



Pontificia Universidad
JAVERIANA
Cali

Asignación Inteligente de Precios a Artículos de Calzado en una Compañía de Retail

Gabriel Orozco Castaño

Código: 8975858

Proyecto Aplicado para optar al título de magíster en Ciencia de Datos

Director

Mg. David Arango Londoño

Pontificia Universidad Javeriana, Cali

Facultad de Ingeniería y Ciencias

Maestría en Ciencia de Datos

Santiago de Cali, Colombia

Junio 11 de 2025

TABLA DE CONTENIDO

Introducción	10
1. Definición del Problema	12
1.1. Planteamiento del Problema	12
1.2. Formulación del Problema	14
1.2.1. Pregunta del Objetivo General	14
1.2.2. Preguntas Principales de los Objetivos Específicos	14
2. Objetivos	15
2.1. Objetivo general	15
2.2. Objetivos Específicos	15
3. Marco Teórico y Antecedentes	16
3.1. Marco Teórico	16
3.2. Antecedentes	19
4. Preparación de Datos	23
5. Identificación y Análisis de Indicadores Clave	27
6. Evaluación del Impacto de la Variación de Precios sobre la Demanda	32
6.1. Selección de los Modelos de Aprendizaje Automático	34
6.1.1. Regresión Logística	34
6.1.2. Random Forest	35
6.1.3. XGBoost	35
6.2. Implementación de los Tres Modelos de Aprendizaje Automático	36
6.3. Evolución de Métricas Clave de Desempeño Comercial en la Categoría Escolar	49
6.4. Aplicación de Pruebas ADF y KPSS	51
6.5. Selección de Modelo de Series de Tiempo	54
6.5.1. Promedio Móvil	54
6.5.2. Holt-Winters	55
6.5.3. Arima	55
6.5.4. Sarima	56
6.5.5. Arimax	56
6.4.6. Prophet	57
6.6. Comparación de resultados de los modelos de series de tiempos	57

7. Desarrollo de un Modelo Estadístico para la Optimización Estratégica de Precios	64
7.1. Análisis de Correlaciones de Variables	64
7.2. Características y Resultados de los Modelos Arimax y Prophet.....	67
7.3. Selección e Implementación del Modelo Prophet	69
7.4. Análisis de Elasticidad para los Tres Artículos Seleccionados	78
7.5. Estrategias de optimización de precios	80
8. Evaluación del Rendimiento y Precisión del Modelo de Aprendizaje Estadístico	84
8.1. Prueba de Estrés por Volumen de Datos	84
8.2. Prueba de Sensibilidad a la Volatilidad.....	84
8.3. Prueba de Manejo de Datos Faltantes	85
8.4. Prueba de Eficiencia Computacional	85
8.5. Prueba de Determinismo del Modelo	85
8.6. Prueba de Integración de Regresores Externos.....	86
8.7. Prueba de Reentrenamiento con Datos Nuevos	86
9. Conclusiones y Trabajos Futuros	88
9.1. Conclusiones	88
9.2. Trabajos Futuros.....	89
10. Referencias Bibliográficas	91

LISTA DE FIGURAS

Figura 1. <i>Participación venta por gran categoría</i>	25
Figura 2. <i>Comparación precio promedio 2023</i>	25
Figura 3. <i>Comparación precio promedio 2024</i>	25
Figura 4. <i>Métricas importantes para el negocio</i>	28
Figura 5. <i>Correlación de variables para el año 2023</i>	28
Figura 6. <i>Correlación de variables para el año 2024</i>	28
Figura 7. <i>Evolución de la venta par (2023 & 2024)</i>	29
Figura 8. <i>Evolución de la venta neta (2023 & 2024)</i>	29
Figura 9. <i>Evolución de la venta margen nominal (2023 & 2024)</i>	29
Figura 10. <i>Evolución de la venta margen porcentual (2023 & 2024)</i>	30
Figura 11. <i>Evolución del precio promedio neto (2023 & 2024)</i>	30
Figura 12. <i>Evolución venta pares por gran categoría – 2023</i>	32
Figura 13. <i>Evolución venta neta por gran categoría - 2023</i>	33
Figura 14. <i>Evolución venta margen nominal por gran categoría - 2023</i>	33
Figura 15. <i>Evolución venta pares por gran categoría – 2024</i>	33
Figura 16. <i>Evolución venta neta por gran categoría - 2024</i>	33
Figura 17. <i>Evolución venta margen nominal por gran categoría - 2024</i>	34
Figura 18. <i>Comparativo métricas</i>	40
Figura 19. <i>Análisis QQ de la variación venta par</i>	44
Figura 20. <i>Análisis QQ de la variación precio promedio</i>	44
Figura 21. <i>Curva ROC para Regresión Logística</i>	45
Figura 22. <i>Curva ROC para Random Forest</i>	45

Figura 23. <i>Curva ROC para XGBoost</i>	46
Figura 24. <i>Análisis QQ de la variación venta par escolar</i>	49
Figura 25. <i>Análisis QQ de la variación precio promedio escolar</i>	49
Figura 26. <i>Evolución venta pares escolar (2023 & 2024)</i>	50
Figura 27. <i>Evolución comparativa entre venta pares y precio promedio bruto del stock</i>	50
Figura 28. <i>Evolución comparativa entre venta pares y porcentaje descuento</i>	51
Figura 29. <i>Evolución comparativa entre venta pares y stock pares</i>	51
Figura 30. <i>Autocorrelación (ACF) & Autocorrelación parcial (PACF)</i>	52
Figura 31. <i>Comparación de los modelos de predicción de series de tiempo</i>	58
Figura 32. <i>Pronóstico y errores residuales ARIMAX</i>	61
Figura 33. <i>Pronóstico y errores residuales PROPHET</i>	61
Figura 34. <i>Comparación venta real vs modelo: Artículo 1</i>	71
Figura 35. <i>Comparación venta real vs modelo: Artículo 2</i>	71
Figura 36. <i>Comparación venta real vs modelo: Artículo 3</i>	71
Figura 37. <i>Proyección venta 2025: Artículo 1</i>	72
Figura 38. <i>Proyección venta 2025: Artículo 2</i>	72
Figura 39. <i>Proyección venta 2025: Artículo 3</i>	73
Figura 40. <i>Residuales: Artículo 1</i>	73
Figura 41. <i>Residuales: Artículo 2</i>	74
Figura 42. <i>Residuales: Artículo 3</i>	74
Figura 43. <i>Distribución de los residuos: Artículo 1</i>	74
Figura 44. <i>Distribución de los residuos: Artículo 2</i>	74
Figura 45. <i>Distribución de los residuos: Artículo 3</i>	75

Figura 46. <i>Comparativo de simulación de precios</i>	82
Figura 47. <i>Tipo de prueba y resultado</i>	87

LISTA DE TABLAS

Tabla 1. <i>Variables de la base de datos</i>	23
Tabla 2. <i>Componentes de la base de datos</i>	26
Tabla 3. <i>Tabla con indicadores importantes</i>	27
Tabla 4. <i>Transformación de la base de datos - unificación de artículos</i>	37
Tabla 5. <i>Indicadores para cálculo de elasticidad</i>	38
Tabla 6. <i>Resultados generales Regresión Logística - predicción elasticidad</i>	39
Tabla 7. <i>Resultados generales Random Forest - predicción elasticidad</i>	39
Tabla 8. <i>Resultados generales XGBoos - predicción elasticidad</i>	39
Tabla 9. <i>Comparativo métricas 2</i>	40
Tabla 10. <i>Matriz de confusión (Regresión Logística, Random Forest, XGBoost)</i>	41
Tabla 11. <i>Visualización de la base de datos con agrupamiento en school</i>	47
Tabla 12. <i>Resultados Regresión Logística total categoría escolar</i>	47
Tabla 13. <i>Resultados Random Forest total categoría escolar</i>	47
Tabla 14. <i>Resultados XGBoost total categoría escolar</i>	47
Tabla 15. <i>Comparación métricas entre Regresión Logística, Random Forest y XGBoost</i>	48
Tabla 16. <i>Prueba ADF y KPSS</i>	52
Tabla 17. <i>Prueba ADF con tendencia</i>	54
Tabla 18. <i>Cuadro conceptual comparativo modelos de series de tiempo</i>	58
Tabla 19. <i>Comparación de las métricas de los modelos de series de tiempo</i>	59
Tabla 20. <i>Prueba de validez Arimax</i>	60
Tabla 21. <i>Prueba de validez Prophet</i>	60
Tabla 22. <i>Incorporación de variables exógenas</i>	63

Tabla 23. <i>Incorporación de semana estacionaria</i>	64
Tabla 24. <i>Correlación de las variables seleccionadas</i>	65
Tabla 25. <i>Características de modelos Arimax y Prophet</i>	68
Tabla 26. <i>Comparativo de los resultados de los modelos</i>	69
Tabla 27. <i>Estadística de los residuos nominales: Artículo 1</i>	76
Tabla 28. <i>Top 10 semanas con mayor diferencia: Artículo 1</i>	76
Tabla 29. <i>Estadística de los residuos nominales: Artículo 2</i>	76
Tabla 30. <i>Top 10 semanas con mayor diferencia: Artículo 2</i>	76
Tabla 31. <i>Estadística de los residuos nominales: Artículo 3</i>	77
Tabla 32. <i>Top 10 semanas con mayor diferencia: Artículo 3</i>	77
Tabla 33. <i>Resultados del cálculo de elasticidad de la demanda</i>	79

RESUMEN

El proyecto tuvo como objetivo analizar el impacto de las variaciones de precios sobre la demanda de una empresa del sector calzado, buscando fortalecer su competitividad y sostenibilidad. Inicialmente, se propuso diseñar un algoritmo de optimización de precios basado en la elasticidad precio-demanda, sin embargo, el análisis de los datos históricos reveló que los cambios de precios no se realizaron bajo condiciones estadísticamente controladas, lo que impidió calcular una elasticidad confiable. Ante esta limitación, se redirigió el enfoque metodológico hacia la proyección de la demanda utilizando variables clave como el stock disponible, el precio promedio, los descuentos aplicados, la semana estacionaria y el comportamiento semanal de las ventas. El estudio se centró en productos de la categoría escolar. Aunque durante el desarrollo del estudio se exploraron distintos modelos de machine learning y de series de tiempo, ARIMAX y Prophet ofrecieron resultados satisfactorios en términos de precisión y ajuste a los datos. Tras una comparación detallada entre ambos, se optó por el modelo Prophet, ya que presentó un mejor rendimiento y un comportamiento más estable, lo que lo hace más adecuado para el contexto analizado. Este modelo permitió simular escenarios futuros de ventas para el año 2025, proporcionando a la empresa una herramienta predictiva robusta que mejora la toma de decisiones. Además, se identificaron oportunidades para optimizar la política de precios mediante la implementación de un sistema de asignación de precios más estructurado, lo que podría mejorar la relación entre el precio y la demanda, optimizando la estrategia comercial de la empresa.

Palabras clave: indicadores de gestión, elasticidad de precio de la demanda, asignación de precios, modelamiento estadístico, diseño de algoritmo, modelo de aprendizaje estadístico.

INTRODUCCIÓN

La competitividad en el mercado está estrechamente ligada a la capacidad de las empresas para comprender y gestionar de manera eficiente la relación entre los precios de sus productos y la demanda de los consumidores. Este proyecto se enfoca en analizar dicha relación en una empresa específica del sector calzado, revisando diversas categorías de productos, pero con un enfoque final en los artículos tipo escolar. Para ello, se implementó el modelo de predicción Prophet, que permitió estimar la demanda futura en función de variables claves, brindando a la organización una herramienta estratégica para anticiparse a las dinámicas del mercado y tomar decisiones informadas en un entorno cada vez más competitivo.

Al centrarse en una empresa dedicada a la producción y comercialización de calzado, se aplicaron métodos estadísticos avanzados que permitieron identificar diversos factores internos relevantes para la gestión comercial. No obstante, la ausencia de una estructura sólida para la asignación de precios limitó la posibilidad de evaluar de manera confiable el impacto de la elasticidad precio-demanda. Esto se debe a que el cálculo de la elasticidad se realizó agrupando en un solo conjunto a todos los artículos tipo escolar, lo que arrojó un valor extremo, consecuencia del método actual “empírico” utilizado para fijar los precios, el cual no cuenta con el respaldo de un análisis estadístico riguroso. Por otro lado, al realizarse un análisis detallado de nueve artículos agrupados según tres grupos específicos: “venta alta y constante”, “venta intermedia y menos constante” y “venta baja y no constante”, se obtuvieron resultados más estables en aquellos productos con un volumen de venta alta y constante.

A pesar de esta limitación, los análisis permitieron evidenciar la falta de un sistema estructurado de asignación de precios, lo cual representa una valiosa oportunidad para que la

organización implemente procesos estadísticos adecuados. Esto facilitaría una política de precios más estratégica sustentada en proyecciones de demanda.

La elasticidad precio es una métrica clave en economía que revela la sensibilidad de los consumidores ante cambios en los precios de los productos. Un análisis de este tipo permite identificar aquellos productos cuyo consumo es más susceptible a las variaciones de precio, proporcionando información crucial para la toma de decisiones estratégicas.

Una vez identificado el obstáculo relacionado con la dificultad de analizar de manera profunda y confiable la elasticidad precio-demanda, como resultado de la ausencia de un proceso estructurado para la asignación de precios, se desarrolló un modelo estadístico de predicción de la demanda utilizando *Prophet*, una herramienta de modelado de series temporales desarrollada por Facebook. Este modelo incorporó variables claves asociadas al comportamiento del consumidor, tales como el stock disponible, el precio promedio, los descuentos aplicados, la semana estacionaria y el comportamiento semanal de las ventas, con un enfoque específico en la categoría escolar.

1. DEFINICIÓN DEL PROBLEMA

Una correcta asignación de precios a los productos que participan en un mercado competitivo no solo garantiza la continuidad de las empresas en el tiempo, sino que también permite maximizar la rentabilidad, optimizar el flujo de efectivo e impulsar el crecimiento en las ventas. Además, influye directamente en la retención de clientes y en la captación de nuevos, al ofrecer un valor ajustado a sus necesidades y preferencias.

Sin embargo, en la empresa donde se desarrolló este proyecto, no existía un proceso estructurado ni un modelo estadístico que facilitara una asignación de precios fundamentada en datos objetivos. Esta carencia limitaba la capacidad de la organización para evaluar de forma analítica el precio actual de sus productos y su relación con variables clave que inciden en el comportamiento de la demanda.

1. 1. Planteamiento del Problema

El principal problema radica en la falta de aplicación de técnicas analíticas y modelos que permitan establecer precios de manera estratégica, considerando las variables más relevantes tanto a nivel interno como del entorno económico. En este caso, la ausencia de un enfoque científico en el análisis de datos revela que la empresa no está aprovechando su información histórica para comprender las tendencias de un mercado en constante cambio. Esta carencia representa uno de los principales desafíos para las organizaciones en el siglo XXI (Barbosa Moyano, 2020); ya que limita la capacidad de identificar las preferencias de los clientes y otros factores claves que influyen directamente en la demanda y en la fijación de precios.

Por otro lado, la asignación de precios basada en suposiciones puede conducir a decisiones erróneas que afecten negativamente la estabilidad financiera y comercial de la organización, así como su posicionamiento de marca. Esta situación se vuelve especialmente crítica si se considera

que “las tácticas de precios promocionales son determinantes en las decisiones de compra” (Peceros y Salsedo, 2022, p. 10), lo que resalta la necesidad de implementar estrategias de precios respaldadas por datos y análisis rigurosos.

En la actualidad, la carencia de análisis científico en el proceso de asignación de precios impide unos niveles óptimos de competitividad en el mercado, ya que, al no contar con el precio correcto, en el artículo correcto, puede generar disminución en las ventas, impactando de manera negativa el flujo de caja, la rentabilidad de la empresa y otros indicadores de alto nivel para la organización, como es el caso del envejecimiento de la mercadería, baja rotación de ventas y exceso de inventarios. También es muy importante considerar la importancia de la fijación de precios respecto a las ganancias que la organización desea obtener, debido a que

Tienen mucha relación con la gestión de ingresos, ya que viene a ser el proceso final para realizar la venta efectiva del producto. Además, a partir de los costos del producto, la empresa decide cuánto más desea ganar por la venta del producto, siempre y cuando compita de manera eficiente en el mercado donde se desenvuelve. (Guzmán Llontop, 2020, p. 23)

Para abordar esta problemática, resulta fundamental implementar un sistema de asignación de precios basado en análisis científicos y datos precisos, lo cual permite optimizar la competitividad de la empresa y fortalecer su desempeño financiero.

Ajustar los precios en función de la demanda permite una fijación más precisa y alineada con los objetivos de rentabilidad, lo que facilita el incremento de ingresos y respalda la toma de decisiones estratégicas informadas. Este enfoque no solo mejora el flujo de caja, sino que también fortalece la posición competitiva y el valor de marca, contribuyendo al crecimiento sostenido y a la sostenibilidad a largo plazo de la organización.

El análisis de datos permite identificar patrones, lo cual constituye un “mecanismo

necesario para distinguir unas cosas de otras, relacionar cosas semejantes, formar grupos de cosas, describir objetos, tomar y explicar decisiones” (Carrasco Ochoa y Martínez Trinidad, 2011, p. 5). A través de este enfoque, es posible establecer relaciones que optimicen la fijación de precios, lo que se traduce en una mejora del desempeño financiero de la empresa. Además, facilita la asignación adecuada de precios basada en las variables más relevantes definidas por la organización. Esto no solo proporciona un entendimiento profundo del proceso y enriquece el conocimiento interno, sino que también permite implementar mejoras operativas que fortalecen la competitividad en el mercado.

1.2. Formulación del Problema

A continuación, se muestran las preguntas a responder:

1.2.1. Pregunta del Objetivo General

¿Cómo a través de los datos históricos de una empresa minorista, es posible diseñar un modelo analítico que permita generar adecuadas estrategias comerciales a partir de la asignación inteligente de precios?

1.2.2. Preguntas Principales de los Objetivos Específicos

1. ¿Cuáles son los indicadores que se consideran más influyentes en la gestión de precios?
2. ¿Cómo se puede medir mediante un análisis estadístico la elasticidad de precios y su impacto en la demanda?
3. ¿Cómo se puede diseñar un modelo de aprendizaje estadístico de optimización de precios que integre la elasticidad de precios y los indicadores más claves de la compañía para lograr una asignación inteligente de precios?
4. ¿Cuál es el método más adecuado para verificar la calidad de las pruebas del modelo de aprendizaje estadístico y cómo asegurar que no estén sesgadas?

2. OBJETIVOS

2.1. Objetivo general

Diseñar un modelo analítico a través de una revisión detallada estadística de los datos de una empresa minorista de calzado ubicada en la ciudad de Manizales, con el propósito de implementar una estrategia de asignación inteligente de precios.

2.2. Objetivos Específicos

1. Analizar los indicadores más importantes para la compañía, como: niveles de stock, rotación de inventarios, envejecimiento de la mercadería, entre otros.
2. Realizar un análisis estadístico de elasticidad de precios para profundizar en el entendimiento del impacto que la variabilidad de los precios ejerce sobre la demanda de los productos.
3. Desarrollar un modelo de aprendizaje estadístico para la optimización de precios que integre la elasticidad de precios y otras variables críticas para su fijación estratégica.
4. Validar el rendimiento y la precisión del modelo de aprendizaje estadístico, mediante una serie de pruebas permitiendo determinar su idoneidad con el fin de garantizar su implementación.

3. MARCO TEÓRICO Y ANTECEDENTES

La revisión de los modelos estadísticos aplicables a la fijación eficiente de precios resulta fundamental para el éxito de este proyecto. A continuación, se presentan una serie de temas claves, como la elasticidad precio-demanda, el comportamiento de la oferta y la demanda, y los indicadores más importantes de rendimiento (KPI), entre otros, que aportan significativamente al enfoque teórico y a la implementación práctica de la investigación.

3.1. Marco Teórico

Se exponen los aportes teóricos que fundamentan el objeto de estudio, como resultado de las investigaciones previas:

- **Producto:** es cualquier bien tangible o intangible ofrecido en el mercado para satisfacer una necesidad o deseo. En marketing, incluye atributos físicos, beneficios funcionales y simbólicos (Kotler y Armstrong, 2017).
- **Mercado:** se entiende como el conjunto de compradores y vendedores que interactúan para intercambiar bienes y servicios (Mankiw, 2018).
- **Consumidor:** El consumidor es la persona o grupo que adquiere productos o servicios para satisfacer sus necesidades (Schiffman y Kanuk, 2010).
- **Oferta:** es la cantidad de bienes o servicios que los productores están dispuestos a vender en el mercado a distintos precios durante un período determinado (Samuelson y Nordhaus, 2010).
- **Demanda:** es la cantidad de un bien que los consumidores desean y pueden adquirir a diferentes niveles de precio en un periodo específico (Mankiw, 2018).
- **Precio:** es el valor monetario que se asigna a un bien o servicio. Es un elemento clave en la mezcla de marketing y en la percepción del valor del producto (Kotler y Keller, 2016).

- Asignación de precios: Es el proceso mediante el cual se determina el valor de adquisición de un bien o servicio, considerando costos, competencia, percepción del consumidor y estrategias comerciales (Nagle *et al.*, 2016).
- Elasticidad precio de la demanda: mide la sensibilidad de la cantidad demandada ante cambios en el precio del producto (Varian, 2014).
- Fórmula elasticidad precio de la demanda: $((\% \text{ cambio en la cantidad demandada}) / (\% \text{ cambio en el precio}))$ (Mankiw, 2018).
- Estimación de la demanda: es la proyección de la cantidad que se espera que los consumidores compren en un periodo determinado, útil para planificar producción y ventas (Churchill & Iacobucci, 2010).
- Inventario de artículos: es el conjunto de bienes disponibles que una empresa posee para su venta o transformación en un momento determinado (Heizer y Render, 2014).
- KPI (Key Performance Indicators): Son métricas cuantificables que permiten evaluar el desempeño de una organización o de procesos específicos.

La ciencia de datos se ha consolidado como una herramienta valiosa para el modelado organizacional, ya que permite comprender de forma más profunda y precisa la realidad de una empresa. Su aplicación facilita el análisis de grandes volúmenes de información, la identificación de patrones y tendencias, y la generación de insights claves que fortalecen la toma de decisiones estratégicas. Como lo señalan García *et al.* (2018): “Los modelos son las relaciones, reglas, patrones y resúmenes extraídos tras el análisis de los datos. Gracias a esta extracción se obtiene el conocimiento útil que se estaba buscando. Estos modelos pueden ser tanto descriptivos como predictivos” (p. 31). De este modo, la ciencia de datos no solo aporta valor analítico, sino que también habilita capacidades predictivas fundamentales para la mejora continua y la

competitividad empresarial. A través de la aplicación de métodos estadísticos y algoritmos avanzados, es posible procesar grandes volúmenes de información, descubrir patrones ocultos y prever tendencias futuras, lo que brinda a las empresas la capacidad de tomar decisiones más informadas y estratégicas. Los modelos generados a partir del análisis de datos ya sean descriptivos o predictivos, permiten extraer relaciones y patrones claves que proporcionan el conocimiento necesario para abordar desafíos y aprovechar oportunidades. Este enfoque basado en datos optimiza la toma de decisiones, mejorando la eficiencia y competitividad de la organización en el mercado.

Teniendo en cuenta la importancia de encontrar el proceso correcto que ayude a determinar la elasticidad de precio de la demanda, se consultan diferentes métodos de fijación de precios, siendo el más utilizado el “análisis marginalista” o también llamado “método de elasticidad demanda-precio”; por otro lado, también se evidencian estudios con el “método del valor percibido” y el “método del valor de uso”. El método marginalista:

Parte del conocimiento de la función de demanda y de costos, postulando que la demanda es una función del precio y que los costos son conocidos. Basándose en estas premisas y mediante la utilización de la optimización matemática, se puede obtener el nivel de precios mediante el cual se maximizan bien sea los beneficios o los ingresos. (Guerrero-Garzón, 2011, p. 15)

Teniendo en cuenta lo anterior, el método de elasticidad demanda-precio aplicado a la asignación de precios, determina el precio óptimo de un producto dentro del mercado, considerando que el precio se establece en el punto en donde la oferta y la demanda se igualan, sin dejar a un lado los costos marginales de producción y la utilidad marginal para los consumidores.

Para este análisis, se calculó la elasticidad precio de la demanda mediante las tasas de variación porcentual entre semanas consecutivas en las ventas y los precios. Este enfoque empírico proporciona una estimación de cómo responde la demanda ante cambios pequeños en el precio

dentro de períodos cortos, lo que resulta especialmente útil en contextos de comercialización dinámica. Según Economipedia (2024), la elasticidad precio de la demanda se define como el cambio porcentual en la cantidad demandada dividido entre el cambio porcentual en el precio del mismo bien. Este método es aplicable a series temporales de datos históricos para evaluar la sensibilidad del mercado frente a variaciones de precios (Question Pro, 2023).

Por otro lado, para efectos del presente trabajo, es importante reconocer la importancia de la Elasticidad cruzada de la demanda, la cual:

Es la relación que guarda el cambio de precio de un bien, respecto a la cantidad demandada de otro bien. Se define como el cambio porcentual en la cantidad demandada del primer bien, dividido entre el cambio porcentual del precio del segundo bien. (Ortiz Lara, 2020, p. 19)

En términos algebraicos se define el coeficiente de ECD de dos bienes A y B como:

$$\varepsilon_{(A,B)} = \frac{\Delta x_A/x_A}{\Delta P_B/P_B}$$

Ecuación 1 - Fórmula para cálculo de elasticidad de precios

Dónde:

$\varepsilon_{(A,B)}$: elasticidad cruzada de los bienes A y B

$\Delta x_A/x_A$: porcentaje de variación de la demanda del bien A

$\Delta P_B/P_B$: porcentaje de variación del precio del bien B

Si bien este trabajo no profundiza en el uso práctico de la elasticidad cruzada de la demanda, se reconoce su valor teórico para identificar relaciones entre productos, particularmente en mercados donde existen bienes sustitutos o complementarios. Este tipo de análisis permite anticipar cómo las variaciones de precio en un producto pueden afectar la demanda de otro, siendo una herramienta estratégica para la toma de decisiones comerciales más complejas.

3.2. Antecedentes

El análisis de la elasticidad precio de la demanda ha demostrado ser una herramienta de

gran utilidad para las empresas en el proceso de fijación de precios, ya que permite obtener una visión más amplia y precisa sobre cómo los cambios en los precios afectan la demanda de los productos. Según los estudios consultados, cuando se trabaja con un portafolio amplio de artículos, resulta apropiado recurrir a técnicas de agrupación propias de la estadística multivariada, las cuales permiten clasificar variables o individuos en función de su similitud. En este sentido, “las técnicas de agrupación y particularmente el análisis de conglomerados contribuyen enormemente al trabajo desarrollado” (Guerrero-Garzón *et al.*, 2012, p. 33). Por tanto, para la ejecución de este proyecto, se consideró indispensable la aplicación de clústeres de artículos basados en variables relevantes como categoría y procedencia de fabricación, lo cual permitió estructurar un panorama más homogéneo. Esta organización facilitó el desarrollo de un análisis estadístico más robusto, orientado a identificar la elasticidad de precios en grupos de artículos con características comunes.

Resulta de gran interés poder revisar proyectos en donde se ha implementado la elasticidad del precio. En el siguiente caso se presenta la aplicación de un modelo econométrico diseñado para analizar el comportamiento del consumo en función de variables como el precio, el ingreso y otros factores externos a lo largo del tiempo. El objetivo no es calcular directamente la elasticidad en su forma clásica, definida como el cambio porcentual en la cantidad demandada dividido entre el cambio porcentual en el precio, sino estimar los efectos marginales de estas variables sobre el consumo, mediante el siguiente modelo: $C_t = \beta_0 + \beta_1 P_t + \beta_2 Y_t + \delta Z_t + u_t$

Donde β_0 es el intercepto; β_1 y β_2 representan los parámetros estimados de la elasticidad precio de la demanda y la elasticidad ingreso en el largo plazo, respectivamente; δ es el coeficiente de la temperatura, (Z_t) y u_t es un término de error. La elasticidad en el largo plazo fue definida como el cambio porcentual en el consumo dado un cambio permanente en los precios o en el ingreso. La elasticidad en el corto plazo es el cambio porcentual en el consumo después del primer mes del cambio en el consumo o en el precio. (Moreno-Aguilar *et al.*, 2021, p. 577)

A continuación, se muestra una descripción más detallada de cada una de las partes que componen la fórmula:

C_t : Esta es la variable dependiente, que representa el consumo del bien o servicio en cuestión en el tiempo t . En términos simples, esta variable captura cuánto se consume de un producto dado.

β_0 : Este es el intercepto o la constante del modelo. Representa el valor del consumo C_t cuando todas las demás variables (precio, ingreso, etc.) son cero.

$\beta_1 P_t$: Aquí P_t es el precio del bien o servicio en el tiempo t , y β_1 es el coeficiente asociado a este precio. El valor de β_1 indica cómo cambia el consumo del bien en respuesta a un cambio en su precio. Este es uno de los componentes clave para medir la elasticidad precio de la demanda, ya que la elasticidad se calcula como el cambio porcentual en la cantidad demandada dividido entre el cambio porcentual en el precio. El coeficiente β_1 te da una idea directa de la sensibilidad del consumo respecto al precio.

$\beta_2 Y_t$: Aquí Y_t representa el ingreso de los consumidores en el tiempo t , y β_2 es el coeficiente que indica cómo varía el consumo del bien a medida que cambia el ingreso. Si β_2 es positivo, significa que, a medida que el ingreso de los consumidores aumenta, también aumenta el consumo del bien (lo que sugiere que el bien podría ser normal).

∂Z_t : Z_t son otras variables que podrían influir en el consumo, como factores externos o condiciones del mercado que no se han especificado explícitamente en la ecuación, pero que se consideran importantes. ∂ es el coeficiente que mide el efecto de estas variables en el consumo.

u_t : Este es el término de error o residual que captura las influencias no observadas o no incluidas en el modelo. Puede deberse a factores imprevistos o errores de medición en los datos.

Por otro lado, teniendo en cuenta la investigación realizada sobre “Elasticidades de

demanda por manzanas chilenas en el mercado de la unión europea: una estimación econométrica” de Cerda *et al.* (2004); en el cual la aplicación de medidas estadísticas como R^2 , las tablas de Durbin-Watson y la revisión de heterocedasticidad, entre otros; llegan a la conclusión de:

La demanda por manzanas chilenas en la UE es inelástica con respecto al ingreso y al precio relativo. En el primer caso, la baja elasticidad ingreso obtenida sugiere que las manzanas son un bien normal, lo cual significa que la demanda aumenta a medida que se incrementa el ingreso, aunque menos que proporcionalmente. (Cerda *et al.*, 2004, p. 407)

De esta forma, es posible encontrar diferentes trabajos investigativos referentes a la asignación de precios, a través del método de elasticidad de la demanda cuyo proceso se ha convertido en un recurso esencial para las empresas en cuanto a la ejecución de estrategias de fijación de precios, ya que se ofrece una comprensión profunda de cómo las variables en los precios inciden en la demanda de los productos, procedimiento que resulta esencial y enriquecedor para el proyecto que se desea desarrollar.

A través de la aplicación de técnicas avanzadas de estadística es posible realizar una segmentación efectiva de los datos, obteniendo un análisis más detallado y preciso, proporcionando un análisis robusto que conlleve a identificar la elasticidad del precio, considerando factores como venta, precio, stock, descuento, etc.

4. PREPARACIÓN DE DATOS

Los datos trabajados en el actual proyecto aplicado fueron tomados de una compañía de Retail dedicada a la fabricación y comercialización de calzado. La base cuenta con datos semanales desde el 1 de enero del 2023 al 31 de diciembre del 2024, lo que permite analizar dos años completos de comportamiento comercial.

La organización compartió la base de datos en formato Excel, la cual fue importada a Python para llevar a cabo todos los procesos de análisis exploratorio y modelado.

Con el fin de conocer los diferentes datos de la compañía, se realizó una revisión de la base de datos existente, para lo cual, se cuenta con 19 variables y 431.807 registros (Tabla 1). Estos datos son los más representativos para la organización, ya que brinda información detallada de los diferentes artículos, su comercialización y periodicidad.

Tabla 1. *Variables de la base de datos*

#	Column	Non-Null	Count	Dtype
0	Año	431807	non-null	int64
1	Semana	431807	non-null	int64
2	Articulo	431807	non-null	object
3	NOMBRE GRAN CAT	431807	non-null	object
4	NOMBRE CATEGORIA	431807	non-null	object
5	EdadSemanas	431807	non-null	int64
6	NombreMarca	431807	non-null	object
7	PCOSTO	431807	non-null	int64
8	PPUBLICO	431807	non-null	int64
9	Procedencia	431807	non-null	object
10	Vtapar	431807	non-null	int64
11	Vtaneto	431807	non-null	int64
12	Mgenvta	431807	non-null	int64
13	VtaCsto	431807	non-null	int64
14	Stkprs	431807	non-null	int64
15	StkCosto	431807	non-null	int64
16	StkPublicosin IVA	431807	non-null	int64
17	VtaBruta	431807	non-null	int64
18	Descuento	431807	non-null	int64

Una vez revisada la base de datos, se hace necesario realizar un análisis exploratorio, con el fin de conocer los datos y el tipo, como:

- Año: 2023 – 2024.
- Semana: 1 – 52.
- Artículo: códigos alfanuméricos de 8 dígitos.
- Nombre Gran Cat: 11 tipos de gran categoría de la A a la K.
- Nombre categoría: 44 tipos de categorías.

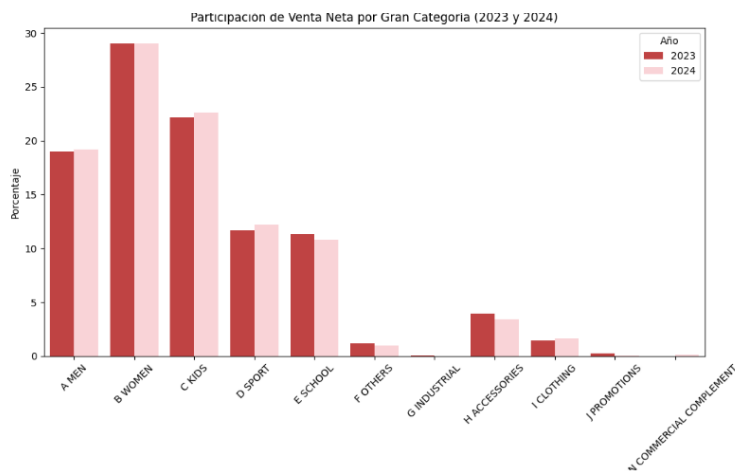
- EdadSemanas: corresponde a la edad de cada artículo al cierre de cada semana.
- Nombre marca: corresponde a un atributo, lo que indica la marca al que está asignado cada artículo.
- PCOSTO: costo unitario de los artículos al cierre de cada semana.
- PPUBLICO: es el precio público con IVA de los artículos al cierre de cada semana.
- Procedencia: se muestra el tipo de sourcing de cada artículo, siendo M = Manisol, N = Nacional, I = Importado.
- Vtapar: venta de los artículos en cada semana.
- Vtaneto: venta turnover sin IVA.
- MgenVta: margen nominal de cada artículo.
- VtaCsto: costo de la venta de los artículos.
- Stkprs: stock pares de los artículos al cierre de cada semana.
- StkCosto: valor de costo del stock.
- StkPublicosin IVA: valorización del stock.

La base de datos fue sometida a un proceso riguroso de exploración, revisión y análisis, con el fin de garantizar la ausencia de datos nulos, así como la verificación de la integridad y coherencia de la información para su adecuado procesamiento. Como resultado de esta revisión, no fue necesario aplicar técnicas de imputación, ya que no se identificaron valores atípicos o faltantes atribuibles a errores en la base de datos o a manipulaciones. Además, se observó un incremento en las ventas durante las primeras y últimas semanas de cada año, lo cual se atribuye a un patrón de tendencia característico del negocio, relacionado con la demanda de productos escolares y a las ventas de temporada de navidad respectivamente.

Una vez realizada la revisión exploratoria, se observaron categorías de artículos que no

tienen una alta participación dentro de la venta (Figura 1), por lo tanto, se excluyen del presente trabajo, con el fin de incorporar solo aquellas categorías que le agregan valor a la organización, las cuales son: A MEN (hombre), B WOMEN (dama), C KIDS (niños), D SPORT (deporte), E SCHOOL (escolar).

Figura 1. *Participación venta por gran categoría*



La Figura 2 y la Figura 3 presentan una comparación entre el ‘precio promedio bruto’, correspondiente al precio de venta sin aplicar descuentos, es decir, el valor de lista, y el ‘precio promedio’, que representa el precio real promedio de venta luego de aplicar los descuentos.

Figura 2. *Comparación precio promedio 2023*

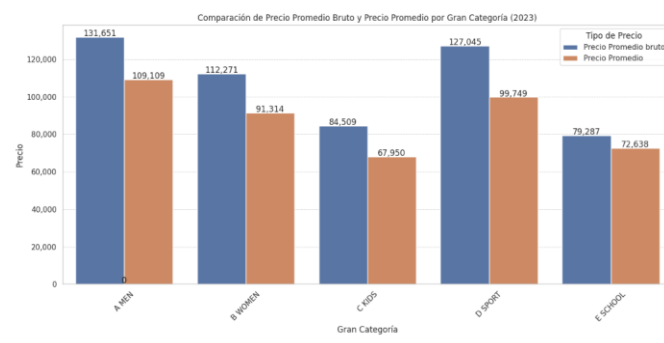
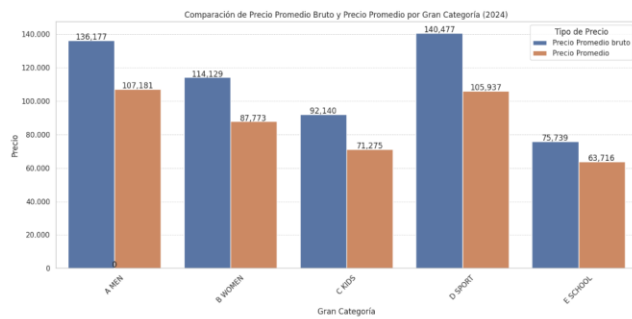


Figura 3. *Comparación precio promedio 2024*



A continuación, se muestra un ejemplo de los datos con los que cuenta el dataset:

Tabla 2. Componentes de la base de datos

Año	Semana	Artículo	NOMBRE GRAN CAT	NOMBRE CATEGORIA	EdadSemanas	NombreMarca	PCOSTO	PPUBLICO	Procedencia	Vtapar	Vtaneto	MgonVta	VtaCsto	Stkprsr	StkCosto	StkPublicosin IVA	VtaBruta	Descuento
0	2023	1	001000JS	C KIDS 21 PREWALKER GIRLS D +	78	BUBBLEGUMMERS	11888	900	M	8	2423	-92449	98072	10	118846	7563	6090	-5627
1	2023	1	001000JA	C KIDS 20 PREWALKER BOYS D +	100	BUBBLEGUMMERS	12497	4900	M	1	757	-11739	12496	0	0	0	4118	-3261
2	2023	1	001000JS	C KIDS 20 PREWALKER BOYS D +	43	BUBBLEGUMMERS	19068	34900	M	1	12822	-2546	18068	7	105478	208294	29228	-16806
3	2023	1	001000KL	C KIDS 21 PREWALKER GIRLS D +	2	BUBBLEGUMMERS	13392	29900	M	0	0	0	0	10	133920	281240	0	0
4	2023	1	001000KW	C KIDS 20 PREWALKER BOYS D +	8	BUBBLEGUMMERS	12888	24900	M	0	0	0	0	33	428208	967818	0	0

5. IDENTIFICACIÓN Y ANÁLISIS DE INDICADORES CLAVE

Una vez se tuvo el acercamiento con el área de merchandising de la compañía, se generaron aquellos indicadores clave (Tabla 3) que son fundamentales para el control del negocio y que aportan significativamente a la realización de estrategias y toma de decisiones, los cuales fueron anexados en el dataset:

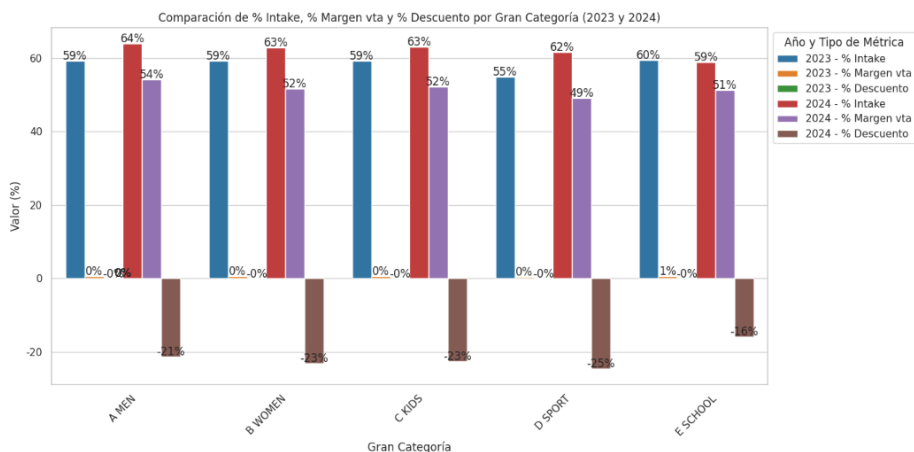
- Precio promedio bruto: corresponde al precio estimado para la venta, sin descuentos.
- Precio promedio: es el precio a lo que realmente fue vendido el artículo, aplicado los descuentos.
- % Margen vta: equivale al margen porcentual obtenido a través del precio de venta real.
- % Descuento: es el % de descuento otorgado; es decir, la diferencia entre el precio promedio bruto vs el precio promedio.
- % Intake: corresponde al % de margen que se pretendía obtener con el precio promedio bruto.

Tabla 3. *Tabla con indicadores importantes*

	AÑO	NOMBRE	GRAN CAT	Precio Promedio bruto	Precio Promedio	% Margen vta
0	2023	A	MEN	131651.0	109109.0	0.5
1	2023	B	WOMEN	112271.0	91314.0	0.5
2	2023	C	KIDS	84509.0	67950.0	0.5
3	2023	D	SPORT	127045.0	99749.0	0.4
4	2023	E	SCHOOL	79287.0	72638.0	0.6
5	2024	A	MEN	136177.0	107181.0	54.2
6	2024	B	WOMEN	114129.0	87773.0	51.7
7	2024	C	KIDS	92140.0	71275.0	52.1
8	2024	D	SPORT	140477.0	105937.0	49.1
9	2024	E	SCHOOL	75739.0	63716.0	51.2
		% Descuento	% Intake			
0		-0.2	59.3			
1		-0.2	59.3			
2		-0.2	59.3			
3		-0.2	54.9			
4		-0.1	59.5			
5		-21.3	63.9			
6		-23.1	62.8			
7		-22.6	63.0			
8		-24.6	61.6			
9		-15.9	58.9			

Se realizaron una serie de gráficas que permitan tener una visión más detallada del Data Set y del negocio:

Figura 4. Métricas importantes para el negocio



Por otro lado, también se consideró la importancia de verificar la correlación entre las diferentes variables del Data Set, permitiendo identificar la alta relación entre algunas de ellas:

Figura 5. Correlación de variables para el año 2023

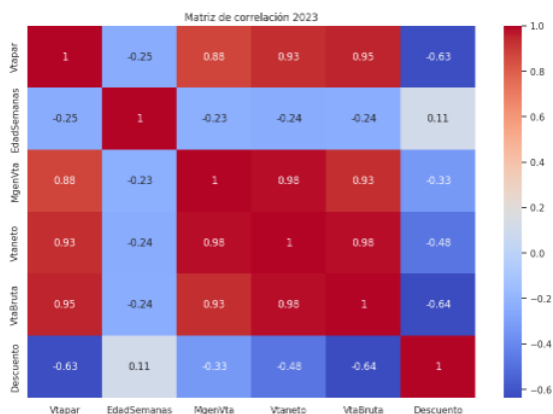
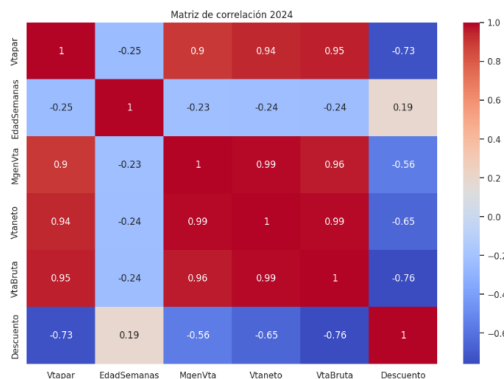


Figura 6. Correlación de variables para el año 2024



De acuerdo con los análisis de correlación presentados en la Figura 5 (año 2023) y la Figura

6 (año 2024), se observa un comportamiento similar en las relaciones entre las variables para ambos años. En particular, se evidencia que un incremento en la venta de los artículos se asocia positivamente con aumentos en el margen de venta, la venta neta y la venta bruta. Asimismo, se identifica una correlación moderada con los descuentos, lo que sugiere que mayores descuentos pueden generar un aumento en las ventas. No obstante, este crecimiento no es necesariamente proporcional, lo que indica que el efecto de los descuentos presenta rendimientos decrecientes. Por tanto, se hace evidente la necesidad de desarrollar estrategias de fijación de precios más inteligentes, que permitan incentivar la demanda sin recurrir a descuentos excesivos o innecesarios.

Teniendo en cuenta las gráficas que se presentan a continuación, se evidencia un comportamiento similar en las ventas a lo largo de los diferentes años, tanto en la venta pares (Figura 7), como en la venta neta (Figura 8) y venta margen (Figura 9). A partir de la observación, se puede inferir que las ventas experimentan incrementos en ciertos períodos.

Figura 7. *Evolución de la venta par (2023 & 2024)*



Figura 8. *Evolución de la venta neta (2023 & 2024)*



Figura 9. *Evolución de la venta margen nominal (2023 & 2024)*



Por otro lado, tanto el porcentaje del margen (Figura 10) como el precio promedio de la venta (Figura 11), no cuentan con un patrón similar para los diferentes años, lo que podría indicar que independiente del precio de venta o la utilidad percibida, esto no influye en el comportamiento de la venta.

Figura 10. *Evolución de la venta margen porcentual (2023 & 2024)*

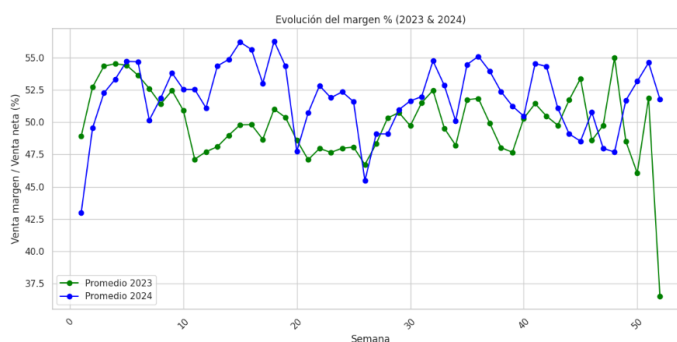
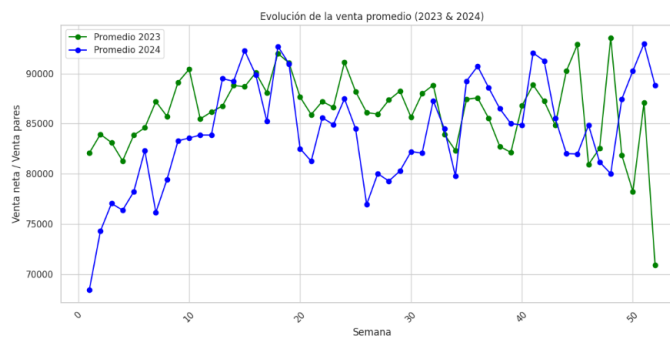


Figura 11. *Evolución del precio promedio neto (2023 & 2024)*



A partir del análisis de las gráficas anteriores, se observa que, si bien las ventas de pares presentan un patrón claro y con tendencias definidas para los años 2023 y 2024, no se identifica un comportamiento consistente en el margen porcentual ni en el precio promedio. Esta falta de uniformidad sugiere que los precios no están siendo asignados de manera estratégica ni alineados

con las dinámicas del mercado, lo cual podría estar afectando la rentabilidad y eficiencia comercial de la empresa.

6. EVALUACIÓN DEL IMPACTO DE LA VARIACIÓN DE PRECIOS SOBRE LA DEMANDA

Al analizar las correlaciones entre las distintas variables correspondientes a los años 2023 y 2024, se identifican aquellas que presentan una alta correlación entre sí, lo que sugiere una correlación significativa en su comportamiento.

- Venta par: corresponden a la cantidad de pares vendidos.
- Venta neta: se refiere a lo venta en dinero sin IVA.
- Margen de venta: es la utilidad bruta (venta neta – costo).
- Descuento: corresponde al descuento realizado (descuento + venta neta = venta bruta).
- Stock pares: cantidad de pares de calzado en inventario.
- Precio promedio: el valor promedio del precio (venta neta / venta par).

El comportamiento de las variables venta pares, venta neta y venta margen durante 2023 y 2024 se muestran de la Figura 12 a la Figura 17, segmentado por las principales categorías. Las gráficas reflejan tendencias marcadas, influenciadas por eventos como la temporada escolar (semana 1 a 8), el día de la madre (semana 18 y 19) y navidad (semanas 48 a 52), entre otros.

Gráficas con el comportamiento del 2023:

Figura 12. Evolución venta pares por gran categoría – 2023

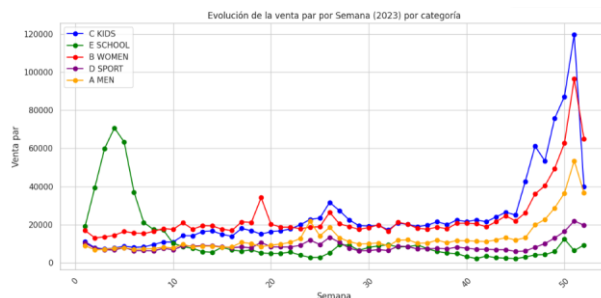


Figura 13. Evolución venta neta por gran categoría - 2023

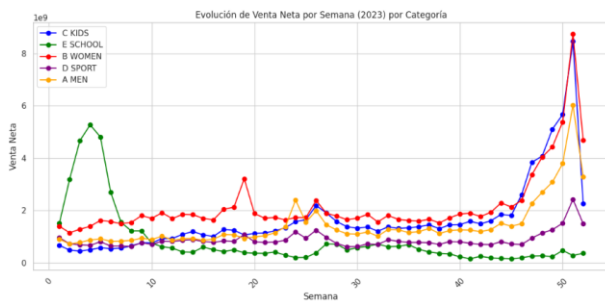
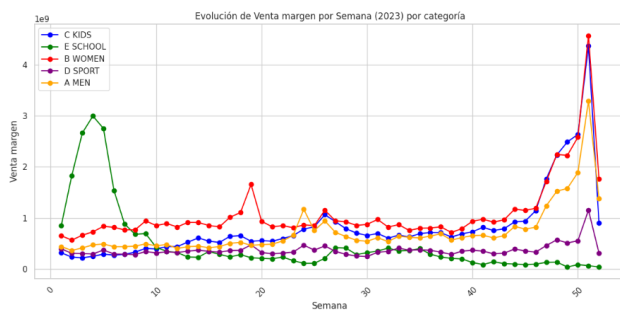


Figura 14. Evolución venta margen nominal por gran categoría - 2023



Gráficas con el comportamiento del 2024:

Figura 15. Evolución venta pares por gran categoría – 2024

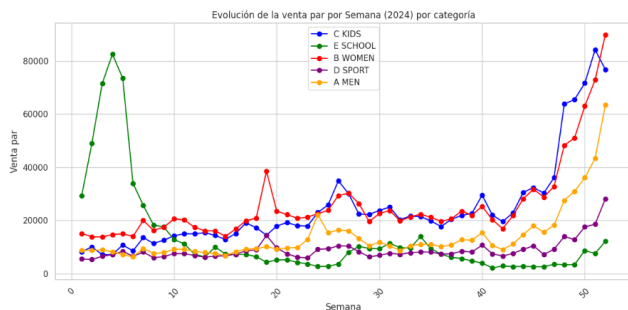


Figura 16. Evolución venta neta por gran categoría - 2024

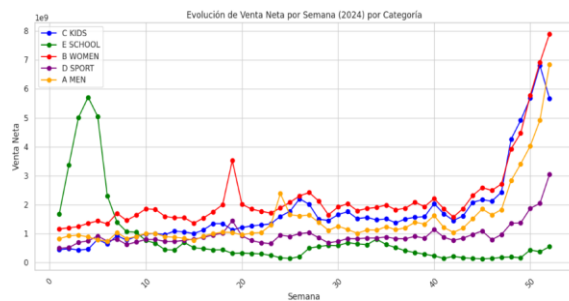
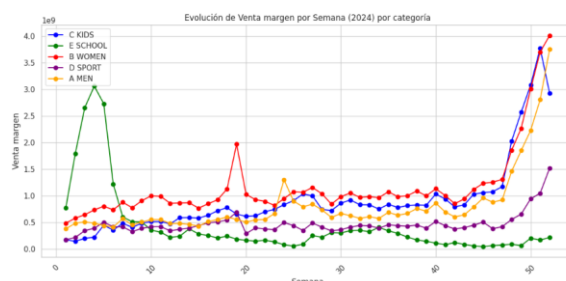


Figura 17. Evolución venta margen nominal por gran categoría - 2024



Para llevar a cabo el presente trabajo, se aplicaron diferentes modelos de aprendizaje automático con el objetivo de analizar la elasticidad de precios en la categoría de calzado. Cada modelo ofrece ventajas particulares que pueden enriquecer el presente objetivo, permitiendo un enfoque de mejor análisis con el fin de evaluar el comportamiento de compra en función del precio.

6.1. Selección de los Modelos de Aprendizaje Automático

6.1.1. Regresión Logística

La regresión logística puede utilizarse en este caso para modelar la probabilidad de un evento binario, como es el caso de la decisión de compra o no compra de un artículo de calzado, en función de variables independientes, tales como el precio y las características del producto. Aunque este método se emplea tradicionalmente para eventos binarios, también puede adaptarse para problemas de clasificación múltiple, lo cual resulta útil al modelar diferentes categorías de resultados. Una de sus principales ventajas frente a otros enfoques estadísticos es la capacidad de incorporar variables explicativas continuas y múltiples factores simultáneamente. Esto permite analizar el efecto combinado de distintas variables sobre la respuesta, considerando posibles interacciones y covarianzas entre ellas, lo que proporciona una comprensión más completa del fenómeno estudiado (Sperandei, 2013). Además, la regresión logística admite la inclusión de variables categóricas, lo que facilita la modelización de atributos como marca y tipo de producto, lo que influye en la decisión de compra. A través de la interpretación de los coeficientes del

modelo, también es posible estimar la elasticidad-precio de la demanda; es decir, el grado de sensibilidad de los consumidores ante variaciones en el precio, lo cual es fundamental para optimizar estrategias comerciales.

6.1.2. *Random Forest*

El Random Forest es un algoritmo de aprendizaje automático basado en un conjunto de árboles de decisión, utilizado para realizar predicciones en diversos contextos, incluyendo la estimación de la elasticidad-precio de la demanda. En este caso, resulta especialmente útil para anticipar la probabilidad de compra o la cantidad demandada en función de variables como el precio y las características del producto. Una de sus fortalezas radica en su capacidad para modelar relaciones complejas y no lineales, capturando patrones de comportamiento que podrían pasar desapercibidos en modelos más simples. Cada árbol dentro del modelo es entrenado sobre una muestra aleatoria generada mediante la técnica de Bootstrap, lo cual permite que una misma observación pueda aparecer en múltiples árboles. Las observaciones que no son utilizadas en el entrenamiento de un árbol, conocidas como out-of-bag, se emplean posteriormente para validar el modelo, lo que incrementa su fiabilidad sin necesidad de un conjunto de prueba adicional (Espinosa-Zúñiga, 2020). La predicción final del modelo se obtiene mediante el ensamblaje de los resultados individuales de todos los árboles, lo cual reduce el sobreajuste y mejora la capacidad de generalización. Además, el modelo permite identificar la importancia relativa de cada variable explicativa, brindando información clave para entender qué factores, como el precio, la categoría o los descuentos, inciden más significativamente en las decisiones de compra.

6.1.3. *XGBoost*

Es un algoritmo avanzado de aprendizaje automático basado en árboles de decisión, ampliamente utilizado para tareas de clasificación y regresión. En el contexto de la elasticidad de

precios, este modelo puede predecir la probabilidad de compra o la cantidad demandada en función de variables como el precio, promociones, características del producto y otros factores relevantes, proporcionando una visión más precisa del comportamiento del consumidor ante fluctuaciones en los precios. Una de sus principales fortalezas radica en su capacidad para modelar relaciones no lineales complejas entre variables, lo que es esencial cuando la respuesta del consumidor no sigue un patrón lineal simple. Además, XGBoost se construye mediante un enfoque secuencial de boosting, lo que significa que ajusta un conjunto de árboles de forma iterativa, corrigiendo errores en cada paso para mejorar progresivamente el desempeño del modelo (Chicaiza Yugcha *et al.*, 2024). Este mecanismo le permite ser altamente eficiente tanto en el tiempo de entrenamiento como en la gestión de grandes volúmenes de datos, lo que lo convierte en una herramienta robusta en escenarios con múltiples atributos de producto.

6.2. Implementación de los Tres Modelos de Aprendizaje Automático

Los tres modelos de aprendizaje automático fueron implementados con el objetivo de evaluar su efectividad en la predicción de elasticidad o inelasticidad de un artículo. Para ello, fue necesario unificar los distintos artículos, consolidándolos en grandes categorías. Esta agrupación permitió generar registros únicos basados en la combinación de año, semana y grandes categorías, facilitando así el análisis a partir de valores agregados. Esta estructura de datos (Tabla 4) contribuye a una mejor generalización de los modelos y a una interpretación más consistente de los patrones de elasticidad en función del comportamiento agregado del consumidor.

Tabla 4. Transformación de la base de datos - unificación de artículos

	Año	Semana	NOBRE GRAN CAT	Vtapar	Vtaneto	Mgenvta	VtaCsto	Stkprs	Stkcosto	StkPublicosin IVA	...	precio_promedio_vta	margen_venta	descuento	intake_stk	intake_vta
1	2023	2	0	6770	716579213	363813035	352766178	158590	8.628397e+09	2.039563e+10	...	105846	50.8	-15.7	57.7	58.5
2	2023	3	0	6981	776681964	413558647	363123317	180711	1.010599e+10	2.330412e+10	...	111256	53.2	-12.7	56.6	59.2
3	2023	4	0	7579	871602506	472168852	399433654	203254	1.141756e+10	2.645981e+10	...	115002	54.2	-11.6	56.8	59.5
4	2023	5	0	7867	914420268	490195458	424224810	213497	1.193729e+10	2.803324e+10	...	116234	53.6	-11.5	57.4	58.9
5	2023	6	0	7018	816745679	436961734	379783945	208918	1.166722e+10	2.739765e+10	...	116378	53.5	-11.0	57.4	58.6

Posteriormente se agregan cinco columnas al dataset (Tabla 5), con las cuales se desea conocer la elasticidad de los precios:

Variacion_precio_promedio_bruto_stk: consiste en identificar la variación porcentual que ha tenido el promedio del stock durante los diferentes períodos, utilizando la siguiente fórmula: $((\text{precio promedio bruto stock período actual} - \text{precio promedio bruto stk período anterior}) / \text{precio promedio bruto stk período anterior})$.

- Variacion_vtapar: $((\text{vtapar período actual} - \text{vtapar período anterior}) / \text{vtapar período anterior})$.
- Variacion_de_las_variaciones: $\text{variacion_vtapar} / \text{variacion_precio_promedio_bruto_stk}$.
- Tipo de elasticidad: se determina el tipo de elasticidad teniendo en cuenta las siguientes condiciones:
 - $(\text{variacion_de_las_variaciones} > 1; \text{"Demanda elástica"})$.
 - $(\text{variacion_de_las_variaciones} < 1; \text{"Demanda inelástica"})$.
 - $(\text{variacion_de_las_variaciones} = 1; \text{"Demanda unitaria"})$.
- Eficiencia: corresponde a una codificación binaria, donde 0 indica demanda elástica y 1 representa demanda inelástica o demanda unitaria.

Tabla 5. Indicadores para cálculo de elasticidad

Variacion_precio_promedio_bruto_stk	Variacion_vtapar	Variacion_de_las_variaciones	Tipo de elasticidad	Eficiencia
-0.45	-0.81	1.80	Demanda elástica	0
-0.12	-18.42	153.50	Demanda elástica	0
1.26	42.44	33.68	Demanda elástica	0
-0.88	-10.51	11.94	Demanda elástica	0
0.01	25.21	2521.00	Demanda elástica	0
...
-3.29	8.45	-2.57	Demanda inelástica	1
-0.10	7.01	-70.10	Demanda inelástica	1
-2.05	10.34	-5.04	Demanda inelástica	1
0.75	0.07	0.09	Demanda inelástica	1
1.10	0.03	0.03	Demanda inelástica	1

Las variables para tener en cuenta para los diferentes modelos son:

Variable objetivo:

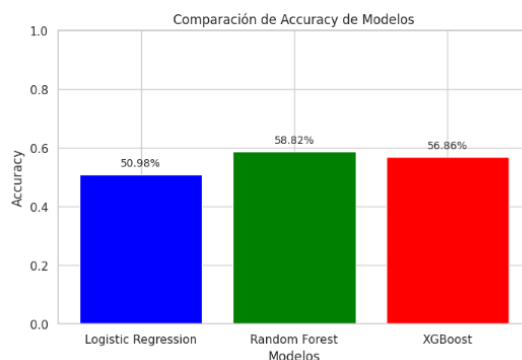
- Eficiencia: corresponde a la identificación de demanda inelástica = 1; demanda elástica = 0.

Variables predictoras:

- Precio Promedio bruto stock: es el valor a lo que la empresa desea vender el artículo (sin iva).
- Stkprs: es la cantidad de artículos en stock con los que se cuentan para la venta.
- % Descuento: el descuento realizado (diferencia entre el precio promedio bruto y precio promedio).
- Vtapar: corresponde a la venta realizada.

Una vez ejecutados los modelos de Regresión Logística, Random Forest y XGBoost, se presentan los resultados obtenidos (de Tabla 6 a Tabla 8). Cada uno de estos modelos se entrenó bajo las mismas condiciones experimentales con el fin de asegurar la comparabilidad de los resultados.

Para el proceso de entrenamiento y validación, el conjunto de datos original fue dividido

Figura 18. Comparativo métricas

Se hace una revisión detallada de cada una de las métricas (Tabla 9); con el cual se puede observar la eficiencia en precisión, recall y F1-score para los tres modelos.

Tabla 9. Comparativo métricas 2

	Model	Precision	Recall	F1-score
0	Logistic Regression	0.45652	0.45652	0.45652
1	Random Forest	0.52273	0.50000	0.51111
2	XGBoost	0.55556	0.54348	0.54945

Al incluir los resultados de accuracy, podemos hacer una comparación más completa entre los modelos:

- Random Forest tiene un accuracy de 58.82%, lo que lo posiciona como el modelo con el mejor desempeño en términos generales para la clasificación correcta de las observaciones. Sin embargo, su precisión (52.27%) y recall (50%) son inferiores en comparación con XGBoost, lo que sugiere que, aunque el Random Forest tiene un accuracy superior, su desempeño es menos equilibrado entre las clases.
- XGBoost presenta un accuracy de 56.86%, lo cual es inferior al Random Forest, pero destaca por su mejor balance entre precisión (55.56%) y recall (54.35%), con un F1-score de 54.95%, que es el más alto entre los tres modelos. Esto indica que XGBoost es más equilibrado, mostrando un rendimiento más confiable en términos de precisión y recall.

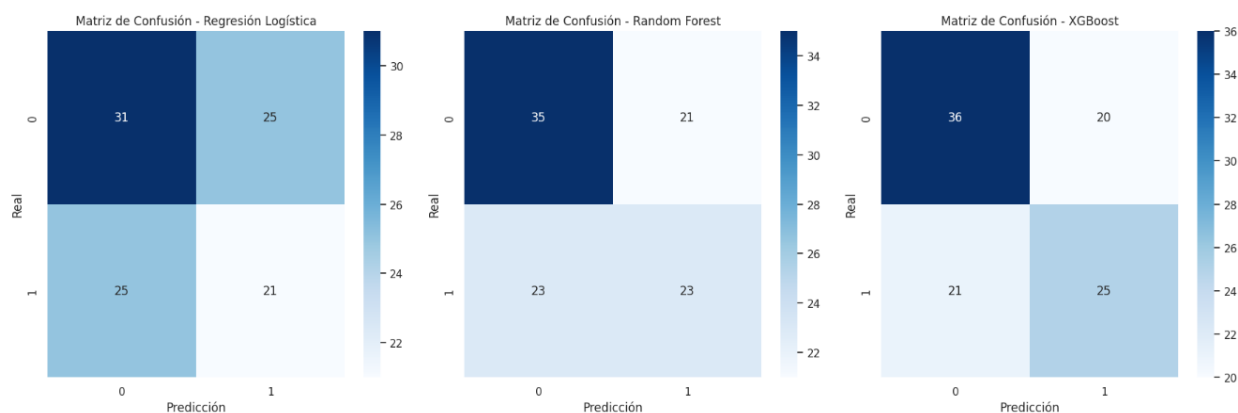
- Regresión Logística tiene el accuracy más bajo, 50.98%, así como su precisión (45.65%). Lo que indica que este modelo está haciendo muchos falsos positivos. Aunque su recall (45.65%) es relativamente alto, su F1-score de 45.65% es el más bajo de los tres, lo que sugiere que el modelo favorece la predicción de la clase positiva apalancado por la métrica precisión.

En resumen, XGBoost es el modelo más equilibrado y confiable, con un buen rendimiento tanto en precisión como en recall, seguido de Random Forest, que tiene un mejor accuracy pero un desempeño menos equilibrado. La Regresión Logística es la menos recomendable debido a su bajo precisión y F1-score, a pesar de su relativamente alto recall.

Matriz de confusión:

Tras analizar métricas como accuracy, precisión, recall y F1-score, se utilizó la matriz de confusión para profundizar en el desempeño de los modelos. Esta herramienta permite identificar errores específicos, como la tendencia a cometer más falsos positivos o negativos, y ofrece una visión más completa sobre cómo cada modelo se comporta frente a los datos. Este análisis es clave para evaluar su capacidad de adaptación y generalización. Teniendo en cuenta los resultados de la Tabla 10, abordaremos los análisis para cada uno de los tres modelos de machine learning:

Tabla 10. *Matriz de confusión (Regresión Logística, Random Forest, XGBoost)*



Regresión Logística:

El modelo identifica 21 verdaderos positivos, aunque hay 25 falsos negativos, lo que sugiere margen de mejora en la detección de casos positivos. También se registran 25 falsos positivos, lo que indica cierta tendencia a clasificar algunos negativos como positivos. Por otro lado, acierta en 31 verdaderos negativos, lo que muestra que el modelo tiene capacidad para distinguir entre clases

Random Forest:

El modelo identifica 23 verdaderos positivos, una cifra ligeramente menor que en la regresión logística, lo que indica que detecta menos casos positivos. Presenta 21 falsos positivos, un valor algo más bajo que en el otro modelo, aunque aún hay margen para reducir este tipo de error. También se registran 23 falsos negativos, lo que sugiere que hay oportunidades para mejorar la detección de positivos. Por otro lado, logra 35 verdaderos negativos, lo que muestra una buena capacidad para reconocer correctamente varios casos negativos.

XGBoost:

Identifica 25 verdaderos positivos, una cifra similar a la de Random Forest y algo menor que la regresión logística, lo que muestra un desempeño aceptable en la detección de casos positivos. Con 20 falsos positivos, presenta el valor más bajo entre los tres modelos, lo que sugiere una mayor precisión al clasificar instancias negativas. Registra 21 falsos negativos, una cifra moderada que indica un buen equilibrio en la identificación de positivos. Además, con 36 verdaderos negativos, alcanza el mejor resultado en esta categoría, lo que refleja una sólida capacidad para reconocer correctamente los casos negativos. A lo largo del trabajo se continuará analizando en mayor profundidad el comportamiento de cada modelo.

El análisis evidencia diferencias claras entre los modelos evaluados en cuanto a su

capacidad para predecir correctamente la eficiencia, entendida como la identificación de demanda elástica (0) o inelástica (1). Utilizando variables claves como el precio promedio bruto del stock, el stock disponible, el porcentaje de descuento y la venta realizada (v_{tapar}), se entrenaron modelos para predecir esta condición. La regresión logística mostró limitaciones importantes, mientras que Random Forest ofreció un desempeño aceptable. XGBoost se perfila como el modelo más preciso y equilibrado en esta comparación preliminar. Para el desarrollo de los modelos, se tiene en cuenta la ‘variación de las variaciones’, de acuerdo con lo informado en el punto anterior, en donde si el resultado es >1 sería “Demanda elástica”; si el resultado es <1 se categoriza como “Demanda inelástica” y si es $=1$ el resultado debe ser “Demanda unitaria”.

Una vez realizado el gráfico Q-Q para las variables venta par y precio promedio bruto stock (Figura 19 y Figura 20), se evidencian algunos datos atípicos; sin embargo, la gráfica continúa con el patrón esperado. Tras una revisión detallada de estos datos atípicos, se concluye que corresponden a información real del comportamiento de las ventas, derivada de las diferentes tendencias comerciales a lo largo del año, como la temporada escolar, el Día de la Madre, Navidad, entre otras. Por lo tanto, estos datos atípicos deben ser conservados, ya que reflejan situaciones reales del mercado que son esenciales para modelar adecuadamente las proyecciones y garantizar la precisión de las predicciones, dado que contribuyen a capturar las fluctuaciones naturales en las ventas.

Figura 19. *Análisis QQ de la variación venta par*

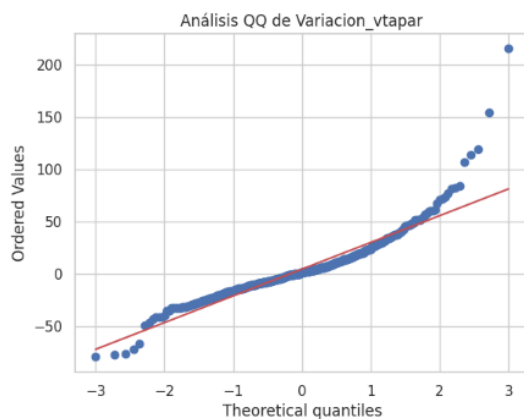
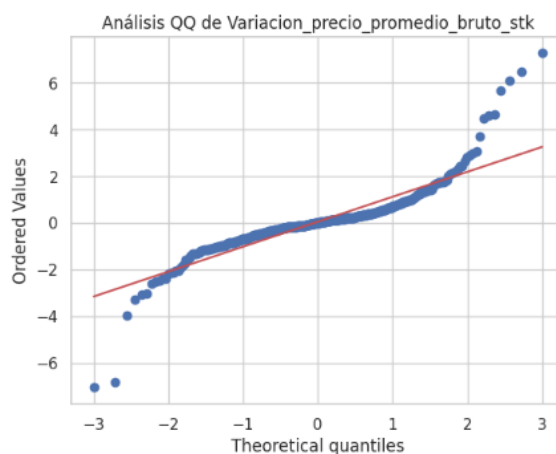


Figura 20. *Análisis QQ de la variación precio promedio*



Tras confirmar que los valores atípicos observados en el análisis Q-Q corresponden a eventos comerciales reales y no a errores de medición, se procedió a evaluar el desempeño de los modelos de clasificación mediante métricas específicas. En este caso, se utilizó la curva ROC (Receiver Operating Characteristic) para analizar la capacidad discriminativa de los modelos binarios empleados.

La regresión logística (Figura 21) presentó un área bajo la curva (AUC) de 0.5, lo que indica un rendimiento equivalente al azar, ya que su curva se aproxima a una línea diagonal sin una clara separación entre las clases.

Por el contrario, los modelos Random Forest (Figura 22) y XGBoost (Figura 23)

obtuvieron un AUC de 0.6, lo que sugiere un rendimiento ligeramente superior al azar en la predicción de la variable 'Eficiencia', en donde 1= demanda inelástica y 0 = demanda elástica. Aunque sus curvas ROC no se alejan significativamente de la diagonal, muestran una desviación positiva que indica una capacidad de discriminación modesta pero superior a la regresión logística.

En conclusión, si bien ninguno de los modelos alcanza un nivel de desempeño alto según la métrica AUC, Random Forest y XGBoost demuestran una capacidad predictiva superior al modelo de regresión logística, sugiriendo que pueden ofrecer una base más sólida para tareas de clasificación en contextos con características similares.

Figura 21. Curva ROC para Regresión Logística

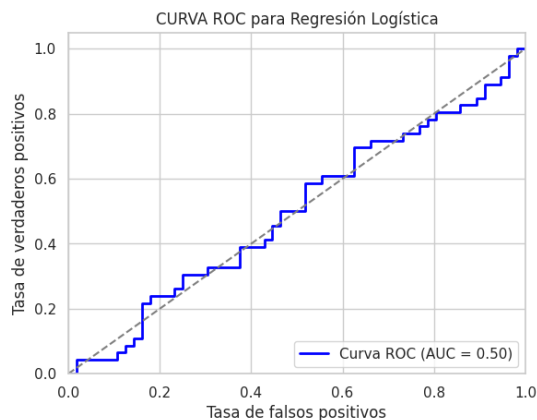


Figura 22. Curva ROC para Random Forest

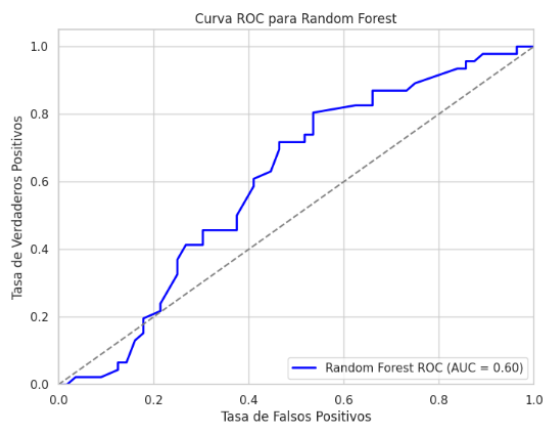
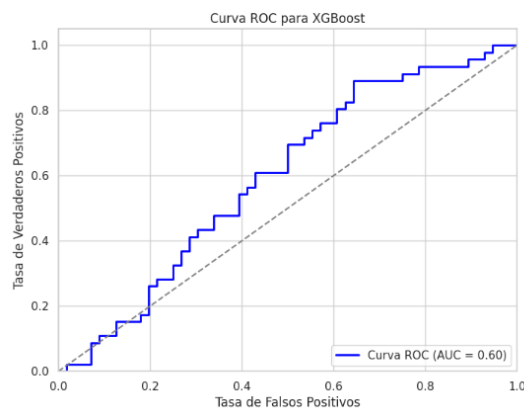


Figura 23. Curva ROC para XGBoost

Dada la dificultad que enfrentan los modelos para realizar predicciones precisas, es relevante destacar que en la organización objeto de estudio, los precios no se determinan según la ley de oferta y demanda; en cambio, la fijación de precios se basa principalmente en datos históricos, márgenes de utilidad mínimos establecidos por la organización y descuentos aplicados según las necesidades comerciales, sin el soporte de un análisis estadístico exhaustivo. Esta falta de un enfoque analítico riguroso limita la capacidad predictiva de los modelos y resalta la necesidad de una evaluación más detallada. Por esta razón, en el presente trabajo se ha optado por enfocar el análisis en la categoría ESCOLAR, por lo tanto, se agrupan los artículos de esta categoría (Tabla 11), con el objetivo de estudiar más a fondo su proceso de asignación de precios. Esta elección se basa en los siguientes motivos:

- Es una categoría que concentra una alta proporción de productos fabricados internamente por la organización y presenta una menor participación de productos importados en comparación con el resto de las categorías.
- Posee incremento de venta a comienzo de cada año debido a la temporada escolar.
- Al ser la mayoría de los productos fabricados dentro de la empresa, es importante tener una planeación exhaustiva que permita garantizar que los artículos que se produzcan tengan un

porcentaje de venta asegurada, permitiendo mantener un adecuado flujo de caja en la organización.

Se llevan a cabo los procesos de Regresión Logística, Random Forest y XGBoost para agrupar los artículos de la categoría escolar. Esto se realiza a través una única base de datos que combina el año, la semana y el total de artículos escolares, como se muestra a continuación:

Tabla 11. Visualización de la base de datos con agrupamiento en school

Año	Semana	NOMBRE	GRAN	CAT	Vtapar	Vtano	ngemta	Vtcosto	stkprs	stkcosto	stkpublicación	IWA	...	precio_promedio_vta	margen_venta	descuento	intake_stk	intake_vta	variacion_promedio_bruto_stk	variacion_vtapar	variacion_de_las_variaciones	Tipo de elasticidad	eficiencia
2023	2		4	39419	3187961024	1830471883	1357489141	450559	1.482416e+10	3.642463e+10	...	80873	57.4	-4.2	59.3	59.2	-1.01	106.87	-105.81	Demanda inelastica	1		
2023	3		4	89937	4664942857	2666281924	1998660933	417875	1.358861e+10	3.343999e+10	...	77830	57.2	-6.4	59.4	59.5	-1.01	82.05	-81.53	Demanda inelastica	1		
2023	4		4	70693	8272344696	2997302274	2275042422	356964	1.166329e+10	2.861356e+10	...	74580	56.8	-6.8	59.2	59.8	0.17	17.95	105.59	Demanda elastica	0		
2023	5		4	63266	4809345018	2754091469	2055233549	311910	1.008471e+10	2.478228e+10	...	76017	57.3	-5.5	59.3	59.6	-0.88	-10.51	11.94	Demanda elastica	0		
2023	6		4	37008	2693832817	1536266814	1157566003	285898	9.224820e+09	2.268651e+10	...	72790	57.0	-6.2	59.3	59.7	-0.13	-41.50	319.23	Demanda elastica	0		

Con base en el enfoque descrito y una vez delimitada la categoría ESCOLAR como objeto de análisis, se procedió a aplicar los modelos de clasificación: Regresión Logística (Tabla 12), Random Forest (Tabla 13) y XGBoost (Tabla 14), utilizando una base de datos unificada que integra información relevante de los productos. Resultados de Regresión Logística:

Tabla 12. Resultados Regresión Logística total categoría escolar

```
array([1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0])
0.47619047619047616
```

Resultado Random Forest:

Tabla 13. Resultados Random Forest total categoría escolar

Accuracy del Random Forest: 0.5238095238095238

Resultado XGBoost:

Tabla 14. Resultados XGBoost total categoría escolar

Accuracy del XGBoost: 0.7142857142857143

Resumen de las métricas:

En general, los tres modelos muestran desempeños distintos al evaluar las métricas (Tabla 15), empezando por la Regresión Logística, el cual tiene un resultado de accuracy de 0.48, lo que indica que acierta en menos de la mitad de los casos, mostrando dificultades para generalizar

correctamente. Su precisión es de 0.55 y el recall de 0.50, lo que significa que identifica algunos casos positivos con cierta precisión. El F1-score de 0.52 refleja un desempeño intermedio entre precisión y recall, y el ROC AUC de 0.47 sugiere que apenas logra distinguir entre clases.

El Random Forest mejora levemente en todas las métricas, con un accuracy de 0.52, una precisión de 0.60, y un recall de 0.50. Esto muestra que es un poco más confiable que la Regresión Logística, especialmente en cómo identifica los casos positivos, aunque sigue siendo limitado. Su F1-score de 0.55 indica un equilibrio algo mejor, y el ROC AUC de 0.53 muestra una ligera mejora en la capacidad para discriminar entre clases.

Por último, XGBoost se destaca entre los tres. Con un accuracy de 0.71, es el modelo que más acierta. Además, alcanza una precisión y un recall, ambos de 0.75, lo que significa que no solo predice bien los positivos, sino que también captura la mayoría de ellos. Su F1-score de 0.75 confirma este buen equilibrio, y el ROC AUC de 0.71 muestra una buena capacidad para diferenciar entre clases.

Tabla 15. Comparación métricas entre Regresión Logística, Random Forest y XGBoost

	Regresión Logística	Random Forest	XGBoost
Accuracy	0.476190	0.523810	0.714286
Precision	0.545455	0.600000	0.750000
Recall	0.500000	0.500000	0.750000
F1-Score	0.521739	0.545455	0.750000
ROC AUC	0.472222	0.527778	0.708333

En resumen, aunque todos los modelos presentan oportunidades de mejora, XGBoost se destaca como el más efectivo en este caso, superando a los demás en todas las métricas evaluadas. Su rendimiento es especialmente relevante cuando se busca identificar correctamente los casos positivos, lo que lo convierte en una herramienta valiosa para el análisis de elasticidad. Adicionalmente, se llevó a cabo un análisis Q-Q quantile-quantile (Figura 24 y Figura 25) con el fin de evaluar el comportamiento de las variables venta par y precio promedio bruto stock;

permitiendo verificar la normalidad de las distribuciones y la idoneidad de los modelos aplicados.

Figura 24. *Análisis QQ de la variación venta par escolar*

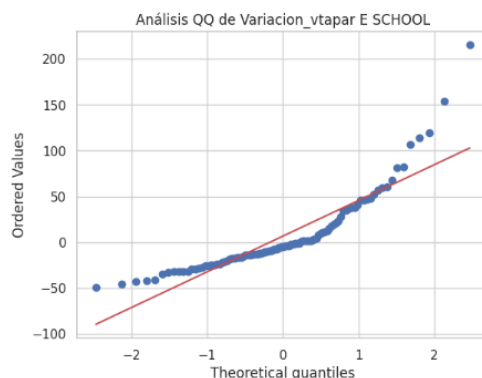
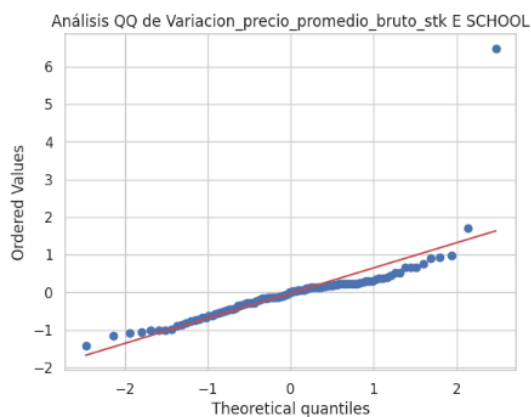
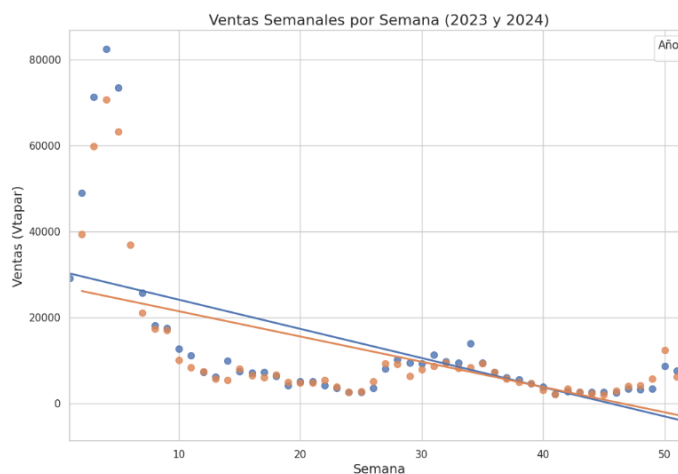


Figura 25. *Análisis QQ de la variación precio promedio escolar*



6.3. Evolución de Métricas Clave de Desempeño Comercial en la Categoría Escolar

Debido a que los resultados de Regresión Logística, Random Forest y XGBoost están por debajo de lo esperado, se decide evaluar si la serie de tiempos es estacionaria con el fin de incorporar posibles modelos de series de tiempo. Con la Figura 26 se muestra el comportamiento de la venta pares en función de las semanas, tanto para 2023 como 2024, evidenciando que, para ambos años, se presentan un incremento en las ventas en los meses de enero y febrero, esto debido a la temporada escolar.

Figura 26. *Evolución venta pares escolar (2023 & 2024)*

Así como se evidencia en los gráficos presentados a continuación, no se observa una relación consistente entre las ventas y variables como el precio promedio bruto del stock, el precio promedio neto y los descuentos aplicados. En la Figura 27, se aprecia que no existe una tendencia clara ni sostenida en la asignación de precios, lo que indica una posible falta de alineación con el comportamiento de la demanda. Este mismo patrón se repite en la Figura 28, donde los descuentos, calculados como la diferencia entre el precio promedio del stock y la venta neta, presentan un comportamiento atípico a partir de finales de 2023, sugiriendo que se han aplicado sin considerar criterios estadísticos sólidos. Finalmente, la Figura 29 muestra una estabilidad en el precio promedio bruto del stock a lo largo de ambos años, a pesar de las variaciones en las ventas pares, las cuales se comportan con parámetros similares para los dos años.

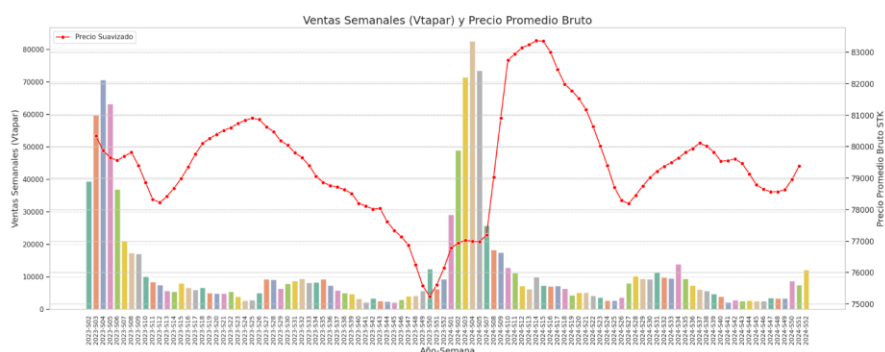
Figura 27. *Evolución comparativa entre venta pares y precio promedio bruto del stock*

Figura 28. Evolución comparativa entre venta pares y porcentaje descuento

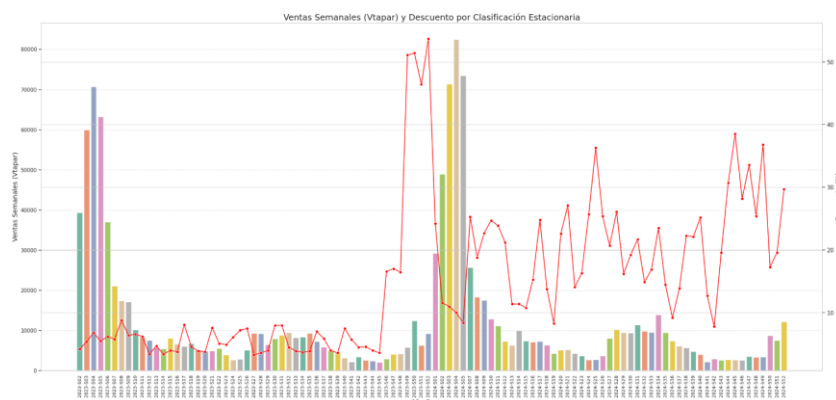
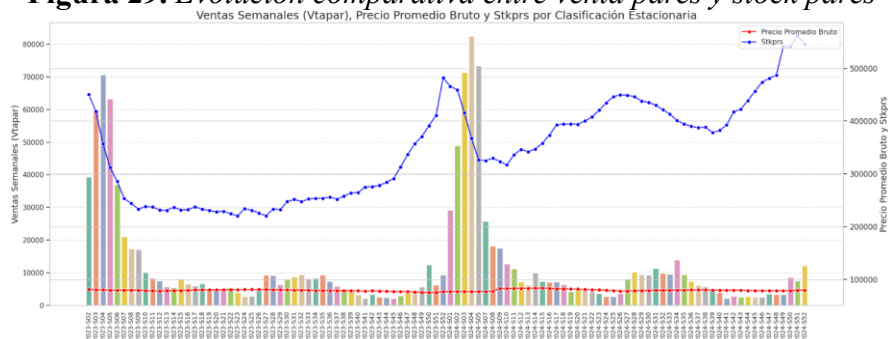


Figura 29. Evolución comparativa entre venta pares y stock pares



6.4. Aplicación de Pruebas ADF y KPSS

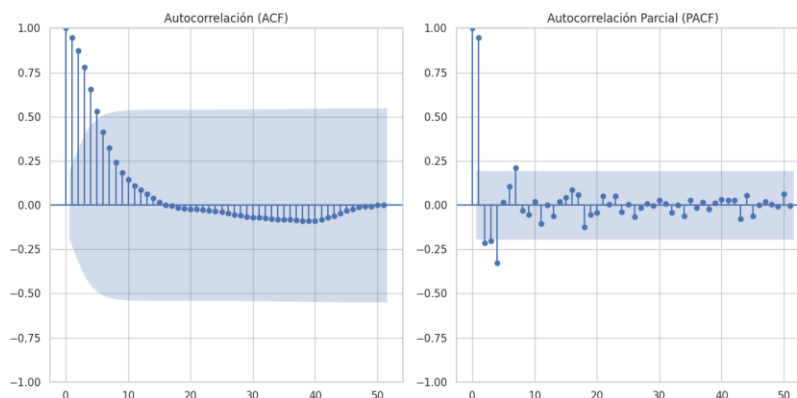
Para garantizar la validez de los modelos de series temporales aplicados en este estudio, es fundamental determinar si los datos presentan un comportamiento estacionario; es decir, si sus propiedades estadísticas (media, varianza y autocorrelación) se mantienen constantes a lo largo del tiempo. La estacionariedad es una condición clave para muchos métodos de modelado y predicción. Con el objetivo de evaluar esta característica, a continuación, se aplican las pruebas de raíz unitaria ADF (Augmented Dickey-Fuller) y KPSS (Kwiatkowski-Phillips-Schmidt-Shin) (Tabla 16), las cuales permiten diagnosticar de forma complementaria la presencia o ausencia de estacionariedad en la serie de tiempo analizada.

Tabla 16. Prueba ADF y KPSS

Prueba Dickey-Fuller Aumentada (ADF):
 Estadístico ADF: -2.2491649754869396
 Valor p: 0.18888745093347215
 Número de rezagos: 1
 Cantidad de observaciones usadas: 100
 Valores críticos:
 1%: -3.497501033
 5%: -2.89090644
 10%: -2.5824349
 Conclusión: La serie NO es estacionaria según la prueba ADF.

Prueba KPSS:
 Estadístico KPSS: 0.6957213747217902
 Valor p: 0.013934420479837254
 Valores críticos:
 10%: 0.347
 5%: 0.463
 2.5%: 0.574
 1%: 0.739
 Conclusión: La serie NO es estacionaria según la prueba KPSS.

Además, se analizaron las funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF), con el fin de explorar la estructura interna de la serie y determinar la presencia de dependencias temporales (Figura 30). Estas herramientas gráficas complementan las pruebas de estacionariedad, al permitir identificar patrones de comportamiento en los rezagos y orientar la selección adecuada de parámetros para los modelos temporales.

Figura 30. Autocorrelación (ACF) & Autocorrelación parcial (PACF)

Ambas pruebas coinciden en que la serie no es estacionaria:

- ADF no rechaza la presencia de una raíz unitaria (lo que sugiere no estacionariedad).
- KPSS rechaza la estacionariedad (confirmando que la serie no es estacionaria).

El análisis inicial de estacionariedad mediante las pruebas Dickey-Fuller Aumentada

(ADF) y KPSS indicó que la serie temporal original no era estacionaria, lo que sugiere la presencia de tendencia o cambios en la varianza a lo largo del tiempo. Esta falta de estacionariedad puede afectar negativamente la estabilidad y precisión de modelos como Arima, Sarima y Arimax, ya que requieren series estacionarias para hacer predicciones confiables. En estos casos, es necesario transformar la serie para estabilizar su media y varianza antes del modelado. Para corregir este problema, se exploró la posibilidad de aplicar diferenciación (restar cada valor con el anterior) para eliminar la tendencia; sin embargo, debido a la cantidad limitada de observaciones disponibles, esta estrategia no fue considerada viable, en su lugar, se optó por modelar explícitamente la tendencia como una variable exógena dentro del modelo Arimax, lo que permitió capturar la estructura temporal sin necesidad de diferenciar la serie. Adicionalmente, se aplicó una transformación logarítmica para estabilizar la varianza. Finalmente, tras estos ajustes, se volvió a aplicar la prueba ADF, la cual confirmó que la serie transformada ahora se podía considerar como estacionaria.

Después de realizar la prueba ADF con la inclusión de una constante y una tendencia lineal ($\text{regression}='ct'$), el modelo considera que la serie puede presentar una tendencia determinista; es decir, un patrón o comportamiento que es predecible y sigue una regla o fórmula específica, sin incertidumbre ni variabilidad aleatoria. El resultado arrojó un estadístico ADF de -3.914 y un p-valor de 0.012 (Tabla 17), lo cual permite rechazar la hipótesis nula de presencia de raíz unitaria al nivel del 5%. Esto indica que, bajo este modelo, la serie de tiempo es estacionaria alrededor de una tendencia lineal; es decir, sus fluctuaciones ocurren en torno a una tendencia estable a lo largo del tiempo.

Tabla 17. Prueba ADF con tendencia

Estadístico ADF con tendencia: -3.914, p-valor: 0.012

La diferencia con las pruebas iniciales podría deberse a que, al incorporar una tendencia al análisis (mediante la opción `regression='ct'` en la prueba ADF), se logró capturar mejor el comportamiento determinista de la serie de tiempo. Esto permitió identificar con mayor precisión su naturaleza estacionaria, algo que no se había conseguido al omitir dicha tendencia.

Con el objetivo de evaluar las ventajas y desventajas de distintos modelos de series temporales, como Promedio Móvil, Holt-Winters, Arima, Sarima, Arimax y Prophet, se presenta una descripción de cada uno de ellos:

6.5. Selección de Modelo de Series de Tiempo**6.5.1. Promedio Móvil**

El modelo de Promedio Móvil es uno de los enfoques más simples y utilizados en el análisis de series temporales para suavizar fluctuaciones y resaltar tendencias a largo plazo. Este modelo se basa en calcular el promedio de un número fijo de observaciones anteriores para predecir el valor de la serie en el siguiente punto temporal. Tal como se describe:

El método de los promedios móviles utiliza el promedio de los k valores de datos más recientes en la serie de tiempo como el pronóstico para el siguiente período. El término "móvil" indica que, a medida que se dispone de una nueva observación, esta reemplaza a la más antigua en el cálculo, generando así un nuevo promedio. (Villarreal, 2016, p. 18)

Esto implica que el promedio se ajusta continuamente, lo que permite al modelo adaptarse a cambios recientes en los datos. Si bien es efectivo para capturar patrones estacionales y eliminar ruido, el modelo de Promedio Móvil tiene limitaciones, ya que no es adecuado para datos con tendencias fuertes o comportamientos no lineales, y su capacidad predictiva disminuye a medida que la complejidad de la serie aumenta.

6.5.2. Holt-Winters

El modelo Holt-Winters es una extensión del suavizamiento exponencial que permite capturar de forma más precisa series temporales que presentan tanto tendencia como estacionalidad. Este enfoque combina tres componentes: el nivel, la tendencia y la estacionalidad. Cabe señalar que, cuando la estacionalidad es constante e independiente del nivel, es factible utilizar el método aditivo de Holt-Winters. No obstante, cuando la estacionalidad crece de manera proporcional al nivel, se recomienda emplear el método multiplicativo de Holt-Winters (MMHW), el cual es la base del presente estudio, ya que permite pronosticar series de tiempo con este tipo de comportamiento (Banda Ortiz y Garza Morales, 2014). Holt-Winters ajusta sus componentes mediante coeficientes de suavizamiento, lo que lo convierte en una herramienta útil para pronósticos en contextos cíclicos, aunque su desempeño puede verse afectado por series muy volátiles.

6.5.3. Arima

El modelo Arima (AutoRegressive Integrated Moving Average) es uno de los enfoques más utilizados para el análisis y pronóstico de series temporales univariantes. Combina tres componentes clave: autorregresivo (AR), que se basa en los valores pasados de la serie; integrado (I), que aplica diferencias sucesivas para eliminar la tendencia y lograr la estacionariedad; y promedio móvil (MA), que incorpora los errores residuales de predicciones anteriores. La notación Arima(p, d, q) refleja la cantidad de términos de cada componente, y su correcta especificación requiere el análisis de autocorrelación y pruebas de raíz unitaria. Como señalan los autores del material de Econometría II, una vez lograda la estacionariedad mediante diferenciación, las predicciones se realizan sobre los valores originales de la serie, ya que son estos los que tienen sentido práctico para la toma de decisiones. Por lo tanto, es importante considerar esta

transformación al interpretar los resultados, especialmente en lo que respecta a los subíndices y las estructuras de las fórmulas utilizadas (De Arce y Mahía, 2001). Este enfoque ofrece gran flexibilidad para modelar patrones temporales complejos, aunque requiere un proceso riguroso de diagnóstico y validación.

6.5.4. *Sarima*

El modelo Sarima es una extensión del modelo Arima que se utiliza cuando los datos muestran patrones estacionales, es decir, cuando los valores de la serie siguen un ciclo que se repite con regularidad (como las ventas de un producto cada temporada). Sarima toma en cuenta tanto las fluctuaciones normales de la serie como esos patrones estacionales. En lugar de solo trabajar con los datos actuales y pasados, Sarima también se enfoca en cómo las observaciones se relacionan con los valores de un período anterior, dependiendo de la estacionalidad de los datos. El modelo tiene dos partes principales: una para manejar las fluctuaciones generales de los datos y otra para captar los patrones estacionales. Esto hace que SARIMA sea muy útil cuando hay ciclos repetitivos en los datos a lo largo del tiempo, como en las ventas o en la demanda de productos. Así, este modelo permite hacer predicciones más precisas, teniendo en cuenta las variaciones estacionales (Suescún Díaz, 2020).

6.5.5. *Arimax*

El modelo Arimax es una variante del modelo Arima que incorpora variables exógenas para mejorar las predicciones. Este modelo consta de tres componentes principales:

La primera es auto regresiva que relaciona la serie con observaciones de períodos anteriores de ella misma. La segunda parte son medias móviles (MA) que relaciona la serie como una función de una sucesión de errores correspondientes a períodos anteriores ponderados. La tercera parte (X) son las observaciones anteriores de una serie exógena. (Avellaneda *et al.*, 2020, p. 28)

Esta estructura le permite al modelo ARIMAX capturar tanto la dinámica interna de la

serie de tiempo como el impacto de factores externos que pueden influir en su comportamiento. Al integrar estas variables exógenas, se logra una representación más completa del fenómeno analizado, lo que mejora la precisión de los pronósticos en contextos donde la serie está afectada por eventos o indicadores externos. Esto lo convierte en una alternativa útil en escenarios comerciales, donde el entorno influye directamente sobre los resultados observados.

6.4.6. Prophet

Prophet es un modelo de series temporales desarrollado por Facebook que se utiliza principalmente para predecir datos con patrones estacionales y tendencias no lineales. Este modelo es especialmente útil para manejar “sets de datos en donde haya observaciones nulas, faltantes, grandes outliers, cambios de tendencia importantes” (López Sáez, 2020, p. 29). De este modo, Prophet destaca por su capacidad para abordar una amplia variedad de desafíos en las series temporales; su flexibilidad para adaptarse a situaciones complejas lo convierte en una opción valiosa, especialmente en contextos donde se producen cambios significativos o no lineales, como en el caso de eventos que alteran el comportamiento de una serie a lo largo del tiempo. Debido a su estructura aditiva y su capacidad para modelar estacionalidades y tendencias de manera eficiente, Prophet se posiciona como una herramienta accesible y potente para realizar pronósticos precisos en escenarios variados.

6.6. Comparación de resultados de los modelos de series de tiempos

Con base en la explicación y características de cada uno de los modelos mencionados, a continuación, se presenta un cuadro comparativo (Tabla 18) que resume las principales características y ventajas de los modelos de series temporales evaluados, lo que permitirá visualizar sus diferencias y determinar cuál es el más adecuado para el análisis en este estudio.

Tabla 18. Cuadro conceptual comparativo modelos de series de tiempo

Modelos de series de tiempo	¿Qué es?	¿Requiere estacionariedad?	¿Maneja tendencia?	¿Maneja estacionalidad ?	¿Soporta variables exógenas?	¿Maneja changepoints?
Promedio móvil	Suaviza la serie con el promedio de los últimos N valores	No	No explícitamente	No	No	No
Holt-Winters	Extiende suavizamiento exponencial: con tendencia (Holt) y estacionalidad (Winters)	No	Sí	Sí	No	No
Arima	Modelo clásico que combina autorregresión + promedios móviles + diferenciación	Sí, tras diferenciar	Sí	No	No	No
Sarima	Extiende Arima con estacionalidad explícita	Sí, tras diferenciar	Sí	Sí	No	No
Arimax	Igual que Arima pero incorpora variables externas	Sí, tras diferenciar	Sí	No	Sí	No
Prophet	Modelo activo: tendencia + estacionalidad + changepoints	No	Sí (automática)	Sí (automática)	Sí (limitado)	Sí (automático)

A lo largo del presente trabajo, se han identificado y analizado patrones de venta similares para los años 2023 y 2024, destacándose un incremento notable entre las semanas 1 a la 8, asociado a la temporada escolar. Considerando que los datos corresponden a una serie de tiempo estacionaria (al menos alrededor de una tendencia), se procede a comparar los valores reales con las predicciones generadas por distintos modelos: Promedio Móvil, Holt-Winters, ARIMA, SARIMA, ARIMAX y Prophet (Figura 31), utilizando las variables descuento, stock pares y precio promedio stock; obteniendo los siguientes resultados:

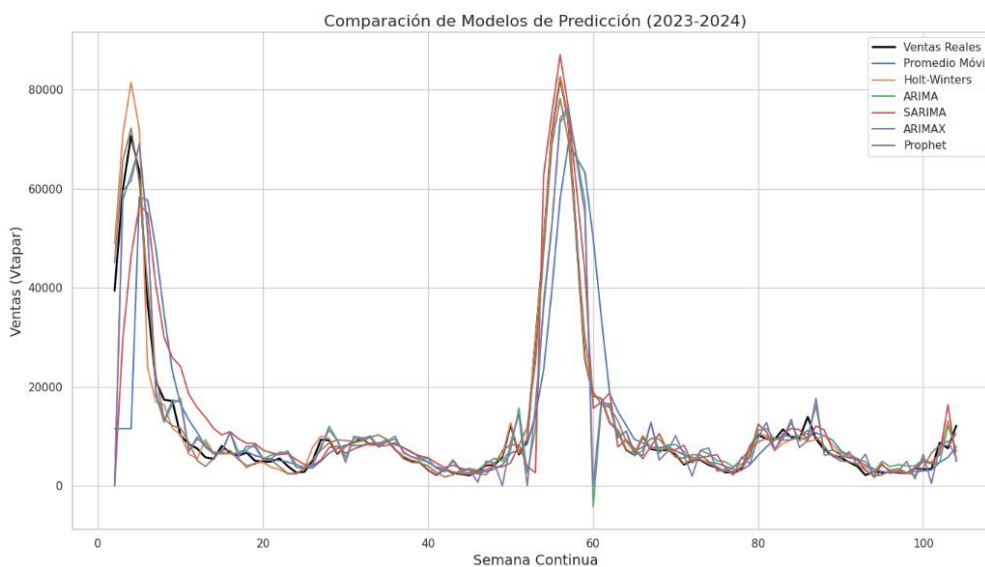
Figura 31. Comparación de los modelos de predicción de series de tiempo

Tabla 19. Comparación de las métricas de los modelos de series de tiempo

	Promedio Móvil	Holt-Winters	ARIMA	SARIMA	ARIMAX	Prophet
MSE	134112447.31	7926742.58	47797107.96	61864521.10	44778293.65	3605236.86
RMSE	11580.69	2815.45	6913.55	7865.40	6691.66	1898.75
MAE	5147.17	1351.24	3555.47	4293.21	3568.13	1389.74
R ²	0.50	0.97	0.82	0.77	0.83	0.99
MAPE	31.15	13.78	35.43	39.20	36.74	17.29

Para todos los modelos de series de tiempo evaluados en este estudio, se utilizó un enfoque consistente de validación, mediante la partición del conjunto de datos en dos segmentos: un 80% destinado al entrenamiento del modelo y un 20% reservado para la prueba. Esta estrategia asegura condiciones equitativas de comparación entre los modelos, permitiendo que cada uno sea ajustado sobre la misma base histórica y evaluado con datos no vistos previamente. Con ello, se busca medir de forma objetiva la capacidad predictiva de cada enfoque y garantizar la solidez de los resultados obtenidos.

Al analizar los resultados de rendimiento de los modelos evaluados (Tabla 19), se destaca el modelo ARIMAX, que alcanza un R² de 83.4% y un MAPE de 36.74%. Esto indica que ARIMAX logra explicar aproximadamente el 83.0% de la variabilidad en las ventas de pares, en función de las variables independientes incluidas en el modelo. Sin embargo, el MAPE sugiere que las predicciones presentan en promedio un error relativo del 36.74% respecto a los valores reales, lo que indica una precisión moderada.

Por otro lado, el modelo Holt-Winters muestra un rendimiento notablemente superior en términos de precisión, con un R² de 97.1% y un MAPE de 13.78%, lo que lo convierte en una

opción muy competitiva para predecir la serie.

Finalmente, el modelo Prophet ofrece un buen desempeño general, con un R^2 de 98.7% y un MAPE de 17.29%, lo que sugiere que logra un excelente ajuste a los datos y mantiene un nivel de error aceptable en las predicciones.

A pesar del buen rendimiento del modelo Holt-Winters, este modelo fue descartado para el análisis final debido a sus limitaciones; ya que requiere que la serie temporal presente una estacionariedad clara, una tendencia regular y que no está influenciada por variables exógenas.

En el caso de la serie temporal analizada, se observan variaciones influenciadas por factores externos como descuentos, precios y niveles de inventario. Debido a esta naturaleza, es necesario emplear un modelo que permita incorporar variables explicativas dentro de su estructura. Modelos como ARIMAX y Prophet ofrecen esta capacidad, lo que los convierte en opciones adecuadas para capturar de forma más precisa el comportamiento de la serie. A partir de un análisis particular y detallado de cada uno, se definirá cuál de estos modelos se ajusta mejor a las características del problema estudiado.

Con el fin de evaluar el desempeño de los dos modelos seleccionados, a continuación, se presentan las pruebas de validez aplicadas a ARIMAX(Tabla 20) y Prophet (Tabla 21):

Tabla 20. Prueba de validez Arimax

Métricas de rendimiento del modelo ARIMAX:
 MSE: 21776287.66
 RMSE: 4666.51
 MAE: 2417.96
 R^2 : 0.92
 MAPE: 19.88%

Tabla 21. Prueba de validez Prophet

Métricas de rendimiento del modelo Prophet:
 MSE: 3685236.86
 RMSE: 1898.75
 MAE: 1389.74
 R^2 : 0.99
 MAPE: 17.29%

Como se observa en los siguientes gráficos, los errores residuales son más altos al inicio de la serie temporal. Esto se debe a la limitada disponibilidad de datos en las primeras

observaciones, que para este caso corresponde a las primeras semanas del 2023, lo que reduce la capacidad del modelo para generar predicciones precisas. A medida que se acumula más información con el paso del tiempo, el modelo logra ajustarse mejor, y las predicciones se vuelven progresivamente más precisas. En el modelo ARIMAX (Figura 32) presenta una mayor acumulación de errores residuales en comparación con Prophet (Figura 33), lo que indica un menor nivel de ajuste del modelo ARIMAX a los datos analizados.

Figura 32. Pronóstico y errores residuales ARIMAX

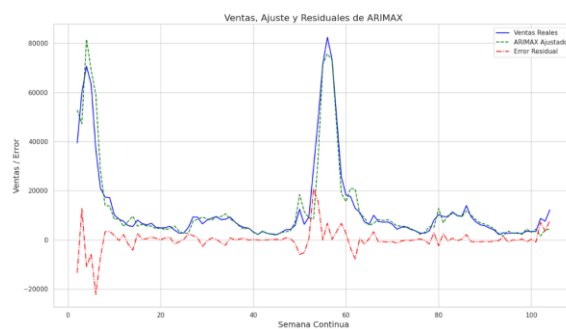
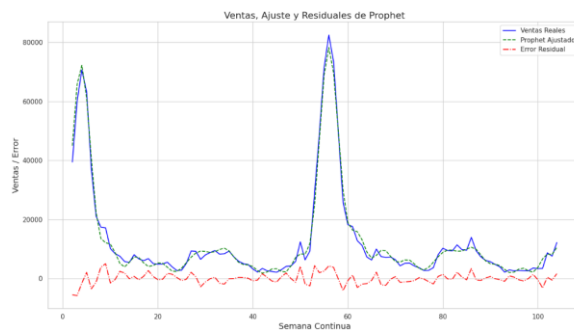


Figura 33. Pronóstico y errores residuales PROPHET



Dado que ambos modelos, ARIMAX y Prophet, presentan resultados favorables en las métricas de evaluación, siendo Prophet superior, se procede a revisar las características específicas de cada uno para comprender mejor sus diferencias y comportamientos.

Ventajas Arimax:

- Ideal cuando las variables independientes tienen un impacto significativo en la serie temporal.
- Al haber variables importantes como precio promedio bruto stock, descuento y stock pares, ARIMAX permite modelar explícitamente estas relaciones y estimar su influencia en la venta de pares.
- Maneja datos estacionarios mejor que Prophet.
- Dado que la serie originalmente no era temporal y fue transformada para su análisis, el modelo ARIMAX podría resultar más adecuado en caso de que se haya requerido diferenciarla para alcanzar la estacionariedad.
- Captura mejor la dependencia en el tiempo.
- ARIMAX asume que la venta de pares depende de sus valores pasados y de las variables independientes.

Limitaciones Arimax:

- No maneja bien las tendencias y estacionalidades si no se especifican correctamente.
- Requiere que la serie sea estacionaria, lo que puede ser un problema si hay cambios estructurales en la tendencia.

Ventajas Prophet:

- Maneja automáticamente tendencias y estacionalidades sin necesidad de hacer diferenciación.
- Al tener una serie temporal de 104 semanas, Prophet puede detectar patrones estacionales semanales o anuales sin necesidad de transformar manualmente los datos.
- Puede incluir variables exógenas (regresores).
- Prophet permite agregar precio, descuento y stock como regresores adicionales, aunque su

impacto en la predicción no es tan explícito como en ARIMAX.

- Maneja bien datos irregulares o con cambios estructurales en la tendencia.
- Si hay rupturas en la tendencia (por ejemplo, un cambio en la estrategia de ventas), Prophet ajusta automáticamente estos cambios.

Limitaciones de Prophet:

- No modela la relación entre la venta de pares y las variables exógenas de manera tan precisa como ARIMAX.
- Puede sobre ajustar si hay demasiado ruido en los datos o si las variables exógenas no tienen una relación clara con la venta de pares.

A ambos modelos se anexaron indicadores económicos (Tabla 22) con el fin de ver su desempeño al incorporar inflación, tasa de interés y TRM. También se creó una nueva variable llamada “Semana Estacionaria” (Tabla 23): Inflación: de acuerdo con lo informado por el Banco de la República.

- Tasa de interés: de acuerdo con lo informado por el Banco de la República.
- TRM: de acuerdo con lo informado por el Banco de la República.
- Semana Estacionaria: de acuerdo con las gráficas de venta pares, se determina agregar como semanas estacionarias de la 1 a la 8.

Tabla 22. Incorporación de variables exógenas

Año,Mes	Categorico,Mes,	TRM	Inflacion,Tasa de interes
0	2023,Enero,1,	4,694.93	*,1.8%,12.8%
1	2023,Febrero,2,	4,751.63	*,1.7%,13.0%
2	2023,Marzo,3,	4,812.61	*,1.1%,13.0%
3	2023,Abril,4,	4,831.80	*,0.8%,13.3%
4	2023,Mayo,5,	4,807.52	*,0.4%,13.3%

Tabla 23. Incorporación de semana estacionaria

	Semana (fecha)	Semana del Año	Semana Estacionaria
51	2024-01-07	1	1
0	2023-01-08	1	1
52	2024-01-14	2	1
1	2023-01-15	2	1
53	2024-01-21	3	1

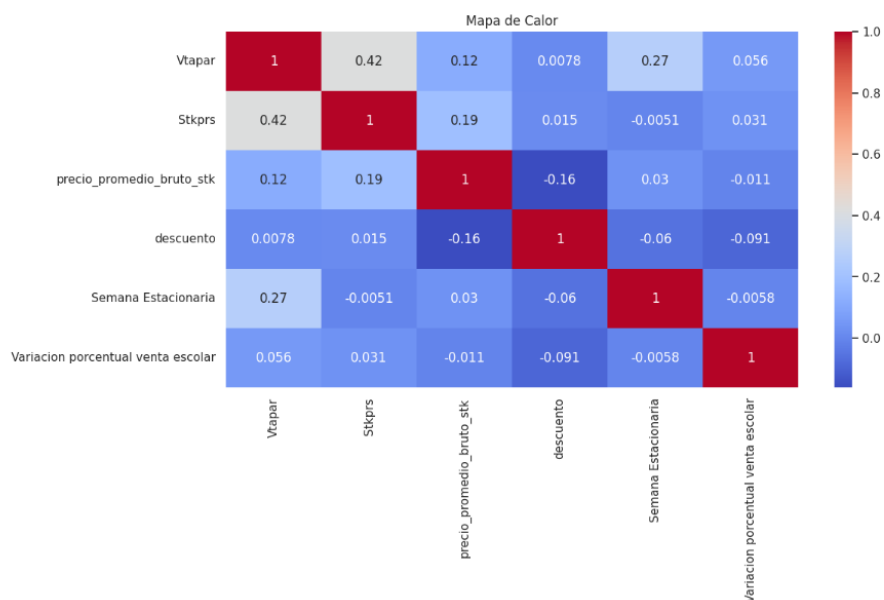
Sin embargo, una vez revisado las nuevas variables incorporadas en los modelos de series de tiempo, arroja que inflación, tasa de interés y TRM no influyeron en mejorar el desempeño de ambos. Por otro lado, “Semana Estacionaria” si les aportó una mejora significativa a los modelos.

7. DESARROLLO DE UN MODELO ESTADÍSTICO PARA LA OPTIMIZACIÓN ESTRATÉGICA DE PRECIOS

7.1. Análisis de Correlaciones de Variables

Una vez incorporada la variable “Semana Estacionaria” al análisis, se realizó una revisión de la correlación entre las variables seleccionadas. Este paso es particularmente importante para el modelo ARIMAX, donde una alta colinealidad entre regresores puede generar problemas de multicolinealidad, afectando la estabilidad y la interpretación de los coeficientes del modelo. Aunque en el caso de Prophet la sensibilidad a la multicolinealidad es menor, también es adecuado evitar regresores redundantes o muy correlacionados, ya que pueden introducir ruido en las predicciones o dificultar la detección de patrones reales.

La matriz de correlación (Tabla 24) muestra que no existen correlaciones fuertes entre las variables utilizadas, lo cual es favorable en la utilización de ambos modelos. Esto sugiere que cada regresor aporta información complementaria, sin redundancias significativas.

Tabla 24. *Correlación de las variables seleccionadas*

Como se observa en la Tabla 24, ninguna de las correlaciones entre regresores supera el umbral crítico de 0.8, por lo que se considera adecuado incluir estas variables en ambos modelos.

En ARIMAX, esto permitirá estimar de forma precisa el efecto de cada variable sobre las ventas, sin riesgo de distorsión por multicolinealidad. En cuanto a Prophet, la incorporación de estos regresores permitirá capturar efectos externos relevantes (como descuentos, precios o comportamiento estacional), complementando su capacidad automática de modelar tendencia y estacionalidad.

A continuación, se hace una explicación más profunda que ayude a tener un mayor entendimiento de las diferentes variables:

- **Stkprs:** representa el inventario disponible en términos de pares. El stock resulta ser un factor clave en las ventas, ya que, si no hay suficiente inventario, las ventas estarán limitadas, independientemente de la demanda. En cuanto a la elasticidad, un inventario muy alto no necesariamente se traduce en mayores ventas si la demanda es baja. Cuando

se cuenta con temporadas de alta demanda, como es el caso de regreso a clases, un inventario adecuado es crucial para maximizar las ventas.

- Precio promedio bruto stk: corresponde al precio promedio de los productos en stock; es decir, la valorización del stock. Si se cuenta con precio promedio bruto alto, puede indicar productos top, lo que podría generar una reducción en las ventas si los clientes buscan alternativas más económicas. En términos generales si se combina un precio bruto alto con grandes descuentos, esto puede generar un incremento en las ventas debido a una percepción de oferta.
- Descuento: es el % de descuento realizado a la venta de los productos (diferencia entre precio promedio bruto del stock y el precio promedio de la venta real). Los descuentos suelen estimular la demanda, especialmente cuando se tiene menor elasticidad de precios. Los grandes descuentos pueden incrementar las ventas a corto plazo; sin embargo, una permanencia de descuentos llega a ocasionar problemas en el margen y en la marca. En cuanto a la segmentación de clientes, los consumidores llegan a ser más sensibles al precio, por lo tanto, tienden a responder a los descuentos.
- Semana Estacionaria: es una variable que indica la tendencia alcista de venta en algunas semanas, es decir, aquellas con alta demanda. La venta de calzado escolar se concentra en semanas específicas, previas al regreso a clases (semana 1 a la 8). Las estrategias de descuento y precios deben ser diferentes en semanas de alto incremento en la venta.
- Variación porcentual venta escolar: ya que no se cuenta con un histórico de los descuentos realizados para las diferentes categorías, se hace necesario agregar un indicador adicional, el cual consiste en la variación de la venta de cada semana a nivel de pares confrontado con la variación de venta del total de las categorías; si ambas tienen un incremento o

disminución en la variación de la venta pares, se asume que fue impactado por una promoción generalizada, como es el caso de: pague 2 lleve 3, segundo par con el 50% de descuento, segundo par con el 30% de descuento, etc. En conversación con el área de merchandising, se recomendó llevar el detalle del histórico de las promociones, lo cual comenzaron a realizar a partir del mes de enero de 2025. Esto será crucial para próximas adaptaciones del modelo.

- Vtapar (dependiente): Es la variable objetivo que representa las ventas en términos de pares vendidos. Esta tiene una relación con las variables anteriores, como Stkprs, ya que define el límite de lo que se puede vender; precio_promedio_bruto_stk el cual influye en la percepción de valor y la competitividad del producto; así mismo descuento, el cual estimula las ventas, especialmente en semanas claves; también semana estacionaria, define cuando se concentran las ventas y cuando son más sensibles a las estrategias de descuentos y precios.

7.2. Características y Resultados de los Modelos Arimax y Prophet

Después de entender en detalle las variables claves que afectan el comportamiento de las ventas en términos de pares (Vtapar), es fundamental evaluar las herramientas analíticas utilizadas para modelar y predecir dicho comportamiento; en este contexto, referente a los dos modelos a utilizar: ARIMAX y Prophet, ambos permiten capturar la dinámica temporal de las ventas, incorporando tanto estacionalidades como variables externas que influyen en la demanda.

A continuación, se muestra un cuadro comparativo (Tabla 25) entre los modelos Arimax y Prophet en cuanto a sus cualidades, características y capacidades:

Tabla 25. Características de modelos Arimax y Prophet

Característica	ARIMAX	Prophet	Serie de tiempo que se está trabajando	Mejor modelo
Relación con variables externas	Lineal	Flexible, puede manejar relaciones no lineales	Tiene tendencia, La serie de tiempos en su conjunto no es considerada estacionaria	Prophet
Patrones	Tendencia y estacionalidad lineales	Estacionalidades múltiples y no lineales	Cuenta con estacionalidad al inicio de cada año.	Arimax, Prophet
Manejo de datos atípicos	Puede verse afectado por valores atípicos	Muy robusto frente a valores atípicos	Cuenta con valores atípicos reales en el stock y precio promedio, esto como resultado del mismo negocio.	Prophet
Facilidad de uso	Requiere más conocimiento técnico para ajustar parámetros	Fácil de usar y con menos ajustes necesarios	N/A	Prophet
Interpretación	Directa (coeficientes de las variables)	Menos directa, pero flexible	N/A	Arimax
Modelado de estacionalidades	Puede manejar estacionalidades sencillas	Excelente para estacionalidades múltiples y complejas	El modelo bajo los test no es estacionario, sin embargo si cuenta con tendencias de ventas e incrementos en las temporadas alta.	Prophet

Al comparar los modelos ARIMAX y Prophet, se observa que Prophet presenta ventajas importantes en contextos con estacionalidades múltiples, valores atípicos y relaciones no lineales entre variables externas, lo cual se alinea mejor con las características de la serie de tiempo analizada. Aunque ARIMAX ofrece una interpretación más directa a través de sus coeficientes y es adecuado para identificar tendencias lineales, su sensibilidad a los valores atípicos y mayor complejidad técnica lo hacen menos práctico frente a Prophet.

Considerando que Prophet es más robusto, se procedió a evaluar su desempeño frente al modelo ARIMAX en escenarios específicos. Para ello, se seleccionaron nueve artículos representativos de la categoría escolar, clasificados en tres grupos según su comportamiento histórico de ventas (tres artículos por cada grupo): “venta alta y constante”, “venta intermedia y menos constante” y “venta baja y no constante”. Esta segmentación permitió aplicar ambos modelos a diferentes contextos comerciales y operativos, pudiendo comparar su capacidad

predictiva, evaluando así cuál de los dos se adapta mejor a cada patrón de venta. Detallando los resultados en la Tabla 26.

Tabla 26. *Comparativo de los resultados de los modelos*

Métricas	ARTÍCULOS CON VENTA ALTA Y CONSTANTE					
	ARTÍCULO 1		ARTÍCULO 2		ARTÍCULO 3	
	Arimax	Prophet	Arimax	Prophet	Arimax	Prophet
MSE	38342.83	3286.03	22539.21	7257.12	10754.64	1970.93
RMSE	195.81	57.32	150.13	85.19	103.70	44.40
MAE	119.21	44.16	91.11	60.96	60.48	27.64
R2	0.90	0.99	0.87	0.96	0.81	0.96
MAPE	43.87%	25.43%	34.79%	29.79%	34.15%	23.59%

Métricas	ARTÍCULOS CON VENTA INTERMEDIA Y MENOS CONSTANTE					
	ARTÍCULO 4		ARTÍCULO 5		ARTÍCULO 6	
	Arimax	Prophet	Arimax	Prophet	Arimax	Prophet
MSE	383.32	118.44	292.00	51.95	645.72	158.27
RMSE	19.58	118.44	17.09	7.21	25.41	12.58
MAE	13.51	8.50	11.49	5.04	17.41	8.69
R2	0.74	0.92	0.85	0.97	0.87	0.97
MAPE	33.60%	22.49%	65.97%	30.96%	61.63%	35.41%

Métricas	ARTÍCULOS CON VENTA BAJA Y NO CONSTANTE					
	ARTÍCULO 7		ARTÍCULO 8		ARTÍCULO 9	
	Arimax	Prophet	Arimax	Prophet	Arimax	Prophet
MSE	52.45	34.04	51.40	18.92	91.33	53.09
RMSE	7.24	5.83	7.17	4.35	9.56	7.29
MAE	5.46	4.20	4.69	3.15	5.98	5.06
R2	0.64	0.76	0.71	0.89	0.56	0.74
MAPE	inf%	inf%	inf%	inf%	inf%	inf%

7.3. Selección e Implementación del Modelo Prophet

Tras la ejecución de los modelos ARIMAX y Prophet, se observa que Prophet presenta un mejor desempeño en términos de precisión y ajuste general. En particular, los tres primeros artículos, que corresponden a los de mayor volumen de ventas muestran resultados significativamente más sólidos bajo el modelo Prophet. Esto sugiere que, para el tipo de productos analizados, especialmente aquellos con “venta alta y constante”, Prophet se adapta mejor al comportamiento de la demanda. Por lo tanto, el modelo Prophet ha sido seleccionado para este trabajo por su mayor capacidad de generar estimaciones precisas y consistentes, especialmente en artículos con venta alta y constante, lo que lo convierte en la opción más adecuada para este estudio. Además, se caracteriza por ser un modelo intuitivo, robusto y automatizable, lo que

refuerza su utilidad en entornos reales de análisis de series temporales.

Si bien el objetivo inicial planteaba el diseño de un algoritmo de optimización de precios basado en la elasticidad, tras el análisis de los datos disponibles se identificó que los cambios de precios históricos no se realizaron bajo condiciones estadísticas controladas, por lo que no es posible calcular una elasticidad precio-demanda robusta ni confiable. En consecuencia, el enfoque se ajustó para mantener la esencia del objetivo, priorizando la proyección de la demanda en función de variables claves como el stock, el precio promedio del stock, el descuento aplicado, la semana de estacionalidad y la variación semanal de la venta escolar. Estas variables fueron incorporadas al modelo de predicción Prophet como regresores exógenos, lo que permitió simular escenarios futuros y analizar el impacto de las condiciones de mercado sobre la demanda estimada.

Con base en este enfoque desarrollado con Prophet, se implementó un modelo de predicción de demanda que estima las ventas semanales considerando las variables claves. Al integrar estos factores, el modelo genera proyecciones para el período 2023–2024 y permite extender las estimaciones para todo el año 2025, incluso con la ausencia de datos operativos futuros. Esta metodología ofrece una herramienta robusta para anticipar la demanda, evaluar el impacto de distintas condiciones comerciales y respaldar decisiones estratégicas de planificación, sin la necesidad de contar con una estimación explícita de la elasticidad precio-demanda. Además, el modelo emplea un enfoque de estacionalidad en modo multiplicativo, lo que significa que las componentes estacionales, tanto anuales como semanales, se modelan como factores que multiplican la tendencia base, en lugar de sumarse a ella. Este tipo de modelo es especialmente útil cuando las variaciones estacionales o los efectos externos cambian de magnitud según el nivel de la serie, como ocurre con las ventas de la serie de tiempos que se está abordando, la cual cuenta con un incremento proporcional más al inicio de cada año debido a la temporada escolar.

Con el propósito de evaluar el desempeño del modelo Prophet, seleccionó como referencia el grupo de artículos catalogados como “venta alta y constante”, los cuales mostraron los mejores resultados en la prueba comparativa previa. Este análisis permite comparar las proyecciones generadas por el modelo con las ventas reales observadas, facilitando así la medición de su precisión. A continuación, se presentan las cifras reales correspondientes a los años 2023 y 2024, junto con las proyecciones del modelo Prophet para su contraste:

Figura 34. Comparación venta real vs modelo: Artículo 1

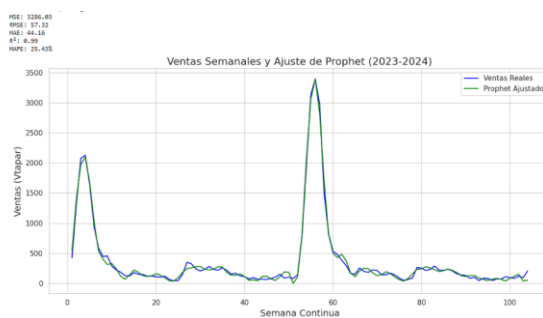


Figura 35. Comparación venta real vs modelo: Artículo 2

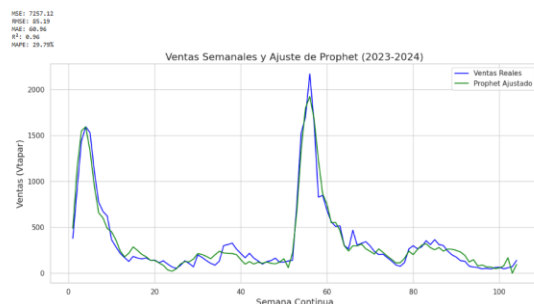


Figura 36. Comparación venta real vs modelo: Artículo 3



Con base en la información histórica de ventas de los años 2023 y 2024, y al ser utilizada

bajo el modelo prophet se pudo proyectar de manera semanal la demanda de pares vendidos para el año 2025 (Figura 37, Figura 38, Figura 39). Este modelo incorporó variables exógenas relevantes como stock disponible, descuentos, precio promedio del inventario, estacionalidad y variación en ventas escolares, permitiendo capturar de forma más precisa los patrones de comportamiento. La implementación técnica validó la integridad de los datos, ajustó el modelo bajo un esquema de estacionalidad multiplicativa y generó predicciones consistentes tanto en el histórico como en el horizonte futuro. La proyección de ventas para el año 2025 evidenció un comportamiento lógico y estructurado, alineado con las tendencias observadas en años anteriores, lo cual valida el uso del modelo como herramienta de apoyo para la planificación comercial y operativa.

Figura 37. *Proyección venta 2025: Artículo 1*

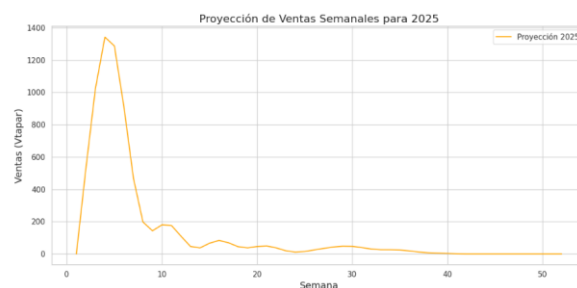
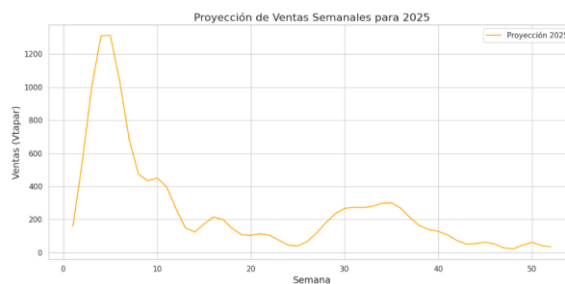


Figura 38. *Proyección venta 2025: Artículo 2*



Figura 39. Proyección venta 2025: Artículo 3



Al analizar los residuales de los tres modelos prophet (Figura 41, Figura 42), se observa que estos fluctúan alrededor del eje cero sin mostrar una tendencia definida, lo cual sugiere que el modelo captura adecuadamente la dinámica de las ventas. Los valores residuales se mantienen en un rango moderado, con picos negativos y positivos; esta dispersión controlada y simétrica indica que los errores de predicción no presentan sesgos importantes, lo que refuerza la validez del modelo en términos de estabilidad y precisión para este caso de estudio.

A nivel nominal, los residuos tienden a incrementarse durante las primeras semanas del año 2024, lo cual se explica por la marcada tendencia de incremento en la venta asociada a la temporada escolar. Para profundizar en el análisis comparativo de los tres modelos estudiados, es crucial examinar el comportamiento de los residuos a lo largo del tiempo, ya que estos reflejan la capacidad de ajuste de cada modelo.

Figura 40. Residuales: Artículo 1

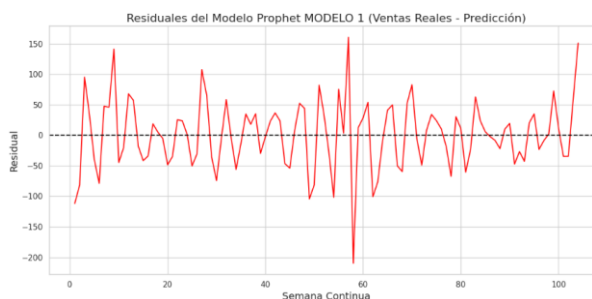


Figura 41. Residuales: Artículo 2

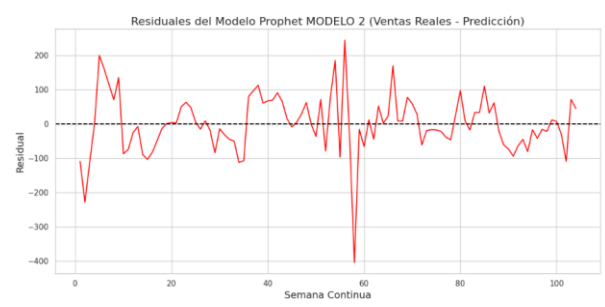
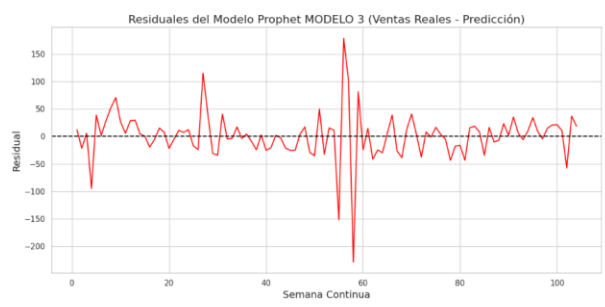


Figura 42. Residuales: Artículo 3



En cuanto a la distribución de los residuos (Figura 43, Figura 44, Figura 45), se observar una forma aproximadamente normal, similar a una campana, lo que sugiere que el modelo ajusta adecuadamente la mayoría de las predicciones, con errores concentrados en torno a cero.

Figura 43. Distribución de los residuos: Artículo 1

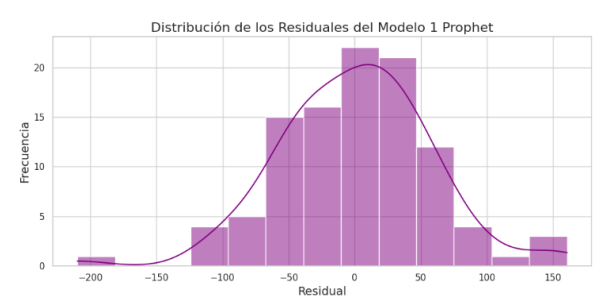


Figura 44. Distribución de los residuos: Artículo 2

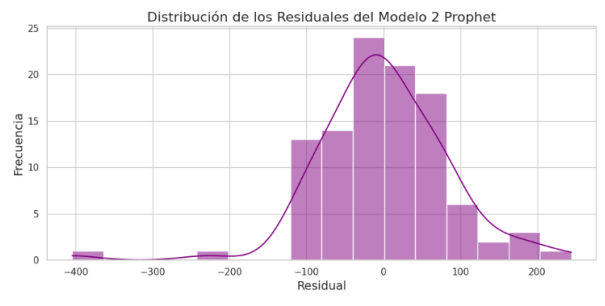
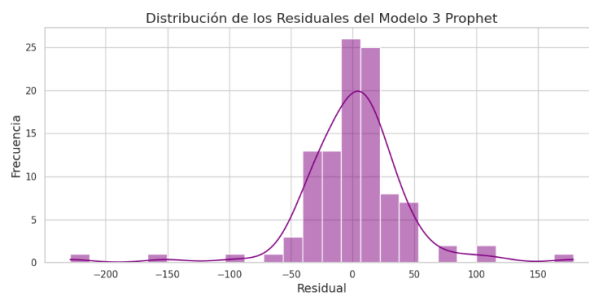


Figura 45. *Distribución de los residuos: Artículo 3*

En comparación con los otros modelos, el Modelo 3 se destaca por su consistencia y precisión, obteniendo los mejores resultados tanto en la gráfica de residuos como en la distribución, lo que refuerza su robustez y capacidad de ajuste a las variaciones de los datos.

Para profundizar en la evaluación de la precisión y la robustez de los modelos aplicados, los tres artículos fueron analizados mediante el cálculo de los residuos nominales; es decir, la diferencia directa entre las ventas reales y las predicciones, así como sus estadísticas descriptivas y las 10 discrepancias más significativas entre los valores observados y los estimados. Este enfoque cuantitativo permite una evaluación más precisa del rendimiento del modelo Prophet en diferentes contextos de demanda, facilitando la identificación de los períodos en los que se presentan mayores desviaciones, lo cual es clave para detectar patrones no capturados o situaciones atípicas en el comportamiento de las ventas.

A partir del análisis general realizado y con el fin de evaluar con mayor detalle el desempeño de cada enfoque aplicado, a continuación, se presenta una revisión individual de los modelos.

Modelo aplicado al artículo 1: logra un rendimiento aceptable, con una mediana de error muy cercana a cero y una media positiva leve, lo cual sugiere que en promedio predice con precisión razonable (ver Tabla 27 y Tabla 28). Además, su rango intercuartílico (IQR) de 69.72 indica una dispersión moderada. Aunque presenta algunos picos con errores altos (como -209 y +160 unidades), estos se concentran en semanas específicas influenciadas por la tendencia alta.

Tabla 27. Estadística de los residuos nominales: Artículo 1

Semana	Año	Vtapar	Prophet_Adjusted	Residuo_Nominal
6	2024	1494	1703.786600	-209.786600
5	2024	2973	2812.413203	160.586797
52	2024	205	53.994143	151.005857
9	2023	455	313.827747	141.172253
1	2023	428	540.003101	-112.003101
27	2023	352	244.492880	107.507120
49	2023	87	191.571823	-104.571823
2	2024	1955	2056.844483	-101.844483
10	2024	384	484.707274	-100.707274
3	2023	2075	1979.817144	95.182856

Tabla 28. Top 10 semanas con mayor diferencia: Artículo 1

```

Estadísticas de los residuos nominales MODELO 1:
count    104.000000
mean      0.746893
std       57.596610
min      -209.786600
25%     -34.879335
50%       2.756640
75%      34.838347
max       160.586797
Name: Residuo_Nominal, dtype: float64

Rango intercuartílico (IQR): 69.72

```

Modelo aplicado al artículo 2: si bien presenta una mayor dispersión de errores y valores extremos más pronunciados (con un mínimo de -404 y un máximo de +244 unidades), destaca por su capacidad de captura de tendencias generales (ver Tabla 29 y Tabla 30). La cercanía de su media y mediana a cero demuestra que el modelo no tiende a sobreestimar ni subestimar sistemáticamente, lo que es una señal positiva. Su IQR más alto (98.62) sugiere que la aplicación de prophet a este artículo resulta ser más sensible a variaciones abruptas.

Tabla 29. Estadística de los residuos nominales: Artículo 2

Semana	Año	Vtapar	Prophet_Adjusted	Residuo_Nominal
6	2024	828	1232.864281	-404.864281
4	2024	2168	1923.751034	244.248966
2	2023	919	1147.209605	-228.209605
5	2023	1529	1329.500712	199.499288
2	2024	1524	1338.718314	185.281686
14	2024	467	297.890410	169.109590
6	2023	1104	943.816909	160.183091
9	2023	621	486.226228	134.773772
7	2023	772	656.982424	115.017576
38	2023	327	214.359564	112.640436

Tabla 30. Top 10 semanas con mayor diferencia: Artículo 2

```

Estadísticas de los residuos nominales MODELO 2:
count    104.000000
mean     -0.104644
std       85.601209
min     -404.864281
25%     -45.020972
50%       0.126154
75%      53.594733
max       244.248966
Name: Residuo_Nominal, dtype: float64

Rango intercuartílico (IQR): 98.62

```

Modelo aplicado al artículo 3: se posiciona como el más equilibrado en términos de estabilidad. A pesar de tener también valores extremos, estos son menos frecuentes y su dispersión es menor (IQR = 37.88). La desviación estándar más baja (44.60) confirma su consistencia (ver (Tabla 31y Tabla 32). Este modelo destaca especialmente en productos con comportamiento continuo y predecible, ya que su capacidad para ajustarse suavemente a la tendencia real se refleja en residuos más controlados y predicciones más cercanas a la realidad.

Tabla 31. Estadística de los residuos nominales: Artículo 3

Semana	Año	Vtapar	Prophet_Adjusted	Residuo_Nominal
6	2024	579	807.725430	-228.725430
4	2024	1529	1350.300570	178.699430
3	2024	1078	1229.497418	-151.497418
27	2023	285	169.804704	115.195296
5	2024	1267	1165.683895	101.316105
4	2023	314	408.695959	-94.695959
7	2024	631	549.522262	81.477738
9	2023	259	188.228504	70.771496
50	2024	41	98.603956	-57.603956
8	2023	250	197.590767	52.409233

Tabla 32. Top 10 semanas con mayor diferencia: Artículo 3

```

Estadísticas de los residuos nominales - MODELO 3:
count    104.000000
mean      1.004657
std       44.598744
min      -228.725430
25%      -21.148054
50%       4.175617
75%      16.733092
max       178.699430
Name: Residuo_Nominal, dtype: float64

Rango intercuartílico (IQR): 37.88

```

Si bien el modelo Prophet aplicado a los tres artículos muestra niveles aceptables de rendimiento, sus resultados varían según los valores de venta y la constancia a lo largo de las semanas. En el primer artículo, el modelo destaca por su precisión promedio y comportamiento estable, siendo útil para escenarios con estacionalidad moderada. En el segundo artículo, los resultados son más dispersos, pero el modelo sigue siendo eficaz en contextos con alta volatilidad, gracias a su sensibilidad a cambios abruptos. Finalmente, en el tercer artículo, el modelo sobresale por su regularidad y bajo nivel de error, lo que lo convierte en una opción confiable para productos con demanda continua y predecible.

Aunque los resultados obtenidos con el modelo Prophet son en general satisfactorios y evidencian un buen desempeño en la predicción de ventas, es importante reconocer que aún existen limitaciones que afectan su precisión. Uno de los factores claves es la ausencia de variables exógenas relacionadas con el detalle cualitativo de las promociones de descuento. Si bien se dispone de información sobre los descuentos aplicados durante los años 2023 y 2024, no se cuenta con datos sobre el tipo específico de promociones ni las fechas exactas de su implementación. Esta limitación es crucial considerando que la información está agregada a nivel semanal, cuando en la práctica pueden coexistir múltiples promociones en distintos días de una misma semana. Esta falta de granularidad dificulta al modelo capturar con exactitud las variaciones inducidas por campañas promocionales, lo que podría haber contribuido a ciertas imprecisiones, particularmente en periodos donde estas estrategias tuvieron un efecto notable sobre la demanda.

7.4. Análisis de Elasticidad para los Tres Artículos Seleccionados

Con el objetivo de asegurar que el modelo ofrezca resultados eficientes, se realizan una serie de pruebas complementarias. Aunque no es posible determinar con certeza el grado de elasticidad del artículo, debido a que los precios durante los años 2023 y 2024 fueron definidos de forma discrecional por la organización, sin seguir un criterio matemático o sistemático, se procede a calcular la elasticidad precio de la demanda como un ejercicio exploratorio. Este análisis no busca establecer relaciones causales estrictas, sino más bien identificar posibles patrones de comportamiento entre variaciones de precio y cantidades vendidas, que puedan aportar información valiosa para futuras decisiones comerciales o de modelamiento. De este modo, se incorpora una mirada más amplia sobre la sensibilidad de la demanda, aún en contextos donde las decisiones de la asignación de precios no han sido controladas de forma técnica.

Para este estudio, se desarrolló una función que calcula la elasticidad precio de la demanda;

esta métrica permite cuantificar cómo varían las ventas ante cambios en el precio, mediante la comparación de semanas consecutivas y el cálculo de la variación porcentual tanto en ventas como en precios. A partir de ello, se obtiene una elasticidad promedio que indica si la demanda es sensible o no a dichas variaciones. Este cálculo de elasticidad fue aplicado a los tres artículos con “venta alta y constante” (Tabla 33), dando como resultado diferentes resultados dependiendo del artículo:

Tabla 33. Resultados del cálculo de elasticidad de la demanda

ARTÍCULOS CON VENTA ALTA Y CONSTANTE					
Artículo 1		Artículo 2		Artículo 3	
-1,05	Elasticidad negativa	-2,72	Elasticidad negativa	0,55	Elasticidad positiva

Teniendo en cuenta los resultados obtenidos, se observa que dos de los artículos presentan elasticidad negativa, lo cual en principio indicaría que un aumento en el precio está asociado con una disminución en las ventas, especialmente en el caso del Artículo 2, con una elasticidad de -2.72, lo que sugiere una aparente alta sensibilidad al precio. Por otro lado, el Artículo 3 muestra una elasticidad positiva de 0.55, lo que en una lectura inicial podría interpretarse como un efecto de "prestigio", donde un mayor precio podría estar asociado con una percepción de mayor valor. Sin embargo, dado el contexto previamente mencionado, con precios determinados por decisiones internas de liquidación y ventas altamente influenciadas por la estacionalidad, estos resultados deben interpretarse con cautela. La falta de control sobre los factores externos puede estar generando relaciones poco representativas o engañosas entre precios y pares vendidos, lo que limita la validez del análisis desde un punto de vista económico riguroso.

La alta variabilidad observada en los resultados puede atribuirse a la naturaleza de los artículos analizados, ya que corresponden a productos escolares cuya demanda está fuertemente condicionada por la temporada, concentrándose en semanas específicas del año. Asimismo, debe

considerarse que los precios reales e históricos del año 2023, pero aún más del año 2024, no fueron definidos a partir de criterios estadísticos ni mediante procesos de optimización, sino que los descuentos aplicados respondieron a decisiones tácticas internas, enfocadas principalmente en la evacuación de inventario. Esta estrategia, orientada por necesidades operativas más que por señales del mercado, pudo haber generado distorsiones en la relación entre precio y demanda, alterando la respuesta natural del consumidor y dificultando así un análisis fiable de la elasticidad desde una perspectiva económica rigurosa.

7.5. Estrategias de optimización de precios

Para fortalecer la utilidad del modelo Prophet aplicado en este trabajo, se desarrolló una metodología de análisis complementaria que incluye el cálculo dinámico de la elasticidad precio de la demanda y un proceso de optimización de precios. A diferencia de la elasticidad tradicional, que se calcula como la razón entre el cambio porcentual en la cantidad demandada y el cambio porcentual en el precio entre dos periodos consecutivos; es decir, comparando la venta actual frente a la anterior y el precio actual frente al anterior, esta se asume como un valor único y estático, y parte del supuesto de que la relación entre precio y ventas es constante en el tiempo. Sin embargo, en contextos comerciales reales, especialmente en sectores con promociones tácticas, estacionalidades marcadas o decisiones internas que afectan el precio, este enfoque puede resultar limitado o poco representativo.

Con el objetivo de potenciar el uso del modelo Prophet como herramienta predictiva y de apoyo a la toma de decisiones, se implementaron dos metodologías complementarias orientadas a la estimación del precio óptimo que maximiza los ingresos proyectados para los tres artículos en estudio.

En primer lugar, se implementó un enfoque de optimización de precios mediante

simulación con modelo Prophet multivariado. Este enfoque utiliza un modelo previamente ajustado con variables exógenas, como el stock disponible, el precio promedio, los descuentos aplicados, la semana estacionaria y el comportamiento semanal de las ventas; con el fin de generar un conjunto de escenarios en los que el precio se modifica sistemáticamente dentro de un rango dinámico, definido por el comportamiento histórico. Para cada escenario, se recalculan las predicciones de ventas y los ingresos proyectados, seleccionando finalmente el precio que maximiza los ingresos esperados.

El segundo enfoque corresponde al modelo de optimización de precios mediante simulación lineal, caracterizado por su simplicidad y rapidez. A partir del precio promedio histórico, se define un rango fijo de variación y, dentro de ese intervalo, se sustituyen sistemáticamente los valores del precio en el conjunto de datos. Para cada escenario, se generan nuevas predicciones de demanda con el modelo previamente ajustado, y se calculan los ingresos estimados multiplicando el precio por las ventas proyectadas.

A continuación, se muestra los resultados de los dos enfoques aplicados a cada uno de los artículos seleccionados:

Figura 46. Comparativo de simulación de precios

Optimización de precio aplicado al artículo 1			
Escenarios	Pares vendidos	Ingreso	Precio Promedio
Información real (2023 y 2024)	39.026	3.450.564.565	88.417
Optimización de precios con simulación multivariado	58.476	5.120.453.175	87.565
Optimización de precios con simulación lineal	39.379	3.776.009.648	95.888

Optimización de precio aplicado al artículo 2			
Escenarios	Pares vendidos	Ingreso	Precio Promedio
Información real (2023 y 2024)	35.687	2.090.666.676	58.583
Optimización de precios con simulación multivariado	47.803	3.292.115.499	68.868
Optimización de precios con simulación lineal	35.736	2.313.291.712	64.733

Optimización de precio aplicado al artículo 3			
Escenarios	Pares vendidos	Ingreso	Precio Promedio
Información real (2023 y 2024)	20.467	1.755.271.994	85.761
Optimización de precios con simulación multivariado	28.186	2.535.604.807	89.959
Optimización de precios con simulación lineal	20.491	1.928.420.174	94.110

El análisis de los resultados de los tres artículos (Figura 46), indica que, desde una perspectiva lógica y considerando las ventas históricas, el enfoque de simulación lineal podría ser el más adecuado. Aunque ajusta los precios a niveles superiores a los reales, logra mantener las ventas de pares muy cercanas a las observadas en la realidad, lo que tiene una justificación más coherente en el mercado. La simulación multivariada, en cambio, sobreestima los ingresos al presentar una relación ilógica entre el precio y las ventas, ya que no tiene sentido que un precio más alto incremente las ventas. En este contexto, la simulación lineal sugiere que no sería necesario aplicar tantos descuentos, ya que los artículos podrían venderse sin rebajas, lo cual no solo optimiza los ingresos, sino que también favorece el valor de la marca. Al no depender de descuentos constantes, la marca logra mantener su percepción y su posicionamiento en el mercado, lo que hace que el enfoque de simulación lineal sea, por lo pronto, el más adecuado.

Es importante tener en cuenta que ambos enfoques de simulación, tanto el multivariado

como el lineal, se construyen sobre la información histórica disponible, por lo que su validez depende en gran medida de la calidad los datos reales utilizados. En particular, si los precios o descuentos registrados históricamente fueron definidos sin una revisión estadística adecuada o sin criterios técnicos claros, existe el riesgo de que los modelos, especialmente el multivariado, aprendan patrones o relaciones incorrectas. Esto podría llevar a sobreestimar la demanda o los ingresos proyectados, ya que el modelo se apoya en la coherencia entre las variables exógenas y los comportamientos pasados. Por otro lado, el enfoque de simulación lineal, al no modelar relaciones complejas y limitarse a sustituir sistemáticamente los precios dentro de un rango fijo, es menos sensible a posibles errores o distorsiones en los datos históricos. Esto lo convierte en una alternativa más robusta y confiable en contextos donde se sospecha que las decisiones comerciales previas no estuvieron respaldadas por un análisis riguroso.

8. EVALUACIÓN DEL RENDIMIENTO Y PRECISIÓN DEL MODELO DE APRENDIZAJE ESTADÍSTICO

Una vez completada la etapa de análisis aplicado, centrada en la proyección de ventas y la optimización de precios, el enfoque del estudio se dirige ahora hacia la validación técnica del modelo. El objetivo de esta fase es evaluar de manera rigurosa el rendimiento del modelo Prophet en términos de precisión, estabilidad y robustez, mediante una serie de pruebas específicas. Estas pruebas permitirán determinar la idoneidad del modelo para su implementación como herramienta de apoyo en la toma de decisiones comerciales.

8.1. Prueba de Estrés por Volumen de Datos

Con el fin de evaluar la capacidad del modelo Prophet ante grandes volúmenes de datos, se realizó una prueba de estrés replicando el conjunto de observaciones (Figura 47). Esta simulación permitió comprobar su comportamiento sin comprometer la estructura ni la calidad de la información. El modelo se ajustó correctamente, demostrando robustez computacional y escalabilidad, aspectos clave para su implementación en entornos productivos con historiales extensos o datos en flujo continuo.

8.2. Prueba de Sensibilidad a la Volatilidad

Para evaluar la adaptabilidad del modelo ante datos inciertos, se simuló una volatilidad moderada añadiendo ruido gaussiano (media cero, desviación estándar de 10) a la serie original. Esto emula fluctuaciones comunes por promociones, estacionalidad o cambios en la demanda. El modelo Prophet se entrenó sin errores, demostrando su capacidad para generar predicciones confiables incluso en contextos inestables.

8.3. Prueba de Manejo de Datos Faltantes

Para evaluar la robustez del modelo Prophet ante datos incompletos, se simuló un 20% de valores nulos asignados aleatoriamente en la serie temporal (Figura 49). Aun con esta pérdida, el modelo se entrenó sin errores, demostrando su capacidad para operar eficazmente en contextos reales con registros faltantes por errores, inconsistencias o ausencia de datos.

8.4. Prueba de Eficiencia Computacional

Con el fin de evaluar la eficiencia computacional del modelo Prophet, se midió el tiempo de entrenamiento utilizando una serie temporal compuesta por 104 observaciones, correspondientes a un único artículo durante un periodo de dos años. El resultado fue un tiempo de entrenamiento de aproximadamente 0.05 segundos, lo que evidencia una alta velocidad de procesamiento en contextos de baja volumetría de datos.

8.5. Prueba de Determinismo del Modelo

Continuando con el proceso de pruebas, se evaluó la propiedad de determinismo del modelo Prophet, entendida como la capacidad del modelo de generar resultados idénticos al ser entrenado múltiples veces con los mismos datos y parámetros. Para ello, se entrenaron dos instancias independientes del modelo con idéntica configuración y conjunto de datos. Los resultados predichos (\hat{y}) en ambos casos fueron exactamente iguales, lo que confirma que Prophet se comporta de manera determinística en este contexto. Esta característica es fundamental para asegurar la reproducibilidad de los resultados, especialmente en aplicaciones analíticas donde la trazabilidad del proceso es crítica.

8.6. Prueba de Integración de Regresores Externos

Con el propósito de validar la capacidad del modelo Prophet para incorporar variables explicativas externas, se realizó una prueba utilizando un regresor adicional simulado. Este regresor binario fue integrado al modelo mediante el método `add_regressor`. El modelo se entrenó exitosamente, sin errores de procesamiento, lo que demuestra que Prophet permite la inclusión de variables adicionales que pueden capturar factores exógenos relevantes para la predicción. Esta funcionalidad amplía el potencial analítico del modelo y lo habilita para escenarios de mayor complejidad estructural.

8.7. Prueba de Reentrenamiento con Datos Nuevos

Con la intención de comprobar la capacidad del modelo Prophet para ser reentrenado con datos nuevos, se desarrolló una prueba de actualización incremental. Para ello, se simuló observaciones adicionales que fueron concatenadas al conjunto original. El modelo fue ajustado nuevamente con esta serie extendida, sin presentar errores en el proceso de entrenamiento. Este resultado valida la viabilidad del uso de Prophet en contextos operativos donde los datos se actualizan periódicamente, como en entornos de monitoreo de la demanda.

Con base en las distintas pruebas que se realizaron (Figura 47), se concluye que el modelo Prophet cuenta con las condiciones técnicas necesarias para ser implementado en el entorno empresarial ya que ofrece precisión, estabilidad y adaptabilidad. El modelo fue capaz de procesar conjuntos de datos más grandes sin presentar fallas, lo que confirma que puede usar alta cantidad de registros de datos. También mantuvo predicciones estables incluso cuando se introdujeron cambios aleatorios en los datos, lo que demuestra su buen desempeño en situaciones de variabilidad. Frente a datos faltantes, Prophet logró entrenarse sin problemas, lo que es importante

considerando que en la práctica los registros no siempre están completos.

En cuanto a rendimiento, el modelo mostró un tiempo de entrenamiento muy corto, lo cual permite aplicarlo incluso cuando se necesite ejecutar varios modelos al mismo tiempo. Prophet también pudo incorporar variables adicionales, lo que permite incluir factores externos que influyen en la demanda. Por último, se verificó que puede actualizarse con nueva información sin inconvenientes, lo que es fundamental para trabajar con datos que se actualizan constantemente.

Figura 47. *Tipo de prueba y resultado*

Tipo de prueba	Resultado
Prueba de estrés por volumen de datos	Modelo con muchos datos pasó la prueba
Prueba de sensibilidad a la volatilidad	Prueba con precio volátil: Ok
Prueba de manejo de datos faltantes	Modelo con datos faltantes pasó la prueba
Prueba de eficiencia computacional	Tiempo de entrenamiento: 0.05255126953125 segundos
Prueba de determinismo del modelo	Prueba de entrada de datos: superada
Prueba de integración de regresores externos	Prueba de nuevos regresores: superada
Prueba de reentrenamiento con datos nuevos	Prueba de reentrenamiento: superada

En conjunto, estas pruebas muestran que el modelo basado en Prophet es una herramienta confiable y versátil para realizar pronósticos de series de tiempo en el contexto empresarial para el cual fue aplicado.

9. CONCLUSIONES Y TRABAJOS FUTUROS

9.1. Conclusiones

Este proyecto tuvo como objetivo principal analizar la relación entre las variaciones de precios y la demanda de productos en una empresa del sector calzado, con el fin de mejorar su competitividad y sostenibilidad. A través de un enfoque metodológico que inicialmente se centró en la elasticidad precio-demanda, se identificó que los datos históricos no permitían calcular una elasticidad confiable debido a la falta de condiciones estadísticas controladas en los cambios de precios. Ante esta limitación, el enfoque se redirigió hacia la proyección de la demanda utilizando variables claves como el stock disponible, el precio promedio, los descuentos aplicados, la semana estacionaria y el comportamiento semanal de las ventas, con enfoque en los productos de la categoría escolar. Tras evaluar diversos modelos, se determinó que el modelo Prophet, integrado con regresores exógenos, proporcionó las mejores estimaciones de demanda, especialmente para aquellos productos con venta alta y constante. Este modelo permitió simular escenarios futuros de ventas para el año 2025, brindando a la empresa una herramienta predictiva eficaz para la toma de decisiones informadas y para adaptarse mejor a las dinámicas del mercado. Aunque el proyecto enfrentó limitaciones relacionadas con la falta de un sistema estructurado basado en la asignación inteligente de precios, los resultados obtenidos destacan la importancia de implementar un enfoque estadístico adecuado en la fijación de precios, lo que podría mejorar significativamente la estrategia comercial de la organización. En resumen, la herramienta desarrollada ofrece una base sólida para anticipar la demanda y tomar decisiones más acertadas en un entorno competitivo cada vez más dinámico.

9.2. Trabajos Futuros

El presente estudio ha proporcionado valiosos conocimientos sobre la relación entre los precios y la demanda de artículos de calzado para una compañía de Retail. Sin embargo, existen varias limitaciones y áreas que pueden ser mejoradas en futuros trabajos. En primer lugar, la imposibilidad de calcular una elasticidad precio-demanda confiable, debido a la falta de una metodología robusta para la asignación de precios subraya la necesidad urgente de desarrollar una estructura sólida y estadísticamente validada para la asignación de precios. Este es un paso primordial con el cual la empresa podrá tomar decisiones más informadas y precisas en cuanto a la estrategia de precios. En este sentido, resulta crucial que la empresa avance hacia el desarrollo de una estructura formal para la asignación de precios, basada en criterios estadísticos y análisis de datos, en lugar de continuar con el enfoque empírico actual. Esto permitiría establecer precios de forma más coherente y alineada con el comportamiento real de la demanda, lo cual es fundamental para aplicar posteriormente modelos de optimización más avanzados. Una vez implementada esta base estructurada, será posible explorar técnicas más sofisticadas, como algoritmos de machine learning, para ajustar los precios de manera dinámica y estratégica, respondiendo con mayor precisión a las condiciones cambiantes del mercado.

Además, se recomienda que en el futuro se amplíe el modelo de optimización de precios a otras categorías de productos. Esto implicaría ajustar el análisis y la estrategia según las características particulares de cada grupo, adaptando el modelo para analizar artículos de diferentes segmentos, como ropa, accesorios entre otros. Cada segmento podría presentar patrones de demanda, elasticidad de precios y factores de mercado específicos que requerirán un enfoque diferenciado.

En un horizonte temporal más cercano, también se podrían explorar nuevas variables

exógenas que mejoren la predicción, como los cambios en las preferencias de los consumidores o las dinámicas del mercado, lo que permitiría anticipar cambios en la demanda y ajustar la estrategia de precios de manera continua y adaptable. En un horizonte de tiempo más lejano, la automatización de estos procesos mediante algoritmos de machine learning podría ser una herramienta clave para actualizar dinámicamente las estrategias de precios en tiempo real, lo que optimizaría la toma de decisiones a nivel empresarial y respondería con mayor rapidez a las fluctuaciones del mercado.

Finalmente, este trabajo establece una base sólida para futuras investigaciones que aborden la mejora de la asignación de precios y la proyección de la demanda, ayudando a la empresa a responder de manera más efectiva a las demandas del mercado.

10. REFERENCIAS BIBLIOGRÁFICAS

- Avellaneda, J., Ochoa, C., y Figueroa-García, J. (2012). Comparación entre un sistema neuro difuso auto organizado y un modelo ARIMAX en la predicción de series económicas volátiles. *Ingeniería*, 17(2), 26-34. <https://core.ac.uk/download/pdf/229149522.pdf>
- Banda Ortiz, H. y Garza Morales, R. (2014). Aplicación teórica del método Holt-Winters al problema de credit scoring de las instituciones de microfinanzas. *Mercados y Negocios*, 15(2), 5-22. <https://dialnet.unirioja.es/servlet/articulo?codigo=5811252>
- Barbosa Moyano, I. C. (2020). *Importancia del análisis de datos como método para minimizar los riesgos de un emprendimiento* [trabajo, Universidad Militar Nueva Granada]. <https://repository.umng.edu.co/server/api/core/bitstreams/f43de1ec-a82d-4911-9412-7aa37479dfa6/content>
- Carrasco Ochoa, J. A. y Martínez Trinidad, J. F. (2011). Reconocimiento de patrones. *Komputer Sapiens*, 2(3). https://www.researchgate.net/profile/Horacio-Carvajal-2/publication/281423031_Mineria_de_datos_en_el_ambiente_bancario/links/55e6705508aebdc0f58bb4d3/Mineria-de-datos-en-el-ambiente-bancario.pdf
- Cerda, A., Lobos, G., Kufferath, E. y Sánchez, F. (2004). Elasticidades de demanda por manzanas chilenas en el mercado de la Unión Europea: una estimación econométrica. *Agricultura técnica*, 64(4), 399-408. https://oes.chileanjar.cl/files/V64I4A08_es.pdf
- Chicaiza Yugcha, O. F., Martínez Guaman, C. J., Orozco Manobanda, I. A. y Arellano Castro, Ángel D. (2024). Previsión del consumo eléctrico en el cantón Salcedo mediante técnicas de aprendizaje automático. *Revista Odigos*, 5(1), 9-24. <https://doi.org/10.35290/ro.v5n1.2024.1134>

Churchill, G. A., & Iacobucci, D. (2010). *Investigación de mercados*. Cengage.

De Arce, R. y Mahía, R. (2001). ADE – Econometría II. Modelos de Series Temporales. En

Modelos Arima. <https://gc.scalahed.com/recursos/files/r161r/w24503w/Box-Jenkins.PDF>

Economipedia. (2024). *Elasticidad precio de la demanda*.

<https://economipedia.com/definiciones/elasticidad-precio-de-la-demanda.html>

Espinosa-Zúñiga, J. J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base

de solicitudes de tarjetas de crédito. *Ingeniería, investigación y tecnología*, 21(3).

<https://www.scielo.org.mx/scielo.php?pid=S1405->

[77432020000300002&script=sci_arttext](https://www.scielo.org.mx/scielo.php?pid=S1405-77432020000300002&script=sci_arttext)

García, J., Molina, J. M., Berlanga, A., Patricio, M. A., Bustamante, Á. L. y Padilla, W. R.

(2018). *Ciencia de datos: Técnicas analíticas y aprendizaje estadístico: un enfoque*

práctico. Alfaomega.

Guerrero-Garzón, P. P. (2011). *Metodología para la fijación de precios de repuestos en una*

empresa del sector automotriz mediante la utilización de la elasticidad precio – demanda

[tesis de maestría, Universidad Nacional de Colombia].

<https://repositorio.unal.edu.co/bitstream/handle/unal/7852/822038.2011.pdf?sequence=1>

[&isAllowed=y](https://repositorio.unal.edu.co/bitstream/handle/unal/7852/822038.2011.pdf?sequence=1&isAllowed=y)

Guerrero-Garzón, P. P., Hernández Losada, D. F. y Díaz Monroy, L. G. (2012). Metodología para

la fijación de precios mediante la utilización de la elasticidad precio-demanda. Caso tipo:

repuestos del sector automotor. *Apuntes del CENES*, 9-36.

<https://revistas.uptc.edu.co/index.php/cenes/article/view/13>

Guzmán Llontop, A. R. (2020). *Investigación de modelos matemáticos de optimización de*

rentabilidad a través de la asignación de precios para una entidad financiera

- emprendimiento* [trabajo de bachiller, Pontificia Universidad Católica del Perú].
<https://tesis.pucp.edu.pe/server/api/core/bitstreams/571b71a3-b57a-4a1e-9e6a-b94b59e2b582/content>
- Heizer, J. y Render, B. (2014). *Administración de operaciones*. Pearson.
- Kotler, P. y Armstrong, G. (2017). *Principios de marketing*. Pearson.
- Kotler, P. y Keller, K. L. (2016). *Dirección de marketing*. Pearson.
- López Sáez, J. I. (2018). *Análisis de series de tiempo. Pronóstico de demanda de uso de aeropuertos en Argentina* [tesis de especialización, Instituto Tecnológico de Buenos Aires]. <https://ri.itba.edu.ar/server/api/core/bitstreams/46f84c43-9130-49c0-bf7e-4b41ecb051ef/content>
- Mankiw, N. G. (2018). *Principios de economía*. Cengage Learning.
- Moreno-Aguilar, L. A., Guerrero-López, C. M., Colchero, M. A., Quezada-Sánchez, A. D., y Bautista-Arredondo, S. (2021). Elasticidad precio y elasticidad ingreso de la demanda de cerveza en México. *Salud Pública de México*, 63(4), 575-582.
<https://doi.org/10.21149/12026>
- Nagle, T., Hogan, J., & Zale, J. (2016). *The Strategy and Tactics of Pricing*. Routledge.
- Ortiz Lara, A. (2020). *Modelo de mercado para la elasticidad de precios de artículos de una cadena departamental* [tesis de pregrado, Universidad Autónoma de México].
<https://ru.dgb.unam.mx/bitstream/20.500.14330/TES01000840089/3/0840089.pdf>
- Peceros, L. y Salcedo, M. K. (2022). Estrategia de asignación de precios y posicionamiento en el establecimiento AGRIFER INMAR en la provincia de Andahuaylas - 2022. [tesis de pregrado, Universidad César Vallejo]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/93919>

Question Pro. (2023). *Elasticidad de precios: Qué es, tipos y cómo calcularla*.

<https://www.questionpro.com/blog/es/elasticidad-de-precios/>

Samuelson, P. A. y Nordhaus, W. D. (2010). *Economía*. McGraw-Hill.

Schiffman, L. y Kanuk, L. (2010). *Comportamiento del consumidor*. Pearson.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica*, 24(1), 12-

18. <https://hrcak.srce.hr/115732>

Suescún Díaz, A. F. (2019). *Modelo Sarima para el pronóstico del nivel del río Magdalena a la altura del municipio de Barrancabermeja* [tesis de especialización, Fundación Universitaria los Libertadores].

<https://repository.libertadores.edu.co/server/api/core/bitstreams/e823e6b9-33fa-48e8-9c4d-ae75023c2b93/content>

Varian, H. R. (2014). *Microeconomía intermedia*. Antoni Bosch.

Villarreal, F. (2016). *Introducción a los Modelos de Pronósticos*. Universidad Nacional del Sur.

https://www.matematica.uns.edu.ar/uma2016/material/Introduccion_a_los_Modelos_de_Pronosticos.pdf