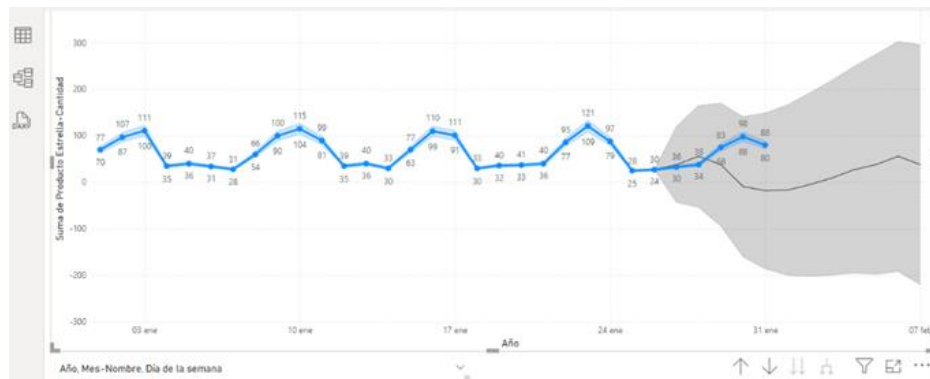
La alta varianza en estos periodos indica que es mucho más difícil predecir las ventas.

Modelos de Predicción

A continuación, se realiza de forma resumida una explicación de las herramientas citadas en el análisis de datos y que tendrán aplicación en el presente capítulo:

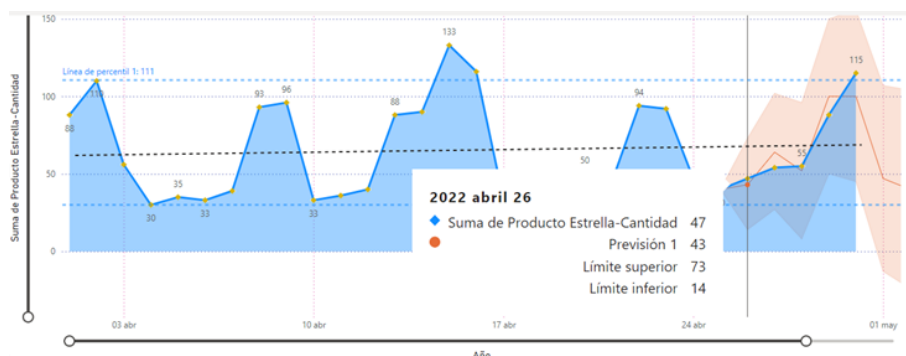
Figura 4

Previsión de Enero de 2021



En la anterior grafica se puede observar la previsión para 12 días, 5 puntos hacia atrás, 9 puntos de estacionalidad y con un intervalo del 95% de confianza en la configuración. Se puede apreciar que trabajar los datos con un rango de tiempo más corto permite identificar o hacer a la idea del comportamiento del producto vendido de una forma más clara.

En la figura 5 se puede observar los datos de enero de 2022 con características algo similares al periodo anterior del 2021; en esta imagen al situarse en uno de los puntos de previsión es posible identificar el rango mínimo, rango máximo, la cantidad vendida del producto real y la previsión del producto generado por Power BI. Es de notar que la diferencia entre lo real vs lo previsto presenta una brecha corta.

Figura 5*Previsión de Enero de 2022*

En la figura 6 se puede observar un patrón de comportamiento más ajustado de la previsión frente al comportamiento real, dado que la tendencia sigue un comportamiento de tipo estacional por los valles y picos que se muestran tanto en el dato real como en el de la previsión.

Figura 6*Previsión de Enero de 2023*

Luego del análisis de datos con una herramienta de inteligencia de negocios como Power BI para la previsión del producto de venta; fue necesario ampliar la aplicación de otras herramientas de predicción como lo son ARIMA, SARIMAX y PROPHET.

Para el modelo de predicción ARIMA se obtienen los datos con index por fecha como se muestra en la tabla 6.

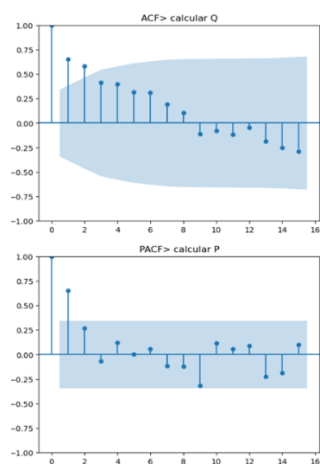
Tabla 6*Datos con Index por Fecha*

	Month_Year	Sales
0	2021-01-01	1980
1	2021-02-01	1746
2	2021-03-01	2127
3	2021-04-01	2147
4	2021-05-01	2199
5	2021-06-01	2173
6	2021-07-01	1701

Nota. Esta tabla muestra las cantidad de producto vendido en cada mes.

Esto es un paso indispensable antes de crear el modelo ARIMA, ya que requiere que el Index de los datos sean la fecha correspondiente.

Se grafica el ACF y PACF y como resultado se obtienen los gráficos que se muestran en la figura 7.

Figura 7*Gráficos de ACF y PACF*

Se determina el valor p , q para el modelo ARIMA. El primer rezago muestra una autocorrelación muy alta, lo que indica una fuerte dependencia entre observaciones consecutivas. Después del primer rezago, la función de autocorrelación decae exponencialmente hacia cero, lo que sugiere un componente MA (Media Móvil).

Gráfico PACF (Partial Autocorrelation Function)

El primer rezago muestra un pico significativo, lo que confirma la relación directa entre una observación y la anterior. Después del primer rezago, la función de autocorrelación parcial decae rápidamente hacia cero, indicando que no hay relaciones directas significativas con rezagos más lejanos.

En base a la interpretación de los gráficos ACF y PACF, se puede proponer los siguientes valores iniciales para los parámetros del modelo ARIMA:

$p = 1$: Dado que el primer rezago en el PACF es significativo y los siguientes decaen rápidamente, sugiere que un término autorregresivo de orden 1.

$q = 1$: El decaimiento exponencial en el ACF después del primer rezago sugiere que un término de media móvil de orden 1.

Determinar el valor de d para el modelo ARIMA.

El parámetro d en un modelo ARIMA representa el número de diferencias necesarias para hacer que una serie temporal sea estacionaria.

Se realiza la prueba ADF y como resultado se obtiene el ADF que se muestra en la tabla 7, donde se evidencia que la serie inicialmente no es estacionaria ya que el valor estadístico ADF no es menor que el valor crítico 0.05.

Tabla 7*Resultado Prueba ADF*

ADF Statistic	p-value
-2.195674567434072	0.20781908336693272

Nota. En la tabla se evidencia el ADF

Se calcula la diferencia de la serie $d=1$ y se vuelve a hacer la prueba ADF y como resultado se obtiene el ADF que se muestra en la tabla 8.

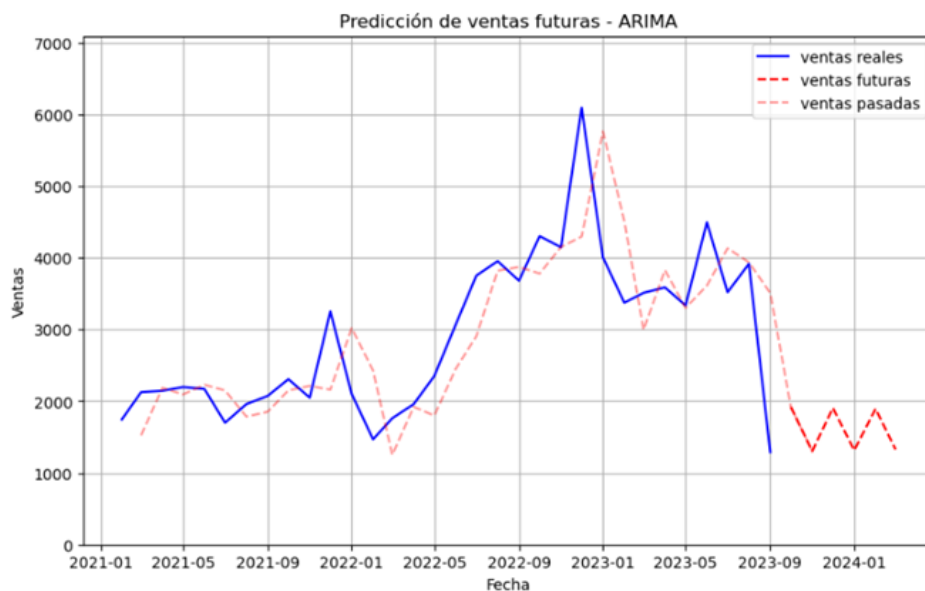
En la tabla 8 se evidencia que la diferencia $d=1$ es estacionaria ya que el valor estadístico ADF es menor que el valor crítico 0.05, por ende, al crear el modelo ARIMA es necesario especificar que el parámetro $d=1$.

Tabla 8*Resultado de la Nueva Prueba*

ADF Statistic	p-value
-6.704264931921631	3.820249313157915e-09

Nota. En la tabla se evidencia el ADF corregido.

Se calcula el modelo ARIMA con los parámetros p, d, q especificados, se calculan las predicciones pasadas con los datos reales y se calculan las predicciones futuras de los próximos 6 meses, generando un gráfico final para observar el resultado como se muestra en la figura 8.

Figura 8*Modelo ARIMA*

Este modelo presenta una predicción que se ajusta muy bien a los datos reales, mostrando una correlación significativa con las fluctuaciones observadas. Sin embargo, se evidencia un sobreajuste, donde el modelo captura no solo la tendencia general de las ventas, sino también el ruido aleatorio y las variaciones estacionales de los datos. Este sobreajuste se manifiesta en la capacidad del modelo para replicar los picos y valles en las ventas, lo que puede ser problemático para la generalización a datos futuros. Además, la predicción muestra un retardo de un mes, lo que implica que las ventas del mes actual se ven influenciadas por las ventas del mes anterior, un comportamiento esperado en series temporales. Aunque este modelo logra un buen ajuste, su susceptibilidad al sobreajuste sugiere que podría no ser el mejor para hacer predicciones más allá de los datos históricos.

Se realiza el análisis de sensibilidad variando los parámetros p , q . En este caso se calculan los parámetros de p , q según el menor AIC. Aunque un modelo con más parámetros

podría tener un mejor ajuste a los datos, su AIC será más alto debido a la penalización. Esto ayuda a evitar el sobreajuste.

Como resultado se obtiene el p , q como se muestra en la tabla 9.

Tabla 9

Resultados del Código del p , q

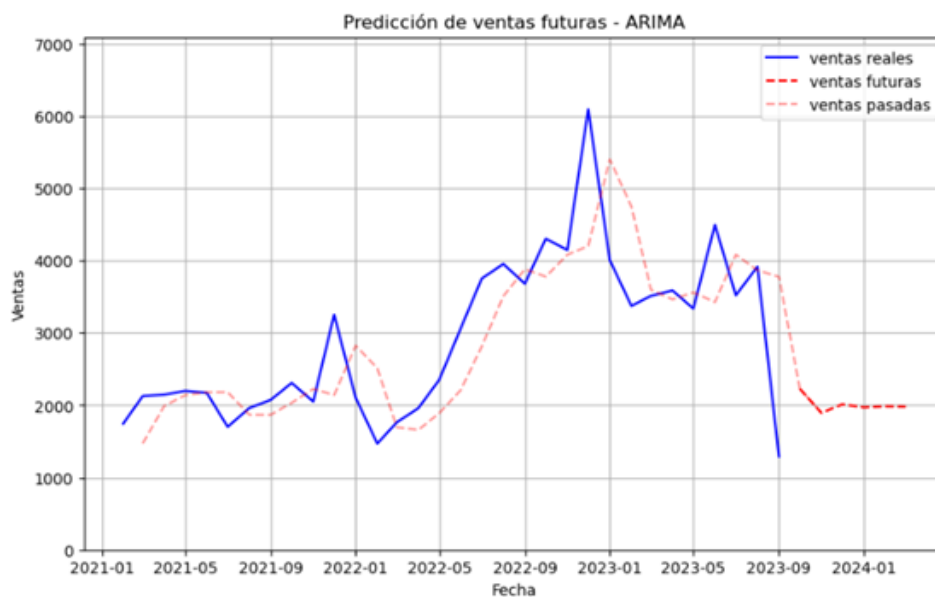
P	q	Menor AIC
1.0	0.0	508.58933862765304

Nota. En la tabla se evidencia Disponibilidad de tiempo.

Se calcula que los parámetros adecuados son $p=1$ y $q=0$ y se grafica el resultado de este modelo ARIMA p, d, q (1,1,0), como resultado se obtiene el grafico que se muestra en la figura 23, donde se pueden apreciar las ventas reales, ventas futuras y ventas pasadas.

Figura 9

Modelo ARIMA



Con un parámetro $q=0$, este modelo tiende a suavizar los picos y las caídas en las ventas, proporcionando una proyección más estable y menos susceptible a las fluctuaciones repentinas. Aunque el ajuste es menos preciso en comparación con el modelo 1, este modelo ofrece una visión más conservadora de las ventas futuras, lo que puede ser ventajoso en contextos donde se busca estabilidad y menos variabilidad en las proyecciones. A diferencia del modelo ARIMA $p,d,q (1,1,1)$, el AIC del modelo 2 es menor, lo que sugiere que, aunque proporciona proyecciones más suaves, logra un ajuste más eficiente en términos de calidad predictiva, al equilibrar mejor el ajuste a los datos y la complejidad del modelo.

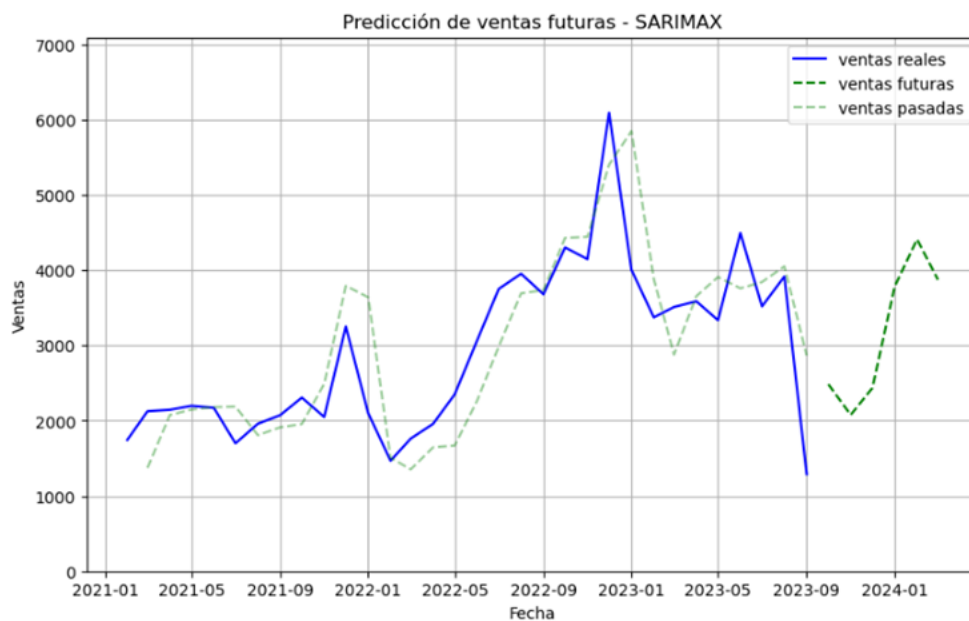
Creación del Modelo de Predicción SARIMAX

Dado que el modelo SARIMAX tiene la opción de agregar variables exógenas y frecuencia de estacionalidad, utilizamos el mismo proceso del modelo ARIMA, donde agregaremos los parámetros estacionales P, D, Q, s, y la variable exógena.

Se determina la variable exógena, esta variable exógena que se identifica en esta serie de tiempo muy evidente es los incrementos de ventas en los meses de diciembre y enero, es por eso por lo que se agrega una nueva columna al dataframe identificando si es diciembre o enero con un 1, y los meses restantes serán 0. Con esta variable exógena podemos definir una estacionalidad de ventas en esta época.

Dado que se definen los meses de diciembre y enero, vamos a trabajar con una frecuencia de estacionalidad de 10 a 11 según los resultados.

Se genera el modelo SARIMAX $p, d, q, P, D, Q, s (1,1,0,1,1,0,10)$, se calcula el modelo SARIMAX con los parámetros p, d, q, P, D, Q, s especificados, se calculan las predicciones pasadas con los datos reales y se calculan las predicciones futuras de los próximos 6 meses, generando un gráfico final para observar el resultado como se observa en la figura 10.

Figura 10*Modelo SARIMAX*

Este modelo entrega un resultado muy ajustado a los datos reales y con una respuesta estacional proporcional de 10 meses en diciembre y enero, generando un excelente resultado con los datos anteriores. Se puede observar la predicción de los próximos 6 meses como se incrementa sabiendo que se acerca diciembre y enero. Sin duda es una muy buena predicción si tenemos en cuenta esta variable exógena.

Creación del Modelo de Predicción PROPHET

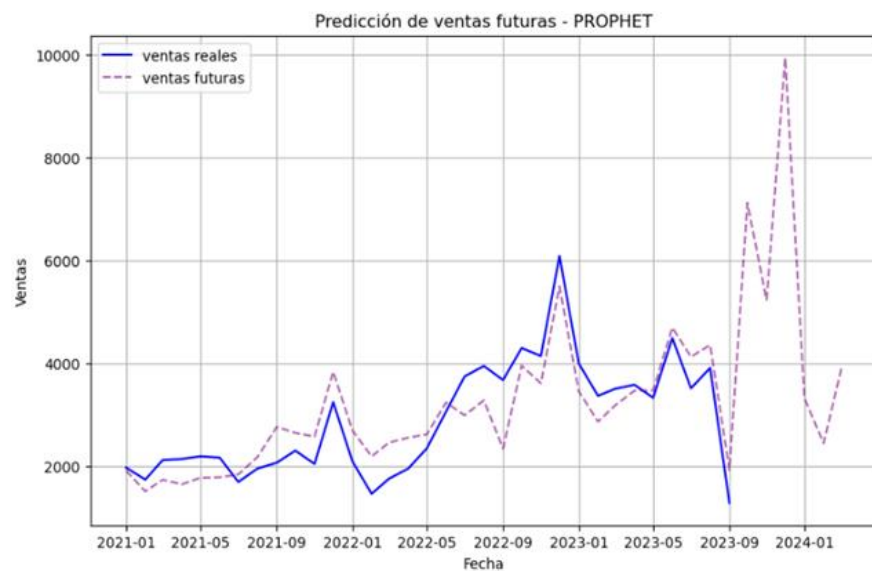
Para trabajar con el modelo de predicción Prophet será necesario reestructurar de nuevo el dataframe con los datos de fecha y ventas, renombrándolos como 'month_year': 'ds', 'sales': 'y' y como resultado se obtiene la reestructuración del dataframe para PROPHET como se muestra en la tabla 10.

Tabla 10*Reestructuración del Dataframe para PROPHET*

	ds	Y
0	2021-01-01	1980
1	2021-02-01	1746
2	2021-03-01	2127
3	2021-04-01	2147
4	2021-05-01	2199

Nota. Esta tabla muestra los datos organizados por meses para la reestructuración del dataframe para PROPHET.

Se puede ver que ahora se tienen dos columnas principales llamadas *ds* y *y* respectivamente. Se genera el modelo PROPHET y como resultado se obtiene la generación del modelo que se muestra en la figura 11.

Figura 11*Modelo PROPHET*

Este modelo presenta un ajuste más marcado en comparación con las predicciones anteriores, mostrando variaciones en su rendimiento: en ciertos períodos, las predicciones se sitúan por encima de los valores reales, mientras que en otros momentos se encuentran por debajo. Además, se observa una clara estacionalidad en los meses de diciembre y enero. Sin embargo, las proyecciones futuras son notablemente agresivas, ya que el modelo mantiene un patrón alcista que refleja la estacionalidad de los años anteriores.

Criterios de Selección del Mejor Modelo

A la hora de seleccionar un modelo de predicción para nuestros datos de ventas, es crucial establecer criterios que aseguren no solo la precisión de las predicciones, sino también la robustez y la capacidad de generalización del modelo. Para este análisis, hemos considerado los siguientes criterios:

Grado de ajuste: Buscamos un modelo que no esté excesivamente ajustado a los datos históricos. Un modelo muy ajustado puede capturar el ruido en lugar de las verdaderas tendencias, lo que lleva a una disminución de su rendimiento en datos futuros. Por lo tanto, es preferible un modelo conservador que ofrezca predicciones más estables y generalizables como se muestra en la tabla 11.

Tabla 11

Resultados Criterio 1

ARIMA	SERIMAX	PROPHET
Poco ajustado	Medio ajustado	Muy ajustado

Nota. En esta tabla se muestran los resultados del criterio 1 que define si la respuesta del modelo según su ajuste.

Criterios de selección de modelos: Se utilizan métricas de Error Cuadrático Medio (RMSE), calculando, usando la función $\text{np.sqrt}(\text{mean_squared_error}(\text{var1}, \text{var2}))$, como se muestra en la tabla 12.

Tabla 12

Resultados Criterio 2

ARIMA	SERIMAX	PROPHET
820	668	527

Nota. En esta tabla se muestran los resultados del criterio 2 que define si el error cuadrático medio.

Selección del Modelo de Predicción

El modelo SARIMAX se destaca porque combina un ajuste moderado (evitando tanto el sobreajuste como el subajuste) y una precisión sólida (con un RMSE razonablemente bajo). Aunque Prophet tiene el RMSE más bajo, su ajuste excesivo puede llevar a predicciones menos confiables en condiciones futuras que difieren de los datos históricos.

Por lo tanto, SARIMAX es la mejor opción para predecir ventas futuras porque:

Captura adecuadamente las estacionalidades y patrones sin ajustarse demasiado a las variaciones del pasado.

Mantiene un buen equilibrio entre precisión y capacidad de generalización, asegurando predicciones más robustas a largo plazo.

Predicción de Demanda de Ventas Basada en el Modelo SARIMAX

Este informe presenta los resultados de la predicción de demanda de ventas para los próximos seis meses, empleando el modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors). A partir del análisis de las ventas observadas entre

2021 y 2023, se identificaron tendencias clave que pueden orientar la toma de decisiones estratégicas.

En particular, los meses de diciembre y enero destacan por registrar los volúmenes de ventas más altos del periodo analizado. En diciembre de 2022, las ventas alcanzaron un máximo de 6,095 unidades, lo que representa un aumento significativo en comparación con otros meses. Este comportamiento puede atribuirse a factores como las festividades, promociones de fin de año y el comportamiento del consumidor, que suele concentrar sus compras en esta época. En enero, las ventas también son altas, aunque muestran un leve descenso en comparación con diciembre.

Por otro lado, se observó una caída notable en las ventas a partir de septiembre de 2023, cuando las cifras descendieron a 1,292 unidades, marcando el nivel más bajo en el periodo analizado. Aunque las causas de esta disminución no son completamente claras, podrían estar relacionadas con factores macroeconómicos, cambios en las tendencias de consumo o una menor efectividad de las campañas de marketing.

El año 2022, en general, fue especialmente positivo, mostrando un crecimiento sostenido en las ventas y un pico destacado en diciembre. Las cifras de este año superaron consistentemente las de 2021, lo que sugiere una recuperación en la demanda o una mejora en las estrategias de ventas y marketing implementadas por la empresa.

Sistema de Análisis de Series Temporales

Para facilitar la comprensión del sistema de predicción de ventas desarrollado, resulta fundamental representar su estructura mediante un diagrama paso a paso. La complejidad de los procesos involucrados, que abarcan desde la preparación de los datos hasta la evaluación y selección del mejor modelo de predicción, puede resultar difícil de interpretar solo a través de

una descripción escrita. El diagrama permite sintetizar de forma visual los pasos clave, mejorando la claridad y accesibilidad del sistema, permitiendo que las empresas optimicen sus estrategias en función de pronósticos confiables y efectivos.

Conclusiones

Este proyecto permitió generar información útil y estructurada a partir de los datos proporcionados, facilitando un análisis más profundo del comportamiento del producto estudiado. Si bien no se midió directamente el impacto comercial de las estrategias propuestas, los resultados obtenidos evidencian el potencial del enfoque aplicado para apoyar la toma de decisiones. Además, la metodología desarrollada puede adaptarse y aplicarse a otros productos que presenten patrones de comportamiento similares, lo cual representa una oportunidad para comerciantes, emprendedores y negocios que gestionan líneas de productos destinados al consumo masivo.

Los análisis realizados muestran que las ventas del producto estrella presentan una ligera estacionalidad, influenciada por factores como festividades y fluctuaciones propias de temporadas altas y bajas. Este comportamiento subraya la importancia de incorporar estas características en los modelos de predicción para optimizar su precisión. Asimismo, la variabilidad en las ventas mensuales, reflejada en la desviación estándar y el rango de los datos, destaca la necesidad de considerar factores externos como promociones, ajustes de precios y competencia al evaluar la exactitud de los modelos.

La inclusión de variables exógenas, como la estacionalidad específica de los meses de diciembre y enero, mejoró considerablemente la precisión del modelo SARIMAX. Esto pone de manifiesto la relevancia de realizar un análisis detallado de las características de los datos y su impacto en los resultados. Aunque SARIMAX fue el modelo más adecuado en este caso, es importante señalar que las predicciones siempre conllevan cierto nivel de incertidumbre y están sujetas a cambios en el entorno o eventos inesperados.

Por otro lado, Power BI fue una herramienta fundamental para explorar y visualizar los datos de forma clara y accesible. Su utilidad radica en facilitar la identificación de patrones y tendencias relevantes para el análisis. No obstante, la disminución en las ventas durante septiembre de 2023 requiere una investigación exhaustiva para determinar sus causas y diseñar estrategias correctivas que permitan prevenir situaciones similares en el futuro.

El sistema propuesto se caracteriza por su flexibilidad, ya que puede adaptarse a distintos tipos de series temporales e incorporar nuevas variables exógenas. Esto lo convierte en una solución versátil y aplicable a diversos contextos empresariales, con el potencial de generar valor en diferentes industrias.

Recomendaciones

Es importante investigar eventos especiales como promociones, cambios en el menú o competencia local que puedan haber influido en las ventas, así como analizar la correlación con variables externas como clima, eventos deportivos o indicadores económicos. Segmentar los datos por criterios específicos (días de la semana o tipo de cliente) puede revelar patrones útiles. Además de ARIMA, SARIMAX y Prophet, se sugiere explorar modelos avanzados como RNN o bosques aleatorios, así como modelos híbridos que combinen estadística y aprendizaje automático.

Para evaluar y mejorar los modelos, utiliza validación cruzada, intervalos de confianza y monitoreo regular para reentrenarlos según sea necesario. También es clave integrar el sistema de predicción con la planificación empresarial, desarrollar alertas para desviaciones significativas y presentar resultados claros mediante visualizaciones accesibles. Por último, se recomienda crear una interfaz intuitiva para facilitar el uso del sistema a usuarios no técnicos.

Referencias

- Angulo, D., & Ruiz, J. (2023). Aplicacion de modelos de series temporales para la predicción de la demanda de energía eléctrica en Colombia. *Revista de Ingenieria Industrial*, 44(2), 195-205.
- Arredondo, R. &. (2013). *Métodos de pronóstico: aplicaciones en la empresa*. Londres: Pearson.
- Avendaño, K. A., & Vega, V. (2020). *Diseño de un modelo de pronósticos de demanda para mejorar el stock de una empresa importadora de productos para laboratorio clínico*. Guayaquil: ESPOL.
- Biswal, A. (17 de octubre de 2024). *Simplilearn.¿Qué es Power BI? Arquitectura y características explicadas*: <https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi>
- Boada, A. J. (2017). Sistema de proyección de la demanda. Caso práctico de predicción automatizada en empresas de venta por catálogo. *Revista Perspetiva Empresarial*, 23-41.
- Box, G. P., Jenkins, G. M., & Reisel, G. C. (2015). *Análisis de series temporales: pronósticos y control*. Wiley.
- Contreras, A., Atziry, C., Martínez, J. L., & Sánchez, D. (2016). Análisis de series de tiempo en los pronósticos de la demanda de almacenamiento de productos perecederos. *Estudios Gerenciales*, 387-396.
- Cruz, R. S., & Correa, C. (2017). Previsión de demanda intermitente con métodos de series de tiempo y redes neuronales artificiales: Estudio de caso. *Dina*, 9-16.
- Fernández, R., Vilalta, J. A., & Quintero, A. (2019). Una revisión crítica sobre modelos de predicción para la demanda turística. *TURYDES*, 12.

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Pronóstico: principios y práctica*. Melbourne, Australia.: Otextos.
- Lizana, M. (4 de Septiembre de 2023). *medium.com*. Análisis predictivo en Power BI. Pronóstico. Medio.: <https://medium.com/@akanemar/an%C3%A1lisis-predictivo-en-power-bi-forecast-3f37dbc47021>
- Llanes, R. P., & Benítez, N. G. (2020). Una revisión bibliográfica sobre modelos para predecir las producciones de leche. *Revista Ingeniería Agrícola*, 10.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2020). *Pronóstico: Métodos y aplicaciones (5.a ed.)*. Hoboken,: John Wiley & Sons.
- Microsoft. (25 de 04 de 2024). *Tutorial: Creación de un modelo de Machine Learning en Power BI*. learn.microsoft.com: <https://learn.microsoft.com/es-es/power-bi/connect-data/service-tutorial-build-machine-learning-model>
- Molina, K. G. (2020). *Implementacion de un modelo analitico para la predicción de la venta del portafolio de productos OTC de un laboratorio farmacéutico*. Salgolquí.
- Montgomery, D. C., & Johnson, L. A. (2021). *Pronóstico y análisis de series temporales*. Hoboken, Nueva Jersey: Wiley.
- Moreno, R. V. (2022). Pronóstico de la producción de leche en Panamá mediante el uso de series de tiempo. *Visión Antataura*,, 138-156.
- Pankratz, A. (1991). *Pronóstico con modelos univariados de Box-Jenkins: conceptos y casos*. Nueva York: Wiley.
- Rivero, A., Arciniegas, S., & Fernández, M. (2022). Un modelo para predecir la demanda en farmacias. *Redmarka*, 1-14.

- Saldarriaga, E. A., Álvarez, M., & Álvarez, M. (2023). *Análisis de reducción del tiempo del plan de demanda en Pintuco Medellín*. Medellín: Corporación Universitaria Minuto de Dios.
- Sepúlveda, C. A., & Benavides, M. T. (2023). *Análisis predictivo de demanda de servicios bajo series temporales*. Medellín: Universidad de Antioquia.
- Taylor, J. (2018). *Pronóstico: principios y práctica*. Melbourne: OTexts.
- Tsay, R. (2005). *Análisis de series de tiempo: sus aplicaciones*. Boston, Massachusetts: Thomson Learning.
- Velasteguí, M. Y., & Argoti, C. M. (2023). *Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en el cantón Huaca Ecuador*. Tulcán, Ecuador: Universidad Politécnica Estatal del Carchi.