

Escuela de Negocios
Tipo de documento: Tesis de maestría



Master in Management + Analytics

Desarrollo e implementación de un sistema automatizado de pronóstico de demanda dentro del sector de la venta de neumáticos

Autoría: Montamat, Matías Exequiel

Año: 2025

¿Cómo citar este trabajo?

Montamat, M. (2025) "*Desarrollo e implementación de un sistema automatizado de pronóstico de demanda dentro del sector de la venta de neumáticos*". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella.

<https://repositorio.utdt.edu/handle/20.500.13098/13746>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



UNIVERSIDAD
TORCUATO DI TELLA

Master in Management + Analytics

***Desarrollo e implementación de un sistema automatizado
de pronóstico de demanda dentro del sector de la venta de
neumáticos***

Un enfoque estratégico para la mejora del negocio: Comparación de técnicas de ML para la mejora en la precisión del modelo de previsión de demanda y la optimización de la gestión de inventarios.

Autor: Montamat, Matías Exequiel

Tutora: Cancelo, Teresa Inés

Resumen.

La predicción precisa de la demanda es un componente clave para la gestión eficiente de la cadena de suministro en cualquier organización. El objetivo principal de esta tesis es desarrollar e implementar una solución basada en técnicas estadísticas y de machine learning para una empresa que actualmente depende en gran medida de la intervención manual de sus propietarios al realizar sus pronósticos de demanda.

Con este fin, se propone comparar diversas metodologías de pronóstico: Media Móvil, ARIMA, SARIMA, Prophet y XGBoost. Estos modelos serán evaluados para determinar su eficacia en la predicción de la demanda en el contexto específico de la empresa.

El propósito es ofrecer una herramienta que no solo aumente la precisión de los pronósticos, sino que también automatice el proceso, contribuyendo a una toma de decisiones más informada, eficiente y estratégica en la planificación de la cadena de suministro y la gestión de inventarios.

Los resultados de esta investigación tienen el potencial de transformar significativamente la eficiencia y rentabilidad de la empresa dentro de un mercado altamente competitivo.

En última instancia, el trabajo pretende ser una contribución innovadora en la elaboración de pronósticos basados en enfoques cuantitativos, representando un avance en la mejora continua de las operaciones de la empresa. Aprovechando la experiencia y el conocimiento interno de la organización, se busca no solo diferenciarla de la competencia, sino también establecer una base sólida para la futura incorporación de tecnologías de información más avanzadas en sus operaciones diarias.

Abstract.

Precise demand forecasting is essential for efficient supply chain management in any organization. The main objective of this thesis is to develop and implement a solution based on statistical and machine learning techniques for a company that currently relies heavily on the manual intervention of its owners in forecasting demand.

This project aims to enhance the forecasting process by comparing different forecasting techniques: Moving Average, ARIMA, SARIMA, Prophet, and XGBoost. These models will be evaluated to determine their effectiveness in predicting the company's demand.

The goal is to provide a tool that not only improves forecast accuracy but also automates the process, supporting more informed, efficient, and strategic decision-making in supply chain planning and inventory management.

The results of this research have the potential to significantly enhance the company's operational efficiency and profitability in a highly competitive market.

Ultimately, this thesis aims to be an innovative contribution to forecasting methodologies based on quantitative approaches, representing a step forward in the continuous improvement of the company's operations.

By leveraging the organization's internal experience and knowledge, the project seeks not only to differentiate the company from its competitors but also to establish a solid foundation for the future integration of more advanced information technologies into its daily operations.

Índice

I. Introducción.	5
1. Contexto.	5
2. Problema.	8
3. Objetivo de la investigación.	9
II. Marco Teórico.....	10
1. Pronósticos.	10
1. Horizonte de tiempo.....	10
2. Nivel.....	11
3. Elaboración.....	12
4. Ejecución.....	13
2. Importancia de la detección de patrones en los datos.	14
3. Modelos Predictivos Cuantitativos.....	15
4. Consideraciones para la selección de modelos predictivos.	26
5. Historia	29
III. Metodología.	31
1. Datos.....	31
1.1 Descripción y estructura del conjunto de datos.....	31
1.2 Procesamiento de datos.....	32
IV. Desarrollo.	34
1. Análisis descriptivo de los datos.....	34
2. Implementación de los modelos de pronóstico.	45
2.1 Modelo Benchmark: Naive Last Value.....	48
2.2 Media Móvil (Moving Average – MA).....	53
2.3 Prophet.....	57
2.4 ARIMA.....	61
2.5 SARIMA.....	65
2.6 XGBoost.....	69
V. Conclusión.	74
VI. Reflexiones finales.....	76
VII. Bibliografía.	78

I. Introducción.

1. Contexto.

Michelin es una empresa francesa reconocida como uno de los mayores fabricantes de neumáticos del mundo, operando durante más de 100 años. La compañía es célebre por la calidad e innovación en la fabricación de neumáticos para una amplia gama de vehículos, que incluyen automóviles, camiones, motocicletas y bicicletas. Además de su actividad principal en la producción de neumáticos, Michelin también está involucrada en el ámbito gastronómico, siendo conocida por otorgar las prestigiosas estrellas Michelin a restaurantes destacados a través de sus guías gastronómicas.

La entidad objeto de estudio es una filial de Michelin ubicada en la provincia de Córdoba, Argentina, con más de 30 años de trayectoria en la industria. Esta filial se dedica principalmente a la comercialización de neumáticos y otros accesorios automotrices, ofreciendo también servicios especializados de alineación y balanceo. La investigación se centrará en el sector de comercialización de neumáticos, que se considera el área de mayor relevancia económica dentro de la organización.

El negocio mantiene conexiones estratégicas con una variedad de clientes, desde conductores individuales hasta empresas grandes de transporte; y establece asociaciones clave con sus proveedores para asegurar la disponibilidad y calidad de sus productos.

El mercado de neumáticos en Argentina ha experimentado una evolución significativa, influenciada por cambios en las tendencias económicas, tecnológicas y de consumo (McKinsey & Wmc, 2023).

Inicialmente, la liberalización económica junto con la apertura de mercados facilitó la introducción de tecnologías avanzadas y la entrada de fabricantes internacionales, lo que incrementó la demanda de neumáticos para automóviles y camiones ligeros.

Posteriormente, la industria se diversificó, adoptando neumáticos radiales de mayor durabilidad, impulsada por políticas de liberalización y tecnologías más eficientes. Con el advenimiento del nuevo milenio, la conciencia ambiental y la eficiencia en el consumo de combustible se convirtieron en prioridades, fomentando una mayor competencia e innovación en el diseño y fabricación de neumáticos, incluyendo la introducción de tecnologías como los sensores de presión de neumáticos (TPMS).

En tiempos más recientes, el mercado ha enfrentado desafíos económicos que han impactado la demanda, destacándose un mayor énfasis en la relación calidad-precio y en la seguridad vial.

Por otro lado, la industria argentina del neumático ha atravesado diversas fases de crecimiento y contracción. Históricamente, el mercado ha estado marcado por la presencia tanto de fabricantes locales como de marcas internacionales. En los últimos años, el sector ha experimentado un resurgimiento, impulsado por el aumento en la producción de automóviles y el crecimiento del parque automovilístico, lo que ha elevado la demanda de neumáticos tanto de fabricantes de equipos originales (OEM) como de reposición.

Actualmente, el mercado de neumáticos en Argentina enfrenta una dualidad de desafíos y oportunidades. La transición hacia vehículos eléctricos, la digitalización de la cadena de suministro y las regulaciones ambientales más estrictas están configurando las tendencias emergentes. Además, la pandemia de COVID-19 ha alterado los patrones de

movilidad y la demanda de neumáticos, obligando a la industria a adaptarse a nuevas realidades.

En cuanto a los proveedores de la empresa, estos también enfrentan la inestabilidad económica del país, lo que dificulta la certeza sobre la recepción de los productos demandados. Los nuevos regímenes de exportación podrían desviar materias primas hacia el exterior, mientras que las restricciones a las importaciones podrían paralizar la producción de ciertos productos.

De acuerdo con las perspectivas futuras del mercado de neumáticos en Argentina (Markets&data, 2024), estudios realizados anticipan un crecimiento continuo impulsado por diversos factores. Uno de estos factores es la recuperación económica; a medida que la economía se estabilice, se espera que el sector automotriz se recupere, lo que, a su vez, incrementará la demanda de neumáticos.

Además, la innovación y la tecnología desempeñarán un papel crucial. Los avances continuos en la tecnología de neumáticos, especialmente en áreas como sostenibilidad y rendimiento, generarán nuevas oportunidades para los fabricantes.

Por último, la expansión del mercado se verá favorecida por el aumento de inversiones por parte de empresas locales e internacionales en instalaciones de producción y redes de distribución, mejorando así la capacidad y el alcance del mercado.

A través de esta introducción, se ha buscado explorar e ilustrar las diversas facetas que presentan considerables desafíos en relación con una estimación precisa de las cantidades de producto demandadas.

2. Problema.

El desafío central abordado por esta tesis radica en la limitación del enfoque actual para el pronóstico de ventas de la empresa. Actualmente, la predicción se basa exclusivamente en las unidades vendidas en los mismos meses del año anterior al que se desea prever (por ejemplo, para proyectar abril de 2023, se utilizan las unidades vendidas en abril de 2022).

Técnicamente, este enfoque se conoce como el modelo "Naive Last Value 12 (mensual)". Aunque este método puede ser viable en condiciones de demanda estable, se revela como insuficiente en un contexto de demanda dinámico y fluctuante, como el que se enfrenta actualmente.

El negocio adquiere neumáticos nacionales e importados para su posterior comercialización a consumidores finales. Actualmente, enfrenta un problema crítico debido a un exceso significativo de inventario en prácticamente todas las líneas de productos. Esta acumulación responde a una desaceleración de las ventas en comparación con períodos anteriores, lo que ha generado un desfase entre la oferta disponible y la demanda efectiva.

Esta situación conlleva elevados costos de almacenamiento, que abarcan tanto los gastos operativos asociados al uso del espacio físico, mantenimiento y seguros, como el riesgo de obsolescencia o deterioro de los neumáticos.

En este contexto, se vuelve imperativo optimizar el proceso de pronóstico de demanda, a fin de alinear de forma más precisa las decisiones de abastecimiento con las necesidades reales del mercado. Mejorar la precisión de las proyecciones permitirá reducir el sobrestock, para minimizar costos innecesarios y planificar de manera más eficiente la gestión integral de la cadena de suministro.

3. Objetivo de la investigación.

El método de pronóstico actual no logra reflejar adecuadamente las tendencias del mercado. Su principal debilidad radica en la escasa utilización de datos históricos, lo que impide capturar una visión más completa y precisa de los patrones de demanda a lo largo del tiempo. Esta falta de profundidad en el análisis deriva en estimaciones incompletas y sesgadas, con consecuencias directas en términos de pérdidas operativas.

El objetivo central de esta investigación consiste en desarrollar y evaluar modelos avanzados de pronóstico de demanda, incorporando técnicas que aprovechen los datos históricos de ventas en horizontes temporales más amplios.

La innovación propuesta radica en la implementación de técnicas de pronóstico más sofisticadas, capaces de capturar patrones complejos y variaciones temporales en la demanda. Estos modelos no solo serán diseñados y desarrollados, sino también sometidos a una evaluación rigurosa, con el fin de validar su eficacia y relevancia práctica en el contexto particular de la empresa.

Una vez desarrollados e implementados los modelos, se planea integrarlos en el proceso de toma de decisiones, para que estos funcionen como un componente esencial previo a la determinación de las cantidades de producto a adquirir.

La solución inicialmente tomará la forma de un código (*script*), diseñado para cargar los datos pertinentes y generar pronósticos de demanda de manera automatizada. Este enfoque proporcionará una herramienta práctica y eficiente para la implementación y prueba inicial de los modelos, ofreciendo una base operativa sólida antes de una eventual integración más completa en el sistema de gestión de la empresa.

II. Marco Teórico.

1. Pronósticos.

Los pronósticos juegan un papel crucial en la planificación y toma de decisiones empresariales, al proporcionar estimaciones sobre valores futuros de variables relevantes. Se pueden definir como "proyecciones fundamentadas en el análisis de datos históricos, modelos matemáticos o la experiencia de expertos en el campo" (Montgomery, 2015). Otra definición acertada es la de "realizar un enunciado sobre el valor futuro de una variable de interés, fundamentado ya sea por el análisis de datos históricos disponibles, el juicio de expertos en el tema o una combinación de ambas cosas" (Montemayor Gallegos, 2014).

1. Horizonte de tiempo.

Se pueden comenzar diferenciando los pronósticos según su horizonte de tiempo, lo que influye en su precisión y enfoque estratégico.

1.1 Corto Plazo

Los pronósticos a corto plazo (hasta 3 meses) son cruciales para la gestión operativa y la planificación táctica, suelen utilizarse para predecir la demanda de productos y gestionar inventarios de manera precisa (Gardner et al., 2018).

1.2 Mediano plazo.

Los pronósticos a mediano plazo (3 meses a 3 años) son trascendentes para la planificación financiera y estratégica (Chatfield, 2016). Estos pronósticos ayudan en la elaboración de presupuestos y en la identificación de oportunidades de crecimiento.

1.3 Largo plazo.

Los pronósticos a largo plazo (más de 3 años) se enfocan en decisiones estratégicas de alto nivel, considerando factores macroeconómicos y tecnológicos, lo que permite a las empresas adaptarse a cambios disruptivos en la industria (Wagner et al., 2019).

2. Nivel.

Los pronósticos son fundamentales para la toma de decisiones en todos los niveles organizacionales, abarcando desde operaciones diarias hasta estrategias a largo plazo (Fildes et al., 2019).

2.1 Operativo.

A nivel operativo, los pronósticos se emplean para la gestión de inventarios, producción, cadena de suministro y atención al cliente. Se enfocan en satisfacer la demanda del mercado de manera eficiente, minimizando costos y maximizando la satisfacción del cliente.

2.2 Táctico.

A nivel táctico, los pronósticos apoyan la implementación de decisiones estratégicas, definiendo políticas y planes detallados en ventas, producción, compras y marketing. Ayudan a optimizar la asignación de recursos y a gestionar operaciones diarias, alineando las decisiones tácticas con los objetivos estratégicos de la organización.

2.3 Estratégico.

A nivel estratégico, los pronósticos guían decisiones de alto nivel para la dirección futura de la organización, considerando cambios en el entorno empresarial. Son cruciales para formular estrategias competitivas e identificar oportunidades de crecimiento a largo plazo.

Pronosticar en niveles operativos, tácticos y estratégicos es esencial para el éxito organizacional, ya que ofrece una visión coherente del futuro y facilita decisiones informadas.

3. Elaboración.

Hanke (2010) distingue el procedimiento de elaboración de un pronóstico, en dos enfoques; uno cualitativo y otro cuantitativo.

3.1 Enfoque Cualitativo.

Los pronósticos cualitativos, también conocidos como " Managerial Judgement" (juicio gerencial), son ampliamente utilizados por gerentes y propietarios de empresas. Estos métodos se basan en la experiencia y el conocimiento del mercado y los productos, sin necesidad de datos históricos. Aunque pueden incorporar elementos cuantitativos básicos, su eficacia disminuye con la complejidad y la cantidad de productos. A pesar de esto, es el método más utilizado según el International Journal of Forecasting (Fildes et al., 2019), basado en investigaciones de Peterson (1993) y McCarthy et al. (2006).

3.2 Enfoque Cuantitativo.

Los métodos cuantitativos representan un enfoque analítico para la elaboración de pronósticos, utilizando modelos computacionales basados en principios matemáticos y estadísticos. Estos métodos permiten analizar datos históricos y proyectar tendencias futuras con precisión, adaptándose a los datos disponibles para replicar la realidad de manera exacta.

Dentro de los métodos cuantitativos, se distinguen los univariados y multivariados. Los métodos univariados se enfocan en una única variable y su relación con valores pasados, siendo adecuados para situaciones donde la variable de interés depende principalmente de su historial, como en series temporales de demanda de productos. En contraste, los métodos multivariados consideran la influencia de múltiples variables en la variable de interés, siendo útiles para pronósticos complejos que se ven afectados por diversos factores, como en el caso de pronósticos económicos impactados por factores macroeconómicos.

4. Ejecución.

Finalmente, los pronósticos pueden clasificarse según la forma en la que se lleven a cabo, ya sea por producto o de manera agregada (Fildes et al., 2019).

4.1 Pronósticos por producto.

Los pronósticos por producto son esenciales en la gestión de la demanda, permitiendo estimar con precisión las necesidades futuras de productos individuales o categorías específicas en el corto plazo. Este enfoque detallado busca minimizar tanto la escasez como el exceso de inventario, optimizando la satisfacción del cliente y la eficiencia operativa.

4.2 Pronósticos agregados.

Los pronósticos agregados son fundamentales para gestionar y comprender la demanda en un nivel macroeconómico. Este enfoque se centra en la estimación de la demanda total de productos o servicios, agrupando múltiples productos o categorías según criterios temporales, geográficos u otros relevantes para la organización.

2. Importancia de la detección de patrones en los datos.

Previo a la selección de un modelo para llevar a cabo pronósticos, se debe realizar un análisis exhaustivo de los datos. Este proceso es crucial para identificar patrones que puedan estar presentes en los datos, lo que a su vez mejora la precisión de los pronósticos. En el análisis de series temporales, es posible reconocer distintos tipos de patrones estructurales que constituyen la base para una modelización y un pronóstico efectivos (Box et al., 2015).

Por un lado, tenemos los patrones horizontales o estacionarios, que se distinguen por la estabilidad de los datos, donde las observaciones se agrupan alrededor de una media constante, sin una tendencia clara, y con variabilidad aleatoria. Identificar la estacionariedad es fundamental, ya que muchos modelos de pronóstico asumen que los datos son estacionarios para poder aplicarse correctamente (Hyndman & Athanasopoulos, 2021).

En contraste, los patrones de tendencia se manifiestan como cambios persistentes en una dirección específica a lo largo del tiempo, indicando períodos de crecimiento o decrecimiento. La identificación de estas tendencias sirve para interpretar la dirección general de los datos y ajustar los modelos predictivos en consecuencia.

Por otro lado, los patrones estacionales reflejan variaciones cíclicas que se presentan en momentos específicos dentro de un período fijo, como días, semanas, meses o años. Estas variaciones suelen estar influenciadas por factores estacionales, como cambios climáticos o eventos recurrentes, y su correcta modelización capta la variabilidad inherente en los datos.

Por último, los patrones cíclicos se diferencian de los estacionales por manifestarse como fluctuaciones en intervalos más largos y menos predecibles. Estos ciclos generalmente resultan de factores como fluctuaciones económicas o eventos históricos. Su identificación facilita la interpretación de las tendencias a largo plazo en los datos.

3. Modelos Predictivos Cuantitativos.

Los métodos predictivos cuantitativos emplean modelos computacionales y técnicas estadísticas para analizar datos históricos y proyectar tendencias futuras con precisión. Existe una gran variedad de modelos predictivos utilizados en la generación de pronósticos; pero la gran diferenciación se da entre los univariados y los multivariados.

3.1 Univariados.

Los métodos de pronóstico cuantitativos univariados se caracterizan por predecir el valor de una variable en función de sus valores pasados, lo que los convierte en modelos de análisis de series de tiempo. Por ejemplo, para prever las ventas mensuales de un producto, se puede considerar su historial de ventas pasado.

Estudios previos han demostrado que estos modelos son altamente efectivos para abordar problemas de predicción en contextos donde se dispone de datos históricos consistentes y donde se requiere una solución de bajo costo computacional (Ali et al., 2009). Se caracterizan por su simplicidad estructural, su facilidad de implementación, y su capacidad para capturar patrones relevantes como tendencias, estacionalidades y ciclos. Además, resultan particularmente útiles en entornos con disponibilidad limitada de información, permitiendo generar pronósticos confiables de manera ágil y eficiente.

3.1.1 Media Móvil (Moving Average - MA).

El modelo de Media Móvil es una técnica ampliamente reconocida en el análisis y pronóstico de series temporales. Su principal objetivo es capturar las tendencias y patrones de variación mediante el cálculo de promedios de observaciones pasadas. Aunque su enfoque es relativamente simple, resulta eficaz y constituye una herramienta valiosa para la predicción de la demanda. (Montgomery, D.C., Jennings, C.L., & Kulahci, 2008).

El funcionamiento del modelo se basa en la determinación de promedios de datos históricos en intervalos de tiempo específicos para proyectar valores futuros. Este modelo se concentra en las tendencias generales de la serie temporal, suavizando las fluctuaciones aleatorias presentes en los datos. Esto permite una identificación más clara de los patrones subyacentes y facilita pronósticos más precisos. Matemáticamente se expresa como:

$$\hat{Y}_t = \frac{1}{n} \sum_{i=1}^n Y_{t-i}$$

Donde \hat{Y}_t es la predicción para el periodo de tiempo t ; Y_{t-i} son las observaciones pasadas hasta n períodos anteriores.

El modelo de Media Móvil predice valores futuros calculando el promedio simple de las observaciones hasta n períodos anteriores. El valor de n se selecciona según la estacionalidad y volatilidad de la serie temporal, y su elección óptima puede determinarse mediante técnicas de validación cruzada o análisis de residuos.

Una de las principales fortalezas de este modelo es su simplicidad y facilidad de implementación. Al calcular promedios de observaciones pasadas, el modelo suaviza las fluctuaciones aleatorias presentes en los datos, facilitando la identificación de tendencias a corto plazo. Esta característica lo hace especialmente útil para pronósticos rápidos y simples en situaciones que requieren resultados inmediatos.

No obstante, este modelo también presenta limitaciones. Una de las más significativas es su incapacidad para capturar relaciones de dependencia entre observaciones pasadas y futuras. Al centrarse exclusivamente en promedios de observaciones pasadas, puede no ser adecuado para series temporales con tendencias complejas o relaciones no lineales entre variables. Además, el modelo puede ser sensible a valores atípicos o fluctuaciones extremas en los datos.

3.1.2 Prophet.

El modelo Prophet, desarrollado por el equipo de científicos de datos de Facebook, es un enfoque aplicado en el análisis de series temporales, diseñado para ofrecer flexibilidad en una variedad de contextos empresariales. Su principal fortaleza radica en su capacidad para adaptarse a diferentes tipos de negocios, convirtiéndolo en una herramienta valiosa tanto para usuarios con experiencia limitada como para aquellos con un conocimiento más profundo en análisis de datos. Prophet está diseñado para ser configurable incluso por personas con conocimientos básicos en la generación de datos, sin necesidad de una comprensión extensa de modelos de series temporales. Esto facilita la utilización del modelo para la toma de decisiones informadas en la predicción de ventas (Taylor & Letham, 2017).

Este modelo se basa en un enfoque descomponible de series temporales (Harvey & Peters, 1990), que integra tres componentes principales: tendencia, estacionalidad y días festivos. La ecuación resultante del modelo es la siguiente:

$$Y(t) = g(t) + s(t) + h(t) + et$$

Siendo $g(t)$ la función de tendencia que modela cambios no periódicos en el valor de la serie de tiempo; $s(t)$ representando cambios periódicos (por ejemplo, estacionalidad semanal, mensual y/o anual); $h(t)$ representando los efectos de días festivos. Finalmente, et es el término de error que representa cualquier cambio idiosincrásico que no se adapte al modelo.

En términos de eficiencia computacional, el ajuste del modelo Prophet se realiza de manera rápida y efectiva. Este modelo no solo captura tendencias no lineales, sino que también incorpora estacionalidades anuales, semanales y diarias, así como efectos de días festivos y

vacaciones. Su rendimiento es óptimo en series temporales con alta estacionalidad y grandes volúmenes de datos históricos.

Además, Prophet muestra una notable robustez frente a datos faltantes, cambios en las tendencias y valores atípicos, lo que le permite mantener un desempeño sólido y confiable incluso en situaciones con incompletitud de datos, alteraciones abruptas en las tendencias históricas de demanda, o anomalías que puedan distorsionar los patrones de comportamiento. El modelo facilita la participación de especialistas en el negocio al permitirles aplicar su conocimiento a través de un conjunto de parámetros y opciones intuitivas, mientras mantiene automatizadas las funciones estadísticas y computacionales (Taylor & Letham, 2017).

Cada componente del modelo puede ser ajustado de acuerdo con las experiencias y preferencias de los usuarios, permitiendo una personalización según las necesidades específicas del contexto (Hills, 2020).

3.1.3 ARIMA.

El modelo ARIMA (Autoregressive Integrated Moving Average) es una técnica estadística utilizada para el análisis y predicción de series temporales. Este modelo combina tres componentes: autorregresivo (AR), integración (I) y media móvil (MA), lo que permite capturar tanto propiedades estacionarias como no estacionarias de una serie temporal (Shumway & Stoffer, 2017).

El modelo autorregresivo (AR) captura la relación lineal entre el valor actual de una serie temporal y las observaciones anteriores, modelando el valor presente como una combinación lineal de valores pasados, ponderados por coeficientes específicos. Este componente permite detectar patrones de autocorrelación en los datos, reflejando la influencia de datos históricos recientes. Se puede expresar matemáticamente como:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

donde y_t es el valor de la serie en el tiempo t , C es una constante, ϕ_1 a ϕ_p son los coeficientes de auto regresión, y ε_t es el término de error aleatorio en el tiempo t .

El término de integración (I) se refiere a la necesidad de diferenciar la serie temporal para que sea estacionaria, es decir, para mantener constante la media y la varianza a lo largo del tiempo. Si una serie temporal no es estacionaria, se aplican diferenciaciones para lograr esta propiedad y asegurar un análisis más fiable.

Por ejemplo, la primera diferencia se puede expresar como:

$$\Delta Y_t = Y_t - Y_{t-1}$$

donde ΔY es la primera diferencia entre el valor de la serie en el tiempo t y el valor de la serie en el tiempo $t - 1$.

Por último, el modelo de promedios móviles (MA) captura la relación lineal entre una observación y el error residual de un modelo de media móvil aplicado a observaciones pasadas. En otras palabras, considera los errores residuales de los pasos de tiempo anteriores para predecir el valor actual. Este componente ayuda a capturar patrones de variación aleatoria en los datos que no pueden explicarse mediante la autocorrelación temporal. Se puede expresar matemáticamente como:

$$y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

donde y_t es el valor de la serie en el tiempo t , C es una constante, θ_1 a θ_q son los coeficientes de la media móvil, y ε_t es el término de error aleatorio en el tiempo t .

Se combinan los componentes AR, I, MA para crear un modelo que puede modelar tanto las tendencias como la variabilidad aleatoria en la serie de tiempo. El modelo ARIMA se puede expresar matemáticamente de la siguiente manera:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Este modelo es ampliamente reconocido y utilizado en el análisis y pronóstico de series temporales, destacándose por su robustez y eficacia en la generación de predicciones precisas, especialmente a corto plazo. Esta capacidad es crucial para obtener estimaciones confiables en horizontes temporales cercanos.

Una de las ventajas significativas del modelo ARIMA es su capacidad para modelar datos con tendencias y estacionalidades, permitiendo capturar y prever estos ciclos con alta precisión. Además, este método es frecuentemente aplicado en el análisis de series temporales mensuales.

3.1.4 SARIMA.

El modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) es una extensión del modelo ARIMA que integra componentes estacionales, lo que lo hace especialmente valioso para analizar series temporales con patrones estacionales recurrentes (Box, Jenkins, Reinsel, & Ljung, 2015).

SARIMA combina los componentes autorregresivos (AR), de integración (I), y de media móvil (MA), y además incorpora un componente estacional (S) que permite modelar variaciones periódicas en los datos. Este componente estacional se define mediante parámetros adicionales: el autorregresivo estacional (P), la diferenciación estacional (D), la media móvil estacional (Q), y s el período estacional. El modelo se expresa matemáticamente como:

$$Y_t = C + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^P \Phi_k Y_{t-ks} + \sum_{l=1}^Q \Theta_l \varepsilon_{t-ls} + \varepsilon_t$$

Donde Y_t es el valor de la serie en el tiempo t , C es una constante, ϕ_i son los coeficientes del componente autorregresivo no estacional de orden p , θ_j son los coeficientes del componente de media móvil no estacional de orden q , Φ_k son los coeficientes del componente autorregresivo estacional de orden P con período estacional s , Θ_l son los coeficientes del componente de media móvil estacional de orden Q con período estacional s , y ε_t es el término de error aleatorio en el tiempo t .

La capacidad del modelo SARIMA para capturar tanto la estacionalidad como los componentes no estacionales permite modelar eficazmente datos con ciclos repetitivos. Este enfoque es esencial para realizar predicciones precisas en contextos donde las fluctuaciones estacionales influyen significativamente en los patrones de comportamiento de la serie temporal.

Su flexibilidad para integrar elementos de autoregresión, diferenciación y promedios móviles en un contexto estacional lo convierte en una herramienta valiosa para ajustar modelos a datos con estructuras temporales complejas.

Esto facilita la obtención de estimaciones más precisas y ajustadas a las características específicas de series temporales con estacionalidad marcada y tendencias prolongadas.

3.2 Multivariados.

Los métodos de pronóstico cuantitativos multivariados se basan en la premisa de que la variable dependiente puede ser proyectada utilizando información de múltiples variables independientes. Estos modelos integran información de diversos factores relevantes, lo que permite enriquecer el análisis y capturar de forma más precisa la dinámica del fenómeno bajo estudio. Por ejemplo, para pronosticar las ventas mensuales de un producto, no solo se tiene en cuenta el historial de ventas, sino también variables como el precio, la estacionalidad o las condiciones económicas.

En función de la forma en que se relacionan las variables independientes con la dependiente, los métodos multivariados pueden clasificarse en lineales y no lineales.

Los métodos lineales se destacan por su sencillez y por la facilidad con la que permiten interpretar los efectos individuales de cada variable explicativa sobre el resultado. La regresión lineal, en sus diferentes variantes, constituye uno de los métodos más utilizados en este enfoque. Sin embargo, esta simplicidad puede resultar una limitación cuando se busca modelar fenómenos complejos o no lineales, donde las interacciones entre variables no se ajustan a una relación proporcional.

En contraste, los métodos no lineales permiten capturar relaciones más complejas entre las variables. Son muy útiles cuando la relación entre las variables es difícil de representar mediante una función lineal. Dentro de este grupo se encuentran los árboles de decisión y las redes neuronales, ampliamente utilizados en entornos de machine learning. Estos modelos suelen ofrecer una mayor capacidad predictiva al adaptarse mejor a los datos, pero son más difíciles de interpretar, lo que puede dificultar la comprensión directa del resultado observado.

3.2.1 Árboles de Decisión.

Los Árboles de Decisión constituyen un algoritmo de aprendizaje automático ampliamente utilizado en tareas tanto de clasificación como de regresión, debido a su capacidad para modelar relaciones complejas entre variables de forma intuitiva y estructurada. Este enfoque se basa en una representación jerárquica que segmenta iterativamente el conjunto de datos en subconjuntos más homogéneos, utilizando criterios de decisión definidos a partir de las características predictoras (Quinlan, 1986).

La construcción de un árbol de decisión comienza con un nodo raíz que abarca el conjunto total de observaciones, y progresa a través de nodos internos donde se realizan pruebas sobre variables seleccionadas según su capacidad para dividir eficazmente los datos. Cada rama representa el resultado de una prueba condicional, y los nodos terminales (hojas) asignan una clase específica o un valor numérico estimado, dependiendo de si la tarea es de clasificación o regresión. La selección de variables en cada nodo se basa generalmente en criterios de ganancia de información, reducción de varianza u otras métricas de impureza que optimizan la partición del espacio de datos.

Una de las principales fortalezas de los árboles de decisión radica en su alta interpretabilidad. Las reglas generadas son fácilmente comprensibles, lo cual resulta especialmente valioso en entornos donde la transparencia en la toma de decisiones es crítica. Además, estos modelos son inherentemente flexibles, permitiendo el tratamiento de variables tanto categóricas como continuas, y presentando una notable tolerancia ante valores faltantes.

No obstante, un desafío recurrente en el uso de árboles de decisión es su tendencia al sobreajuste cuando los árboles crecen sin restricciones, capturando ruido en lugar de patrones

generales. Para mitigar este problema, se recurre a técnicas de poda, que eliminan ramas irrelevantes o poco significativas, y al uso de métodos de ensamblado como los bosques aleatorios (*Random Forests*) o el *Gradient Boosting*, que combinan múltiples árboles con el objetivo de mejorar la precisión y la generalización del modelo (Breiman, 1986).

3.2.2 XGBoost.

XGBoost (eXtreme Gradient Boosting) es un algoritmo de aprendizaje automático basado en árboles de decisión ensamblados de manera secuencial mediante técnicas de boosting. Su eficacia y versatilidad lo han convertido en una de las herramientas más utilizadas en competencias de ciencia de datos y en aplicaciones del mundo real, siendo aplicable a una amplia variedad de problemas más allá del análisis de series temporales (Chen & Guestrin, 2016).

El principio fundamental de XGBoost consiste en construir árboles de decisión de forma iterativa. El primer árbol genera una predicción inicial, y los árboles subsiguientes se entrenan para corregir los errores residuales de sus predecesores. Este proceso continúa hasta alcanzar un modelo ensamblado optimizado, cuya capacidad predictiva supera significativamente la de un único árbol. La optimización del modelo se logra mediante la minimización de una función de pérdida (como el error cuadrático medio, MSE), lo que permite ajustar progresivamente los pesos asignados a las observaciones con mayor error en las etapas anteriores.

Entre sus características más destacadas se encuentra la capacidad para manejar grandes volúmenes de datos, así como su robustez frente a valores nulos o escasos (*sparse data*). Además, XGBoost incorpora técnicas de regularización tanto L1 como L2, lo cual resulta clave para evitar el sobreajuste y asegurar la generalización del modelo a datos no

vistos. Esta combinación de eficiencia computacional, precisión y control sobre el ajuste ha sido fundamental para su éxito en competencias como Kaggle y KDDCup 2015.

XGBoost requiere que todas las variables predictoras sean numéricas. Por lo tanto, en presencia de variables categóricas o de texto, se deben aplicar técnicas de transformación como el *one-hot encoding* para preparar adecuadamente los datos de entrada.

Asimismo, XGBoost ofrece distintos hiperparámetros que pueden ser ajustados para optimizar su rendimiento. Entre los más relevantes se encuentran:

- `nrounds`: especifica el número total árboles que serán ensamblados. A mayor cantidad de árboles, mayor capacidad de ajuste del modelo.
- `max_depth`: define la profundidad máxima para cada árbol. Limitar esta profundidad contribuye a reducir el riesgo de sobreajuste al evitar estructuras demasiado complejas.
- `eta` (learning rate): controla la tasa de aprendizaje del modelo. Una tasa más baja implica que cada árbol realiza ajustes más pequeños, lo cual puede mejorar la generalización si se acompaña de un mayor número de iteraciones.
- `gamma`: determina si una división adicional en un nodo (es decir, crear más ramas en un árbol) vale la pena o no desde el punto de vista de la mejora del modelo.

En conjunto, estas características convierten a XGBoost en una herramienta poderosa y adaptable, capaz de abordar problemas complejos con alta precisión y eficiencia, siendo una opción particularmente atractiva en entornos donde se requiere combinar velocidad de cómputo, escalabilidad y capacidad explicativa.

4. Consideraciones para la selección de modelos predictivos.

Durante el proceso de desarrollo y selección de un modelo predictivo, resulta fundamental considerar aspectos como el riesgo de sobreajuste, el costo computacional y la adecuada implementación de técnicas de validación y optimización. Estos factores son determinantes para garantizar que el modelo no solo ofrezca un alto rendimiento predictivo, sino que también funcione de manera eficiente, escalable y con un nivel adecuado de confiabilidad en sus predicciones (Murphy, 2012).

4.1 Overfitting.

El *overfitting*, o sobreajuste, se produce cuando un modelo se ajusta excesivamente a los datos de entrenamiento, llegando a capturar tanto los patrones reales como el ruido aleatorio, lo que deteriora su capacidad de generalización frente a datos nuevos (Hastie, Tibshirani & Friedman, 2009). Para mitigar este fenómeno, es fundamental limitar la complejidad del modelo mediante estrategias como la regularización, la selección adecuada de variables y el control del número de hiperparámetros entrenables (James et al., 2021).

4.2 Técnicas de validación.

La aplicación de técnicas de validación es un componente esencial en el desarrollo de modelos predictivos, ya que permite evaluar su capacidad de generalización frente a datos no observados. Esta etapa es clave para mitigar el riesgo de sobreajuste y garantizar un desempeño adecuado en contextos reales de aplicación (Kuhn & Johnson, 2013).

Una de las estrategias más utilizadas es la partición simple del conjunto de datos en subconjuntos de entrenamiento y validación, conocida como *holdout validation*. Este método permite estimar el rendimiento del modelo en datos no utilizados durante su ajuste inicial (Bishop, 2006).

Una alternativa más robusta es la *validación cruzada (cross-validation)*, en la que el conjunto de datos se divide en k subconjuntos o *folds*. En cada iteración, uno se utiliza como conjunto de prueba y los restantes para entrenamiento. Este proceso se repite k veces, y los resultados se promedian, proporcionando una estimación más estable y representativa del desempeño del modelo.

Finalmente, la *validación fuera de muestra (out-of-sample validation)* consiste en reservar un conjunto independiente, denominado *test set*, que no participa en ninguna etapa de entrenamiento o calibración. Esta reserva permite realizar una evaluación final objetiva del modelo, asegurando una evaluación imparcial en datos completamente nuevos.

4.3 Técnicas de optimización.

Las técnicas de optimización son fundamentales para seleccionar adecuadamente los hiperparámetros de un modelo predictivo. Los hiperparámetros son configuraciones ajustables del modelo que no se aprenden durante el entrenamiento, pero cuya elección influye directamente en el rendimiento final, afectando tanto la precisión como la capacidad de generalización del modelo (Hastie, Tibshirani & Friedman, 2009).

Entre las técnicas más utilizadas se encuentra la *búsqueda manual*, que, si bien puede resultar laboriosa y menos sistemática, permite al analista adquirir una comprensión más profunda del efecto de cada hiperparámetro sobre el desempeño del modelo.

Una alternativa más estructurada es la *búsqueda en cuadrícula (grid search)*, que evalúa de manera exhaustiva todas las combinaciones posibles dentro de un conjunto definido de valores. Si bien este enfoque garantiza una cobertura completa del espacio de búsqueda, puede implicar un alto costo computacional, especialmente cuando el número de hiperparámetros o sus posibles valores es elevado.

Por otro lado, la búsqueda aleatoria (*random search*) selecciona combinaciones de forma aleatoria dentro del espacio de hiperparámetros. Este método ha demostrado ser más eficiente en muchos contextos, ya que permite explorar regiones relevantes del espacio de búsqueda con menor esfuerzo computacional, especialmente cuando algunos hiperparámetros tienen mayor influencia que otros en el rendimiento del modelo.

4.4 Costo Computacional.

El costo computacional se refiere al conjunto de recursos requeridos para entrenar y ejecutar un modelo predictivo, incluyendo tiempo de procesamiento, uso de memoria y capacidad de almacenamiento. Este aspecto adquiere especial relevancia cuando se trabaja con grandes volúmenes de datos o con algoritmos de elevada complejidad computacional (Domingos, 2012).

Factores como la complejidad algorítmica, el tamaño del conjunto de datos y la eficiencia del código implementado son determinantes en el impacto computacional del proceso. En este sentido, se valoran enfoques que prioricen la eficiencia, como la paralelización de tareas o la utilización de arquitecturas distribuidas, especialmente cuando el modelo debe ser implementado en entornos productivos.

Si bien el costo computacional no suele ser el principal criterio al momento de seleccionar un modelo predictivo, su evaluación permite alcanzar un equilibrio entre la precisión del modelo y su viabilidad operativa. Este balance es esencial para garantizar la escalabilidad y sostenibilidad de las soluciones propuestas en el tiempo.

5. Historia

5.1 Evolución histórica de las técnicas de pronóstico de ventas.

El pronóstico de ventas ha sido una práctica indispensable en la gestión empresarial, proporcionando una base esencial para la toma de decisiones estratégicas. A lo largo de la historia, diversas técnicas y modelos han sido desarrollados y utilizados para predecir la demanda de productos y servicios en distintos sectores empresariales (McCarthy et al., 2006). Estas técnicas han experimentado una evolución significativa, impulsada por avances tecnológicos y cambios en el comportamiento del consumidor.

En las primeras etapas del desarrollo empresarial, los métodos de pronóstico de ventas se basaban principalmente en enfoques cualitativos y cuantitativos simples. Los métodos cualitativos incluían las opiniones de expertos y la investigación de mercado, que se centraban en la intuición y la experiencia para prever la demanda futura. En paralelo, los métodos cuantitativos, como el promedio móvil y el suavizado exponencial, se apoyaban en datos históricos para proyectar tendencias futuras.

Con el avance de la estadística y la mejora en la precisión de los datos, surgieron modelos más sofisticados. Entre estos se destacan la regresión lineal, y los modelos de series temporales. Estos modelos permitieron una mayor precisión en las predicciones al incorporar factores como la estacionalidad y las tendencias cíclicas.

En las últimas décadas, la irrupción del machine learning ha revolucionado el campo del pronóstico de ventas. Algoritmos como las redes neuronales, los árboles de decisión y el aprendizaje profundo han habilitado el análisis de grandes volúmenes de datos y la identificación de patrones complejos que influyen en la demanda.

5.2 Tendencias y enfoques modernos.

Para proporcionar una visión de las técnicas de pronóstico utilizadas hoy en día, se llevó a cabo una exhaustiva investigación de las principales publicaciones en el ámbito de la predicción dentro del sector minorista “Retail”.

En este contexto, se recopiló información relevante proporcionada por una de las competiciones globales más destacadas en este campo: la M5 Competition. Esta competición, en su quinta edición, se centra en la obtención de los mejores resultados en términos de pronósticos, tanto en un enfoque teórico como práctico. Esta competencia se enfocó en el pronóstico de ventas para 42,480 series de tiempo correspondientes a las ventas de Walmart, la mayor empresa minorista del mundo, y ofreció estimaciones precisas sobre la incertidumbre asociada a estos pronósticos (Makridakis et al., 2022).

Tradicionalmente, se ha demostrado que los métodos simples pueden ser igual de efectivos, e incluso a veces más efectivos, que las técnicas más complejas o sofisticadas. No obstante, la última edición de la M5 Competition marcó un hito al ser la primera vez que los cinco primeros lugares fueron ocupados exclusivamente por algoritmos de aprendizaje supervisado. Estos algoritmos superaron a los métodos estadísticos convencionales y a sus combinaciones.

Entre los enfoques más destacados, se incluyen el método de XGBoost (Extreme Gradient Boosting) para generar pronósticos precisos, y las redes neuronales, que imitan el comportamiento del cerebro humano mediante una red de unidades interconectadas, destacándose por su capacidad para captar patrones complejos en los datos.

III. Metodología.

1. Datos.

La información utilizada en este trabajo proviene del sistema de facturación de la empresa, el cual registra de manera sistemática todos los movimientos asociados a la comercialización de neumáticos. Este sistema permite la generación de reportes personalizados según distintos criterios de análisis, garantizando la trazabilidad y la precisión de los datos.

En particular, se utilizarán aquellos reportes que contienen el número exacto de unidades vendidas por producto, con una frecuencia de registro mensual. Estos datos se actualizan regularmente el último día hábil de cada mes, lo que asegura su disponibilidad en tiempo y forma para su análisis.

El período analizado abarca desde el 1 de enero de 2019 hasta el 31 de octubre de 2023, e incluye información correspondiente a 200 modelos distintos de neumáticos.

1.1 Descripción y estructura del conjunto de datos.

El análisis se basa en reportes mensuales que registran las unidades vendidas de cada producto comercializado por la empresa. A modo ilustrativo, se presenta el reporte correspondiente a enero de 2019, en el cual se detallan las variables relevantes para el estudio.

codigo	CAI	descripcion	Cantidad Vendida	Ingresos	precio_unitario	Costo	costo unitario	rentabilidad	%_rentabilidad	%_participacion	mes	ano
1620555P4V	997724	205/55R16 94V PRIMACY 4 XL	102	199516	1956.04	108601.85	1064.72402	90914.14	9.85	11.98	enero	2019
1722545P4V	194560	225/45R17 94W PRIMACY 4 XL	53	102110	1926.60	57859.59	1091.69038	44250.4	4.79	6.22	enero	2019
1519560EXN	893185	195/60R15 88H ENERGY XM2	36	56314	1564.28	29529.8	820.272222	26784.19	2.9	4.23	enero	2019
1826560LTXI	175397	265/60R18 110H LTX FORCE	25	78191	3127.64	43324.18	1732.9672	34866.81	3.77	2.93	enero	2019
1417565F60	15833001	175/65R14 82T F600	25	32437	1297.48	20952.26	838.0904	11484.74	1.24	2.93	enero	2019
1518560EXN	737406	185/60R15 88H ENERGY XM2 XL	21	26485	1261.19	14639.12	697.100952	11845.87	1.28	2.46	enero	2019
1621565LTXI	214154	215/65R16 98T LTX FORCE	20	44854	2242.70	27136.11	1356.8055	17717.88	1.92	2.35	enero	2019
1721555P3V	28677	215/55R17 94V PRIMACY 3	19	45767	2408.79	24148.14	1270.95474	21618.85	2.34	2.23	enero	2019
1518565EXN	86572	185/65R15 88H ENERGY XM2	19	36757	1934.58	19606.97	1031.94579	17150.02	1.85	2.23	enero	2019
1621565PSU	196398	215/65R16 98H PRIMACY SUV	18	46294	2571.89	24539.84	1363.32444	21754.15	2.35	2.11	enero	2019
1722565PSU	834879	225/65R17 102H PRIMACY SUV	18	47369	2631.61	25938.05	1441.00278	21430.94	2.32	2.11	enero	2019
1619555P4V	443149	195/55R16 87V PRIMACY 4	17	44013	2589.00	24669.14	1451.12588	19343.85	2.09	1.99	enero	2019
1726565LTXI	832916	265/65R17 112H LTX FORCE	16	30641	1915.06	16700.26	1043.76625	13940.73	1.51	1.88	enero	2019

Cada fila del reporte representa un producto individual, mientras que las columnas contienen la siguiente información:

- Código: Identificador único del producto dentro del sistema de gestión empresarial.
- CAI: Código adicional interno asignado a cada producto para su trazabilidad.
- Descripción: Denominación comercial y especificación técnica del neumático.
- Cantidad vendida: Total de unidades vendidas del producto en el mes correspondiente.
- Precio unitario: Precio de venta por unidad del producto.
- Costo unitario: Costo de disponer una unidad del producto en stock.
- Ingreso: Importe total facturado por producto, calculado como el precio unitario multiplicado por la cantidad vendida.
- Costo: Costo total asociado al producto, resultante de multiplicar el costo unitario por la cantidad vendida.
- Rentabilidad: Margen bruto del producto, obtenido como la diferencia entre el ingreso y el costo
- % Rentabilidad: Porcentaje de ingreso atribuible al producto respecto al ingreso total.
- % Participación: Proporción de ventas del producto respecto al total de unidades vendidas.
- Mes: Mes correspondiente al registro.
- Año: Año correspondiente al registro.

1.2 Procesamiento de datos.

El conjunto de datos utilizado en este trabajo está conformado por registros mensuales de unidades vendidas por producto, correspondientes al período comprendido entre enero de 2019 y octubre de 2023.

Tras una etapa de reorganización del dataset original, se construyó una base de datos estructurada con 5.663 observaciones y 14 variables. En ella, cada fila representa un producto en un mes específico, incorporando una columna adicional con la fecha de venta expresada en formato mes-año.

Durante el proceso de limpieza, se identificaron valores faltantes en 39 registros. Estos

fueron imputados con el valor cero, bajo el criterio de que su ausencia reflejaba la falta de disponibilidad del producto en el inventario durante ese período.

Asimismo, se detectaron 12 productos que habían sido comercializados inicialmente pero discontinuados con el tiempo. Dado que el objetivo de esta tesis es modelar la demanda de productos actualmente activos, estos fueron excluidos del análisis, resultando en un total de 188 productos vigentes.

A continuación, se realizó un análisis descriptivo exploratorio para comprender la estructura de los datos, identificar patrones relevantes y preparar adecuadamente el terreno para la etapa de modelado.

Posteriormente, se implementaron los distintos algoritmos de predicción propuestos: Media Móvil (MA), Prophet, ARIMA, SARIMA y XGBoost. El desempeño de cada modelo fue comparado con el benchmark actualmente utilizado por la empresa, aplicando métricas estándar de precisión para evaluar su eficacia relativa y su potencial aplicación en el contexto del negocio.

Finalmente, se desarrolló un análisis de optimización del inventario, etapa clave del trabajo, orientada a mitigar riesgos financieros asociados a la inmovilización de capital.

Todo el trabajo analítico y de modelado fue desarrollado utilizando R Studio, aprovechando la flexibilidad y potencia del lenguaje R tanto para la exploración descriptiva como para la implementación de los modelos de pronóstico. Esta herramienta resultó especialmente adecuada por su capacidad para manipular grandes volúmenes de datos, generar visualizaciones de alta calidad y aplicar con facilidad diversas bibliotecas estadísticas y de machine learning, permitiendo así una ejecución eficiente y reproducible de todo el proceso analítico.

Posteriormente, se observa una recuperación gradual en las ventas mensuales, aunque sin alcanzar los niveles previos a la pandemia. Este fenómeno es comprensible y se atribuye, por un lado, a las secuelas del deterioro económico general y, por otro, al fuerte proceso inflacionario que afectó especialmente a los precios del rubro neumáticos.

Por último, el gráfico confirma un patrón estacional en la demanda del negocio. Durante los meses de junio-julio y diciembre-enero se registra un incremento sostenido en las ventas a diferencia del resto del año. Esto se asocia a la planificación de vacaciones, momentos en los que los consumidores suelen revisar o reemplazar los neumáticos de sus vehículos. Además, estas fechas coinciden con períodos de mayor disponibilidad de ingresos, como el aguinaldo o bonificaciones.

En la etapa prepandémica, las ventas en los meses estacionales llegaban a triplicarse respecto al resto del año, pero este patrón se atenuó tras la pandemia debido al aumento de precios y la pérdida de poder adquisitivo, lo que llevó a muchos consumidores a postergar la compra o elegir productos de menor calidad.



Figura 2 - Mapa de calor de la evolución de la cantidad total de ventas mensuales.

El mapa de calor ofrece una representación visual complementaria que permite identificar de forma clara la evolución de las ventas a lo largo del tiempo. En él, cada celda representa un mes de un determinado año, y la intensidad del color indica el volumen de ventas correspondiente. Se evidencia una disminución progresiva en la intensidad del color de izquierda a derecha, lo que refleja la caída sostenida en la cantidad de unidades vendidas en los últimos años.

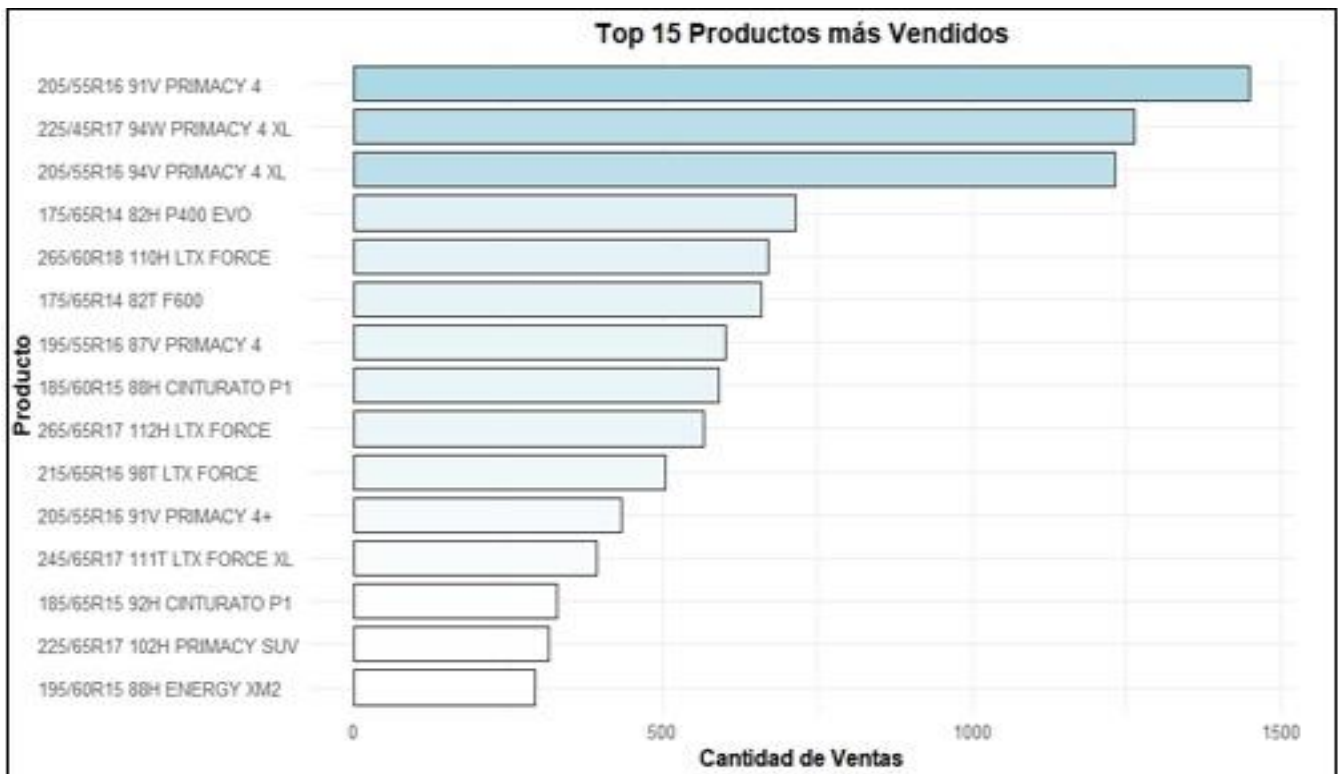


Figura 3 - Ranking de los productos más vendidos.

El gráfico de barras muestra los quince productos con mayor volumen de ventas en la empresa. Se evidencia una marcada disparidad entre el producto más vendido y el que ocupa el decimoquinto lugar en el ranking, lo que evidencia una concentración significativa de la demanda en pocos artículos. La decisión de representar únicamente los quince productos más vendidos responde al objetivo de facilitar la lectura y el análisis visual. Aun así en este subconjunto, ya comienzan a notarse diferencias sustanciales en los niveles de venta.

Si se analizan los 188 productos que conforman el catálogo completo de la empresa, la desigualdad en el volumen de ventas se vuelve aún más evidente. El siguiente gráfico se incluye con el propósito de profundizar sobre este tema y visualizar con mayor claridad la concentración de la demanda.

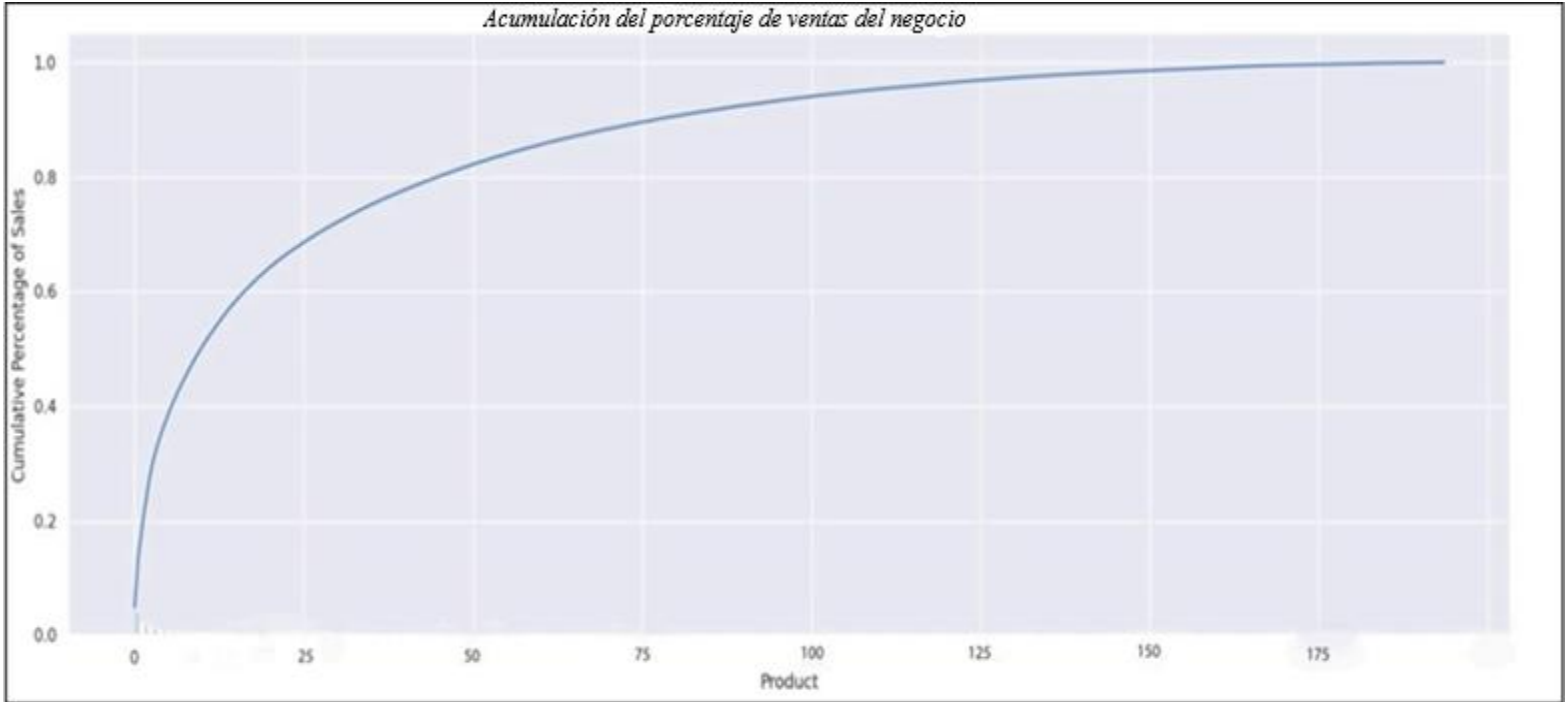


Figura 4 - Acumulación del porcentaje de ventas del negocio.

El gráfico ilustra cómo se distribuye el volumen total de ventas entre los distintos productos del catálogo. Se observa que aproximadamente el 20% de los artículos concentra cerca del 80% de las ventas, lo que revela una marcada asimetría en la demanda. Este comportamiento, consistente con lo analizado previamente en la Figura 3, indica una fuerte concentración de las ventas en un subconjunto reducido de productos.

Sin embargo, analizar únicamente el volumen vendido puede llevar a interpretaciones equivocadas. En particular, podría cuestionarse la conveniencia de incluir o no los productos con escasa rotación al momento de efectuar los pronósticos. Por ello, se vuelve necesario incorporar la dimensión económica, especialmente la rentabilidad.

En este negocio, los productos de mayor rotación suelen estar vinculados a márgenes unitarios bajos, debido a la fuerte competencia existente en el mercado. Por ejemplo, neumáticos como los 205/55 R16, habituales en vehículos de alta circulación (Toyota Corolla, Ford Focus, Peugeot 208), registran altos volúmenes de venta pero con márgenes reducidos. En contraste, productos orientados a segmentos de nicho, como los neumáticos para BMW X6, presentan una demanda mucho menor, pero márgenes significativamente superiores. En términos comparativos, la ganancia obtenida por la venta de dos neumáticos para un BMW X6 puede igualar o incluso superar la obtenida por la venta de cien unidades del modelo más común.

Esta situación resalta la importancia de generar pronósticos que contemplen tanto productos de alta como de baja rotación, ya que estos últimos pueden representar una porción sustancial de la rentabilidad total del negocio.

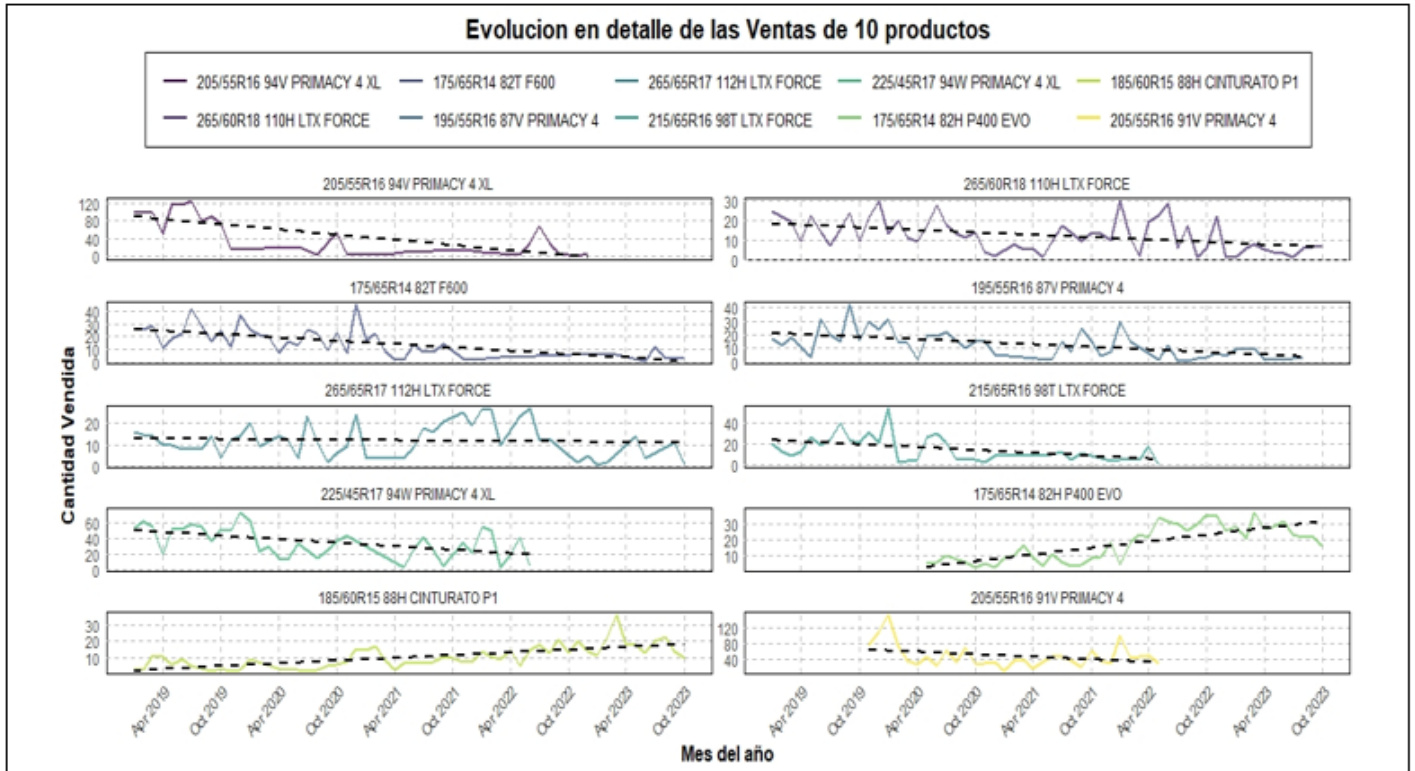


Figura 5 - Evolución detallada de las cantidades vendidas de 10 productos del negocio.

Con el objetivo de profundizar el análisis exploratorio, se seleccionaron aleatoriamente diez productos del catálogo y se analizó la evolución temporal de sus ventas de forma individual. Para facilitar la interpretación visual de las tendencias, se incorporaron líneas de ajuste en cada uno de los gráficos.

En general, se observa una tendencia decreciente en el volumen de ventas para la mayoría de los productos, en línea con lo identificado previamente en la Figura 1. Además, se evidencian diferencias en cuanto a los períodos de comercialización: algunos productos comienzan a registrarse tiempo después de enero de 2019, mientras que otros dejan de venderse antes de octubre de 2023. Esta heterogeneidad motivó la exclusión de 12 productos del análisis, ya que no resulta metodológicamente adecuado pronosticar la demanda de artículos que ya no forman parte de la oferta activa del negocio.

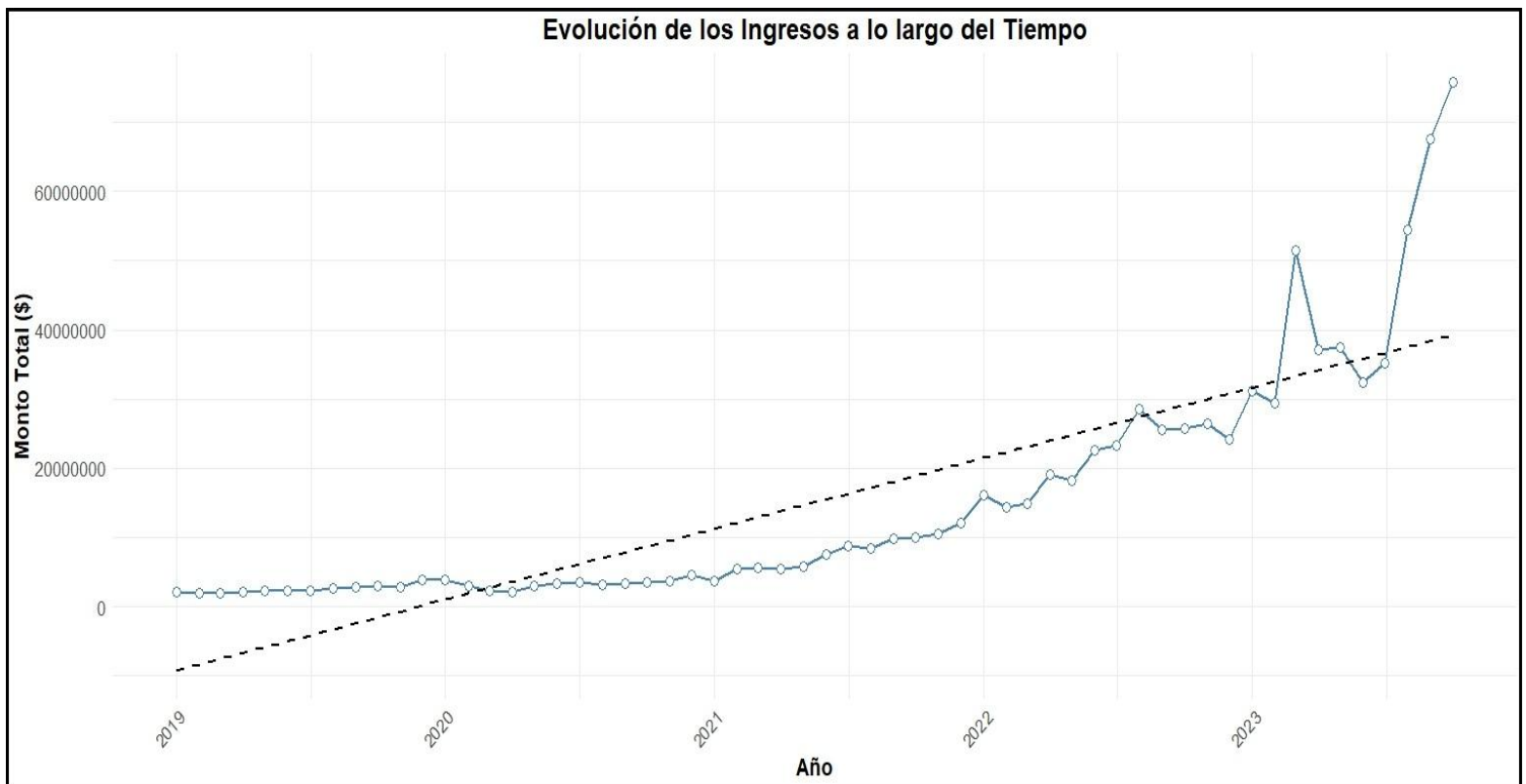


Figura 6 - Ingresos monetarios del negocio.

La figura muestra la evolución mensual de los ingresos totales generados por la venta de todos los productos de la empresa, expresados en pesos argentinos, durante el período comprendido entre 2019 y 2023.

A primera vista, se evidencia una tendencia creciente sostenida en los ingresos, lo cual podría interpretarse como un signo de expansión o un buen desempeño comercial de la empresa. Sin embargo, esta lectura debe abordarse con cautela, ya que los datos están expresados en términos nominales. En un contexto macroeconómico como el de Argentina, marcado por altos niveles de inflación en los años posteriores a la pandemia, parte del aumento observado podría deberse a la pérdida del poder adquisitivo del peso y no necesariamente a un incremento real en las ventas o la rentabilidad.

Por ello, se recomienda interpretar el gráfico con prudencia, ya que el incremento nominal en los ingresos podría reflejar, en gran medida, el efecto de la inflación, más que una mejora real en el desempeño económico de la empresa.

Con el objetivo de resolver esta ambigüedad y obtener una visión más precisa del desempeño económico, se procedió a deflactar la serie de ingresos mensuales, utilizando como base el mes de enero de 2019. Para ello, se empleó el Índice de Precios al Consumidor (IPC) Nivel General elaborado y publicado por el Instituto Nacional de Estadística y Censos (INDEC) de la República Argentina.

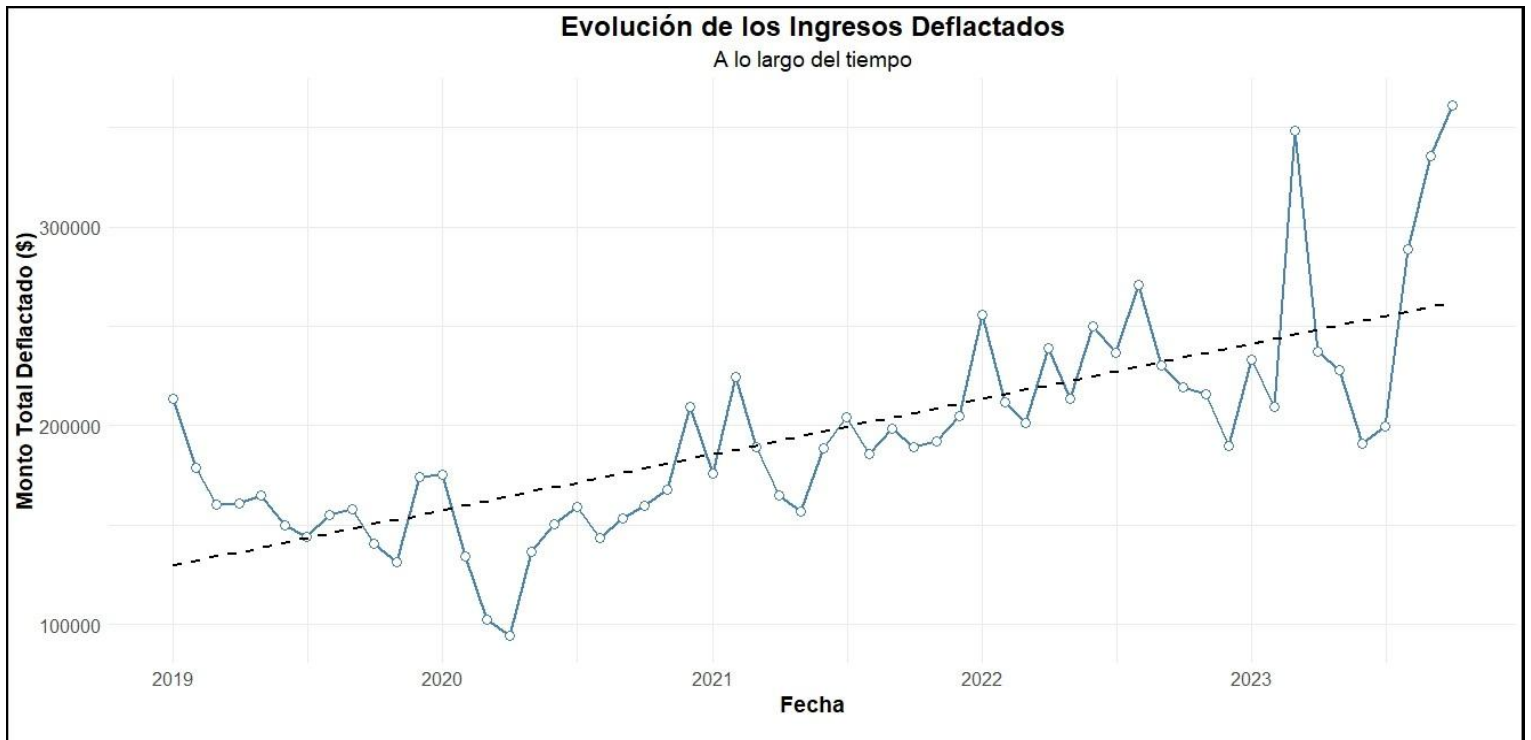


Figura 7- Ingresos monetarios deflactados del negocio.

Al comparar ambos gráficos, se evidencia que la línea de tendencia correspondiente a los ingresos deflactados presenta un crecimiento sostenido, aunque menos pronunciado que la curva de ingresos nominales, lo cual resulta coherente con el contexto inflacionario del período analizado.

No obstante, la tendencia ascendente en términos reales sugiere que el incremento en los ingresos no se debe exclusivamente al efecto de la inflación, sino que también refleja una mejora concreta en el desempeño económico de la empresa. En este sentido, el análisis ajustado por inflación permite obtener una interpretación más precisa y robusta sobre la evolución genuina de la actividad comercial.

Con el objetivo de profundizar este análisis y comprender con mayor detalle el comportamiento de los ingresos reales a lo largo del tiempo, a continuación se presenta su evolución desagregada por año.

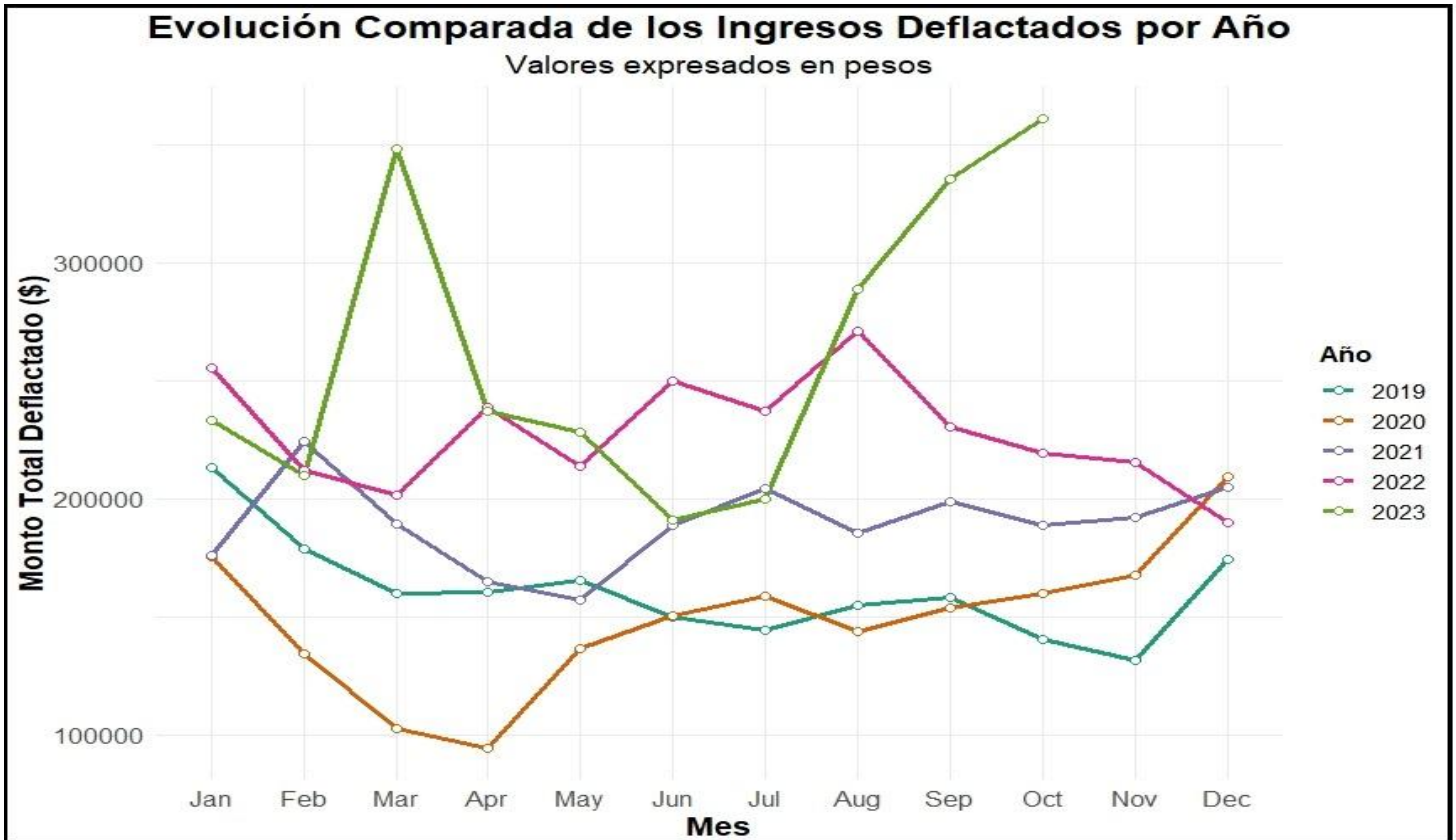


Figura 8 - Ingresos monetarios deflactados del negocio, desagregados por año.

Tal como se mencionó previamente en relación con la estacionalidad del negocio, los ingresos tienden a concentrarse en los períodos de junio-julio y diciembre-enero. Esta regularidad se verifica a lo largo de la mayoría de los años analizados, evidenciando un patrón estacional estable en el comportamiento de la demanda. Sin embargo, en el último año del período analizado, se observa una desviación de esta tendencia, ya que los picos de ingresos se registran en los meses de marzo y octubre.

Con el objetivo de facilitar la interpretación visual y destacar los patrones interanuales, la evolución de los ingresos deflactados se representa en la siguiente figura de forma comparativa por año calendario.

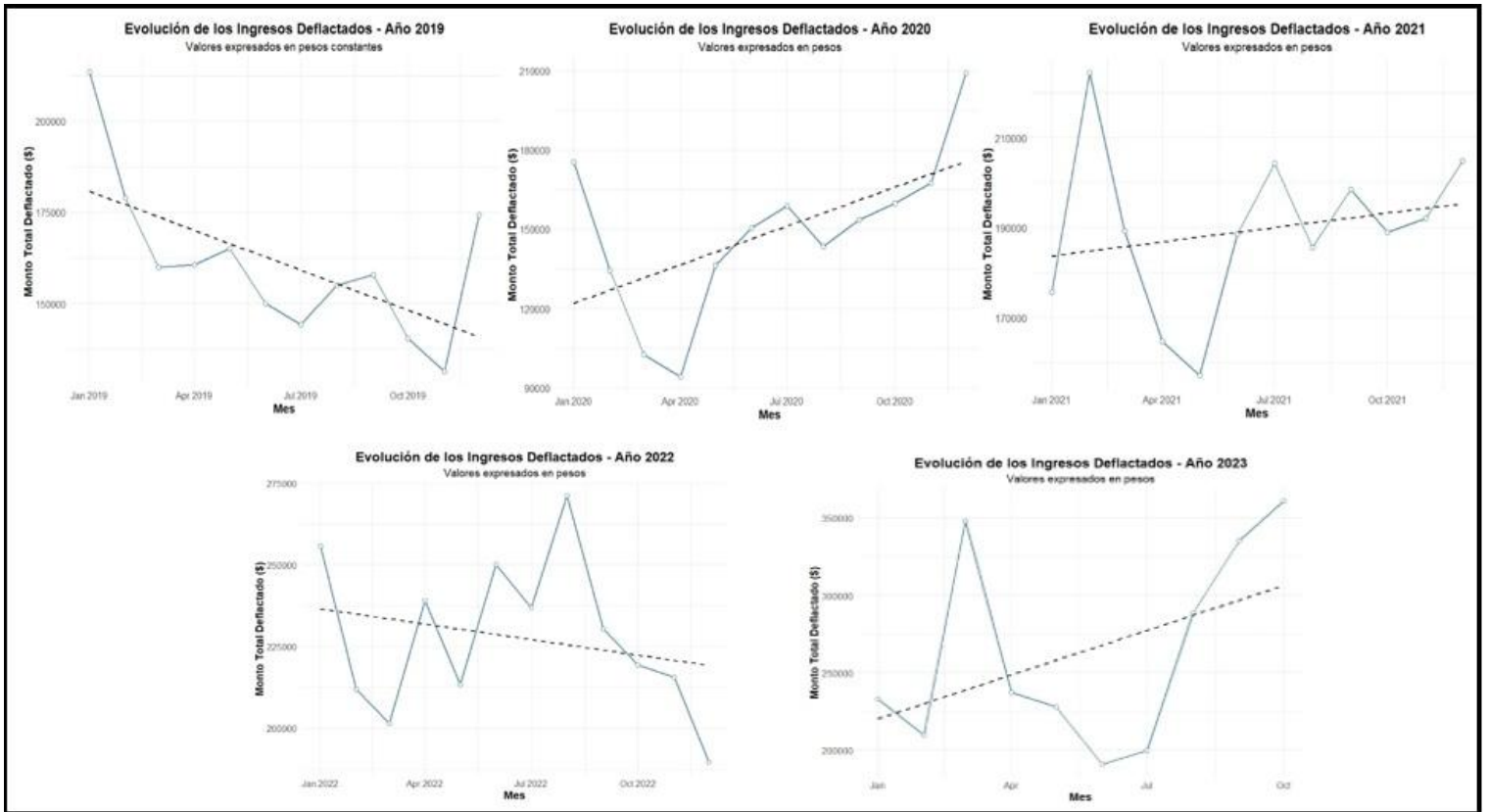


Figura 9 - Ingresos monetarios deflactados del negocio, visualizados por año.

La representación gráfica permite identificar y corroborar con mayor precisión las particularidades de cada período analizado, facilitando la detección de patrones estacionales, tendencias y posibles anomalías en el comportamiento interanual de los ingresos reales.

Para concluir el análisis exploratorio, se examina a continuación la evolución de los precios unitarios correspondientes a algunos de los productos más representativos del portafolio de la empresa. El propósito de este ejercicio es detectar variaciones significativas en sus valores a lo largo del tiempo, las cuales pueden responder a dinámicas propias del mercado, procesos inflacionarios o cambios en la estrategia comercial.

A fin de facilitar la comparación entre productos y mejorar la interpretación visual de la información, los precios fueron transformados mediante logaritmos naturales.

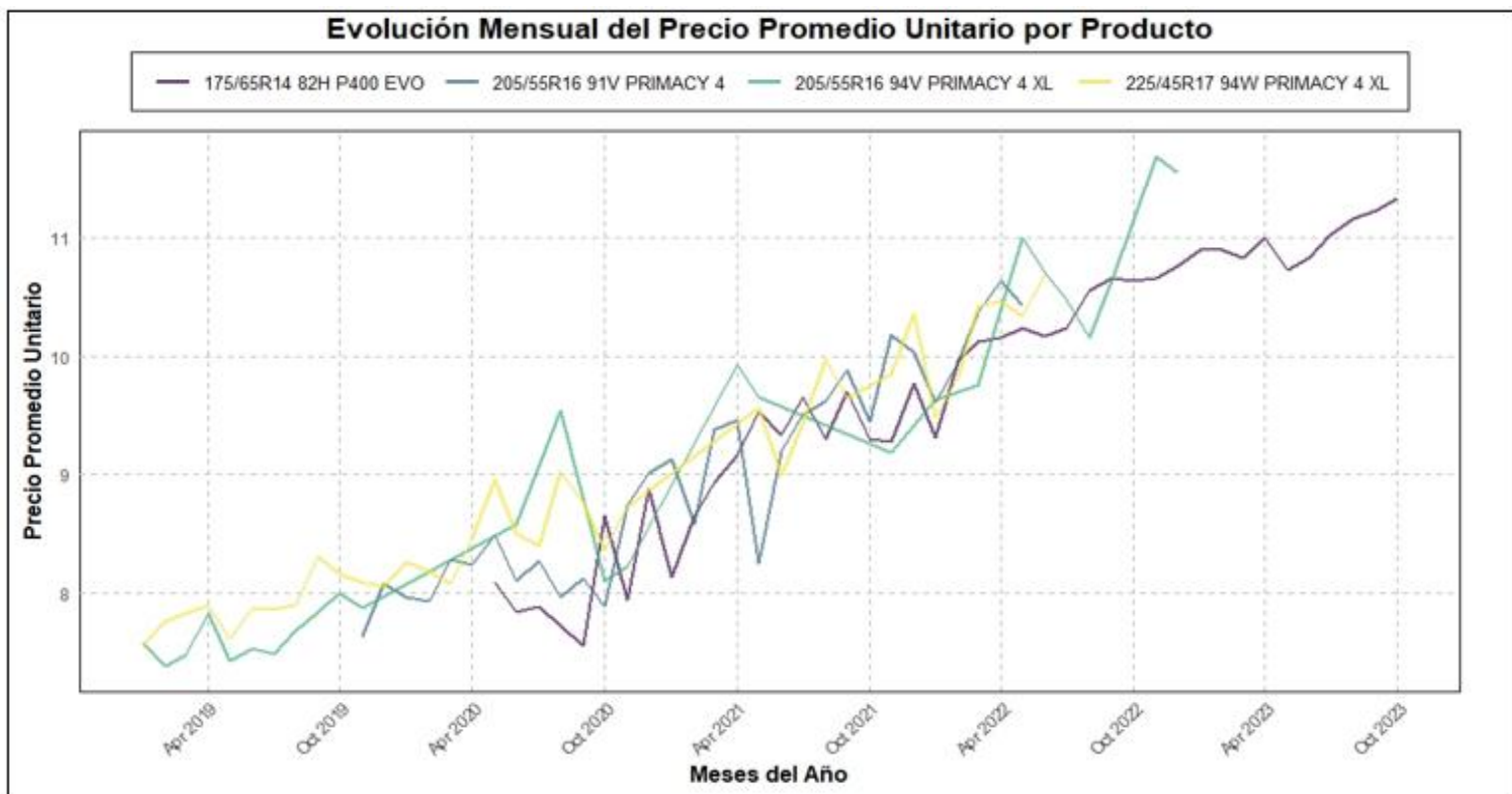


Figura 10 – Evolución del precio promedio unitario de productos del negocio.

El gráfico exhibe una tendencia creciente sostenida en los precios unitarios de los neumáticos, acompañada por una elevada dispersión, atribuible principalmente al contexto inflacionario que atravesó el país durante el período analizado. En este sentido, el marcado incremento en los precios unitarios ha incidido directamente sobre las cantidades vendidas, reflejándose en una contracción de la demanda, tal como fue evidenciado en la Figura 1.

A su vez, el análisis sugiere que el crecimiento de los ingresos empresariales observado en los últimos años no se explica únicamente por un mayor volumen de ventas, sino fundamentalmente por el aumento sostenido en los precios unitarios. Entre 2019 y 2023, los precios unitarios experimentaron un crecimiento exponencial, lo que implica que, aun con una menor cantidad de unidades vendidas, los ingresos actuales pueden alcanzar niveles ampliamente superiores.

2. Implementación de los modelos de pronóstico.

Partición del Dataset.

Con el propósito de implementar y evaluar los modelos predictivos propuestos para la estimación de la demanda de neumáticos, se llevó a cabo una partición estructurada del conjunto de datos mediante la técnica de validación *Holdout Validation*.

Esta metodología consiste en dividir el conjunto de datos históricos en dos subconjuntos disjuntos y consecutivos en el tiempo: uno destinado al entrenamiento de los modelos (*train set*) y otro utilizado exclusivamente para su validación (*validation set*), simulando de forma más realista el flujo temporal de información disponible en un entorno empresarial. La partición elegida se definió de la siguiente manera:

- Train Set: comprende el período desde enero de 2019 hasta marzo de 2023.
- Validation Set: incluye datos correspondientes de abril hasta octubre de 2023.

La elección de una ventana de validación de siete meses respondió al objetivo de evaluar el desempeño de los modelos sobre un período suficientemente representativo, que abarca tanto meses de alta como de baja demanda. De este modo, se obtiene una visión más realista del comportamiento del modelo ante las fluctuaciones propias del negocio.

El *train set* fue utilizado para ajustar los modelos y realizar, en cada caso, la correspondiente calibración de hiperparámetros. Posteriormente, cada modelo fue evaluado sobre el *validation set*, con el fin de estimar su capacidad de generalización y comparar el desempeño relativo entre las distintas alternativas propuestas. Esta estrategia metodológica permite garantizar que la selección de modelos se realice de forma objetiva y sobre la base de evidencia empírica, contribuyendo así a fortalecer la robustez de las conclusiones obtenidas.

Métrica de error.

La evaluación de los modelos predictivos constituye una etapa crítica en todo proyecto de pronóstico, tan relevante como la propia construcción del modelo. Este proceso implica comparar los valores estimados con los reales, y para ello se recurre a métricas cuantitativas que permiten medir objetivamente el grado de precisión alcanzado. La correcta selección de la métrica es fundamental, ya que no solo determina la calidad de las predicciones, sino que también orienta la toma de decisiones en contextos comerciales.

En este caso particular, dado que todos los productos involucrados en el análisis se miden en una misma unidad (cantidades vendidas), pero presentan una marcada heterogeneidad en sus volúmenes de demanda, se optó por utilizar el *Error Porcentual Absoluto Ponderado* (WAPE, por sus siglas en inglés: *Weighted Absolute Percentage Error*). Esta métrica se define como la suma de los errores absolutos entre las predicciones y los valores reales, dividida por la suma de los valores reales de ventas. Se expresa formalmente como:

$$WAPE = \frac{\sum_{t=1}^n |A_t - F_t|}{\sum_{t=1}^n |A_t|}$$

Donde A_t representa el valor real y F_t el valor pronosticado en el tiempo t . Al ponderar el error absoluto por el volumen total de ventas, esta métrica permite una evaluación más equitativa del desempeño del modelo.

El WAPE presenta múltiples ventajas. En primer lugar, facilita la comparación entre productos con distintos niveles de demanda, ya que evalúa el error relativo con respecto a la magnitud de las ventas de cada uno. Además, al ponderar los errores en función de la relevancia comercial de cada producto, proporciona una medida de precisión que refleja adecuadamente su impacto en los resultados del negocio. Su expresión en términos

porcentuales también contribuye a una interpretación intuitiva de la precisión del modelo, permitiendo comparaciones directas con umbrales de aceptación utilizados habitualmente en el ámbito empresarial. Por último, al considerar tanto errores por sobreestimación como por subestimación, el WAPE ofrece una visión integral del desempeño del modelo, lo que lo convierte en una métrica ampliamente adoptada en la industria para evaluar la calidad de los pronósticos.

Criterio de éxito.

El criterio de éxito adoptado en este trabajo se fundamenta en la capacidad de los modelos propuestos para superar el desempeño del modelo actualmente empleado por la empresa, el cual se utiliza como benchmark. La métrica empleada para esta comparación es el *Weighted Absolute Percentage Error* (WAPE). Un modelo será considerado exitoso si logra reducir el WAPE en relación con el benchmark, lo que implicará una mejora en la precisión de las predicciones de demanda. En este sentido, el modelo que obtenga el menor valor de WAPE será identificado como el de mejor desempeño.

Tratamiento del período de pandemia por COVID 19.

El análisis y modelización de series temporales de demanda enfrenta desafíos particulares al incorporar períodos extraordinarios como el acontecido durante la pandemia de COVID-19. Estos eventos disruptivos pueden generar alteraciones estructurales en los patrones históricos, afectando tanto la estacionalidad como la tendencia general de las series.

En este trabajo, se optó por conservar la totalidad del historial de datos disponible entre enero de 2019 y octubre de 2023, incorporando así tanto el período previo como el correspondiente a la pandemia y pospandemia. Esta decisión responde a la necesidad de preservar la continuidad de la serie temporal y capturar adecuadamente la evolución completa

del negocio, sin perder información valiosa sobre los mecanismos de recuperación o adaptación que se dieron tras el evento disruptivo.

A fin de dar tratamiento específico al período pandémico, se incluyó en los modelos una variable indicadora (dummy) correspondiente a los meses comprendidos entre febrero de 2020 y febrero de 2021, periodo en el que se observaron efectos anómalos en la demanda asociados a las restricciones sanitarias y a las alteraciones logísticas y comerciales generalizadas. Esta variable permite al modelo distinguir el comportamiento propio de ese contexto, sin distorsionar la dinámica general de los datos fuera de ese intervalo.

En la presente sección, todas las referencias a WAPE, así como las comparaciones entre los valores reales y pronosticados, se refieren exclusivamente al conjunto de validación, correspondiente al período reservado para la evaluación final de los modelos.

2.1 Modelo Benchmark: Naive Last Value.

Como punto de partida, se adopta como modelo benchmark el Naive Last Value 12 mensual, que representa la estrategia actualmente utilizada por la empresa para proyectar sus volúmenes de venta. Este enfoque consiste en estimar la demanda futura de un determinado mes del año basándose exclusivamente en el valor observado en el mismo mes del año anterior.

Por ejemplo, para predecir las ventas correspondientes a abril de 2024, el modelo toma como estimación directa el volumen registrado en abril de 2023. De este modo, se asume una estacionalidad anual perfecta, sin contemplar posibles efectos de tendencia, ciclos económicos u otras fuentes de variabilidad.

A continuación, se presentan los resultados obtenidos al aplicar este modelo sobre el conjunto de validación, evaluando su desempeño a partir del error WAPE. La tabla que sigue incluye una selección de los primeros quince productos del portafolio (sobre un total de 188), con el objetivo de facilitar la interpretación sin perder representatividad estadística.

Tabla Comparativa - Modelo Benchmark

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	10	0.69
Producto 2	24	9	0.75
Producto 3	23	47	1.06
Producto 4	15	32	1.22
Producto 5	46	21	0.54
Producto 6	44	19	0.59
Producto 7	17	34	1.22
Producto 8	11	25	1.31
Producto 9	37	55	0.51
Producto 10	65	44	0.36
Producto 11	17	44	1.59
Producto 12	70	43	0.39
Producto 13	27	12	0.59
Producto 14	13	35	1.71
Producto 15	38	66	0.88
...
Error Total	0.49

Figura 11 – Modelo Benchmark: WAPE en conjunto de prueba (validation set).

La tabla muestra, para cada producto, la sumatoria de las cantidades reales vendidas en el conjunto de prueba (y_true), la sumatoria de las cantidades predichas por el modelo en el conjunto de prueba (y_baseline), y el error WAPE correspondiente (wape_baseline).

El modelo arrojó un WAPE total de 0.49 al ser evaluado sobre el conjunto de prueba. Este valor indica que, en promedio, las predicciones del modelo presentan un desvío del 49% respecto a los valores reales observados. Este nivel de error es significativo, ya que el rubro

de comercialización de neumáticos implica trabajar con productos de alto costo unitario, rotación estacional y requerimientos logísticos particulares debido a su volumen y almacenamiento.

A continuación, se presenta un análisis detallado del desempeño del modelo benchmark sobre un producto seleccionado aleatoriamente dentro del conjunto de validación. Este mismo producto será utilizado de manera consistente en la evaluación de los demás modelos, con el objetivo de examinar las capacidades predictivas bajo las mismas condiciones.

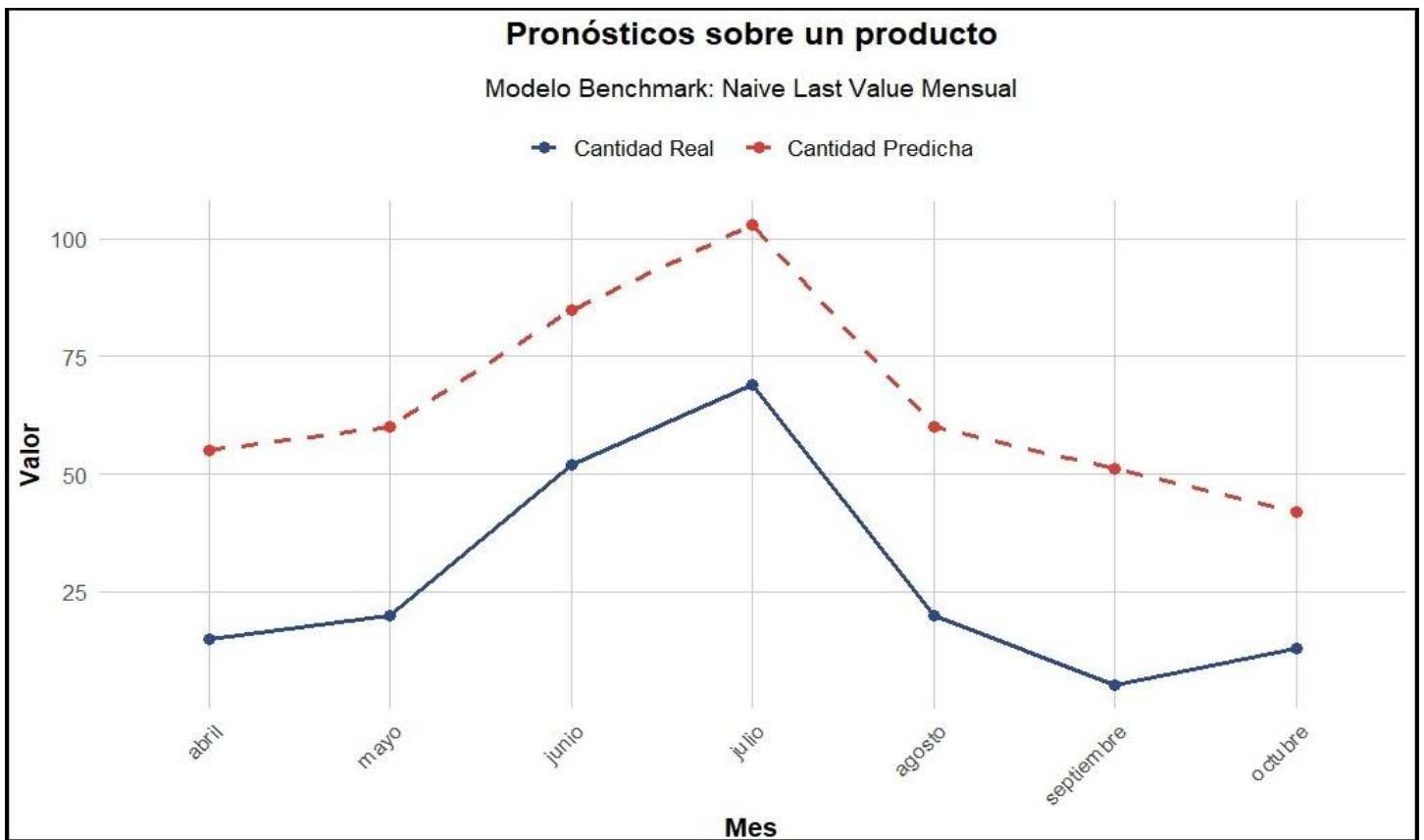


Figura 12 – Modelo Benchmark: Pronóstico sobre producto seleccionado al azar.

Si bien el modelo logra capturar de forma general el patrón estacional de la demanda, producto del conocimiento empírico acumulado por el propietario de la empresa, presenta una tendencia sistemática a la sobreestimación. Las cantidades proyectadas superan de manera consistente a las ventas efectivamente registradas.

Se aprecia que la línea que representa las cantidades predichas se encuentra notablemente por encima de la línea de las cantidades reales de demanda, lo cual evidencia con claridad el bajo nivel de precisión del modelo.

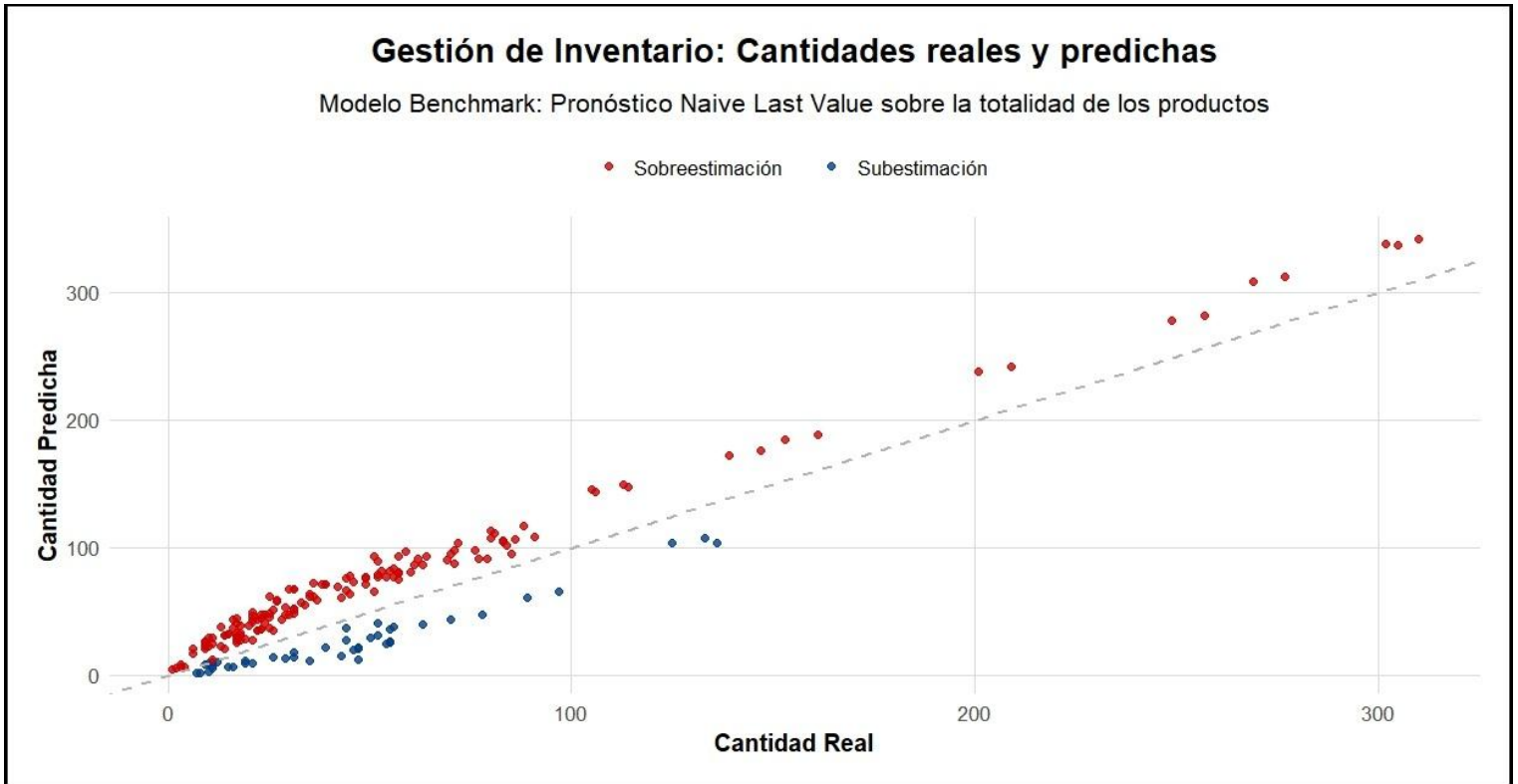


Figura 13 – Modelo Benchmark: Cantidades reales vs cantidades predichas.

Finalmente, se evalúa el desempeño del modelo benchmark sobre la totalidad de los productos de la empresa. El gráfico de dispersión presentado muestra la relación entre las cantidades reales vendidas y las proyectadas por el modelo, permitiendo visualizar el desempeño operativo actual del negocio en términos de gestión de inventario.

Se observa que las predicciones del modelo se desvían de manera significativa respecto a la línea de referencia ($y = x$), lo cual evidencia una alta imprecisión en los pronósticos. Si bien la baja ocurrencia de quiebres de stock podría interpretarse como un aspecto positivo

desde la perspectiva del nivel de servicio, el mantenimiento de un volumen tan elevado de inventario excedente implica una ineficiencia económica considerable.

De los 188 productos evaluados, 145 presentan situaciones de sobreestimación, lo que confirma una tendencia sistemática del modelo a proyectar niveles de demanda superiores a los realmente observados. Esta tendencia no solo se manifiesta con una alta frecuencia, sino que además resulta especialmente preocupante por la magnitud de los desvíos registrados, los cuales generan un impacto económico significativo en términos de inventario excedente.

Como consecuencia, el costo total de almacenamiento de unidades de productos durante el periodo analizado es de \$179.896.105,34. El costo de almacenamiento considerado en este análisis corresponde al costo variable derivado de las unidades acumuladas en exceso de cada producto. Es decir, se refiere a las unidades sobrantes resultantes de la sobreestimación en los pronósticos. Este cálculo se realiza de forma individual para cada producto, dado que cada neumático posee un costo unitario específico. Formalmente, el costo total de almacenamiento se determina mediante la siguiente expresión:

$$Costo = \sum_{t=abril}^{octubre} \sum_{i=1}^{188} \max(0, Cantidad\ Predicha_{i,t} - Cantidad\ Real_{i,t}) \times Costo\ Unitario$$

Donde $Cantidad\ Predicha_{i,t}$ es la demanda pronosticada por el modelo para el producto i en el mes t , $Cantidad\ Real_{i,t}$ es la demanda real del producto i en el mes t ; y $Costo\ Unitario$ es el costo unitario del producto i .

Esta fórmula contabiliza los casos donde la cantidad pronosticada supera a la demanda real, ya que la subestimación no genera costos de almacenamiento adicionales.

2.2 Media Móvil (Moving Average – MA).

Como primera solución al problema identificado en el modelo benchmark, se propone la implementación del modelo de medias móviles (MA). Este enfoque estima la demanda futura a partir del promedio de un número fijo de observaciones pasadas, asignando el mismo peso a cada dato dentro de la ventana seleccionada.

La longitud de la ventana móvil (parámetro k) fue calibrada empíricamente para cada producto, determinando que una ventana de tres períodos ($k = 3$) proporcionaba el mejor equilibrio entre suavización y capacidad de reacción ante cambios recientes en la demanda. Ventanas más largas generan pronósticos excesivamente suavizados, mientras que ventanas más cortas resultan sensibles al ruido y al comportamiento aleatorio de la serie.

Se utilizaron las funciones de la librería *forecast* (Hyndman et al., 2023). Para automatizar el proceso sobre la totalidad de productos, se aplicó la función *ma()* en combinación con las herramientas del paquete *dplyr*, lo que permitió calcular la media móvil simple de cada serie de ventas mensuales de manera sistemática y eficiente. Este procedimiento se ejecutó de forma individual sobre cada producto, respetando la secuencia temporal dentro de cada serie.

Los resultados obtenidos al efectuar los pronósticos sobre el conjunto de prueba fueron los siguientes:

**Tabla Comparativa - Modelo
Moving Average**

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	13	0.53
Producto 2	24	15	0.51
Producto 3	23	41	0.85
Producto 4	15	26	0.88
Producto 5	46	29	0.41
Producto 6	44	28	0.44
Producto 7	17	28	0.73
Producto 8	11	20	0.91
Producto 9	37	48	0.34
Producto 10	65	51	0.25
Producto 11	17	36	1.08
Producto 12	70	53	0.26
Producto 13	27	17	0.41
Producto 14	13	28	1.18
Producto 15	38	57	0.61
...
Error Total	0.41

Figura 14 – Modelo MA: WAPE en conjunto de prueba (validation set).

El modelo permitió generar estimaciones de demanda con una capacidad de ajuste limitada, aunque mejor que el modelo benchmark, ya que arrojó un WAPE total de 0.41. Si bien este enfoque simple suaviza las fluctuaciones aleatorias de la serie y facilita la identificación de patrones de corto plazo, presenta ciertas restricciones al no incorporar dinámicas más complejas ni capturar adecuadamente tendencias o estacionalidades de mayor profundidad.

A pesar de esto, los resultados obtenidos constituyen una mejora inicial en términos de precisión respecto al método actual.

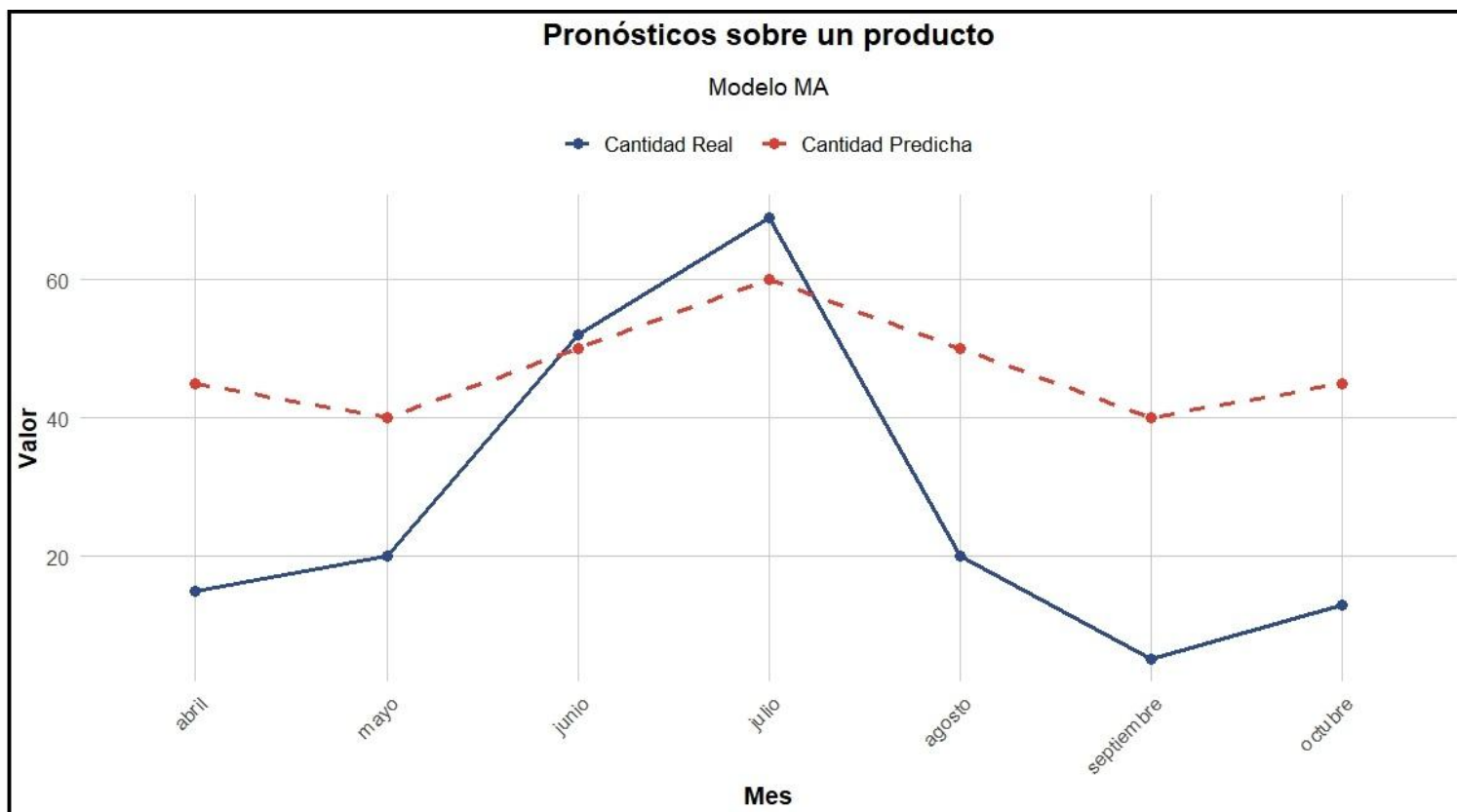


Figura 15 – Modelo MA: Pronóstico sobre producto seleccionado al azar.

Se observa que el modelo no logra adaptarse a las dinámicas propias del negocio. Sus predicciones tienden a converger hacia un promedio histórico reciente, pero sin la sensibilidad necesaria para captar las variaciones significativas a lo largo del tiempo. Como consecuencia, el modelo no consigue anticipar adecuadamente los picos de ventas ni las caídas pronunciadas que suelen registrarse en determinados períodos del año.

Este comportamiento genera pronósticos que resultan útiles como una aproximación general, pero insuficientes cuando se requiere una planificación precisa que contemple las fluctuaciones reales de la demanda.

Gestión de Inventario: Cantidades reales y predichas

Modelo MA: Pronóstico Moving Average sobre la totalidad de los productos

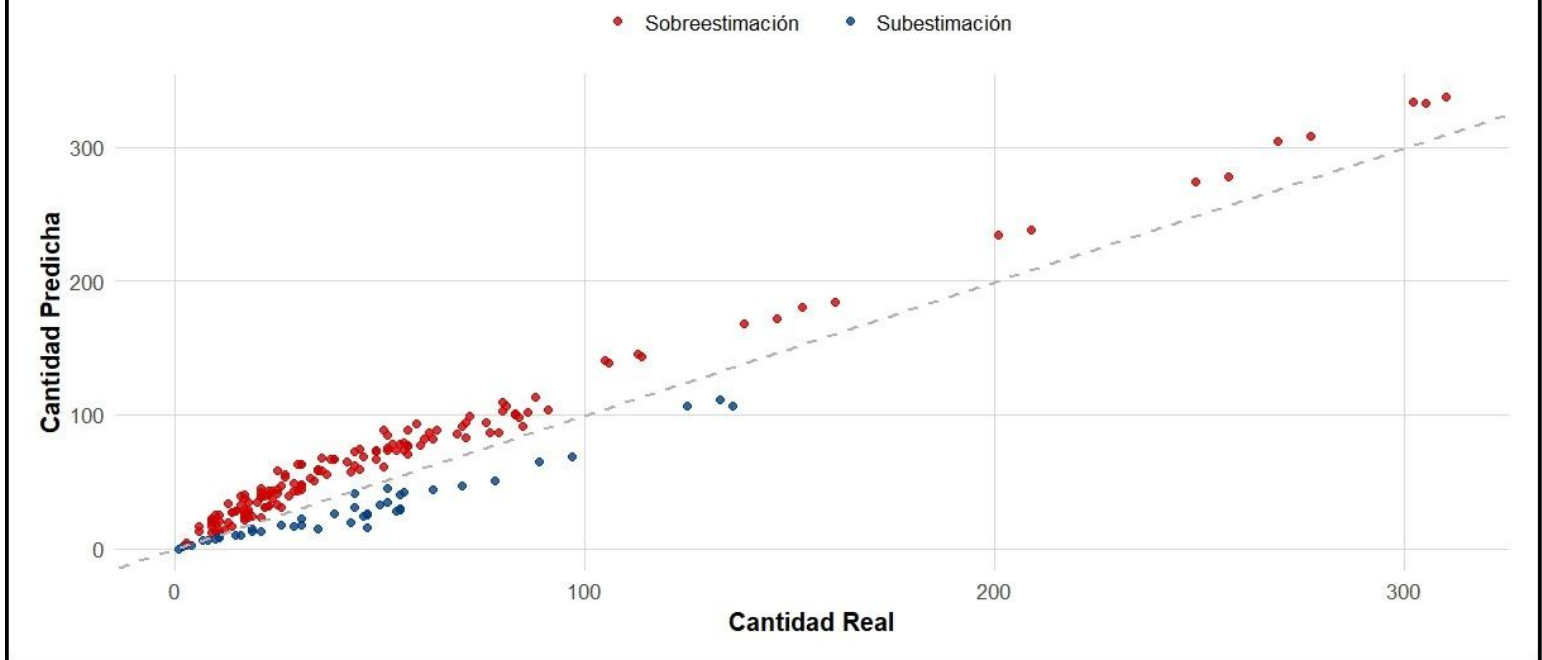


Figura 16 – Modelo MA: Cantidades reales vs cantidades predichas.

Las predicciones obtenidas presentan un nivel considerable de imprecisión, con una marcada dispersión que evidencia la incapacidad del modelo para reflejar adecuadamente la dinámica real de la demanda. La tendencia a sobreestimar se mantiene, con errores de gran magnitud, aunque se advierte una leve atenuación en comparación con el modelo benchmark.

Este comportamiento se traduce en un costo total de almacenamiento de \$155.395.477,98 durante el período analizado, lo que representa una leve mejora respecto al costo generado por el modelo anterior. No obstante, los resultados evidencian que persisten oportunidades de optimización en la gestión de inventario, particularmente en lo que refiere a la reducción de costos asociados al exceso de stock.

2.3 Prophet.

Con el objetivo de explorar un enfoque más flexible y robusto para la predicción de series temporales, se implementó el modelo Prophet, ampliamente reconocido por su capacidad para capturar automáticamente patrones complejos de tendencia y estacionalidad.

Se decidió trabajar con una estrategia de optimización de hiperparámetros para obtener la mejor configuración posible de Prophet en cada caso. Para ello, se configuró el modelo con estacionalidad anual, y se optimizaron tres hiperparámetros clave:

- `changeoint.prior.scale`
- `seasonality.prior.scale`
- `seasonality.mode`

Donde `changeoint.prior.scale` controla la sensibilidad del modelo ante cambios en la tendencia, `seasonality.prior.scale` ajusta la flexibilidad para modelar las estacionalidades, y `seasonality.mode` define si la estacionalidad se comporta de forma aditiva o multiplicativa.

El proceso de ajuste se llevó a cabo mediante una búsqueda aleatoria de hiperparámetros (*random grid*), permitiendo explorar múltiples combinaciones posibles de manera eficiente.

Todo el proceso fue implementado en el lenguaje R, utilizando la librería *prophet* (Taylor & Letham, 2018), junto con los paquetes *purrr* y *dplyr* para automatizar la iteración sobre los productos y sistematizar la evaluación de combinaciones.

Los resultados alcanzados al validar los pronósticos en el conjunto de prueba fueron los siguientes:

**Tabla Comparativa - Modelo
PROPHET**

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	17	0.37
Producto 2	24	20	0.17
Producto 3	23	29	0.31
Producto 4	15	20	0.35
Producto 5	46	39	0.17
Producto 6	44	38	0.15
Producto 7	17	21	0.25
Producto 8	11	16	0.47
Producto 9	37	42	0.15
Producto 10	65	54	0.18
Producto 11	17	24	0.45
Producto 12	70	62	0.13
Producto 13	27	22	0.21
Producto 14	13	19	0.48
Producto 15	38	43	0.15
...
Error Total	0.23

Figura 17 – Modelo Prophet: WAPE en conjunto de prueba (validation set).

El modelo Prophet demostró un desempeño notablemente superior en comparación con los modelos previamente evaluados, posicionándose como la opción más precisa para la estimación de la demanda. En particular, alcanzó un WAPE total de 0.23, lo que representa una mejora sustancial frente a los valores obtenidos por el modelo benchmark (0.49) y el modelo de medias móviles (0.41). Cabe destacar que la optimización de hiperparámetros mediante random grid resultó clave para maximizar la capacidad predictiva del modelo.

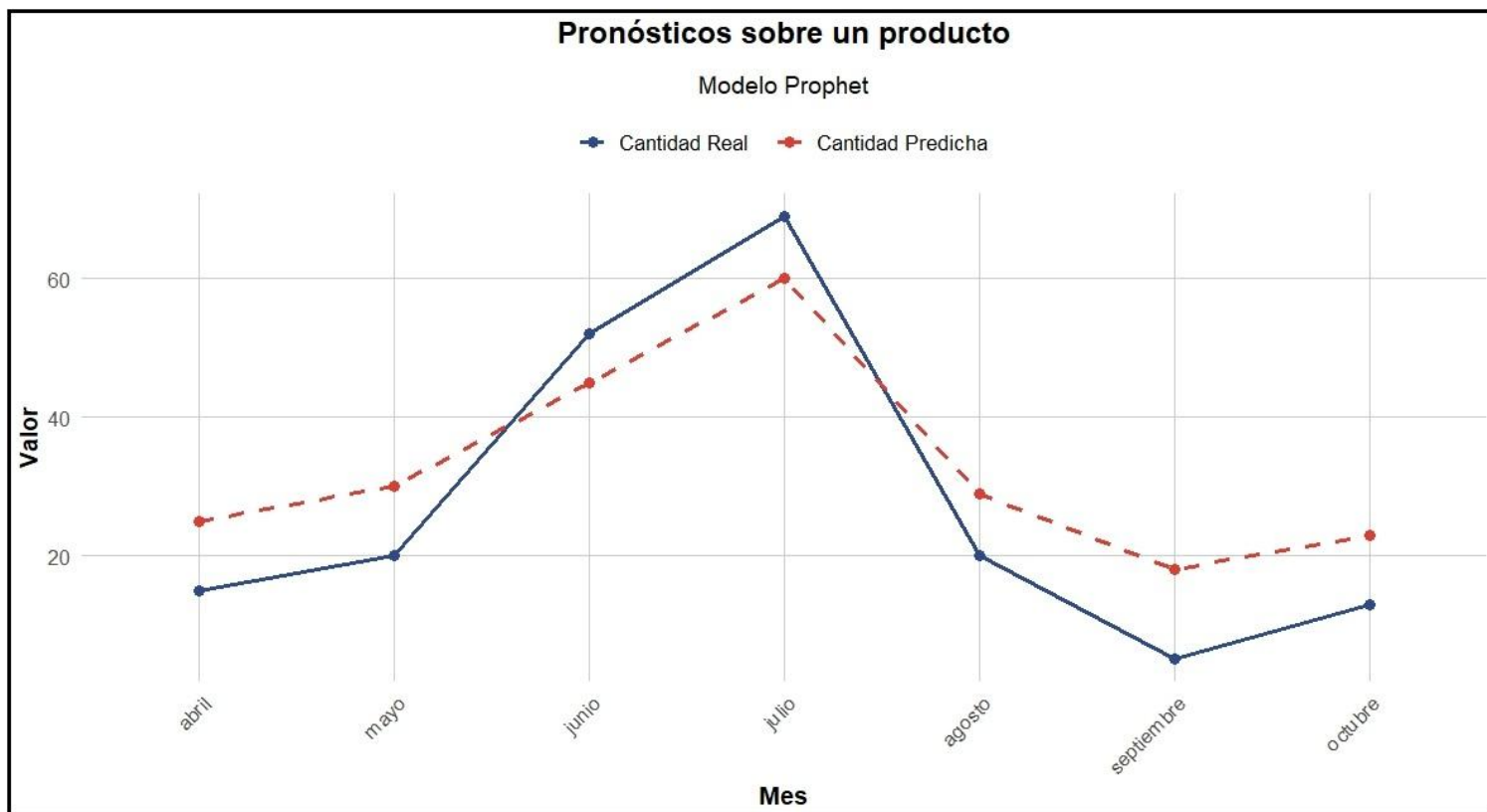


Figura 18 – Modelo Prophet: Pronóstico sobre producto seleccionado al azar.

Se observa que el modelo presenta un ajuste adecuado a los datos reales. Las predicciones generadas por Prophet logran seguir de manera consistente los niveles de ventas observados, capturando correctamente el pico estacional característico del negocio. Si bien la anticipación de estos incrementos en la demanda no es exacta en su totalidad, el modelo muestra una adecuada sensibilidad a las variaciones estacionales.

En relación a las fluctuaciones mensuales, Prophet logra replicar satisfactoriamente la estructura y el patrón de la serie. Estos resultados reflejan que el modelo ofrece un desempeño sólido a la hora de estimar la demanda.

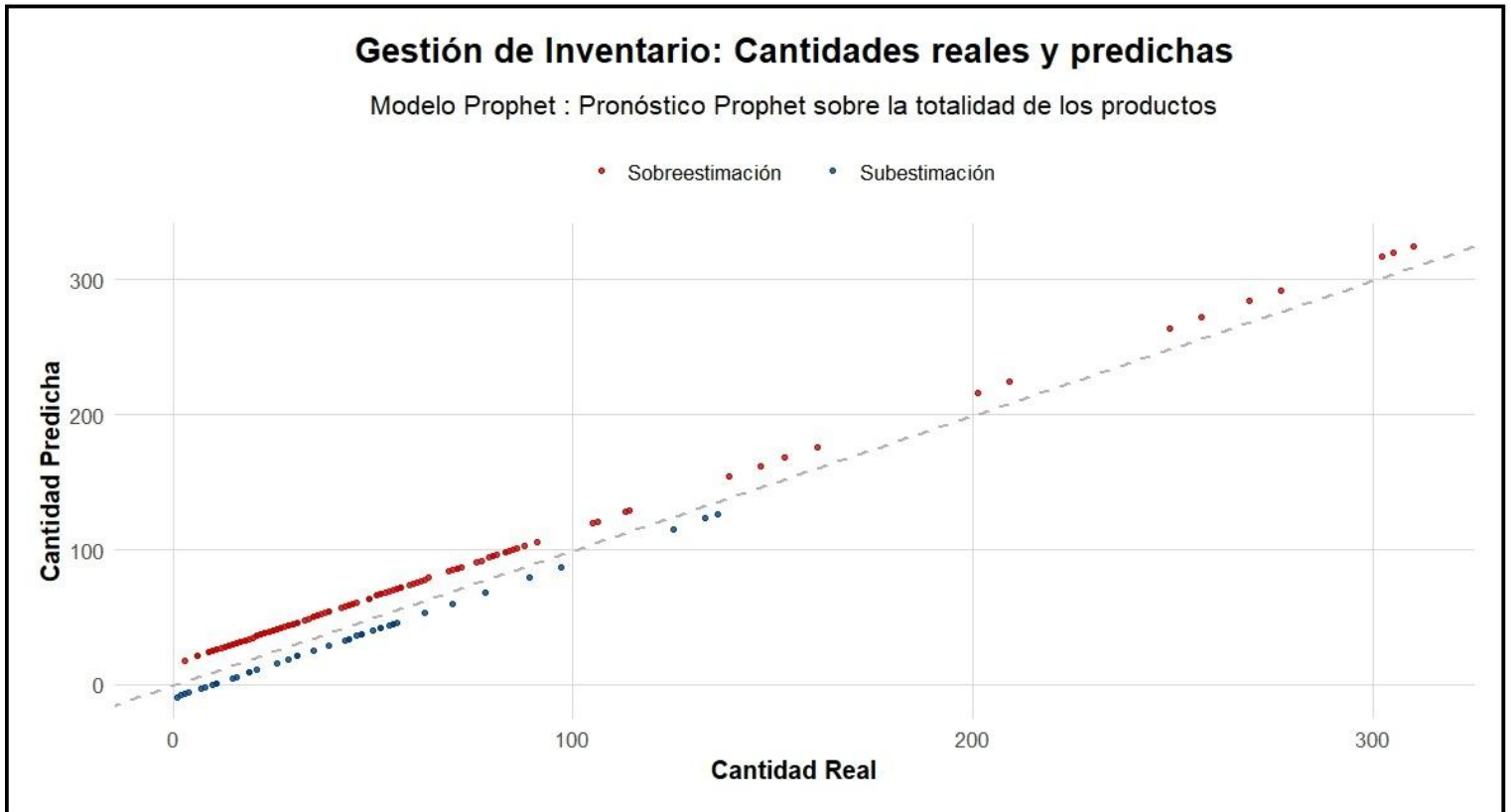


Figura 19 – Modelo Prophet: Cantidades reales vs cantidades predichas.

La implementación del modelo Prophet permitió mejorar significativamente la precisión de las predicciones. Al analizar la gestión de inventario resultante, se observa una notable disminución en la dispersión de los puntos en el gráfico, lo que evidencia que el modelo comienza a capturar de manera más adecuada la dinámica real del negocio.

Si bien persisten las sobreestimaciones, la magnitud de las mismas se reduce considerablemente en comparación con los modelos previamente evaluados. El costo total de almacenamiento se redujo a \$109.632.852,43 lo que representa una mejora sustancial respecto a los escenarios planteados previamente.

Este resultado destaca el valor práctico de Prophet como herramienta para la planificación y toma de decisiones.

2.4 ARIMA.

A continuación, se implementó el modelo el modelo ARIMA (AutoRegressive Integrated Moving Average) para estimar la demanda del negocio. La elección de este enfoque se fundamenta en su capacidad para modelar series temporales no estacionarias mediante la aplicación de diferenciaciones sucesivas, así como en su aptitud para capturar estructuras tanto autorregresivas como de medias móviles, lo que proporciona mayor flexibilidad frente a series con comportamiento dinámico moderado.

El modelo fue calibrado para cada producto utilizando la librería *forecast* (Hyndman et al., 2023) en el entorno R. La selección de los hiperparámetros p, d, q se realizó mediante la función `auto.arima()`. Este procedimiento comienza evaluando si la serie temporal es estacionaria, ya que la aplicación del modelo ARIMA requiere que la serie no presente tendencias ni variaciones de varianza en el tiempo. Para verificar esta condición, el algoritmo aplica un test de estacionariedad (KPSS), determinando si es necesario diferenciar la serie y cuántas diferenciaciones son requeridas para estabilizarla. El número de diferenciaciones necesarias define el valor óptimo del parámetro d .

Una vez estabilizada la serie, el algoritmo continúa con la selección de los parámetros p y q que representan, respectivamente, el número de rezagos autorregresivos y el número de rezagos de la media móvil que mejor describen la estructura temporal de la serie. En lugar de evaluar todas las combinaciones posibles, lo que resultaría en un proceso computacionalmente costoso, el `auto.arima()` inicia con cuatro modelos base y selecciona aquel con el menor valor del criterio de información corregido de Akaike (AICc). A partir de esta selección inicial, el algoritmo ajusta iterativamente los parámetros p y q aumentando o disminuyendo sus valores en una unidad, con el fin de verificar posibles mejoras en la calidad del ajuste. Este

procedimiento continúa hasta que no se detectan mejoras adicionales en el AICc. Este enfoque no solo agiliza el proceso de ajuste, sino que también garantiza la obtención de un modelo parsimonioso y eficaz dentro de la muestra. La metodología seguida se basa en el trabajo de Hyndman y Khandakar (2008), quienes desarrollaron el algoritmo de selección automática de modelos ARIMA en R.

A continuación, se presentan los resultados de los pronósticos obtenidos mediante este modelo.

Tabla Comparativa - Modelo ARIMA

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	14	0.49
Producto 2	24	17	0.38
Producto 3	23	38	0.75
Producto 4	15	24	0.71
Producto 5	46	31	0.37
Producto 6	44	30	0.38
Producto 7	17	25	0.59
Producto 8	11	19	0.81
Producto 9	37	46	0.33
Producto 10	65	48	0.27
Producto 11	17	34	1.01
Producto 12	70	55	0.24
Producto 13	27	18	0.38
Producto 14	13	26	1.05
Producto 15	38	54	0.53
...
Error Total	0.37

Figura 20 – Modelo ARIMA: WAPE en conjunto de prueba (validation set).

El modelo ARIMA alcanzó un WAPE de 0.37, logrando estimaciones de demanda con un nivel de precisión superior al obtenido por los modelos Benchmark y Medias Móviles, demostrando una mejor capacidad de ajuste a los patrones históricos de la serie. No obstante, su desempeño resulta considerablemente inferior al modelo Prophet, que obtuvo un WAPE de 0.23.

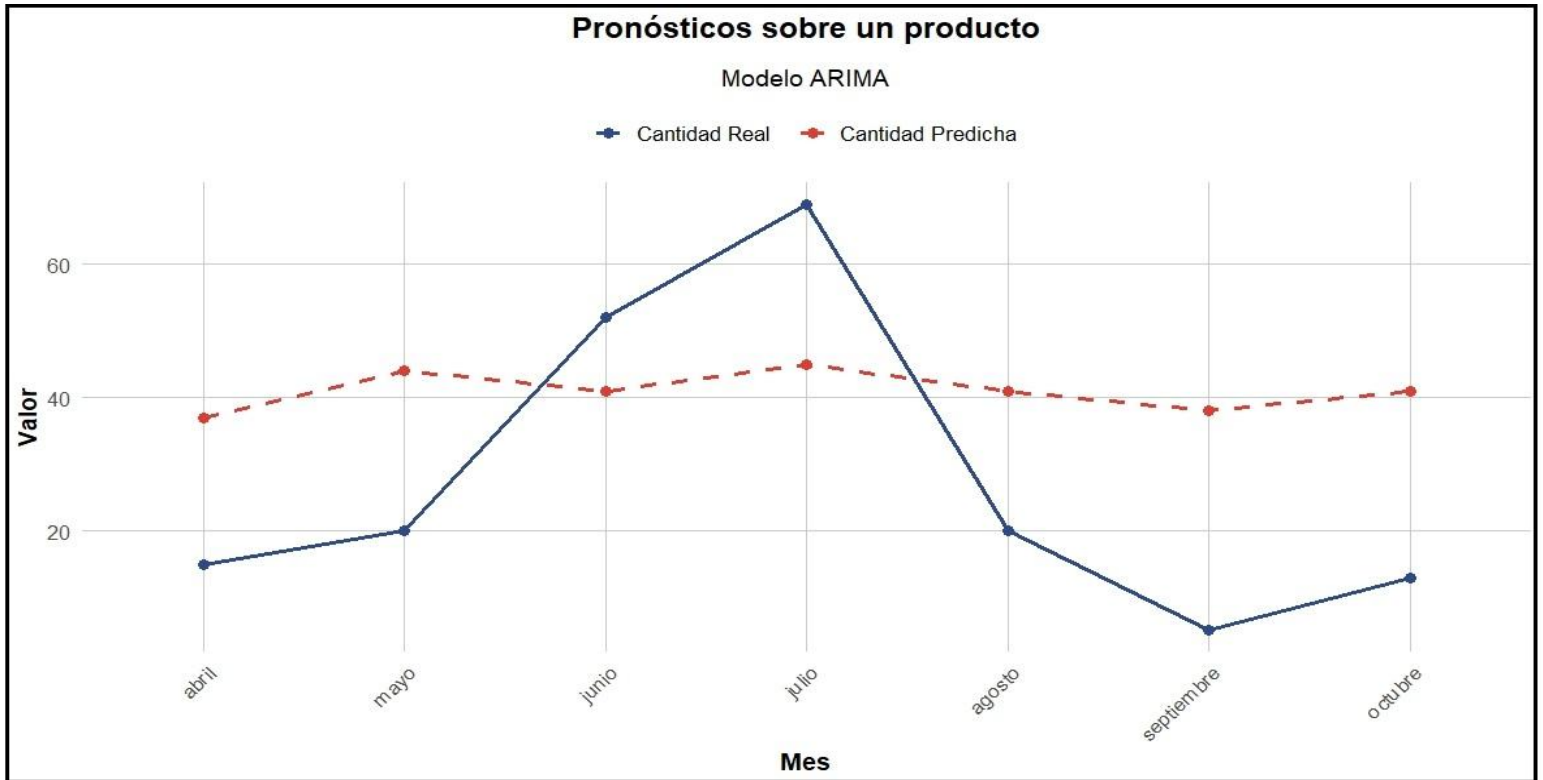


Figura 21 – Modelo ARIMA: Pronóstico sobre producto seleccionado al azar.

La curva de pronóstico tiende a mantenerse prácticamente horizontal, lo que evidencia que el modelo genera estimaciones con escasa variabilidad a lo largo del tiempo. En otras palabras, las estimaciones se mantienen casi constantes durante varios períodos consecutivos, sin responder adecuadamente a las fluctuaciones que presenta la serie temporal. Esta rigidez sugiere que el modelo no logra adaptarse con precisión a los cambios estacionales o a las oscilaciones propias del comportamiento de consumo.

Gestión de Inventario: Cantidades reales y predichas

Modelo ARIMA: Pronóstico ARIMA sobre la totalidad de los productos

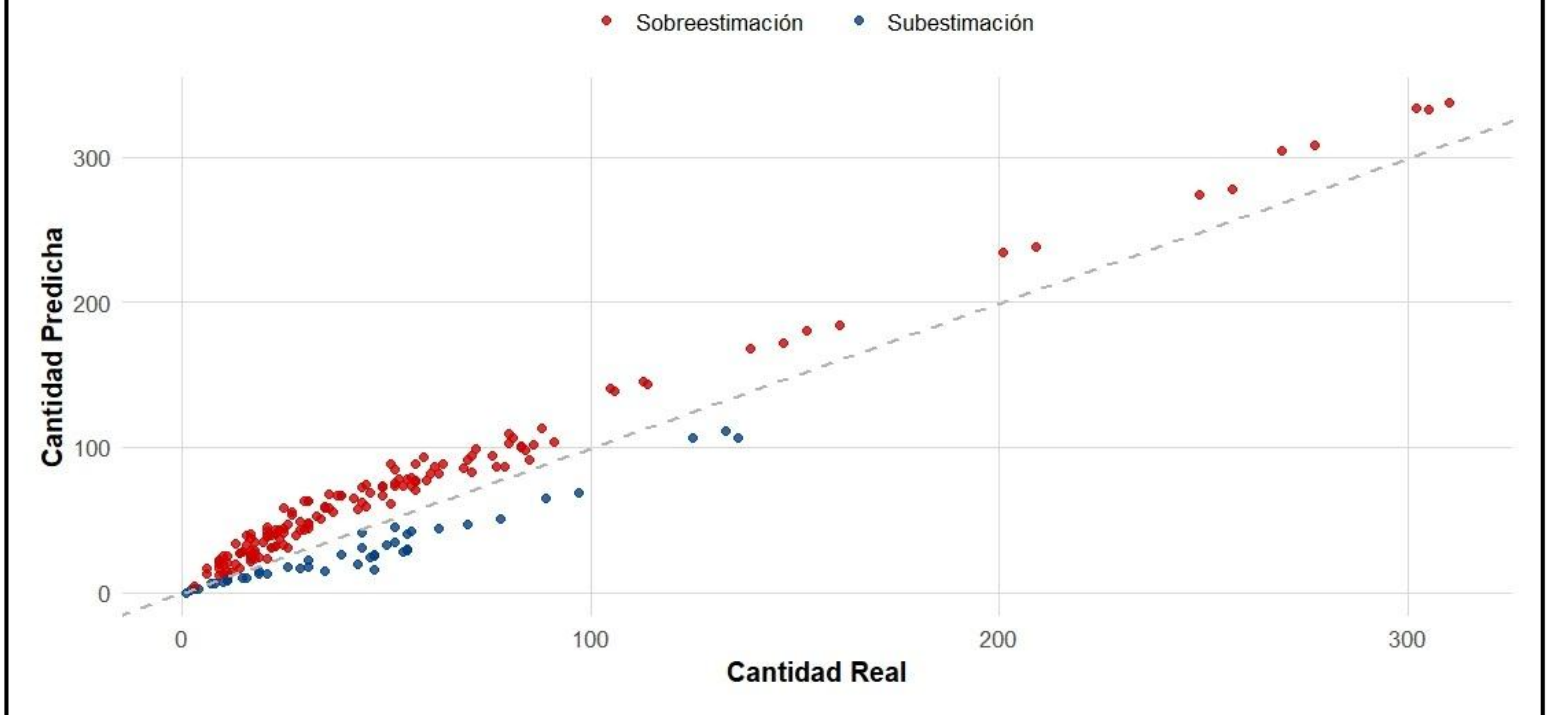


Figura 22 – Modelo ARIMA: Cantidades reales vs cantidades predichas.

Las predicciones continúan exhibiendo un nivel considerable de imprecisión. Las estimaciones distan aún de reflejar adecuadamente las cantidades efectivamente requeridas, lo que evidencia limitaciones en la capacidad del modelo para capturar con precisión la dinámica real de la demanda.

Este desempeño se traduce en un costo total de almacenamiento de \$142.174.533,45 durante el período analizado. Aún se evidencian ineficiencias en la gestión del stock desde una perspectiva económica. En particular, el modelo continúa mostrando una tendencia a sobreestimar la demanda real en múltiples casos, existiendo situaciones donde la discrepancia entre las cantidades pronosticadas y las ventas efectivamente registradas es notoria.

2.5 SARIMA.

Debido al limitado ajuste alcanzado por el modelo ARIMA, y particularmente a su escasa capacidad para capturar la estacionalidad característica del negocio, se decidió implementar el modelo SARIMA (Seasonal AutoRegressive Integrated Moving Average). Este enfoque extiende la estructura del modelo ARIMA al incorporar explícitamente componentes estacionales, lo que permite modelar series temporales que presentan patrones recurrentes en intervalos fijos. De este modo, se espera que el modelo capture con mayor precisión las fluctuaciones cíclicas propias de la demanda, brindando estimaciones más alineadas con el comportamiento real del negocio.

En este caso, la demanda presenta una estacionalidad anual caracterizada por múltiples picos, el primero concentrado en los meses de junio y julio, y el segundo en el período comprendido entre diciembre y enero. Por lo tanto, se configuró la serie temporal con una frecuencia de 12 períodos, lo que permite al modelo reconocer la estructura estacional anual. Esta frecuencia se estableció mediante la función `ts()`, utilizando la librería *forecast* en R. Posteriormente, se aplicó la función `auto.arima()` con el parámetro `seasonal = True` para que el algoritmo busque de manera automática la mejor combinación de parámetros estacionales (P, D, Q, m) junto con los no estacionales (p, d, q), optimizando el ajuste del modelo mediante el criterio AICc.

A continuación, se presentan los resultados de los pronósticos obtenidos mediante este modelo.

Tabla Comparativa - Modelo SARIMA

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	22	0.16
Producto 2	24	21	0.13
Producto 3	23	20	0.15
Producto 4	15	12	0.21
Producto 5	46	45	0.02
Producto 6	44	40	0.11
Producto 7	17	18	0.06
Producto 8	11	14	0.27
Producto 9	37	40	0.09
Producto 10	65	61	0.06
Producto 11	17	21	0.25
Producto 12	70	65	0.08
Producto 13	27	24	0.15
Producto 14	13	16	0.26
Producto 15	38	42	0.13
...
Error Total	0.15

Figura 23 – Modelo SARIMA: WAPE en conjunto de prueba (validation set).

El modelo SARIMA evidenció el mejor desempeño predictivo entre todas las alternativas evaluadas, alcanzando un WAPE total de 0.15. Esta mejora sustancial en la precisión de los pronósticos pone de manifiesto la capacidad superior del modelo para capturar la dinámica real del negocio. La incorporación explícita de los componentes estacionales dentro de su estructura permitió modelar con fidelidad los patrones recurrentes de la demanda, ajustando las estimaciones a los ciclos propios del mercado.

Este resultado destaca la relevancia de considerar la estacionalidad como un factor central en la modelización de la demanda en entornos comerciales con comportamiento cíclico.

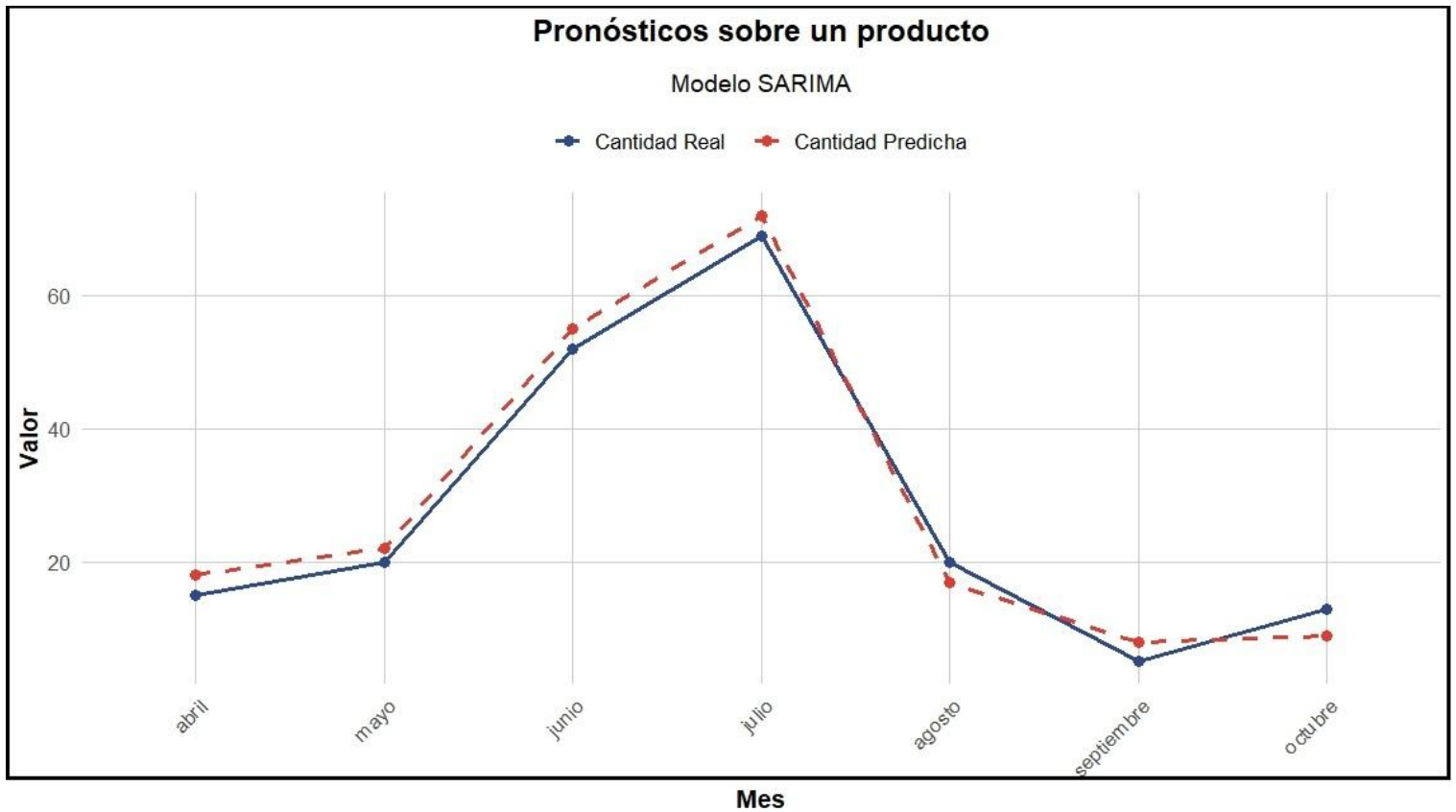


Figura 24 – Modelo SARIMA: Pronóstico sobre producto seleccionado al azar.

Las predicciones generadas por el modelo logran replicar con alta fidelidad los niveles de venta observados, capturando adecuadamente los picos estacionales característicos del negocio. En particular, se destaca su capacidad para anticipar correctamente los incrementos de demanda durante los meses de junio y julio, así como para adaptarse a las variaciones mensuales a lo largo del período de validación.

La mejora en la precisión del modelo no solo se evidencia en la significativa reducción del error WAPE, sino también a nivel cualitativo, a través de la capacidad del modelo para reproducir con solidez tanto la estacionalidad como la variabilidad inherente a la demanda de neumáticos.

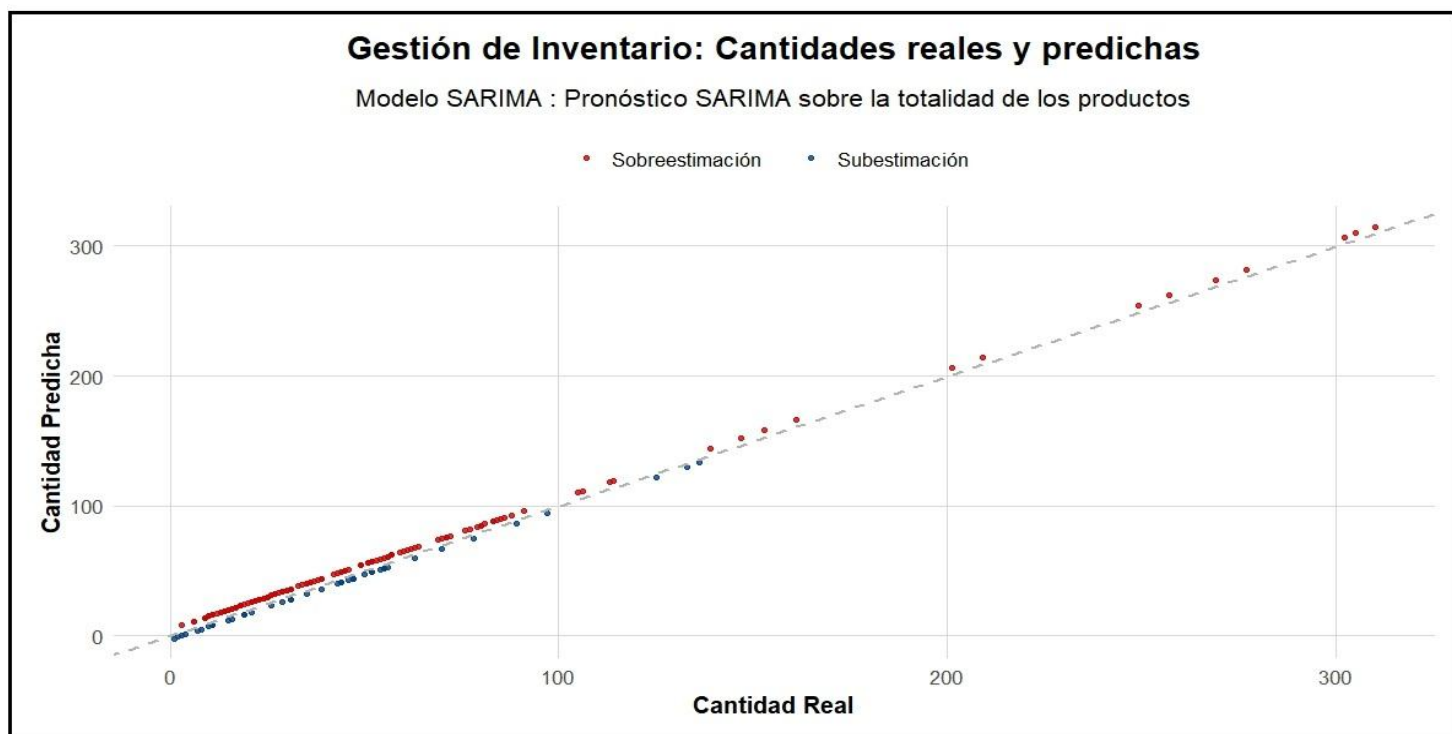


Figura 25 – Modelo SARIMA: Cantidades reales vs cantidades predichas.

Las predicciones generadas por el modelo exhiben un nivel de precisión sustancialmente superior, logrando minimizar de forma efectiva tanto las sobreestimaciones como las subestimaciones de la demanda. Esta mejora se refleja claramente en el gráfico, donde los puntos se ubican con mayor proximidad a la línea de referencia ($y = x$), evidenciando la capacidad del modelo para replicar con fidelidad las cantidades efectivamente demandadas.

Como consecuencia directa, se logró una reducción significativa del volumen de inventario excedente en la mayoría de los productos que previamente registraban pronósticos sistemáticamente por encima de la demanda real.

Esta optimización operativa se tradujo en una disminución de los costos totales de almacenamiento, alcanzando un valor de \$88.174.533,45 durante el período analizado, lo que representa una mejora económica sustancial para la gestión del negocio.

2.6 XGBoost.

Finalmente, se avanzó con la implementación de una estrategia multivariada basada en el modelo XGBoost (Extreme Gradient Boosting). Esta técnica, que se fundamenta en la construcción de árboles de decisión, aplica un enfoque secuencial de boosting con el objetivo de generar modelos más robustos y reducir los errores de pronóstico. Se destaca por su alta capacidad de modelado, su eficiencia computacional y su excelente desempeño al trabajar con series que presentan patrones complejos, no lineales y potencialmente ruidosos.

Con el propósito de maximizar la capacidad predictiva del modelo, se realizó un proceso de feature engineering, orientado a enriquecer la base de datos mediante la incorporación de algunas variables adicionales que pudieran aportar información relevante sobre la estructura temporal y comercial de la demanda. Las variables generadas e incorporadas fueron las siguientes:

- trimestre: variable que identifica el trimestre al que corresponde cada observación, contribuyendo a modelar estacionalidades de mayor periodicidad.
- mayor_demanda: variable binaria que toma valor 1 si el mes corresponde a los períodos de alta demanda estacional (junio-julio o diciembre-enero) y 0 en caso contrario, permitiendo capturar los picos característicos del negocio.
- precio_unitario: variable que refleja la evolución de los precios en el tiempo, permitiendo capturar efectos de sensibilidad a precios.
- costo_unitario: incluida para incorporar información sobre la estructura de costos, que puede incidir en las decisiones comerciales y en la dinámica de ventas.

- *pandemia*: variable indicadora (dummy) que toma valor 1 en los meses comprendidos entre febrero de 2020 y febrero de 2021, y 0 para el resto del periodo.

Adicionalmente, se aplicó la técnica de *one-hot encoding* sobre todas las variables categóricas, con el objetivo de transformarlas en un formato numérico compatible con el algoritmo.

Con el objetivo de maximizar el desempeño del modelo, se llevó a cabo un proceso de optimización de hiperparámetros utilizando la técnica de *random grid*. Los hiperparámetros ajustados durante este proceso fueron los siguientes:

- *nrounds*
- *max_depth*
- *eta*
- *gamma*

Todo el proceso fue implementado en el lenguaje R, utilizando la librería *xgboost* en conjunto con las funciones de los paquetes *caret* y *dplyr* para la manipulación de datos y la optimización de hiperparámetros.

A continuación, se presentan los resultados obtenidos tras la implementación de este modelo:

**Tabla Comparativa - Modelo
XGBoost**

Productos	y_true	y_baseline	wape_baseline
Producto 1	26	20	0.25
Producto 2	24	21	0.13
Producto 3	23	27	0.21
Producto 4	15	18	0.21
Producto 5	46	41	0.11
Producto 6	44	40	0.11
Producto 7	17	21	0.24
Producto 8	11	14	0.27
Producto 9	37	40	0.09
Producto 10	65	57	0.13
Producto 11	17	21	0.24
Producto 12	70	65	0.08
Producto 13	27	24	0.13
Producto 14	13	17	0.31
Producto 15	38	41	0.11
...
Error Total	0.19

Figura 26 – Modelo XGBoost: WAPE en conjunto de prueba (validation set).

El modelo XGBoost evidenció un desempeño muy sólido, alcanzando un WAPE total de 0,19. Si bien el modelo SARIMA se mantiene como el más preciso en términos absolutos, XGBoost destaca por su elevada capacidad de generalización y su flexibilidad para incorporar múltiples variables explicativas. Estas características lo posicionan como una herramienta estratégica de gran valor para la toma de decisiones.

El modelo logró ajustarse con precisión a los patrones históricos de la demanda, minimizando de manera efectiva tanto las sobreestimaciones como las subestimaciones.

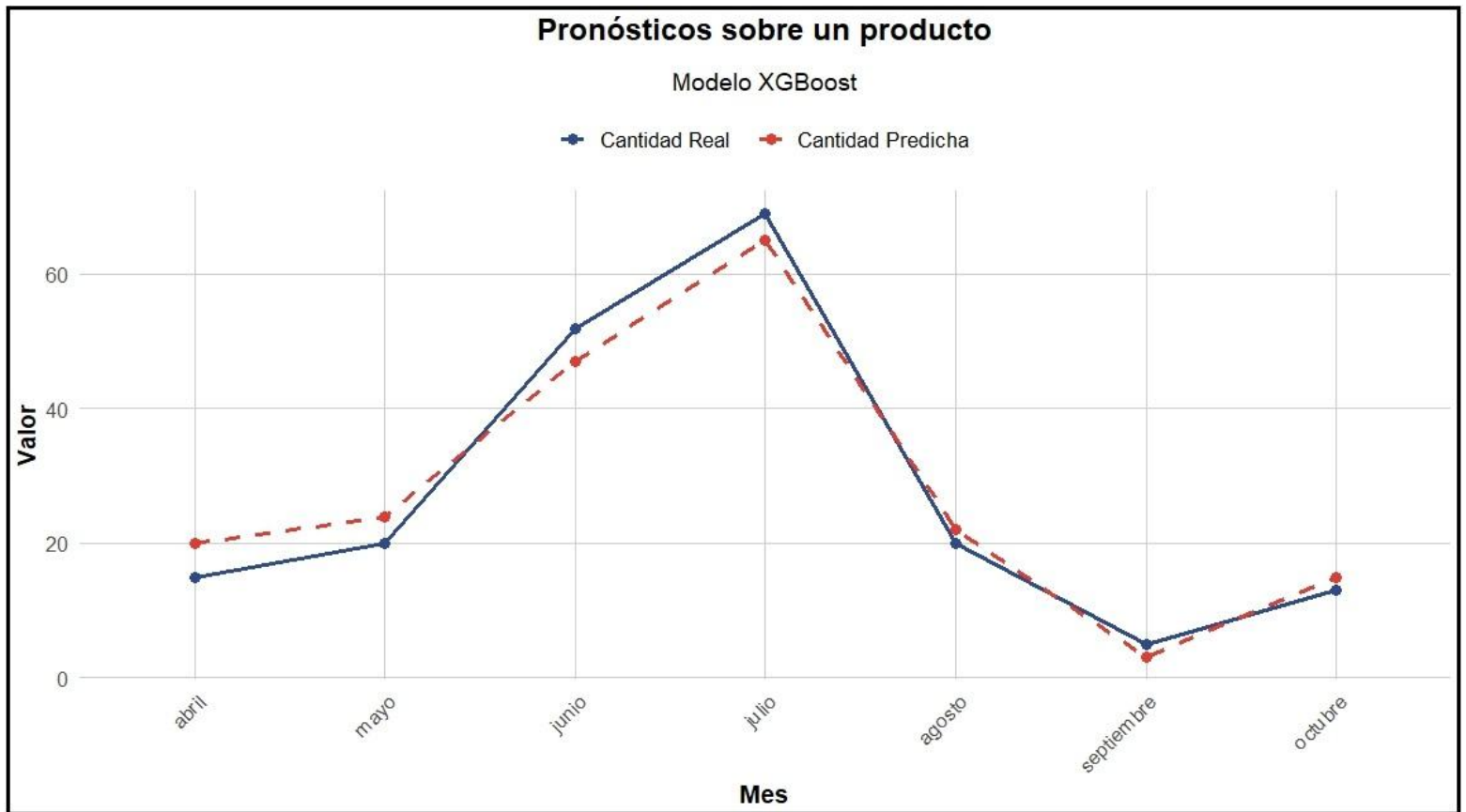


Figura 27 – Modelo XGBoost: Pronóstico sobre producto seleccionado al azar.

El modelo evidencia un desempeño notable al capturar la dinámica de la demanda del negocio. Las predicciones replican con fidelidad los niveles mensuales de venta, anticipando con precisión los picos estacionales característicos, en particular durante los meses de mayor actividad como junio y julio.

Asimismo, el modelo sigue correctamente las variaciones mensuales, tanto los incrementos como las caídas en la demanda, lo cual resulta clave para alinear los niveles de inventario. Respeta de forma consistente la estructura temporal de la serie, capturando no solo las tendencias generales, sino también los matices y la variabilidad propia de la demanda de neumáticos.

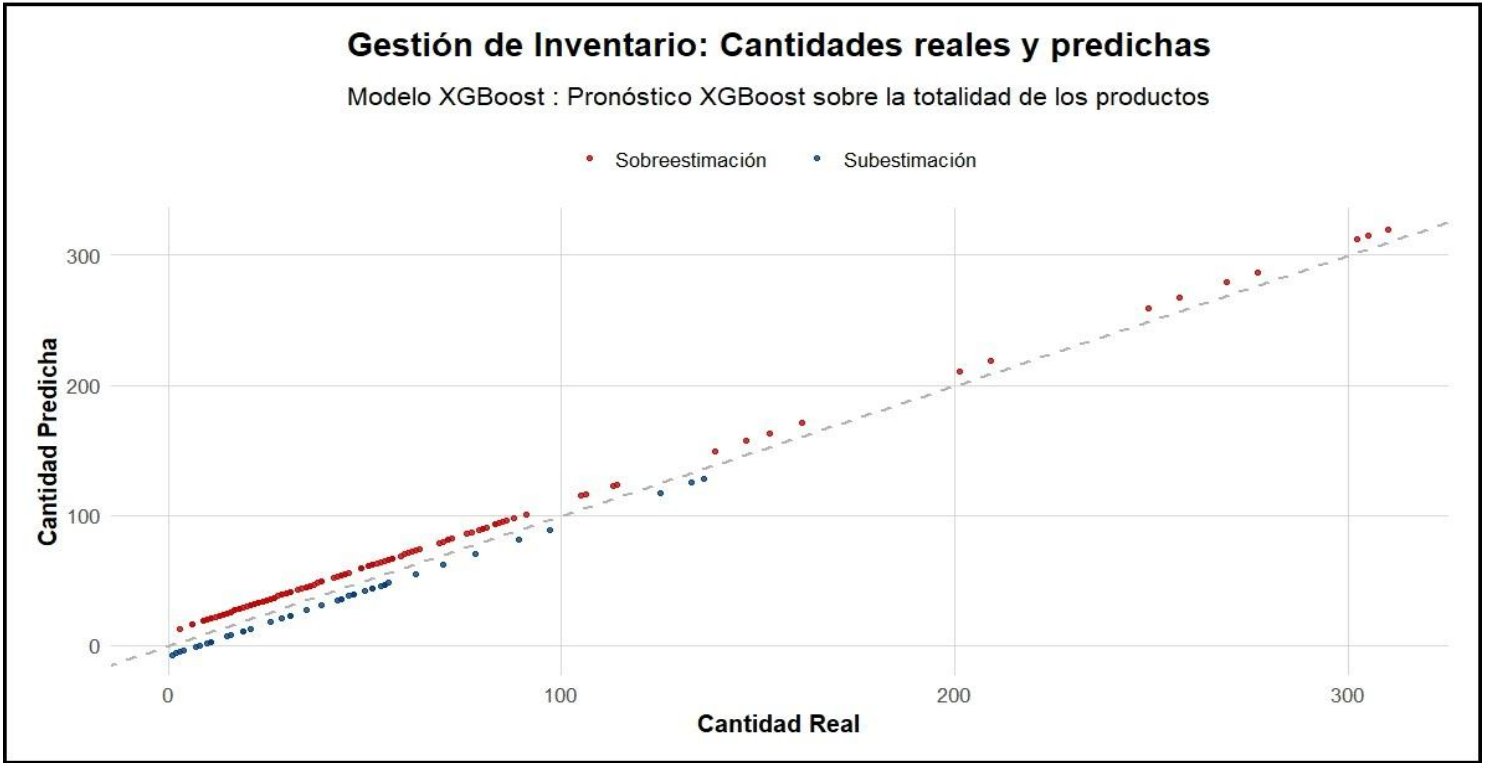


Figura 28 – Modelo XGBoost: Cantidades reales vs cantidades predichas.

Se observa una mejora sustancial en la gestión de inventarios a partir de las predicciones generadas por el modelo, las cuales muestran una sólida alineación con las cantidades realmente demandadas, minimizando de forma precisa tanto las sobreestimaciones como las subestimaciones. Esta precisión se ve reflejada en la alta concentración de puntos en torno a la línea de referencia ($y = x$) en los gráficos de dispersión, lo que evidencia la capacidad del modelo para reproducir de manera confiable el comportamiento real del negocio.

Gracias a este ajuste, los costos totales de almacenamiento se redujeron a \$98.592.485,65 durante el período analizado, generando un impacto económico altamente favorable para la compañía. En este sentido, XGBoost no solo optimiza la eficiencia operativa, sino que también aporta directamente a la rentabilidad mediante una gestión más efectiva.

V. Conclusión.

La presente tesis abordó una problemática crítica en la operación de la empresa analizada, vinculada con la ineficacia del modelo actual de pronóstico de demanda, sustentado en un enfoque manual, intuitivo y carente de respaldo analítico. Esta práctica tradicional ha derivado en una acumulación sostenida de inventario excedente, elevados costos de almacenamiento y una limitada capacidad de respuesta frente a los cambios en la demanda del mercado.

Frente a este escenario, se diseñó e implementó un sistema de pronóstico automatizado basado en la evaluación comparativa de diversas metodologías estadísticas y de machine learning: Media Móvil, Prophet, ARIMA, SARIMA, y XGBoost. Cada modelo fue entrenado y testeado sobre la base de datos histórica de la empresa, considerando las particularidades del sector de comercialización de neumáticos, caracterizado por una demanda estacional, cíclica y sujeta a factores operativos y logísticos relevantes.

Entre todas las alternativas exploradas, el modelo SARIMA demostró el mejor desempeño global, alcanzando un WAPE total de 0.15, el valor más bajo entre los modelos evaluados. Esta superioridad se debe a su capacidad para capturar patrones estacionales de manera explícita, combinando componentes autorregresivos, de medias móviles y diferenciación estacional. En el caso analizado, donde la demanda exhibe picos definidos en los meses de junio-julio y diciembre-enero, SARIMA mostró una adaptación precisa a los ciclos, ajustando con solidez las proyecciones mensuales. A ello se suma su facilidad de implementación, estabilidad operativa y alto nivel de interpretabilidad, cualidades que lo convierten en una herramienta especialmente valiosa para entornos que requieren previsiones confiables y sostenidas en el tiempo.

La aplicación del modelo permite mejorar reducir de forma significativa los costos asociados al exceso de inventario. Si bien el impacto financiero fue el principal foco de análisis, esta tesis promueve la sustitución de un enfoque empírico por una práctica sistemática, replicable y respaldada por evidencia cuantitativa, capaz de integrarse de manera efectiva en los procesos de toma de decisiones estratégicas.

Este caso evidencia que, aun en empresas con estructuras tradicionales, es posible adoptar herramientas analíticas avanzadas con resultados concretos. El uso de modelos predictivos no solo potencia la eficiencia operativa, sino que promueve una cultura organizacional orientada a los datos, sentando las bases para una evolución continua.

Mirando hacia adelante, existen múltiples caminos para profundizar y enriquecer esta estrategia de pronóstico. Por un lado, la incorporación de variables exógenas, como indicadores macroeconómicos, condiciones climáticas o cambios normativos, permitiría capturar con mayor precisión los determinantes de la demanda. Por otro, avanzar hacia esquemas de pronóstico agregado, complementarios a los pronósticos individuales por producto, posibilitaría mejorar la planificación estratégica, facilitar decisiones de compra consolidadas y detectar con mayor robustez las tendencias generales del mercado. Esta aproximación resulta particularmente pertinente en sectores como el analizado, donde la estacionalidad se combina con una amplia variedad de productos y una logística de distribución compleja.

En definitiva, lo desarrollado en esta tesis constituye un punto de partida para una transformación más profunda, en la que los datos ya no se conciben como un insumo estático, sino como un activo estratégico central para impulsar decisiones empresariales más inteligentes, oportunas y efectivas.

VI. Reflexiones finales.

La incorporación de un modelo predictivo riguroso en la gestión de la demanda representa un cambio sustantivo en la forma en que la empresa puede administrar sus recursos, tomar decisiones informadas y adaptarse a contextos de mercado dinámicos. Más allá de brindar una respuesta eficaz a la problemática del exceso de inventario, la solución desarrollada en esta tesis permite vislumbrar oportunidades de transformación en dimensiones clave de la operación, como la política de abastecimiento, la estrategia de almacenamiento y la gestión integral de la cadena de suministro.

Al alinear las decisiones comerciales con patrones reales de consumo, se favorece una asignación más eficiente del capital, se reduce el inmovilizado y se optimiza el uso de la capacidad logística. Estas mejoras no solo impactan en términos operativos, sino que contribuyen directamente a fortalecer la rentabilidad del negocio, consolidando ventajas competitivas tangibles en un sector altamente exigente como es el de la comercialización de neumáticos.

Asimismo, el modelo trasciende su rol técnico para posicionarse como una herramienta estratégica de alto valor. Su capacidad para anticipar la evolución de la demanda brinda soporte a decisiones más complejas, tales como la planificación de precios, la gestión de márgenes y el diseño de ciclos de vida de productos. Este tipo de análisis prospectivo resulta especialmente relevante en mercados caracterizados por una elevada estacionalidad, diversidad de referencias y presiones competitivas crecientes, donde la agilidad y la precisión constituyen pilares diferenciadores.

No obstante, la adopción plena de este nuevo sistema exige superar desafíos organizacionales importantes. Uno de los principales radica en la necesidad de promover un cambio cultural hacia una toma de decisiones basada en datos, en contraposición con prácticas tradicionales fundadas en la intuición o la experiencia individual. Para ello, será crucial generar procesos de formación, documentación y adopción que faciliten la apropiación del modelo por parte de los equipos de trabajo, consolidando así una cultura analítica dentro de la organización.

En segundo lugar, garantizar la vigencia del sistema en el tiempo implicará integrarlo de forma fluida con las plataformas de información existentes, automatizar la captura de datos y desarrollar interfaces que faciliten su uso transversal. La actualización periódica de sus parámetros, así como la revisión de los supuestos que lo sustentan, se volverán elementos clave para preservar su capacidad explicativa frente a cambios en los comportamientos de consumo o en el contexto económico general.

En definitiva, lo abordado a lo largo de esta tesis representa el comienzo de un proceso transformador hacia una gestión empresarial más analítica, estructurada y basada en evidencia. La verdadera fortaleza del enfoque desarrollado no reside únicamente en su precisión predictiva, sino en su capacidad para impulsar una evolución cultural dentro de la organización, fomentando el uso sistemático de los datos como insumo clave para la toma de decisiones. Este trabajo marca un paso firme hacia un modelo de gestión más inteligente y adaptable, capaz de responder con agilidad a los desafíos del entorno y de consolidar mejoras sostenibles en términos operativos y estratégicos.

VII. Bibliografía.

1. Jaimes Campos, D y López Zúñiga, E. (2021). Modelo de Forecast para predecir la demanda semanal de alimentos y bebidas de consumo masivo. Universidad de los Andes. <https://repositorio.uniandes.edu.co/entities/publication/83fbf165-4b6a-4ebd-8100-28745d0b8c0e>
2. Zalazar Hornig, Eduardo Javier & Jose Alexander. (2016). Estimación de demanda utilizando redes neuronales: El caso de una empresa manufacturera en Chile. Science Direct. <http://repositorio.udec.cl/jspui/handle/11594/7221>
3. Hanke, J. (2010). Pronosticos en Los Negocios (9na. ed.). Pearson Educación.
4. Fildes, R., Shaohui, M., & Kolassa, S. (2019, 12). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283-1318. <https://doi.org/10.1016/j.ijforecast.2019.06.004>
5. The Wiley Blackwell Companion to Tourism. Wiley.
6. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
7. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
8. Lew, A. A., Hall, C. M., & Williams, A. M. (2014). Urban Tourism and the Experience Economy. In
9. [https://www.marketsanddata.com/industry-reports/south-america-tyre-market\(2023\)](https://www.marketsanddata.com/industry-reports/south-america-tyre-market(2023))
10. McCarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, 25, 303-324.

[http://refhub.elsevier.com/S0169-2070\(19\)30192-X/sb152](http://refhub.elsevier.com/S0169-2070(19)30192-X/sb152)

11. Peterson, R. T. (1993). Forecasting practices in retail industry. *Journal of Business Forecasting*, 12, 11-14. [http://refhub.elsevier.com/S0169-2070\(19\)30192-X/sb171](http://refhub.elsevier.com/S0169-2070(19)30192-X/sb171)
12. Zukin, S. (1998). Urban Life Styles: Diversity and Standardization in Spaces of Consumption.
13. *Urban Studies*, 35, 825-839. <https://doi.org/10.1080/0042098984574>
14. Ali, O. G., Sayin, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*. doi:10.1016/j.eswa.2009.04.052
15. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Wiley.
16. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
17. Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
18. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). Springer.
19. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
20. Domingos, P. (2012). A few useful things to know about machine learning.
21. Alon, I., Qi, M., & Safdowsky, R. J. (2001, Mayo). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and*
22. *Consumer Services*, 8(3), 147-156. [https://doi.org/10.1016/S0969-6989\(00\)00011-4](https://doi.org/10.1016/S0969-6989(00)00011-4)
23. Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for

- multiple regression and correlations. *Journal of the American Statistical Association*, 80(391), 614-619. <https://doi.org/10.2307/2288477>
24. Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bounds constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16(5) 1190-1208. <https://doi.org/10.1137/0916069>
25. Choi, T.-M., Hui, C.L., Liu, N., Ng, S.-F., & Yu, Y. (2014, Marzo). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, 59, 84-92. <https://doi.org/10.1016/j.dss.2013.10.008>
26. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36, 54-76. doi.org/10.1016/j.ijforecast.2019.04.014
27. Mentzer, J. T., & Cox, J. E. (1984). Familiarity, application, and performance of sales forecasting techniques. *Journal of Forecasting*, 3(1), 27-36. doi.org/10.1002/for.3980030104
28. Prophet | Forecasting at scale. (n.d.). Meta Open Source. Retrieved December 27, 2022, from <https://facebook.github.io/prophet/>