



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Demanda por categoría en retail: integración de eventos comerciales en modelos de
pronóstico.

MATE -214

PROYECTO INTEGRADOR

Previo a la obtención del Título de:

Ingeniero Estadístico y Matemática

Presentado por:

Gerleen Josué Goya Pincay

Kerly Adriana Benavides Salmerón

GUAYAQUIL - ECUADOR

Año: 2025

DEDICATORIA

A mis padres. Sin su esfuerzo y apoyo, este logro no habría sido posible.

Gerleen Goya P

DEDICATORIA

A Jack,

Te extraño cada día.

Gracias por haber sido mi compañero incondicional, mi apoyo en cada madrugada de estudio, y mi impulso constante cuando todo parecía difícil.

Este camino lo recorrimos juntos, y siempre lo llevaré en mi corazón.

A mis padres y hermana, pilares fundamentales en mi vida.

Kerly Benavides S.

AGRADECIMIENTOS

En primer lugar, a Dios, el autor de mi vida, quien me eligió antes de la fundación del mundo, por ser mi guía y fortaleza para continuar, y por los planes que sé que guarda para mí, perfectos en su tiempo.

A mi mamá, Marjorie, mi inspiración para seguir adelante. Por cada madrugada que estuvo conmigo cuando tenía que viajar hacia la universidad. Por acompañarme en cada proceso durante mi formación académica. Quien celebró conmigo mis logros y creyó en mí incluso cuando yo dudaba.

A mi papá, Henry, mi ejemplo de fuerza y superación. Por siempre levantarme el ánimo con su carisma y por nunca faltar ese abrazo de bienvenida todos los días al regresar a casa.

A mis hermanos, Elkin y Zuleyka, por su apoyo incondicional.

A la ESPOL, por brindarme una formación académica integral y por todas las oportunidades de crecimiento profesional y personal.

A mi tutora de tesis, Mariela, por su orientación, paciencia y disposición para guiarnos durante este proceso. Su apoyo tiene gran peso en el desarrollo de este trabajo.

A mis docentes y coordinadores, quienes, con sus conocimientos, aportaron de manera significativa a mi formación.

A Rosemery, por su amistad y sus consejos, que me ayudaron a crecer y a tomar decisiones claves en mi vida.

A Kerly, por su amistad y compañerismo. Por su valioso aporte al desarrollo de este proyecto.

Gracias a Geovanny, Pablo, Abraham, Nicolás, Laura, Hellen, Fabricio y todos quienes me brindaron su amistad sincera, llenando esta etapa con momentos que quedarán marcados en mi memoria y que hicieron que este camino sea más llevadero y alegre.

A todas las personas mencionadas, mi gratitud por acompañarme durante este proceso.

Gerleen Goya P.

AGRADECIMIENTOS

A Dios y a la Virgen María, por darme vida, guiar mis pasos y brindarme fuerza en cada momento de mi camino.

A mi madre, Rosely Salmerón, quien me demuestra su amor en cada detalle.

A mi padre, José Benavides, gracias por estar siempre pendiente de mí, levantándose cada mañana.

Mi eterna gratitud a mis padres por su apoyo incondicional, por la confianza que siempre han tenido en mí y por ser la guía y el ejemplo que iluminan cada paso de mi vida.

A mi hermana, Sheyla Benavides, quien con sus consejos me ha acompañado a lo largo de estos 23 años y se ha convertido en mi modelo a seguir.

A Jack, por enseñarme el valor de la persistencia.

A Hannie, quien en los momentos más difíciles ha logrado sacarme sonrisas y me recibe siempre con alegría. Gracias, loquita.

A mi profesora Elimar Marchán, por su paciencia y comprensión a lo largo de mi formación académica.

A la profesora Jessica Menéndez, por haberme orientado en mi inserción al mundo laboral.

A las profesoras Heydi Roa y Mariela González, gracias por su guía y apoyo durante estos cuatro meses.

Gerleen Goya, mi compañero desde Álgebra Lineal I en el segundo semestre, gracias por tu amistad y apoyo. Te deseo muchos éxitos.

Kerly Benavides S.

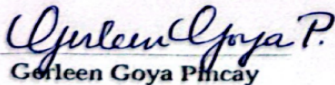
DECLARACIÓN EXPRESA

Nosotros, *Gerleen Josué Goya Pincay* y *Kerly Adriana Benavides Salmerón*, acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales, se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotros durante el desarrollo del proyecto de graduación, pertenecerá de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de nuestra innovación, de ser el caso. En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique los autores que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 29 de mayo del 2025.

 Gerleen Goya Pincay	 Kerly Benavides Salmerón
--	--

EVALUADORES

Heydi Mariana Roa López

PROFESOR DE MATERIA

Mariela Alexandra González Narváez

TUTOR DE PROYECTO

RESUMEN

El sector de la moda está caracterizado por la alta variabilidad en la demanda, fuertemente influenciada por factores externos como los eventos comerciales, las tendencias y promociones. En un entorno altamente competitivo, la previsión de la demanda es clave para la rentabilidad y sostenibilidad de las empresas. Este trabajo busca desarrollar un modelo de pronóstico, incorporando variables exógenas, para generar predicciones confiables, alineando las estimaciones con el comportamiento real del mercado y optimizando las estrategias de adquisición de inventarios, contribuyendo al cumplimiento de los objetivos planificados y reduciendo tanto la escasez de productos como el sobreinventario.

Se compararon modelos de series temporales ARIMA estacional y SARIMAX, utilizando una base de datos de ventas históricas previamente imputada, garantizando la consistencia de las categorías más representativas.

Los resultados evidencian la importancia de realizar pronósticos desagregados por categoría debido a los diferentes patrones de comportamiento que un pronóstico global no captura. En este contexto, los modelos SARIMAX demostraron un mejor desempeño, al incorporar factores externos que mejoraron su capacidad predictiva.

Se concluye que el modelo propuesto permite alinear los objetivos comerciales con el comportamiento del mercado, optimizando los niveles de adquisición de inventarios y fortaleciendo la toma de decisiones en entornos dinámicos.

Palabras Clave: Pronósticos, Retail, Variables exógenas, SARIMAX, Series temporales.

ABSTRACT

The fashion industry is defined by the high variability in demand, highly influenced by external factors such as commercial events, trends, and promotions.

In a highly competitive environment, demand forecasting is a key factor for the company's profitability and sustainability. This investigation aims to develop a forecasting model, incorporating exogenous variables, in order to provide reliable predictions, compared to traditional methods based solely on historical sales data. Through this approach, the aim is to align the estimates with actual market behavior to optimize inventory management strategies. Consequently, it contributes to fulfilling planned goals, mitigating both product shortages and inventory surpluses.

Seasonal ARIMA and SARIMAX models were applied to an imputed historical sales dataset to guarantee the reliability of records.

The results underscore the critical importance of managing forecasts at the category level, as aggregate forecasts fail to capture the distinct behavioral patterns inherent to each category. Within this framework, SARIMAX models demonstrated superior performance compared to traditional approaches by incorporating external variables that enhanced predictive accuracy.

This research concludes that the proposed modeling approach facilitates a more precise alignment of business objectives with actual market demand, thereby optimizing inventory levels and reinforcing strategic decision-making in dynamic and competitive market environments.

Keywords: Forecast, Retail, Exogenous variables, SARIMAX, Time series.

ÍNDICE GENERAL

RESUMEN	I
ABSTRACT	II
ABREVIATURAS	VI
SIMBOLOGÍA	VII
ÍNDICE DE FIGURAS	X
ÍNDICE DE TABLAS	XI
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1 Descripción del problema	2
1.2 Justificación del problema	4
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos	4
1.4 Marco teórico	5
1.4.1 Pronóstico de la demanda en el sector minorista	5
1.4.2 Imputación de datos en series temporales	6
1.4.3 Segmentación mediante agrupamiento: K-means	7
1.4.4 Proceso estocástico	8
1.4.5 Fundamentos de las series temporales	10
1.4.6 Modelos estocásticos univariantes	12
1.4.7 Causalidad y test de Granger	20

1.4.8 Diagnóstico de residuos	22
1.4.9 Supuestos	24
1.4.10 Paradoja de Simpson	26
1.4.11 Métricas de evaluación	27
CAPÍTULO 2	31
2. METODOLOGÍA	31
2.1 Clasificación de categorías por nivel de ventas	32
2.2 Recopilación y estructuración de datos históricos	33
2.3 Tratamiento de datos faltantes e imputación	34
2.4 Integración de eventos comerciales	35
2.5 Análisis y modelado	35
2.5.1 Análisis en Dominio de la Frecuencia	35
2.6 Evaluación de resultados y consideraciones finales	36
CAPÍTULO 3	37
3. RESULTADOS Y ANÁLISIS	37
3.1 Imputación de datos faltantes	37
3.2 Participación por Categoría	39
3.3 Variaciones interanuales	40
3.4 Descomposición y análisis de correlaciones	42
3.5 Pruebas de estacionariedad de las series	47
3.6 Prueba de Causalidad	48
3.7 Ajuste de modelos	50
3.8 Métricas y diagnóstico de residuales	53
3.9 Pronósticos	54

3.9.1 Categoría 1	54
3.9.2 Categoría 2	56
3.10 Categoría 3	58
3.10.1 Categoría 4	60
CAPÍTULO 4	63
4. CONCLUSIONES Y RECOMENDACIONES	63
4.1 Conclusiones	63
4.2 Recomendaciones	65

BIBLIOGRAFÍA

APÉNDICES

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
SKU	Stock Keeping Unit (Unidad de Mantenimiento de Stock)
ARIMA	AutoRegressive Integrated Moving Average (Media Móvil Integrada Autoregresiva)
SARIMA	Seasonal AutoRegressive Integrated Moving Average
RB	Ruido Blanco
DFT	Discrete Fourier Transform (Transformada de Fourier Discreta)
AR	AutoRegresivo
MA	Media Móvil
ARMA	AutoRegresivo de Media Móvil
SARIMAX	Seasonal ARIMA with eXogenous variables (ARIMA Estacional con variables exógenas)
Q-Q plot	Quantile-Quantile plot, gráfico para comparar distribuciones
MAE	Error Absoluto Medio (Mean Absolute Error)
RMSE	Raíz del Error Cuadrático Medio (Root Mean Square Error)
MAPE	Error Porcentual Absoluto Medio (Mean Absolute Percentage Error)
MSE	Error Cuadrático Medio (Mean Squared Error)
RSS	Residual Sum of Squares (Suma de cuadrados de residuos)
SP	Simpsons Paradox (Paradoja de Simpson)

SIMBOLOGÍA

p	Orden del modelo autoregresivo
Y_t	Valor de la serie temporal en el tiempo t
T	Longitud total de la serie temporal
ϕ_i	Parámetros del modelo AR
B	Operador de rezago (Backshift operator)
ACF	Función de Autocorrelación (Autocorrelation Function)
PACF	Función de Autocorrelación Parcial (Partial Autocorrelation Function)
$\phi(B)$	Polinomio autorregresivo en el operador de rezago
$\theta(B)$	Polinomio de media móvil en el operador de rezago
a_t	Ruido blanco o término de error
Δ^d	Operador de diferenciación de orden d
λ_i	Raíces del polinomio característico
I	Integrado (diferenciación para estacionariedad)
p, d, q	Parámetros del modelo ARIMA
P, D, Q	Parámetros estacionales del modelo SARIMA
s	Periodicidad estacional
β	Vector de coeficientes para variables exógenas
X_t	Variables exógenas en el modelo SARIMAX
Y_t	Vector de series temporales en el tiempo t
ε_t	Vector de residuos o errores (ruido blanco)

H_0	Hipótesis nula del test de causalidad de Granger
$\hat{\varepsilon}_t$	Residuos estimados del modelo
$\hat{\rho}_h$	Estimación de autocorrelación en el rezago h
$\hat{Y}_t$	Valor predicho para el tiempo t
n	Número total de observaciones
x_i	Observación en el algoritmo de K-means
μ_j	Centroide del clúster C_j en K-means
C_j	Conjunto de observaciones asignadas al clúster j
K	Número de armónicos o clústeres según el contexto
f	Frecuencia en el análisis espectral
Z	Variable categórica

ÍNDICE DE FIGURAS

Figura 2.1	Diagrama de flujo de la metodología aplicada	31
Figura 2.2	Selección del número de clústers óptimo	33
Figura 2.3	Distribución de Ventas por clúster	33
Figura 3.1	Series de ventas originales e imputadas con HoltWinters, por categoría (20182024).	38
Figura 3.2	Participación anual de ventas por categoría (20182024).	39
Figura 3.3	Ventas anuales agregadas por categoría (20182024).	41
Figura 3.4	Comparación de tendencias de ventas por categoría normalizadas (base 100) (20182025).	42
Figura 3.5	Descomposición de las series de ventas imputadas por categoría.	43
Figura 3.6	Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) por categoría.	44
Figura 3.7	Análisis espectral por Fourier de las series - Categoría 1	45
Figura 3.8	Análisis espectral por Fourier de las series - Categoría 2	46
Figura 3.9	Análisis espectral por Fourier de las series - Categoría 3	46
Figura 3.10	Análisis espectral por Fourier de las series - Categoría 4	47
Figura 3.11	Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 1.	55
Figura 3.12	Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 2.	57
Figura 3.13	Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 3.	59

Figura 3.14 Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 4.	61
---	----

ÍNDICE DE TABLAS

Tabla 3.1	Resultados de pruebas de estacionariedad (ADF y KPSS) por categoría.	47
Tabla 3.2	Pruebas de causalidad de Granger (p-valores) por categoría, campaña y distintos rezagos.	50
Tabla 3.3	Resumen de métricas de desempeño, criterios de información y pruebas de residuos para los modelos SARIMA y SARIMAX por categoría.	54
Tabla 3.4	Pronóstico de la Categoría 1 para las semanas 2025.0012025.026.	56
Tabla 3.5	Pronóstico de la Categoría 2 para las semanas 2025.0012025.026.	58
Tabla 3.6	Pronóstico de la Categoría 3 para las semanas 2025.0012025.026.	60
Tabla 3.7	Pronóstico de la Categoría 4 para las semanas 2025.0012025.026.	62
Tabla A.1	Variaciones porcentuales interanuales de ventas – Categoría 1 (sin imputar)	
Tabla A.2	Variaciones porcentuales interanuales de ventas – Categoría 2 (sin imputar)	
Tabla A.3	Variaciones porcentuales interanuales de ventas – Categoría 3 (sin imputar)	
Tabla A.4	Variaciones porcentuales interanuales de ventas – Categoría 4 (sin imputar)	
Tabla A.5	Variaciones porcentuales interanuales de ventas – Categoría 1	
Tabla A.6	Variaciones porcentuales interanuales de ventas – Categoría 2	
Tabla A.7	Variaciones porcentuales interanuales de ventas – Categoría 3	
Tabla A.8	Variaciones porcentuales interanuales de ventas – Categoría 4	

CAPÍTULO 1

1. INTRODUCCIÓN

En el ámbito minorista, poder predecir con precisión la demanda se ha convertido en un aspecto importante de la planificación empresarial y operativa debido a la competencia cada vez más feroz, que requiere tomar decisiones oportunas. En una época en la que los hábitos de los consumidores son inestables y hay pocos factores de control, como las múltiples influencias del mercado, la economía y las estaciones, basarse en estimaciones incorrectas para la planificación puede dar lugar a una amplia gama de retos: desviaciones de las ventas con respecto a los objetivos, roturas de stock, exceso de stock, exceso de costos por rebajas no planificadas.

Esta situación tiene una relevancia significativa en el sector minorista de la moda, ya que la demanda de un producto puede sufrir frecuentes altibajos, impulsados a su vez por las tendencias del momento, los nuevos anuncios, las campañas de marketing exitosas o cualquier otro evento comercial.

Para hacer frente a estos retos, varios estudios recientes han sugerido diferentes enfoques al problema de aumentar la precisión de los modelos de proyección de la demanda. Estos trabajos han demostrado claramente la importancia tanto del tiempo como de las variables exógenas. En el estudio realizado por Khalid (2024) se presta especial atención al valor potencial de diversos modelos tradicionales y avanzados para distintos horizontes temporales y dinámicas comerciales.

Por otro lado, se ha demostrado que la inclusión de variables exógenas es un procedimiento que mejora la capacidad de previsión. Ghareeb et al. (2023) muestran que cuando se incorporan eventos como vacaciones y promociones en los modelos de demanda, la capacidad de estos últimos

para captar variaciones aumenta y los modelos son más coherentes al entorno real.

En respuesta a esta necesidad, este trabajo presenta la implementación de un modelo de previsión de demanda orientado a categorías en el negocio minorista, incluyendo eventos comerciales como variables exógenas. El objetivo es desarrollar un método predictivo óptimo y alinear los objetivos mensuales con el comportamiento real del mercado. Adicionalmente, se busca fomentar, en cierta medida, la planificación basada en datos.

1.1 Descripción del problema

En el sector minorista, realizar previsiones precisas de la demanda es fundamental para el funcionamiento y la planificación del negocio. En la industria de la moda, los patrones de ventas de categorías como camisetas, camisas, pantalones, chompas y jeans varían en función de las estacionalidades y los eventos comerciales, incluidas las temporadas y las tendencias de moda. Asimismo, estas categorías se dividen a su vez en líneas básicas y de tendencia, cada una con sus propias características y comportamientos únicos.

La empresa tiene dificultades para predecir la demanda de estos productos, especialmente durante meses clave, como aquellos correspondientes a promociones o rebajas. Estos eventos provocan picos de demanda que, si no se modelan adecuadamente, pueden llevar a tomar decisiones empresariales erróneas. Según Chopra (2016), la predicción de la demanda es parte integral de la toma de decisiones en la planificación operativa. Este método posibilita el ajuste de la oferta con la demanda prevista, lo que da lugar a la doble ventaja de mitigar los costos operativos y mejorar la experiencia del cliente.

La falta de un método de predicción preciso puede dar lugar a desviaciones entre los objetivos de ventas y la respuesta real del mercado. Esto supone un riesgo de que los objetivos se fijen demasiado altos, por lo que resultan poco prácticos, o demasiado bajos, comprometiendo la

sostenibilidad de la empresa. Se ha evidenciado que la falta de precisión de las previsiones tiene un impacto directo y a corto plazo en la gestión de inventarios, ya que permite desviaciones en el indicador de rebajas, que se eleva por encima de lo establecido por las políticas internas. Esta situación genera numerosas actividades operativas y financieras para la empresa. Fildes et al. (2019) destacan la importancia de la precisión en la predicción de la demanda con el objetivo de mejorar el rendimiento operativo y la rentabilidad de las cadenas minoristas.

El problema más evidente en el negocio es que no posee un modelo para la previsión de la demanda que tenga en cuenta su estacionalidad, los datos históricos y los eventos comerciales relevantes. La resolución de este problema permitiría establecer objetivos mensuales más precisos en función de la demanda prevista, mejorar la planificación y aumentar la eficiencia comercial del negocio.

El proyecto implica una serie de consideraciones operativas y de costos que deben abordarse. Por lo tanto, en primer lugar, el análisis se limita a las categorías que aparecen en los registros históricos de ventas. En consecuencia, los cambios progresivos en la clasificación de los productos podrían afectar la coherencia del estudio. Además, el rigor de la previsión está condicionado por la disponibilidad y la calidad de los datos, especialmente los relacionados a eventos comerciales pasados.

Otra limitación importante está relacionada con la situación operativa actual de la empresa. Por ello, la intervención directa en su sistema de toma de decisiones o la modificación del proceso resulta inviable sin una prueba preliminar del modelo. La ejecución de este plan está sujeta a la aceptación de los directores comerciales de la empresa.

Desde la perspectiva del cliente, debería ser un modelo que mejore la eficiencia y que no requiera grandes inversiones iniciales en software e infraestructura. Por lo tanto, se dará preferencia a un software estadístico de código abierto que resulte conveniente.

1.2 Justificación del problema

En el entorno minorista actual, altamente competitivo y con una demanda impulsada por las promociones, la presencia de un modelo de previsión sólido y flexible es imprescindible para una buena gestión. Este trabajo de investigación busca proporcionar una solución concreta a la brecha entre los objetivos de ventas mensuales y la demanda real.

A diferencia de los métodos de planificación estándar, que se basan únicamente en ajustes porcentuales o en promedios históricos, el enfoque que se propone recomienda tener en cuenta los patrones estacionales, las tendencias y los eventos comerciales, para realizar previsiones más precisas y flexibles.

En términos empresariales, una previsión de la demanda más precisa permite optimizar la asignación de objetivos de la empresa, así como prever escenarios clave de exceso de oferta o déficit, lo cual aporta a una mejora en la gestión de productos para la venta. Además de ser una solución a un problema operativo específico, el modelo puede extenderse a otras unidades de negocio y replicarse fácilmente en empresas con dinámicas comerciales similares.

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar un modelo de pronóstico de demanda por categoría para una empresa del sector retail, utilizando técnicas de series temporales ajustadas por eventos comerciales, orientado a la mejora de la alineación de objetivos mensuales con el comportamiento real del mercado.

1.3.2 Objetivos Específicos

- Analizar la evolución histórica de la demanda por categoría en el segmento de ropa infantil, identificando patrones de estacionalidad y tendencia.

- Incorporar variables asociadas a eventos comerciales, como campañas promocionales, feriados y temporadas clave, en el modelo de pronóstico.
- Evaluar el desempeño del modelo propuesto mediante métricas estadísticas de precisión y comparar su efectividad frente a métodos tradicionales.

1.4 Marco teórico

1.4.1 Pronóstico de la demanda en el sector minorista

Anticipar la demanda en las empresas del sector retail se constituye como una pieza crucial en las empresas para tomar decisiones más acertadas. Pronosticar el comportamiento del consumo de los clientes, a futuro, es el objetivo principal, para optimizar procesos de compra, abastecimiento, distribución y la gestión del inventario.

Haque et al. (2023) explican que estimar de manera precisa la demanda influye directamente en la rentabilidad del negocio, pues los errores en esta pueden derivar en sobrestock, desabastecimiento o costos relacionados con altas rebajas.

La alta variabilidad del mercado retail es una caracterización importante, derivada de factores como el comportamiento cambiante de los consumidores, eventos imprevistos, como conflictos políticos o económicos, desastres naturales, entre otros, y la estacionalidad. Adicionalmente, Özden Gür Ali et al. (2009) comentan la complejidad que añade al proceso la diversidad de productos que existe en el mercado y la necesidad de realizar estimaciones por categorías o unidades de mantenimiento de stock (SKUs).

En este sentido, los modelos de series temporales se han presentado como herramientas útiles para recoger patrones repetitivos en el tiempo y pronosticar la demanda futura. Entre estos, el SARIMAX, modelo estacional ARIMA con variables exógenas, se destaca por ser utilizado con éxito en combinación con los estudios exógenos de variables como la estacionalidad, promociones,

precio y eventos especiales, entre otros, permitiendo que las predicciones sean más precisas.

En el estudio de Alharbi and Csala (2022) se proporcionó un modelo para prever el consumo de electricidad en Arabia Saudita aplicando variables exógenas, como la meteorología, las tendencias macroeconómicas y el crecimiento poblacional, lo que resultó en una mejora significativa en la precisión de los pronósticos a largo plazo.

Esta metodología es especialmente beneficiosa en el sector minorista, donde la variable dependiente se ve afectada por factores externos, como campañas publicitarias, eventos, temporadas de ventas, condiciones climáticas, entre otros.

Khandelwal et al. (2023) demostraron que el modelo SARIMAX sigue siendo una herramienta fácil y efectiva, frente a modelos como el ARIMA, para predecir las ventas minoristas. Una variante de este modelo incluye efectos exógenos de días festivos y promociones. Tales opciones reducen en gran medida la imprecisión de las previsiones y, por lo tanto, permiten tomar decisiones eficazmente, planificar inventarios y optimizar la cadena de suministro.

1.4.2 Imputación de datos en series temporales

En el contexto del análisis de series temporales, la presencia de datos faltantes no es una novedad. Estos pueden atribuirse a diversas causas, entre ellas errores durante la extracción de los datos, imprecisiones en la ingesta de datos en los sistemas de información o interrupciones en los procesos de recopilación de datos (Khan and Hoque (2020)). La presencia de datos faltantes puede sesgar las conclusiones y comprometer la precisión de las predicciones. Es imprescindible que estos datos se imputen adecuadamente para garantizar la obtención de predicciones coherentes y evitar la aparición de sesgos en los análisis de estimación. Tal y como afirman Box et al. (2015), existen diversos métodos de imputación para tratar las series temporales. La selección del método depende de las características estructurales de los datos y de la naturaleza de los valores perdidos:

- La interpolación lineal es un proceso relativamente sencillo. En los casos en que falta un valor, se aproxima mediante los valores de los puntos vecinos, y el valor resultante se somete a una media ponderada. Sin embargo, este método supone que los valores cambian de forma suave y que la diferencia entre los valores de los puntos adyacentes no es significativa.
- Otra técnica que se puede emplear es la de los modelos estadísticos. En casos de mayor complejidad, se pueden utilizar modelos basados en los disponibles para predecir casos ocultos. Este procedimiento facilita la consideración de la autocorrelación temporal, lo que permite modelar los datos observados en la línea. Hyndman and Athanasopoulos (2018) postulan que esto es especialmente beneficioso cuando la serie temporal presenta dependencias complicadas entre observaciones.
- Además, se ha demostrado que el suavizado exponencial y medias móviles son útiles para la estimación de datos faltantes, especialmente en casos en los que los datos son estacionales o presentan una variación pequeña. La idoneidad de estos enfoques depende del supuesto de que los valores desconocidos dependen de un promedio de las observaciones más recientes.

La idea primaria, tras la imputación, es que la estructura no se vea afectada por los valores imputados, y que esta dependa del conjunto de datos y de lo que haya ocurrido con ellos en los períodos anteriores.

Box et al. (2015) hacen hincapié en que debe mantenerse la continuidad temporal en la imputación, ya que las desviaciones de la estructura pueden tener un impacto significativo en el orden del poder predictivo de los modelos.

1.4.3 Segmentación mediante agrupamiento: K-means

El análisis de agrupamiento, como explica Esri (2025), forma parte de los métodos de aprendizaje no supervisado, cuyo objetivo es clasificar un conjunto de datos en grupos, también

denominados conglomerados, de manera que los elementos de un mismo grupo sean más similares entre sí que los elementos pertenecientes a grupos diferentes. Las categorías de productos que comparten características comunes (por ejemplo, las ventas) pueden ser útiles para simplificar el modelado y refinar la precisión de las predicciones en la planificación de la demanda. Este método ha sido ampliamente estudiado por Lloyd (1982), quien propuso el algoritmo K-means que reduce al mínimo la suma de cuadrados dentro de los grupos.

Uno de los métodos más reconocidos en el análisis de conglomerados es el algoritmo K-means. Este enfoque tiene como objetivo dividir un conjunto de datos en k conglomerados con una variabilidad mínima dentro de cada conglomerado y una variabilidad máxima entre conglomerados. Formalmente, K-means busca minimizar la siguiente función objetivo:

$$\min_C \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (1.1)$$

donde:

- x_i es una observación (en este caso, una categoría de producto con sus características de ventas).
- μ_j es el centroide del grupo C_j .
- $\|x_i - \mu_j\|$ es la distancia euclidiana entre la observación x_i y el centroide del grupo.

1.4.4 Proceso estocástico

Se denomina serie temporal a una realización de un proceso estocástico indexado en el tiempo $\{Y_t : t \in \mathbb{Z}\}$. El análisis de un proceso estocástico en series temporales se enfoca en determinar la estructura y la dependencia temporal de los valores observados, lo que a su vez ayuda a la predicción futura. El concepto de proceso estocástico ha sido ampliamente tratado por

autores como Kotz and Johnson (2000) y Brockwell and Davis (2002), quienes discuten cómo las variables aleatorias se distribuyen a través del tiempo y cómo sus interrelaciones influyen en la modelización.

Estacionariedad

La **estacionariedad** es una propiedad crucial de los procesos estocásticos que debe cumplirse para que los modelos de pronóstico sean válidos. Un proceso es **estrictamente estacionario** si sus propiedades estadísticas se mantienen a lo largo del tiempo. Sin embargo, en la práctica, se trabaja con la **estacionariedad débil**, que implica tres condiciones principales:

- Tanto la **media** como la **varianza** permanecen constantes en el tiempo.
- La **covarianza** entre Y_t y Y_{t+h} depende solo de la diferencia h , no del tiempo t .

Los modelos de series temporales, tales como ARIMA o SARIMA, asumen que las series que se modelan cumplen con la propiedad de estacionariedad, lo que facilita la identificación de patrones temporales y su predicción futura. Según Box et al. (2015), la estacionariedad es fundamental para que los modelos autorregresivos y de media móvil sean efectivos.

Ruido blanco

El **ruido blanco**, como explica Meca and Belmonte (2018), es un proceso estocástico con media cero, varianza constante y sin autocorrelación entre sus observaciones. El cual se denota como

$$a_t \sim \mathcal{N}(0, \sigma^2)$$

Lo que significa que las observaciones no están correlacionadas y su valor futuro no puede predecirse basándose en su pasado.

El ruido blanco puede considerarse como el "residuo" o "error" del modelo, y su análisis es esencial para determinar si el modelo de predicción es eficiente. Brockwell and Davis (2002) afirmó que si un modelo tiene residuos que se comportan como ruido blanco, entonces puede considerarse que se ajusta bien, ya que no presenta patrones no modelados en los datos.

1.4.5 Fundamentos de las series temporales

Desde el punto de vista matemático, una serie de tiempo es una colección de registros ordenados cronológicamente, capturados en periodos de tiempo equidistantes para una o varias características asociadas a una unidad de análisis. Cuando la serie representa una sola característica, se denomina serie temporal univariante o escalar; si representa múltiples características observadas simultáneamente, se habla de una serie temporal multivariante o vectorial.

Formalmente, una serie temporal puede definirse como una secuencia de variables aleatorias:

$$\{Y_t\}_{t=1}^T \quad (1.2)$$

donde Y_t representa la observación en el tiempo t , y T denota la longitud total de la serie.

Serie temporal univariante

En este caso, cada observación Y_t es un valor escalar. Se tiene:

$$Y_t \in \mathbb{R}, t = 1, 2, \dots, T$$

Serie temporal multivariante

En este caso, cada observación Y_t es un vector de dimensión n , representando n variables observadas simultáneamente:

$$Y_t \in \mathbb{R}^n, t = 1, 2, \dots, T$$

El interés del estudio de las series temporales es la evaluación de los cambios a lo largo del tiempo; de esa manera, una herramienta clave en este estudio es la aplicación de la Transformada Discreta de Fourier (DFT) para identificar ciclos, tendencias estacionales y patrones repetitivos que no son evidentes.

Matemáticamente, la DTF expresada como:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i \frac{kn}{N}}, \quad k = 0, 1, \dots, N-1 \quad (1.3)$$

Donde:

- X_k representa la amplitud de la frecuencia k .
- x_n son los valores de la serie original.

Esta descomposición en el dominio de la frecuencia no solo permite identificar componentes periódicas dominantes, es decir patrones cíclicos o repetitivos como estacionalidades semanales, mensuales, etc. donde se puede identificar cuáles son los ciclos más importantes de la señal dependiendo de qué frecuencias tienen mayor peso o amplitud, sino que también facilita la eliminación de ruido, filtrando las estacionalidades que no representan importancia, y la mejora de la señal para propósitos de pronóstico. Con el análisis del dominio frecuencial se logra obtener una señal más limpia, obtener estacionalidades ocultas o reducir el ruido, filtrando las frecuencias no relevantes para el estudio. De esta manera, con patrones bien identificados, el modelo de pronóstico de demanda por categorías puede tener una predicción más acertada y confiable con mejores datos estructurados.

Tendencia

De acuerdo con Rojas (2022), la tendencia no es siempre lineal y representa el comportamiento o movimiento de la serie a largo plazo. Su notación varía según el tipo de serie:

- Para series temporales univariadas

$$T_t, t = 1, \dots, T \quad (1.4)$$

- Para series temporales múltiples

$$T_{i,k}, i = 1, \dots, n; k = 1, \dots, m \quad (1.5)$$

Ciclo

Lee (2020) explica que el componente cíclico son las oscilaciones de los datos a largo plazo. Estos ciclos no varían de la duración ni de los períodos fijos.

1.4.6 Modelos estocásticos univariantes

Los modelos AR, autoregresivos, predicen considerando los valores pasados de las series, mientras el modelo MA, media móvil, representa las series temporales como una combinación lineal de los errores pasados; los cuales dieron lugar al modelo ARMA, autorregresivo de Media Móvil, el cual predice utilizando los errores y sus propios valores del pasado, logrando una predicción y modelamiento de datos que cambian con el tiempo. No obstante, dichos enfoques se basan en la estructura temporal, quedando por fuera del modelo cualquier tipo de factor exógeno Hasan et al. (2022).

En la literatura, se encuentran como primer pronóstico clásico los modelos ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal ARIMA), SARIMAX (Seasonal ARIMA with eXogenous variables) o los modelos de suavizamiento exponencial, los cuales las

empresas suelen utilizar para pronosticar ventas dadas sus bases sólidas bajo un enfoque teórico y su habilidad para capturar patrones de tendencia y estacionalidad.

Si bien Chiba and Luna (2020) sugieren que el modelo ARIMA es efectivo frente a otros modelos más complejos, en especial cuando el modelo se ajusta adecuadamente a la estructura de la serie, su capacidad productiva se ve limitada frente a cambios abruptos por causas externas o eventos comerciales puntuales.

Procesos AR, MA y ARMA

Los métodos Autorregresivos (AR), Medias Móviles (MA) y ARMA (AutoRegresivo de Media Móvil) son elementos básicos en el análisis de series temporales estacionarias. Cada uno de estos modelos define un tipo diferente de dependencia de los valores de la serie a lo largo del tiempo y se utiliza para pronosticar los valores futuros de una serie utilizando valores pasados o el residuo de predicciones anteriores. Introducidos y explorados por Box et al. (2015) y Brockwell and Davis (2002).

Modelo Autorregresivo

Un modelo autoregresivo (AR) de orden p denota el comportamiento actual de una serie temporal Y_t como una combinación de sus p observaciones pasadas. La forma general del modelo AR (p) se expresa como sigue:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t \quad (1.6)$$

$$\phi(B)Y_t = a_t \quad (1.7)$$

Donde:

- $\phi_1, \phi_2, \dots, \phi_p$ son los coeficientes del modelo autoregresivo.

- $\{a_t\}$ es el proceso de perturbación aleatoria, el cual se asume con media igual a cero ($E[a_t] = 0$) y varianza fija en el tiempo ($Var[a_t] = \sigma^2$).

Para que el proceso autorregresivo de orden p sea estacionario, es necesario que las raíces del polinomio característico asociado $\phi(B) = 0$ estén fuera del círculo unitario en el plano complejo, es decir, que $|\lambda_i| > 1$ para todas las raíces λ_i . Esto garantiza que las perturbaciones se disipan con el tiempo y que la serie converge a una media constante (Box et al. (2015)).

Modelo MA (q)

El modelo **MA** de orden q (media móvil) describe, como explica Ibe (2014), el valor de la serie en el tiempo actual como una combinación lineal de sus errores pasados. El proceso MA(q) se expresa a continuación:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1.8)$$

donde $\theta_1, \theta_2, \dots, \theta_q$ son los coeficientes del modelo MA.

La expresión 1.8 puede reformularse utilizando el operador de retardos de la siguiente manera:

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (1.9)$$

$$Y_t = \theta(B) a_t$$

Para que el modelo MA sea invertible (lo cual es análogo a la estacionariedad en modelos AR), se requiere que las raíces del polinomio $\theta(B) = 0$ estén fuera del círculo unitario. Esto asegura una representación única en términos de un proceso AR infinito Brockwell and Davis (2002).

Modelos lineales autoregresivos de medias móviles (ARMA (p, q))

El modelo **ARMA**(AutoRegressive Moving Average), tal como lo expone Rodó (2019), combina las características de los modelos AR y MA. Es uno de los modelos más utilizados en el campo de las series temporales estacionarias y se utiliza cuando la serie presenta tanto dependencia temporal (AR) como choques aleatorios (MA). El modelo ARMA (p, q) se expresa como:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (1.10)$$

$$\phi(B)Y_t = \theta(B)a_t$$

(1.11)

Donde:

- Los parámetros $\phi_1, \phi_2, \dots, \phi_p$ corresponden a los coeficientes del modelo autorregresivo del modelo.
- Los coeficientes $\theta_1, \theta_2, \dots, \theta_q$ representan los parámetros del término de media móvil.
- a_t denota error aleatorio asociado al modelo.

La estacionariedad del modelo ARMA está garantizada si las raíces de $\phi(B)$ se encuentran fuera del círculo unitario, mientras que su invertibilidad requiere que las raíces $\theta(B)$ también cumplan esa condición.

Función de autocorrelación simple y parcial

La **función de autocorrelación**, ACF, por sus siglas en inglés, evalúa la correlación entre los valores de la serie temporal Y_t y sus valores con retardos Y_{t-h} , explica Smith (2024) La ACF ayuda a identificar patrones de dependencia y proporciona información crucial para determinar el orden de los modelos AR y MA. Se calcula como:

$$ACF(h) = \frac{\text{Cov}(Y_t, Y_{t+h})}{\sqrt{\text{Var}(Y_t) \cdot \text{Var}(Y_{t+h})}} \quad (1.12)$$

donde h es el desfase temporal. Además, la ACF es útil para identificar si existe correlación entre los valores pasados y los futuros de la serie temporal.

La **función de autocorrelación parcial (PACF)**, por otro lado, mide la correlación entre Y_t y Y_{t-h} , aislando el efecto de los rezagos intermedios. La PACF es particularmente idónea para determinar el parámetro de orden de los modelos AR, ya que muestra cuántos rezagos de la serie son necesarios para explicar la autocorrelación.

El análisis gráfico de la ACF y PACF permite identificar el orden de los modelos AR y MA. Según Brockwell and Davis (2002), estos gráficos son herramientas fundamentales para la selección de los parámetros óptimos en los modelos ARMA.

Diferenciación e integración

La diferenciación es un paso clave cuando se trabaja con series temporales no estacionarias, y su propósito es hacer que la serie temporal sea **estacionaria**. Un proceso con diferenciación de grado d se aplica de la siguiente manera:

$$\Delta^d Y_t = (1 - B)^d Y_t \quad (1.13)$$

Donde B representa el operador de retardos, de tal manera que $BY_t = Y_{t-1}$ y $\Delta^d Y_t$ es el

valor de la serie después de aplicar la diferenciación d -veces.

Cuando una serie no es estacionaria, se aplica la diferenciación hasta que la serie se vuelva estacionaria. Esto puede implicar uno o más pasos de diferenciación, como lo explica Box et al. (2015) en su trabajo sobre el modelo ARIMA.

Modelo ARIMA

El modelo **ARIMA**, AutoRegressive Integrated Moving Average en inglés, es una generalización de los modelos AR y MA, utilizado con fines de modelado de series temporales no estacionarias. ARIMA es un modelo que integra componentes autoregresivos (AR), de media móvil (MA) y un componente de diferenciación para alcanzar la estacionariedad en la serie temporal.

Según Holm (2024), el modelo ARIMA es una herramienta estadística poderosa para el modelado y pronóstico de series temporales para conjuntos de observaciones que pueden ser mapeados a series de tiempo fijas, el cual está basado en tres componentes esenciales:

- Autoregresivo (AR): Componente fundamental, debido a la consideración de valores pasados de la serie temporal para la predicción de valores futuros.
- Media móvil (MA): Utiliza errores pasados, también conocidos como residuos, del modelo para predecir errores futuros, de ayuda para refinar las predicciones. Morpher (2024) explica que la MA permite la incorporación de la dependencia entre una observación y un error residual de un modelo de media móvil. Caracterizada por utilizar errores de pronóstico pasados para influir en predicciones futuras, ayudando a suavizar el ruido en los datos.
- Integrado (I): Implica la aplicación de la diferenciación a la serie temporal para lograr estacionariedad, es decir, corregir la serie de efectos de tendencias y patrones estacionales que podrían afectar la precisión del pronóstico.

Estructura del modelo ARIMA(p,d,q)

El modelo autorregresivo integrado de media móvil se puede expresar como una combinación de la parte autorregresiva, la media móvil y la diferenciación. La fórmula general es:

$$\phi(B)(1-B)^d Y_t = \theta(B)a_t \quad (1.14)$$

Donde:

- Y_t : Serie original (o transformada)
- a_t : Término de error (ruido blanco)
- B : Operador de rezago ($BY_t = Y_{t-1}$)
- $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$: Componente autorregresiva (AR)
- $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$: Componente de media móvil (MA)
- $(1-B)^d$: Operador diferenciador.

Modelo SARIMA

El modelo SARIMA amplía el enfoque del ARIMA al integrar términos de estacionalidad a los modelos. Según Box et al. (2015), la estacionalidad puede ser capturada mediante la incorporación de componentes estacionales específicos en el modelo.

El modelo se representa como **SARIMA(p,d,q)(P,D,Q)[s]**, en donde:

- **p, d, q** corresponden a los parámetros del componente ARIMA (autoregresivo, diferenciación e integración y media móvil).
- **P, D, Q** son los parámetros estacionales del modelo.

- s representa el periodo estacional.

La fórmula general para el modelo SARIMA es:

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D Z_t = \Theta(B^s)\theta(B)a_t \quad (1.15)$$

Donde, además de los términos anteriores implementados en el modelo ARIMA:

- $\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$: AR estacional de orden P
- $\Theta(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{Qs}$: MA estacional de orden Q
- $(1 - B^s)^D$: Diferenciación estacional de orden D

El componente estacional de un modelo SARIMA está conformado de tres partes:

1. **Estacionalidad en la parte AR (autorregresiva)**: Captura el efecto de los valores pasados que ocurren en el mismo período estacional.
2. **Estacionalidad en la parte MA (media móvil)**: En cuanto a este componente, se basa en modelar el efecto de las perturbaciones en la serie.
3. **Diferenciación estacional**: Este componente elimina la estacionalidad de la serie mediante la diferenciación.

Como Hyndman and Athanasopoulos (2018) afirma, estos modelos permiten ajustar los efectos estacionales, mejorando significativamente las predicciones.

Modelo SARIMAX

El modelo autorregresivo integrado de media móvil, comenzó a ser utilizado desde finales de las décadas de 1980 y 1990, cuando se comprendió que para mejorar la predicción era fundamental incluir información externa, el cual es una evolución del modelo SARIMA. Según Hyndman and Athanasopoulos (2018), esto permite tener en cuenta factores externos que no están reflejados en la

serie temporal original, pero que tienen un impacto significativo en la demanda o el comportamiento de la serie.

El modelo SARIMAX se expresa como:

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^DY_t = \Theta(B^s)\theta(B)a_t + \beta X_t \quad (1.16)$$

Donde X_t representa las variables exógenas y β es el vector de coeficientes que mide el impacto de cada regresor sobre la variable dependiente Y_t .

1.4.7 Causalidad y test de Granger

El análisis de causalidad es una extensión de los modelos de series temporales que busca determinar si una variable X ayuda a predecir el comportamiento futuro de otra variable Y . La prueba estadística conocida como el **test de causalidad de Granger** se utiliza con el fin de determinar si una serie temporal X es útil para pronosticar Y , basándose en el principio de que si X causa a Y en el sentido de Granger, entonces la inclusión de X en un modelo de pronóstico debería mejorar la predicción de Y .

Según Granger (1969), el test de causalidad de Granger se fundamenta en la hipótesis de que una serie de tiempo X puede proporcionar información adicional sobre la predicción de Y que no está contenida en las observaciones pasadas de Y únicamente.

Planteamiento formal

La hipótesis nula del test de causalidad de Granger establece que X no causa a Y , es decir, que los rezagos de X no mejoran la predicción de Y . Formalmente, la hipótesis nula se expresa como:

$H_0 : X$ no causa a Y en el sentido de Granger

El test de Granger consiste en comparar dos modelos:

1. Un modelo de regresión de Y_t sobre sus propios rezagos, es decir, un modelo autoregresivo univariado de Y .
2. Un modelo de regresión de Y_t sobre sus propios rezagos y los rezagos de X_t , que captura la influencia de X en la dinámica de Y .

El estadístico de prueba para comparar ambos modelos es el estadístico F , que se calcula de la siguiente manera:

$$F = \frac{(RSS_{restricted} - RSS_{unrestricted})/k}{RSS_{unrestricted}/(n - k - m)} \quad (1.17)$$

Donde

- $RSS_{restricted}$ es la suma de los errores residuales elevados al cuadrado del modelo restringido, el cual solo incluye los rezagos de Y_t .
- $RSS_{unrestricted}$ es la suma de los cuadrados de los residuos del modelo no restringido, considerando los rezagos de Y_t y X_t .
- k es el conteo de restricciones, es decir, el número de rezagos de X_t .
- n es la cantidad de observaciones
- m es el total de variables en el modelo

En caso de que el estadístico F sea significativamente grande, se descarta la hipótesis nula, lo que sugiere que X causa a Y en el sentido de Granger.

1.4.8 Diagnóstico de residuos

Una vez que se ha ajustado el modelo, es crucial realizar un diagnóstico de los residuos $\hat{\varepsilon}_t$ para verificar si se comportan como una serie aleatoria sin patrones. Esto es fundamental porque un modelo que no ha capturado completamente las dependencias en los datos presentará residuos que tienen una estructura temporal o autocorrelación no modelada. El diagnóstico de residuos incluye varias pruebas que nos permiten evaluar si el modelo ha ajustado adecuadamente la evolución temporal.

ACF de residuos

La ACF mide la correlación entre los residuos $\hat{\varepsilon}_t$ y sus rezagos $\hat{\varepsilon}_{t+h}$. Si el modelo es adecuado, los residuos deben ser **no correlacionados** en el tiempo, es decir, deben comportarse como ruido blanco. Si la ACF muestra picos significativos, esto sugiere que los residuos contienen información no modelada, lo que indica que el modelo podría ser mejorado.

La ACF se calcula como:

$$ACF(h) = \frac{\text{Cov}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+h})}{\sqrt{\text{Var}(\hat{\varepsilon}_t) \cdot \text{Var}(\hat{\varepsilon}_{t+h})}} \quad (1.18)$$

Donde h es el desfase temporal y $\hat{\varepsilon}_t$ son los residuos estimados. Si la ACF exhibe valores casi nulos para todos los rezagos, se puede concluir que los residuos no están autocorrelacionados de manera significativa.

Test de Ljung-Box

El **test de Ljung-Box**, como explica Bobbitt (2020), se utiliza para verificar si los residuos se encuentran autocorrelacionados a varios rezagos. La hipótesis nula de esta prueba postula que los residuos no están autocorrelacionados, lo que indica que el modelo capturó de manera adecuada la

estructura de la serie. El estadístico de la prueba se calcula como:

$$Q = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}_h^2}{n-h} \quad (1.19)$$

donde:

- n corresponde al número de valores observados,
- m se refiere al número de rezagos considerados en el test,
- $\hat{\rho}_h$ es la estimación de la autocorrelación en el rezago h .

Si el valor Q alcanza un valor significativamente grande, indica evidencia suficiente para considerar que existe autocorrelación en los residuos, lo que implica que el modelo debe ser ajustado.

Q-Q plot

El **Q-Q plot** (Quantile-Quantile plot) es un método visual que permite evaluar si la distribución de los residuos se asemeja a una distribución gaussiana estándar (Waples (2024)). Si la distribución es normal, los puntos en el gráfico deberían disponerse sobre una línea recta. Cualquier desviación sugiere que los residuos no son normalmente distribuidos, lo que podría implicar deficiencia en el modelo al ajustarse a los datos.

El Q-Q plot es una prueba visual que ayuda a detectar si los residuos tienen colas pesadas o si la serie temporal tiene no linealidades no modeladas.

Normalidad de los residuos

El supuesto de **normalidad en los residuos** es relevante en numerosos modelos de series temporales, dado que la inferencia estadística sobre los parámetros y los intervalos de confianza dependen de esta condición. Si los residuos no son normales, puede afectar la validez de las pruebas

de significancia y los intervalos de confianza. Se pueden utilizar pruebas como el **test de Shapiro-Wilk** o el **test de Anderson-Darling** para verificar el cumplimiento de esta condición.

El test de **Shapiro-Wilk** y el test de **Jarque-Bera** son ampliamente utilizados para evaluar la normalidad de los residuos. La hipótesis nula es que los residuos siguen una distribución normal. Si el valor p resultante es inferior al umbral estadístico (por lo general $\alpha = 0.05$), existe evidencia suficiente que respalda que los residuos no tienen una distribución normal.

1.4.9 Supuestos

Los modelos de series temporales están basados en algunos supuestos que deben cumplirse si se desean realizar inferencias válidas y pronósticos. Los supuestos serán cruciales en algunos aspectos, incluida la inferencia de los parámetros, la validez del modelo y la interpretación. Algunos de los supuestos clave de los diferentes modelos de pronóstico de series temporales que siempre deben validarse incluyen:

Linealidad

Uno de los supuestos más comunes en los modelos AR, MA, ARMA, SARIMA y SARIMAX es que exista una relación **lineal** entre las variables. La linealidad implica que la relación entre las observaciones actuales y sus rezagos es de naturaleza lineal. En otras palabras, el valor futuro de la serie depende de manera proporcional de los valores pasados de la serie en el caso de los modelos AR o de los errores en el caso de los modelos MA.

El supuesto es crítico para los modelos AR y ARMA que están contruidos basándose en la suposición de que los procesos de series temporales pueden modelarse formando combinaciones lineales de valores rezagados y errores. Según Box et al. (2015) es la propiedad fundamental de los modelos clásicos de series temporales.

Estacionariedad

Rico (2020) explica que la **estacionariedad** implica que las propiedades estadísticas de la serie, como la media, la varianza y la autocovarianza, sean constantes a lo largo del tiempo. Si la serie no es estacionaria, debe transformarse realizando una diferenciación hasta que se alcance la estacionariedad.

Esta hipótesis asegura que los patrones observados en el pasado puedan replicarse en el futuro, lo que hace posible realizar predicciones basadas en datos históricos. Sin embargo, si los datos no son estacionarios y no se transforman adecuadamente, los modelos no serán capaces de capturar correctamente las tendencias y ciclos en la serie temporal, afectando la precisión de las predicciones.

Independencia y homocedasticidad de errores

Otro supuesto importante es que los errores (ε_t) sean **independientes y homocedásticos**. La independencia del error significa que el valor del error en cualquier período de tiempo no debe estar correlacionado con otros errores. Si los errores tienen una correlación positiva o negativa, esto se conoce como autocorrelación en el residuo, lo que sugiere que el modelo no capturó la estructura temporal de la serie, lo que podría dar como resultado una precisión deficiente de la predicción.

La **homocedasticidad** del error implica que los errores tienen la misma variabilidad a lo largo del tiempo. Si los errores tienen diferentes varianzas en diversos períodos, denominada como heterocedasticidad, significa que el modelo no es adecuado para realizar inferencias precisas y que la predicción sería insesgada.

Normalidad de los residuos

En algunos casos, en particular en el contexto de inferencia sobre los parámetros del modelo o la elaboración de intervalos de confianza, se asume que los residuales se distribuyen normalmente. La distribución normal de los residuos es un supuesto común en muchos procedimientos de inferencia, incluso en aquellos que utilizan métodos de máxima verosimilitud.

Si los residuos no siguen una distribución normal, esto puede afectar la validez de los intervalos de confianza y las pruebas de hipótesis sobre los parámetros del modelo. Sin embargo, este supuesto no es siempre esencial para la predicción, ya que los métodos de predicción suelen ser robustos ante ciertas violaciones de la normalidad, como en el caso de los modelos ARIMA.

Independencia de las variables exógenas

Cuando se utilizan factores exógenos en modelos SARIMAX, es crucial suponer que las variables exógenas X_t son independientes de los errores del modelo. En otras palabras, las variables externas utilizadas en el modelo no deben estar correlacionadas con los términos de error (ε_t). Si las variables exógenas están correlacionadas con los residuos, el modelo podría realizar estimaciones sesgadas, lo que afectaría la precisión de los pronósticos.

1.4.10 Paradoja de Simpson

Vellaisamy (2014) explica que la paradoja de Simpson, SP, se da al incorporar una tercera variable, por lo cual los resultados pueden ser, aparentemente, contradictorios.

En el contexto de la modelación de la demanda por categoría en retail, esta paradoja se manifiesta al observar que, aunque la tendencia que se genera de la demanda total es clara, las tendencias individuales, es decir, la demanda clasificada por categoría, no coinciden con la agrupada.

Esta puede ser caracterizada matemáticamente en términos de probabilidades condicionales

y la presencia de una variable de confusión.

Formalmente, como indica Rao and Meer (2016), si denotamos Y como la variable de demanda, X como la variable indicadora de la presencia de un evento comercial, y Z como la variable que representa la categoría del producto, puede ocurrir que:

$$\mathbb{E}[Y \mid X = 1, Z = z] < \mathbb{E}[Y \mid X = 0, Z = z], \quad \forall z \in \mathcal{Z}, \quad (1.20)$$

mientras que en la demanda agregada,

$$\mathbb{E}[Y \mid X = 1] > \mathbb{E}[Y \mid X = 0], \quad (1.21)$$

En conclusión, si el análisis es basado en el resultado de la demanda agrupada, la interpretación puede ser errónea, debido a que la demanda por categoría puede tener un comportamiento diferente, lo cual puede tener un impacto en la distribución, especialmente durante eventos comerciales.

1.4.11 Métricas de evaluación

Es crucial evaluar la capacidad de los modelos de pronóstico para garantizar que las predicciones hechas con base en este sean útiles, consistentes, fiables para la toma de decisiones. En analíticas de series temporales, hay métricas comunes que proporcionan información sobre la precisión de un modelo en pronósticos y su habilidad de ajustarse a los datos históricos.

A continuación, van algunas de las métricas más importantes en el marco de los modelos de pronóstico en series temporales.

Métricas de Error Absoluto

Error Absoluto Medio (MAE)

El MAE cuantifica la magnitud promedio de las diferencias absolutas entre los valores predichos ($\hat{Y}_t$) y los observados (Y_t). Su expresión se detalla a continuación:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (1.22)$$

donde:

- Y_t corresponde al valor observado al tiempo t ,
- $\hat{Y}_t$ es el valor predicho en el mismo período,
- n es el total de observaciones.

El MAE no es difícil de interpretar, ya que da una medida directa del error promedio. Sin embargo, se debe considerar que no penaliza los grandes errores de manera significativa, ya que toma el valor absoluto de las diferencias.

Raíz del Error Cuadrático Medio (RMSE)

La Raíz del Error Cuadrático Medio es una métrica que otorga mayor peso a los errores grandes, ya que calcula la raíz cuadrada del promedio de las diferencias entre los valores reales y pronosticados elevadas al cuadrado. La fórmula para el RMSE es:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (1.23)$$

donde:

- Y_t corresponde que se observa en el tiempo t ,
- $\hat{Y}_t$ es el valor pronosticado en el mismo período t ,
- n es la cantidad de observaciones.

Esta métrica es útil cuando se desea penalizar los grandes errores, y es especialmente valiosa cuando las grandes desviaciones de los valores predichos son más problemáticas. El RMSE es sensible a valores atípicos, lo que puede ser una ventaja o desventaja dependiendo del contexto del problema.

Error Cuadrático Medio (MSE)

El Error Cuadrático Medio es similar al RMSE, pero no toma la raíz cuadrada, lo que hace que la penalización de los grandes errores sea aún mayor. Sea Y_t las observaciones en el tiempo t , $\hat{Y}_t$ los valores predichos en el mismo instante de tiempo y n el número de observaciones. La fórmula para el MSE se define como:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (1.24)$$

Métricas Relativas

Error Porcentual Absoluto Medio (MAPE)

El MAPE cuantifica el error relativo expresado en porcentaje, lo que le convierte en una herramienta para comparar el desempeño de los modelos a través de diferentes escalas o unidades.

La expresión matemática correspondiente es la siguiente:

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (1.25)$$

donde:

- Y_t denota el valor real observado al tiempo t ,
- $\hat{Y}_t$ representa la estimación que generó el modelo para ese mismo período,
- n corresponde al total de observaciones evaluadas.

El MAPE es una de las métricas más populares debido a su facilidad de interpretación, ya

que mide el error como un porcentaje. Sin embargo, su limitación consiste en que no se puede usar cuando $Y_t = 0$, ya que la división por cero es indefinida. Además, puede ser muy sensible a los valores pequeños de Y_t , lo que puede distorsionar los resultados.

CAPÍTULO 2

2. METODOLOGÍA



Figura 2.1.
Diagrama de flujo de la metodología aplicada

El desarrollo metodológico del proyecto se organizó en etapas secuenciales orientadas al análisis y modelado de la demanda semanal por categoría, con base en datos históricos y eventos comerciales. Se aplicaron técnicas de segmentación, tratamiento de series temporales y evaluación estadística con el fin de identificar el modelo de pronóstico más óptimo para cada grupo representativo de los productos (Ver Figura 2.1).

2.1 Clasificación de categorías por nivel de ventas

El proceso metodológico del presente trabajo se distribuyó en fases consecutivas destinadas al análisis y modelización de la demanda a partir de la información histórica y de los eventos comerciales. El modelo de previsión adecuado para cada grupo representativo de productos se determinó basándose en algoritmos de segmentación, análisis de series temporales y consideraciones estadísticas.

Las categorías del segmento de moda se clasificaron primero en función de las ventas acumuladas de los últimos tres años como paso inicial. Para ello, se consideró el uso del enfoque de agrupamiento K-means, que se clasifica como un método de aprendizaje no supervisado para el agrupamiento, en el que un conjunto de observaciones se agrupan en clústeres basados en medidas de similitud. En este caso concreto, se trata de condiciones basadas en las ventas totales por categoría. Se identificó el número adecuado de clústeres (k) observando la suma de cuadrados dentro del grupo (WSS) y el valor de k que reducían sustancialmente la variabilidad dentro de cada grupo. Este análisis se apoyó en una visualización de la WSS cambiante a medida que aumentaba el número de clústeres, de modo que se pudiera encontrar un punto de mejora. Tras un estudio exhaustivo de la información disponible, se determinó que el número óptimo era 4. Esta conclusión también se ve confirmada por el hallazgo de cuatro tipos distintos de comportamiento de ventas (véase Figura 2.2).

A continuación, de cada grupo se seleccionó una sola categoría, favoreciendo a aquellas con el mayor volumen de ventas y el comportamiento temporal más estable (Figura 2.3). Estas categorías se tomaron como referencia para la posterior construcción del modelo de previsión.

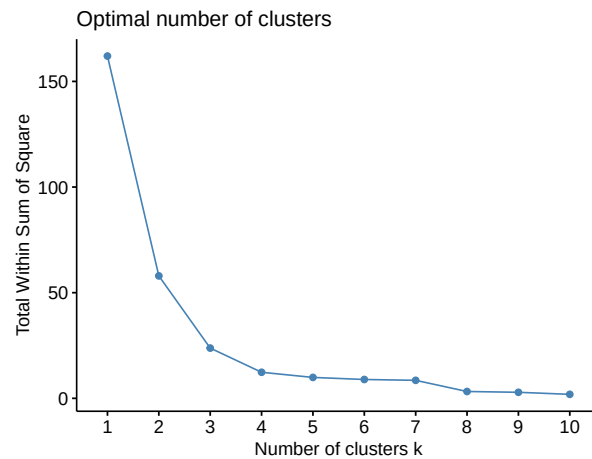


Figura 2.2.
Selección del número de clústers óptimo

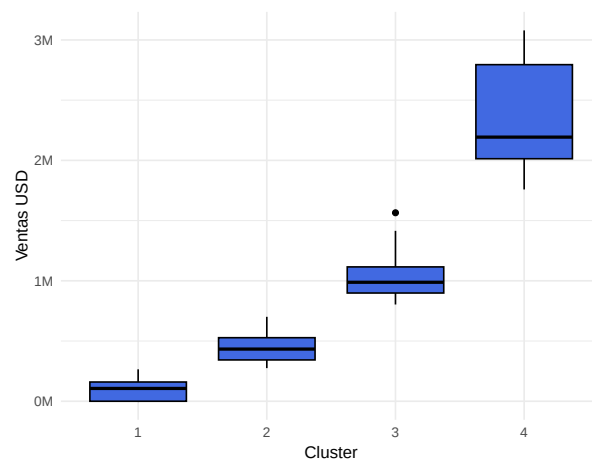


Figura 2.3.
Distribución de Ventas por clúster

2.2 Recopilación y estructuración de datos históricos

Después de seleccionar las categorías, se extrajeron los datos históricos de ventas por semanas desde 2018 hasta la fecha más reciente disponible. Se organizaron en marcos de datos con una observación por cada semana y categoría. Se comprobó la calidad de los datos y se identificaron posibles datos aberrantes.

De esta manera se procedió a trabajar con el conjunto de datos sales.xlsx, que comprende

1.544 registros, donde los importes de ventas en dólares por mes y semana, y seis columnas principales.

De esta manera, se procedió a trabajar con la base de ventas.xlsx que cuenta con 1.544 registros, que incluyen los montos en dólares de las ventas por semana de cada mes, y 6 columnas principales.

En esta base de datos se incluyeron las siguientes variables:

- Año: representa el año natural en el que se registró la venta.
- Mes: representa el período mensual correspondiente de la venta.
- Semana: indica el período semanal asociado a la venta.
- Categoría: clasificación de 4 grupos dentro del segmento de moda.
- Evento Comercial: Variables binarias que indica la ocurrencia del evento comercial.
- Monto de Venta: importe monetario en dólares de las ventas registradas.

La variable Evento comercial se desarrolló mediante la intersección de los conjuntos de datos de ventas con el calendario de eventos comerciales, como se describe más adelante.

2.3 Tratamiento de datos faltantes e imputación

Debido a las características específicas del año 2020, influenciado significativamente por la crisis sanitaria global y el posible cierre de negocios, hubo períodos con ventas anormalmente bajas o sin ventas en absoluto. El análisis de estos casos se realizó de forma individualizada y utilizando métodos de imputación adecuados para garantizar la continuidad y la validez de las series de datos para el análisis de series temporales. Tras este proceso de imputación, se obtuvieron 1.560 registros, sobre los cuales se desarrolló el presente proyecto integrador.

2.4 Integración de eventos comerciales

Al mismo tiempo, se construyó una estructura bien detallada de eventos comerciales relacionados con la empresa, que incluye, por ejemplo, sus campañas y temporadas clave. Este procedimiento se aplicó a las fechas de los eventos y al historial de series de ventas de cada categoría, consolidando esas fechas y creando así variables indicadoras. En particular, las variables indicadoras son iguales a 1 si existió un evento comercial en el período de tiempo y 0 en caso contrario. Finalmente, las variables se incorporaron al modelo como predictores exógenos.

2.5 Análisis y modelado

Después de depurar la información, se realizó un análisis exploratorio de la serie, en el que se reconocieron patrones de estacionalidad, tendencias y posibles problemas de inventario. En este punto, se compararon diferentes métodos de predicción, tanto tradicionales (ARIMA y SARIMA) como aquellos que permiten la incorporación explícita de variables externas (SARIMAX). Se analizó el desempeño del modelo mediante el procedimiento convencional según el error absoluto medio, el error cuadrático medio, el error porcentual absoluto medio, etc. A continuación, se eligió el modelo con el mejor balance entre precisión y simplicidad.

2.5.1 Análisis en Dominio de la Frecuencia

El análisis espectral se realizó mediante la transformada de Fourier de la serie temporal para generar el espectro de frecuencia, lo que permitió descomponer la serie temporal en sus componentes cíclicos básicos. Esto permitió reconocer patrones de ventas que no eran evidentes a simple vista para encontrar la estacionalidad semanal o mensual, que puede estar correlacionada con eventos comerciales.

2.6 Evaluación de resultados y consideraciones finales

El modelo elegido se implementó en las categorías de referencia y se probó su capacidad para generar previsiones que se ajustaran a los datos históricos, así como el impacto de los acontecimientos comerciales. Por último, se debatió la importancia comercial de este enfoque en términos de su posible extensión a otras categorías y de su aplicación práctica en la gestión comercial y la toma de decisiones estratégicas.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

En este capítulo se exhiben los hallazgos obtenidos a partir del análisis de las categorías seleccionadas mediante el método de agrupamiento K-means, dentro del segmento de ropa infantil. En primera instancia, se lleva a cabo un análisis descriptivo de las series históricas de ventas, identificando la presencia de datos faltantes y describiendo su tratamiento mediante técnicas de imputación para garantizar la consistencia de la información. Posteriormente, se examinan los patrones de tendencia y estacionalidad que caracterizan a cada categoría, así como el impacto de eventos comerciales relevantes, como por ejemplo, campañas promocionales, en la dinámica de la demanda. De manera subsecuente, se exponen los resultados de los modelos de pronóstico aplicados, comparando su desempeño mediante métricas estadísticas (MAE, RMSE y MAPE) y destacando la mejora alcanzada con la incorporación de variables exógenas en la predicción, con el fin de aportar una herramienta práctica para la planificación de inventarios y la asignación de objetivos de ventas. Finalmente, se incluye un análisis de viabilidad económica, en el cual se describen los costos asociados al desarrollo de la solución y se evalúa su factibilidad tecnológica y financiera para la empresa.

3.1 Imputación de datos faltantes

Las series originales de ventas presentaban un 18% aproximado de valores faltantes, distribuidos en categorías y períodos distintos. El método de Holt y Winters fue aplicado para así garantizar la continuidad temporal, puesto que resultó apropiado en este contexto al preservar la

tendencia, estacionalidad y nivel.

Las series originales y las imputadas se comparan en las Figuras 3.1a a 3.1d. Después de la corrección, se aprecia que las series mantienen los picos estacionales característicos del mercado, tales como campañas escolares y navideñas, al tiempo que eliminan los tramos con caídas bruscas generadas por los datos faltantes.

Las categorías 1 y 3 presentan patrones estacionales muy bien definidos y se disminuyeron rupturas abruptas sin cambiar su crecimiento estructural. En la categoría 2, la imputación restituyó la continuidad de las series, lo que permitió observar con claridad su tendencia. Por último, la categoría 4 mantiene su estabilidad al conservar las oscilaciones típicas de alta estacionalidad.

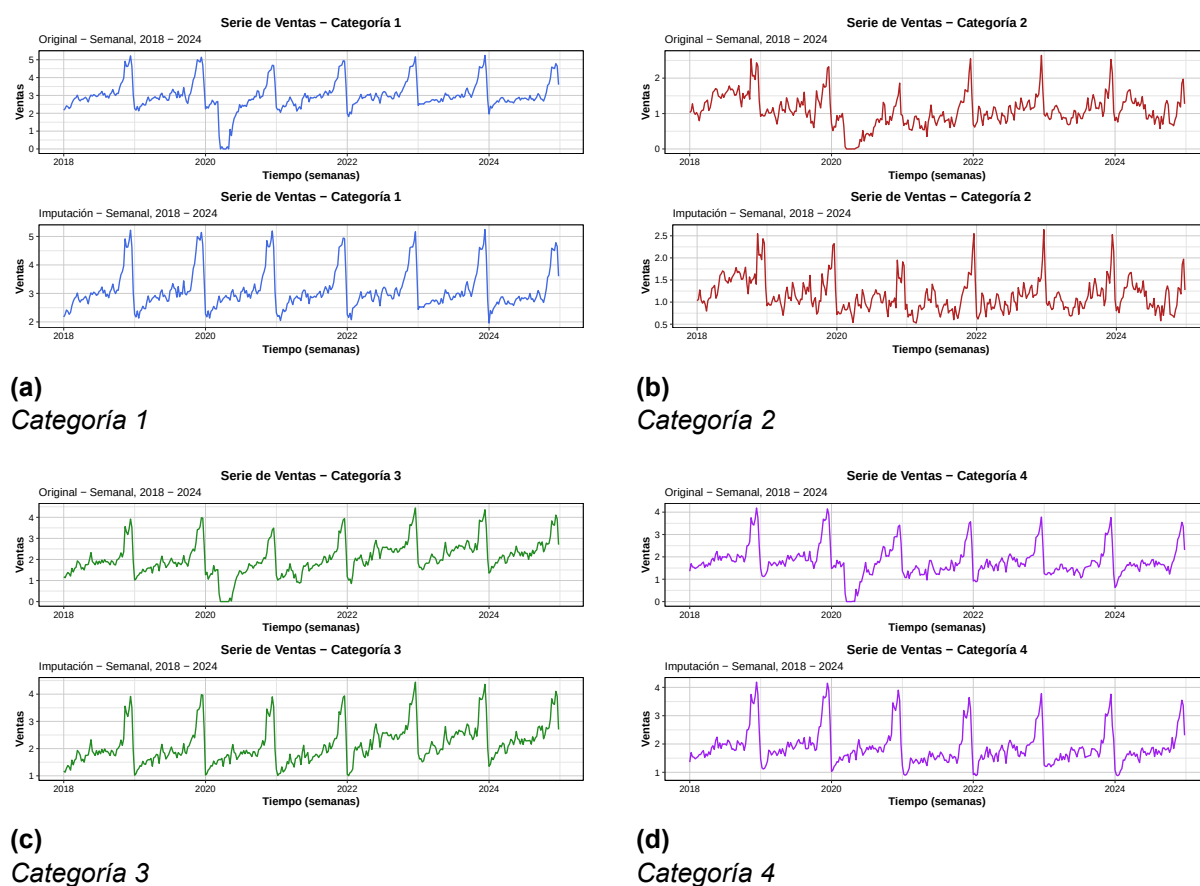


Figura 3.1.
Series de ventas originales e imputadas con Holt–Winters, por categoría (2018–2024).

3.2 Participación por Categoría

Con el objetivo de analizar la composición del portafolio, se calcularon las participaciones relativas de cada categoría respecto al total anual de ventas. En la Figura 3.2 se aprecia que el segmento primordial corresponde a la Categoría 1, cuyo peso constante se mantiene en torno al 37-40% durante todo el periodo. La Categoría 2 exhibe un decrecimiento paulatino de importancia. En contraste, decreció desde un 17.1% en 2018 hasta un 13.5% en 2024. La Categoría 3 aumenta desde 22.7% hasta 29% de forma sostenida, en tanto que la Categoría 4 fluctúa entre 21% y 25% manteniéndose relativamente constante.

Estos resultados ponen en evidencia que, además de las fluctuaciones en ventas totales, el comportamiento de la demanda experimentó cambios estructurales. El crecimiento en la Categoría 3 sugiere una inclinación mayor del mercado por este segmento en particular; mostró un crecimiento sostenido, lo que sugiere que en el futuro podría ganar mayor peso. La disminución observada en la Categoría 2 señala un procedimiento de reemplazo o merma de la competitividad relativa.

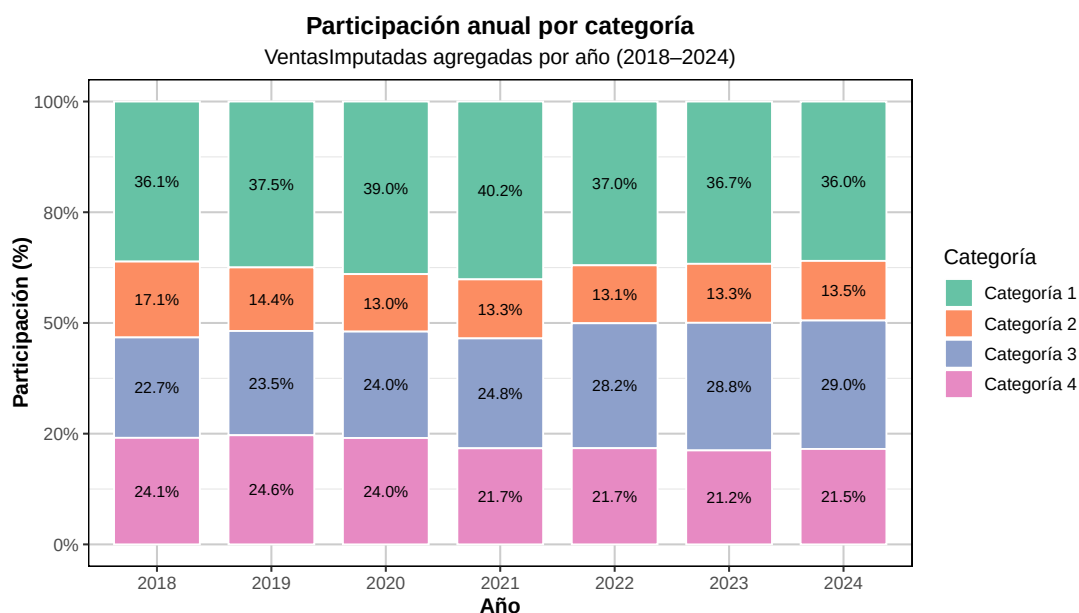


Figura 3.2.
Participación anual de ventas por categoría (2018–2024).

3.3 Variaciones interanuales

El análisis de las variaciones interanuales exhibió comportamientos heterogéneos en las categorías. En la base de datos sin imputar, ciertos períodos exhibían valores desmesurados. Esto era a causa de la existencia de registros ausentes, incluso superiores al 30.000% en la Categoría 4, de acuerdo con lo presentado en el Apéndice A.1. Tras la aplicación del proceso de imputación, fueron corregidas tales distorsiones y se consiguió una visión más coherente sobre la dinámica de las ventas (Véase Apéndice A.5). En la Categoría 1 se evidenciaron cambios moderados con variaciones anuales esperadas. La Categoría 2, como se observa en el Apéndice A.2, en contraste, evidenció disminuciones acumuladas en los primeros años (2018 - 2021), presentando caídas interanuales que lograron sobrepasar el -10%, tal como se evidencia en el Apéndice A.6. La Categoría 3 exhibió un descenso inicial hasta 2020, como se detalla en el Apéndice A.3. Desde el año 2022, ha logrado agregar aumentos interanuales cercanos al 8-10% en los ejercicios recientes, según lo mostrado en el Apéndice A.7. Finalmente, la Categoría 4, luego del proceso de imputación de datos, evidenció una dinámica más constante, como se verifica en el Apéndice A.8, con alteraciones interanuales cerca del 3-4%. Sin embargo, fluctuó con variaciones más abruptas en la versión sin ajustar, tal y como se infiere en el Apéndice A.4.

En la Figura 3.3 se ilustran las ventas anuales consolidadas por categoría que confirman además las tendencias descritas en las variaciones interanuales. La Categoría 2 manifiesta un decrecimiento continuo desde 2018, en tanto que la Categoría 1 conserva niveles sin cambios durante el período. La Categoría 3 exhibe una mejora notoria desde el año 2022. La Categoría 4, por su parte, permanece relativamente constante, presentando oscilaciones moderadas.

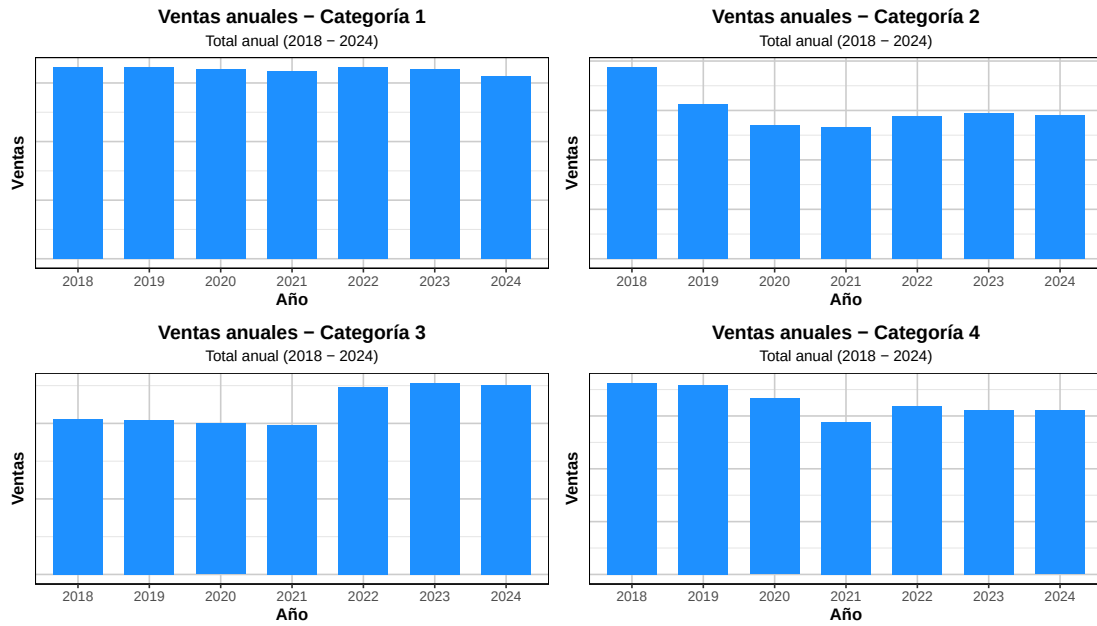


Figura 3.3.
Ventas anuales agregadas por categoría (2018–2024).

De manera complementaria, en la Figura 3.4, se evidencia la evolución relativa de las ventas bajo un esquema de índices normalizados (base 100). Formalmente, sea $i \in \{1, \dots, k\}$ el índice de categoría y t el período. Denotamos por $y_{i,t}$ las ventas de la categoría i en el período t y por $Y_t = \sum_{j=1}^k y_{j,t}$ el total de ventas agregado en t .

Para comparar trayectorias relativas, definimos el índice con base t_0 :

$$I_{i,t} = \frac{y_{i,t}}{y_{i,t_0}} \times 100, \quad t \geq t_0, \quad (3.1)$$

de modo que $I_{i,t_0} = 100$ para toda categoría i .

La categoría 3 presenta un crecimiento sostenido y acumulativo a lo largo del período, llegando a superar el índice 300 en los años más recientes, lo que evidencia una expansión relativa significativa frente al nivel inicial.

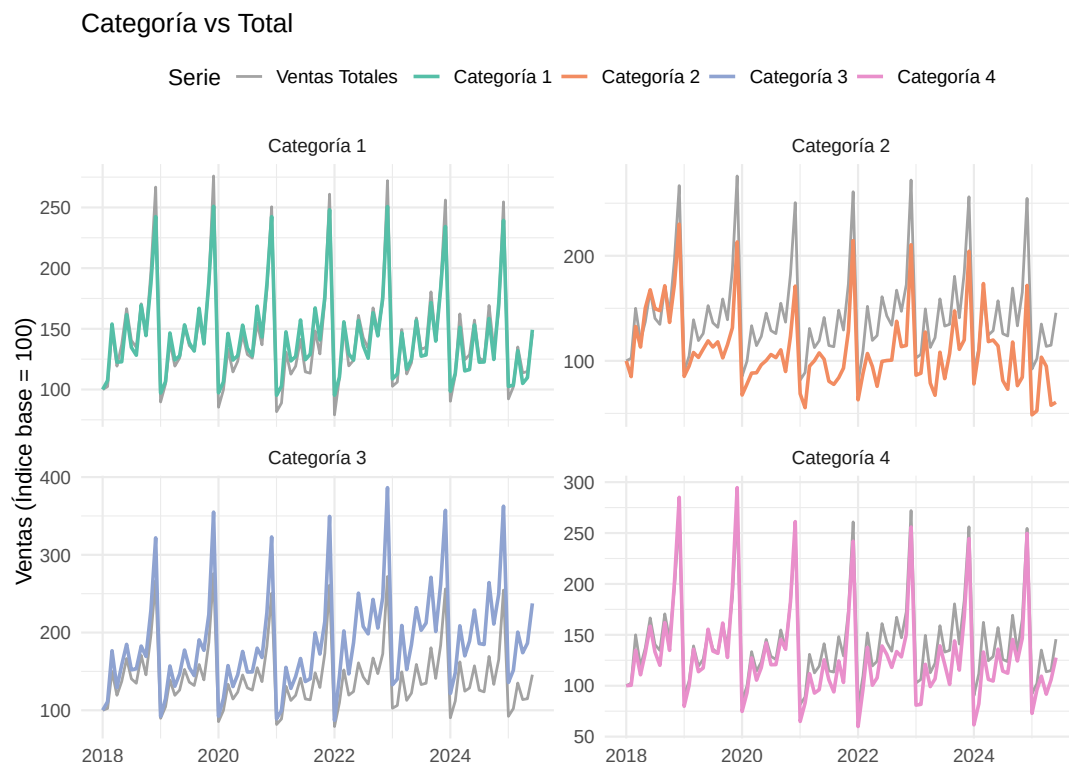


Figura 3.4.
Comparación de tendencias de ventas por categoría normalizadas (base 100) (2018–2025).

3.4 Descomposición y análisis de correlaciones

Se realizó la descomposición de las series de cada categoría para poder identificar de forma clara los componentes de tendencia, estacionalidad y ruido (Figura 3.5). En la Categoría 1, se observa una estacionalidad sumamente marcada con periodicidad anual, reflejada en picos recurrentes que coinciden con meses de oferta, black week y la temporada navideña. No obstante, la tendencia demuestra en los años recientes un comportamiento estable con leves oscilaciones bajas, indicando un mercado estable. Si bien al final del último año se visualiza una tendencia decreciente marcada, esto se relaciona directamente con los ingresos de mercadería, lo que afectó de manera directa al desempeño de las ventas.

La Categoría 2 muestra por su parte un patrón diferente, con una tendencia que es decreciente. Su estacionalidad, si bien está presente, tiene menor amplitud relativa que la

mostrada en la Categoría 1. Esto coincide con su participación por año en los comparativos interanuales que gradualmente se pierde. Por el contrario, en la Categoría 3 se aprecia un alza desde 2022 en la tendencia junto a una estacionalidad intensa y bien definida; este comportamiento apoya la hipótesis de que el segmento ha tomado relevancia dentro del portafolio de categorías. La Categoría 4, finalmente, combina una estacionalidad estable junto con una tendencia algo decreciente en el periodo inicial, luego de una fase de recuperación parcial, quedando como un segmento con comportamiento intermedio.

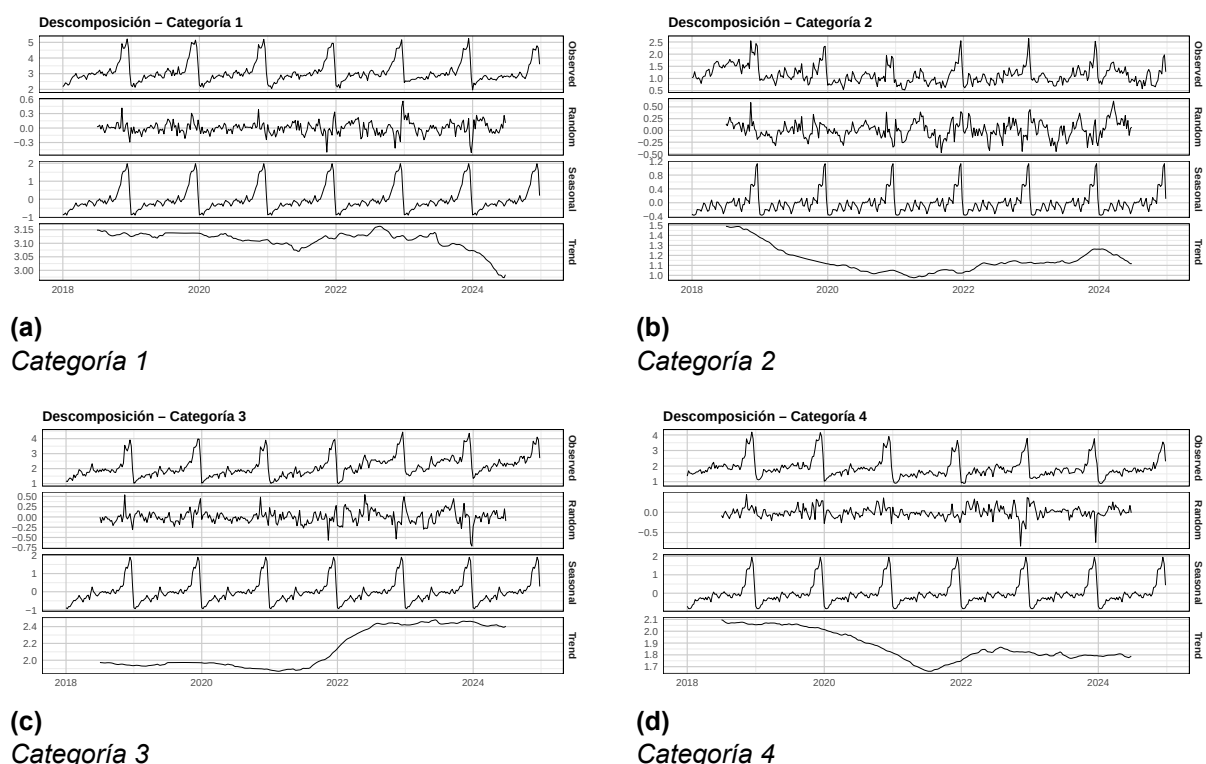


Figura 3.5.
Descomposición de las series de ventas imputadas por categoría.

Mediante el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) se corroboró lo antes mencionado (Figura 3.6). La ACF exhibe picos significativos con rezagos múltiplos de 52 en las Categorías 1 y 3. Esto confirma la presencia de estacionalidad anual. Además, el PACF exhibe un corte para rezagos cortos, lo cual sugiere la inclusión de

términos AR de bajo orden en modelos SARIMA. En la Categoría 2, que es coherente con una serie de menor intensidad estacional y mayor volatilidad, la ACF presenta correlaciones más débiles y dispersas; el PACF, sin embargo, mantiene señales de dependencia en rezagos específicos que se deben considerar en la parametrización. La Categoría 4, en cambio, presenta un patrón medio, mostrando picos estacionales claros aunque de menor tamaño que las categorías 1 y 3, lo que señala que la dinámica depende de la estacionalidad y factores externos.

Estos hallazgos sugieren el ajuste diferenciado de modelos estacionales para cada categoría. La atención se centra en capturar la tendencia decreciente tanto en la Categoría 2 como en la Categoría 4 en un balance entre tendencia y estacionalidad, mientras que en las categorías 1 y 3 la estacionalidad domina la dinámica de las series. El comportamiento que se evidenció a través de este análisis es la base para los ajustes de modelos ARIMA estacionales (SARIMA) con variables explicativas.

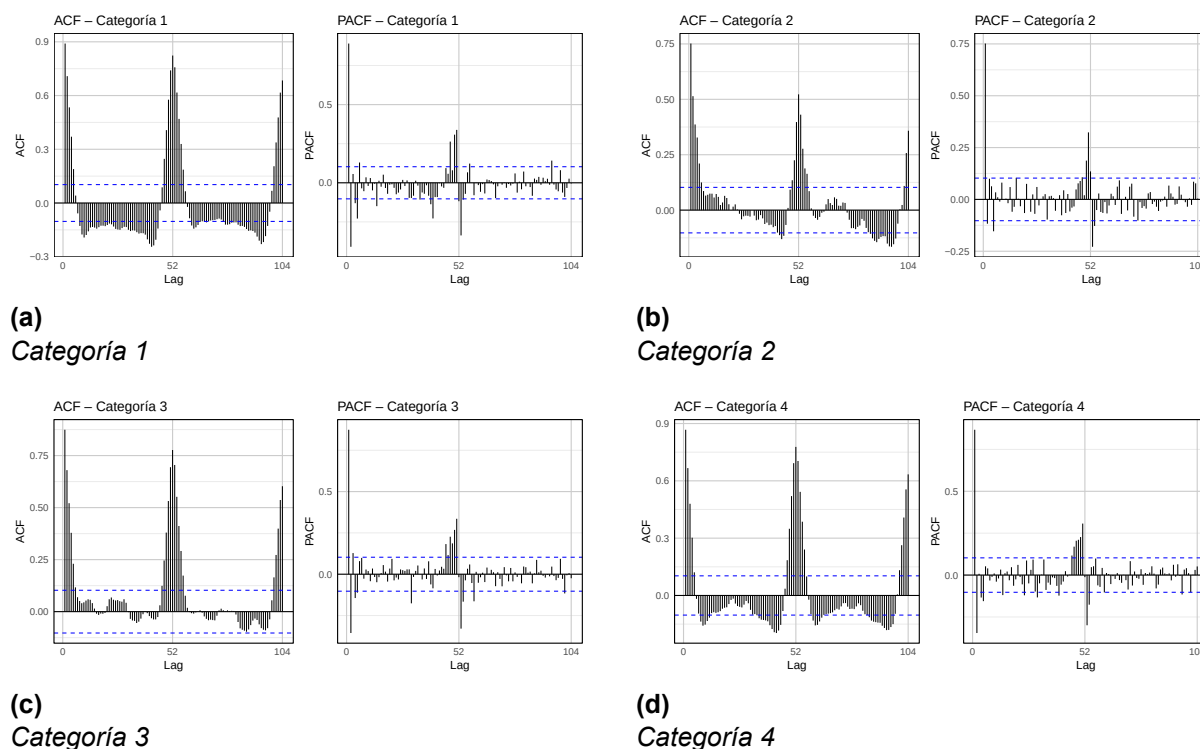


Figura 3.6.

Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) por categoría.

El análisis espectral permitió identificar los principales ciclos que estructuran la dinámica de ventas en cada categoría. En la Figura 3.7, la Categoría 1 presentó picos significativos en torno a 10.3, 13, 17.7, 26 y 55.7 semanas, reflejando la coexistencia de estacionalidad anual, es decir un período de 52 semanas y períodos cortos de entre 2 a 3 meses. La reconstrucción con $K = 8$ armónicos replica de forma adecuada la estructura estacional, conservando los picos de las diversas campañas comerciales.

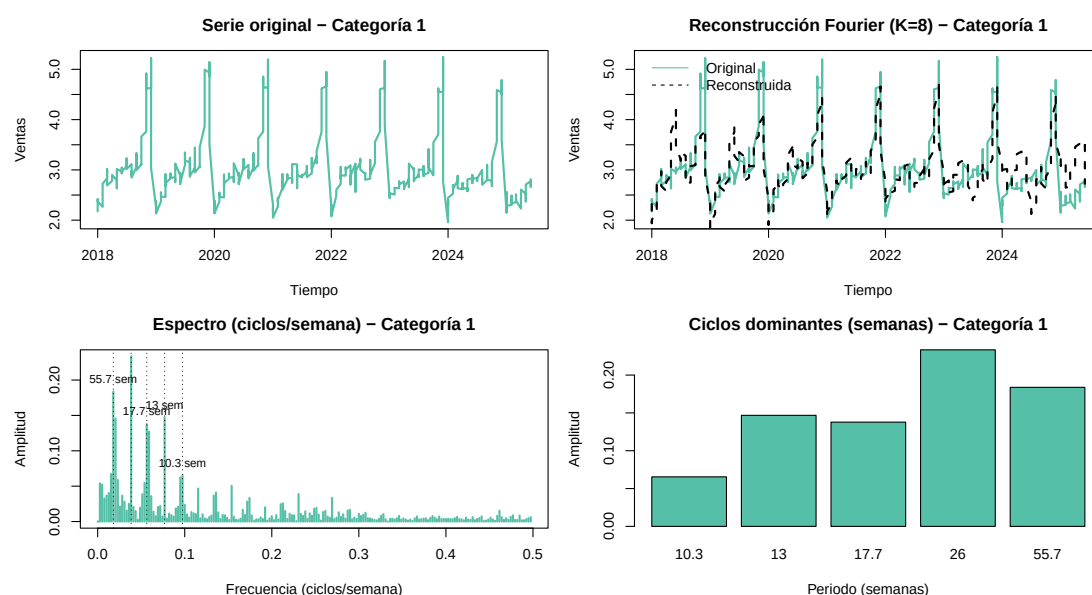


Figura 3.7.
Análisis espectral por Fourier de las series - Categoría 1

Por otro lado, los ciclos de la categoría 2, presentados en la Figura 3.8 exhibieron un comportamiento más diverso, correspondientes a dinámicas de baja frecuencia, sugiriendo una estacionalidad menos marcada, por lo que la reconstrucción con $K = 6$ armónicos logra suavizar la serie y resaltar la tendencia decreciente que se venía observando.

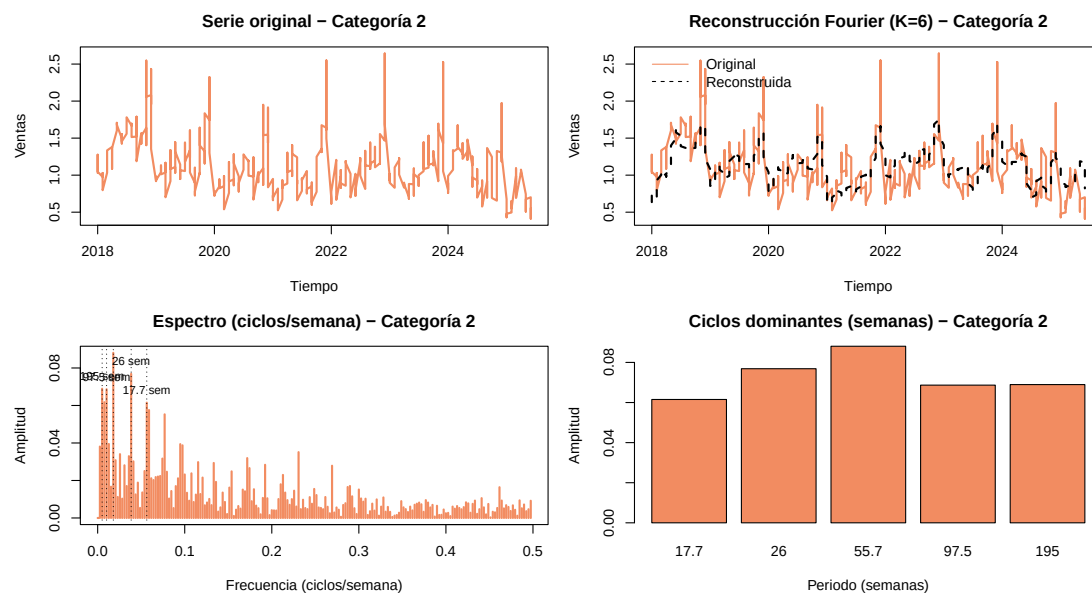


Figura 3.8.
Análisis espectral por Fourier de las series - Categoría 2

En contraste con la categoría 3, la Figura 3.9, en la cual se identificó un patrón estacional muy definido y la serie reconstruida con $K = 8$ se captó la naturaleza repetitiva de las oscilaciones, acorde con la tendencia de crecimiento sostenido evidenciada en los últimos años.

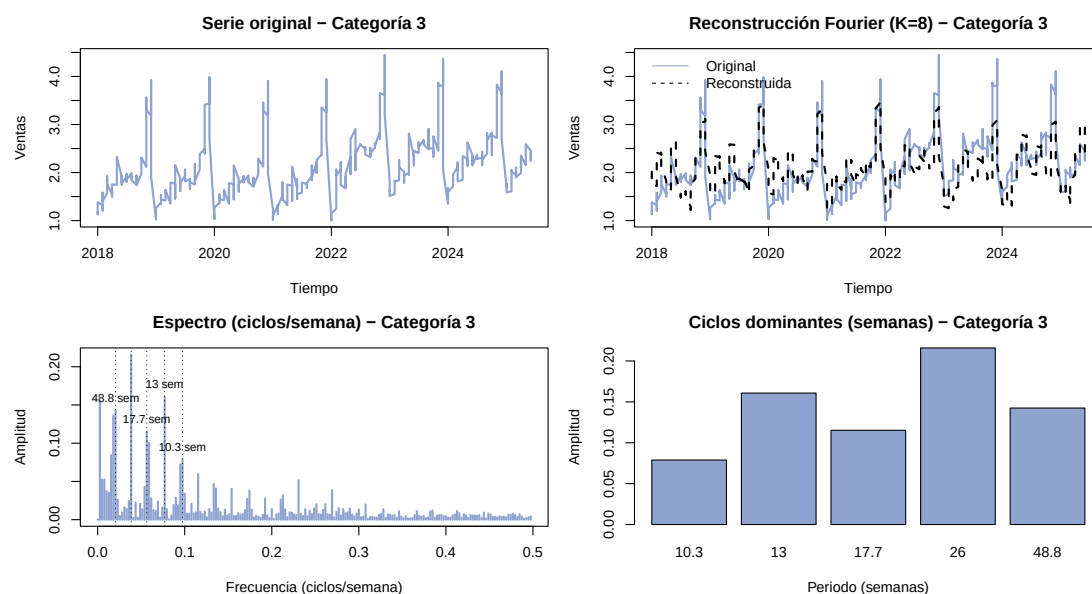


Figura 3.9.
Análisis espectral por Fourier de las series - Categoría 3

Por último, en la Figura 3.10 de la categoría 4, mostró ventas estables y períodos relevantes de 13 y 26 semanas, indicando patrones trimestrales y semestrales. Sin embargo, la amplitud de dichos ciclos es baja, por lo que su impacto en la variación de las ventas es limitado.

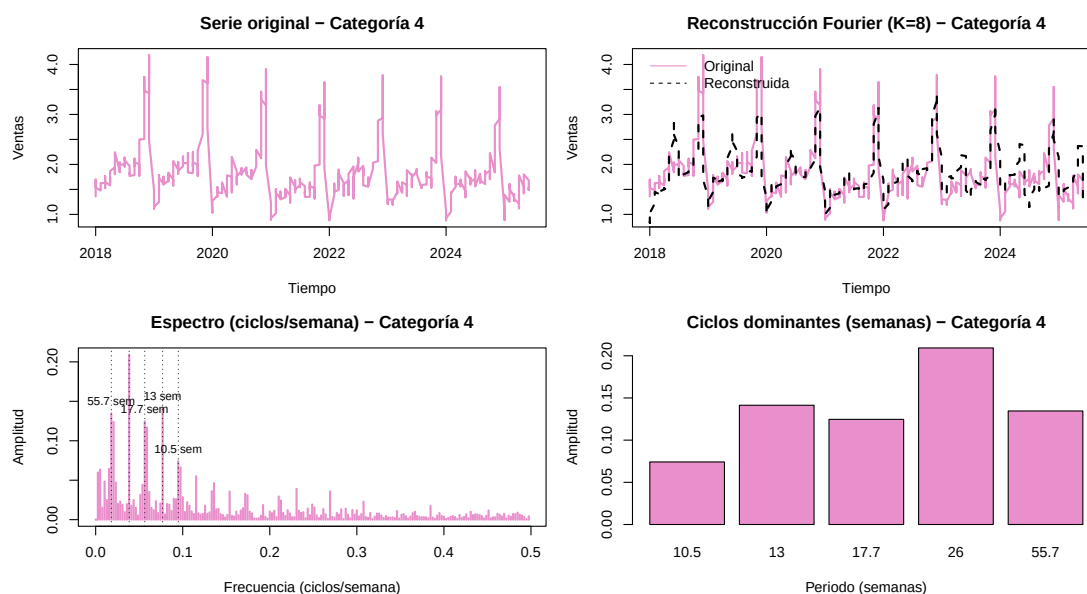


Figura 3.10.
Análisis espectral por Fourier de las series - Categoría 4

3.5 Pruebas de estacionariedad de las series

La evaluación de la estacionariedad se realizó mediante la aplicación del test de Dickey-Fuller Aumentado (ADF), el test KPSS en nivel y el test KPSS con tendencia, de manera complementaria. Los resultados se presentan en la Tabla 3.1.

Tabla 3.1.
Resultados de pruebas de estacionariedad (ADF y KPSS) por categoría.

Categoría	ADF (p-valor)	KPSS-Level (p-valor)	KPSS-Trend (p-valor)	Conclusión
C1	< 0.01	0.10	–	Estacionaria en nivel
C2	< 0.01	0.066	–	Estacionaria (caso limítrofe)
C3	< 0.01	< 0.01	0.10	Estacionaria con tendencia (sugerir $d = 1$)
C4	< 0.01	0.10	–	Estacionaria en nivel

Dentro de las Categorías 1 y 4, el ADF descarta la hipótesis nula de raíz unitaria ($p < 0.01$) y el test KPSS en nivel, por otra parte, no logra rechazar la estacionariedad ($p = 0.10$). Así, ambas

pruebas sugieren que las series tienen nivel estacionario.

Si bien el KPSS en nivel arroja un resultado limítrofe (estadístico 0.426, $p=0.066$), el ADF también rechaza la hipótesis nula de raíz unitaria ($p < 0.01$) en la Categoría 2.

La diferencia es notable, apreciándose sobre todo en la Categoría 3. En tanto que el KPSS a nivel rechaza la hipótesis nula (estadístico 1.334, $p < 0.01$), el ADF señala estacionariedad ($p < 0.01$), concluyendo no estacionariedad. Esta contradicción resulta ser frecuente en series que poseen tendencia determinista. Sin embargo, al implementar la prueba KPSS con la hipótesis nula de estacionariedad en tendencia, dado el valor- p de 0.10 no se puede rechazar la estacionariedad. Esto sugiere que la serie podría considerarse como estacionaria al incorporar un término de tendencia dentro del modelo. Sin embargo, el análisis de correlaciones previo sugiere que se deben aplicar términos de diferenciación a los modelos ajustados.

3.6 Prueba de Causalidad

En la Tabla 3.2 se resumen los hallazgos de las pruebas de causalidad de Granger aplicadas a las cuatro categorías en cada una de las cinco campañas comerciales. Para diferentes rezagos (2 a 12), los p -valores corresponden a los valores reportados, lo cual permite identificar si una campaña genera un efecto relevante en el comportamiento de las ventas.

Para la Categoría 1 se evidencia que las Campañas 4 y 5 ejercen un efecto inmediato y altamente significativo sobre las ventas, ya que sus p -valores casi nulos desde rezagos muy cortos demuestran su impacto directo en el corto plazo. La Campaña 2 muestra un efecto diferido consistente con dinámicas en las que la respuesta de los consumidores se concentra al final de la campaña, pues su significancia solo aparece en rezagos más largos (lag 12). De forma parecida, la Campaña 1 se vuelve así significativa desde rezagos medios y largos, lo cual refleja el peso de la publicidad más intensificada al cierre. Si bien su importancia estratégica radica en criterios de

negocio más que por causalidad estadística, la Campaña 3 no exhibe efectos significativos.

Patrones similares se mantienen en la Categoría 2: las Campañas 4 más 5 influyen de forma significativa desde rezagos cortos, apoyando así su rol como motores inmediatos de la demanda. Se confirma un efecto tardío, evidenciándose la significancia de Campaña 1 en rezagos largos (lag 12). La Campaña 2 presenta un esquema similar. El impacto detectable se observa recién hasta en rezagos extensos. Por su parte, la Campaña 3 no muestra en absoluto una relación causal robusta. Esta relación no tiene significancia estadística.

En la Categoría 3, las Campañas 4 y 5 son las más influyentes nuevamente, con efectos marcadamente persistentes además de p -valores cercanos a cero en todos los rezagos. La Campaña 1 revela significancia desde rezagos intermedios. Esto sugiere un impacto a medio plazo. La Campaña 2 muestra un efecto con retardo también, siendo significativa solo en los rezagos 10 y 12. La Campaña 3 no muestra indicios de causalidad en ningún retraso. Esto difiere con lo observado en otras campañas.

Las Campañas 4 y 5, determinantes muy consistentes del comportamiento de ventas, mantienen un impacto robusto en todas y cada una de las especificaciones de rezago según la evidencia en la Categoría 4. La Campaña 1 presenta significancia a partir de algunos rezagos intermedios, mostrando así un patrón tardío de respuesta. Mientras que la Campaña 3 permanece estadísticamente irrelevante en toda la ventana de rezagos analizada, la Campaña 2 solo alcanza significancia marginal en rezagos más largos, similar al comportamiento en las categorías anteriores.

Este análisis respalda la hipótesis de inclusión de las campañas como regresores exógenos claves para poder capturar efectos diferidos.

Tabla 3.2.*Pruebas de causalidad de Granger (p-valores) por categoría, campaña y distintos rezagos.*

Categoría	Campaña	Lag 2	Lag 4	Lag 6	Lag 8	Lag 10	Lag 12
C1	Campaña1	0.519	0.291	0.141	0.035	0.013	0.001
C1	Campaña2	0.612	0.735	0.874	0.894	0.618	0.016
C1	Campaña3	0.168	0.123	0.253	0.443	0.584	0.697
C1	Campaña4	1.14e-08	1.55e-24	2.31e-40	1.18e-64	2.69e-65	2.00e-66
C1	Campaña5	2.10e-44	4.41e-53	2.14e-49	2.41e-48	5.84e-49	3.34e-48
C2	Campaña1	0.495	0.236	0.400	0.176	0.059	0.008
C2	Campaña2	0.949	0.767	0.900	0.424	0.225	0.011
C2	Campaña3	0.331	0.104	0.200	0.321	0.327	0.456
C2	Campaña4	0.131	2.23e-19	7.16e-34	5.03e-36	3.08e-35	8.33e-34
C2	Campaña5	1.10e-16	2.89e-20	5.34e-20	3.06e-20	6.84e-21	9.20e-20
C3	Campaña1	0.192	0.413	0.561	0.007	6.80e-04	2.69e-04
C3	Campaña2	0.815	0.634	0.808	0.542	0.051	2.64e-04
C3	Campaña3	0.703	0.796	0.792	0.907	0.962	0.985
C3	Campaña4	0.018	1.58e-10	2.84e-49	1.57e-48	2.05e-46	9.65e-45
C3	Campaña5	3.40e-37	5.87e-38	1.13e-36	9.85e-37	3.60e-35	1.12e-34
C4	Campaña1	0.072	0.121	0.198	0.023	0.006	0.002
C4	Campaña2	0.303	0.634	0.770	0.708	0.631	0.029
C4	Campaña3	0.554	0.702	0.861	0.901	0.858	0.891
C4	Campaña4	2.69e-04	1.57e-16	7.86e-46	1.25e-44	1.14e-42	5.37e-42
C4	Campaña5	4.98e-35	8.63e-37	1.32e-35	4.40e-36	7.18e-35	6.64e-35

3.7 Ajuste de modelos

A partir del análisis de los resultados previos, se ajustaron los modelos de series temporales. Dichos modelos integran los términos autorregresivos, además de los de promedios móviles, así como también la diferenciación estacional indispensable. En los casos de los SARIMAX, se incorporaron también variables exógenas junto con funciones de Fourier que hacen posible capturar patrones estacionales de mayor complejidad. Las formulaciones conclusivas de cada modelo, junto a sus coeficientes calculados, se presentan a continuación.

El modelo SARIMA escogido para la **Categoría 1** fue un SARIMA(1,0,0)(1,1,0)₅₂, y su representación matemática se muestra en la Ecuación 3.2.

$$(1 - 0.4530B)(1 + 0.5458B^{52})\nabla_{52}Ventas_t = -0.0004. \quad (3.2)$$

Fue estimado, igualmente, un modelo SARIMAX(2, 1, 3)(1, 0, 0)₅₂ con regresores vinculados a campañas comerciales más componentes de Fourier, expuesto en la Ecuación 3.3.

$$\begin{aligned}
 (1 + 0.3916B + 0.5990B^2)\nabla Ventas_t &= 0.0055 \text{ Campaña}1_t - 0.0282 \text{ Campaña}2_t + 0.0648 \text{ Campaña}3_t \\
 &+ 0.1448 \text{ Campaña}4_t + 0.2937 \text{ Campaña}5_t \\
 &+ \sum_{k=1}^8 \left(\alpha_k \sin \frac{2\pi kt}{52} + \beta_k \cos \frac{2\pi kt}{52} \right) \\
 &+ (1 - 0.1338B - 0.0991B^2 - 0.6691B^3)(1 - 0.3730B^{52})\varepsilon_t
 \end{aligned} \tag{3.3}$$

El modelo SARIMA en la **Categoría 2** es un SARIMA(1, 1, 1)(1, 1, 1)₅₂, mostrado en la Ecuación 3.4, mientras que el modelo SARIMAX, según se aprecia en la Ecuación 3.5, une una estructura SARIMAX(0, 1, 3)(0, 0, 2)₅₂ con términos de Fourier de orden $K = 6$ y regresores exógenos.

$$(1 - 0.4939B)\nabla \nabla_{52} Ventas_t = (1 - 0.8788B)(1 - 0.1392B^{52})(1 - 0.5854B^{52})\varepsilon_t \tag{3.4}$$

$$\begin{aligned}
 \nabla Ventas_t &= -0.0150 \text{ Campaña}1_t - 0.0441 \text{ Campaña}2_t + 0.1595 \text{ Campaña}3_t \\
 &+ 0.2186 \text{ Campaña}4_t + 0.4725 \text{ Campaña}5_t \\
 &+ \sum_{k=1}^6 \left(\alpha_k \sin \frac{2\pi kt}{52} + \beta_k \cos \frac{2\pi kt}{52} \right) \\
 &+ (1 - 0.4408B - 0.2253B^2 - 0.0784B^3)(1 + 0.1810B^{52} + 0.2551B^{104})\varepsilon_t
 \end{aligned} \tag{3.5}$$

El modelo SARIMA se puede sintetizar en la Ecuación 3.6, habiéndose adoptado la forma SARIMA(0, 1, 2)(0, 1, 0)₅₂ para la **Categoría 3**.

$$\nabla \nabla_{52} Ventas_t = (1 - 0.4366B - 0.3705B^2)\varepsilon_t \tag{3.6}$$

Conjuntamente, se ajustó un modelo SARIMAX(3, 1, 1)(1, 0, 0)₅₂ junto con armónicos de Fourier de orden $K = 8$, el cual se representa en la Ecuación 3.7.

$$\begin{aligned}
 (1 - 0.3563B + 0.1626B^2 - 0.0807B^3)\nabla Ventas_t = & 0.0093 \text{ Campaña}1_t - 0.0570 \text{ Campaña}2_t \\
 & + 0.2155 \text{ Campaña}3_t + 0.4382 \text{ Campaña}4_t \\
 & + 0.5204 \text{ Campaña}5_t + \sum_{k=1}^8 \left(\alpha_k \sin \frac{2\pi kt}{52} + \beta_k \cos \frac{2\pi kt}{52} \right) \\
 & + (1 - 0.8845B)(1 - 0.1788B^{52})\varepsilon_t
 \end{aligned} \tag{3.7}$$

Finalmente, el modelo SARIMA para la **Categoría 4** fue un SARIMA(0, 1, 2)(1, 0, 1)₅₂, mostrado en la Ecuación 3.8, mientras que el modelo SARIMAX corresponde a un SARIMAX(0, 1, 2)(1, 0, 0)₅₂ con regresores junto con funciones de Fourier de orden $K = 8$, la representación de lo cual se encuentra en la Ecuación 3.9.

$$(1 - 0.3546B - 0.1898B^2)\nabla_{52} Ventas_t = (1 + 0.9998B^{52})(1 - 0.9512B^{52})\varepsilon_t \tag{3.8}$$

$$\begin{aligned}
 \nabla Ventas_t = & 0.0049 \text{ Campaña}1_t + 0.0228 \text{ Campaña}2_t + 0.1512 \text{ Campaña}3_t \\
 & + 0.3631 \text{ Campaña}4_t + 0.6249 \text{ Campaña}5_t \\
 & + \sum_{k=1}^8 \left(\alpha_k \sin \frac{2\pi kt}{52} + \beta_k \cos \frac{2\pi kt}{52} \right) \\
 & + (1 - 0.6017B - 0.2658B^2)(1 - 0.1255B^{52})\varepsilon_t
 \end{aligned} \tag{3.9}$$

3.8 Métricas y diagnóstico de residuales

La Tabla 3.3 resume, para los modelos SARIMA y SARIMAX dentro de cada categoría, los indicadores de desempeño (MAE, MSE, RMSE, MAPE), los criterios informativos (AIC, BIC, logLik) y los resultados del diagnóstico de residuos. Garantizando una selección informada del modelo para cada categoría.

Para la Categoría 1, SARIMA exhibe errores predictivos más bajos, en tanto que un AIC inferior y superior log-verosimilitud denotan un ajuste estadísticamente más parsimonioso en SARIMAX. Ambos modelos cumplen con la independencia de los residuos.

Las métricas de error son comparables, demostrando SARIMA una leve superioridad en MAPE para la Categoría 2. No obstante, los criterios de información denotan una evidente superioridad del SARIMAX. El **SARIMAX** se muestra preferible gracias a su eficiencia estadística superior, considerando el comportamiento aleatorio de los residuos de ambos modelos.

Con respecto a la Categoría 3, SARIMAX presenta valores de AIC y log-verosimilitud que favorecen de manera convincente, a pesar de que SARIMA tuvo errores absolutos menores. La ligera diferencia observada en los errores no justifica mantener el SARIMA, por lo cual se prefiere el **SARIMAX**, priorizando la mejora observada en parsimonia.

El modelo SARIMA exhibe un error absoluto medio inferior e igualmente un MAPE inferior, denotando de este modo una ventaja leve en precisión. No obstante, el SARIMAX sobresale debido a un AIC notablemente inferior y una log-verosimilitud considerablemente superior. Adicionalmente, presenta un BIC que resulta competitivo. La selección del SARIMAX como modelo final para esta categoría se justifica, dado que ambos modelos satisfacen los supuestos de residuos y la mejora en criterios de información.

Tabla 3.3.

Resumen de métricas de desempeño, criterios de información y pruebas de residuos para los

modelos SARIMA y SARIMAX por categoría.

Categoría	Modelo	MAE	MSE	RMSE	MAPE	AIC	BIC	logLik	JB (p)	LB (p)	DW (p)
C1	SARIMA	0.095	0.021	0.145	3.08%	-246.14	-231.17	127.07	<0.0001	0.691	0.833
	SARIMAX	0.113	0.022	0.150	3.65%	-280.62	-171.58	168.31	<0.0001	0.704	0.888
C2	SARIMA	0.128	0.032	0.179	12.31%	-96.77	-78.07	53.39	<0.0001	0.793	0.834
	SARIMAX	0.143	0.036	0.190	13.23%	-121.58	-32.01	83.79	<0.0001	0.955	0.915
C3	SARIMA	0.140	0.048	0.220	6.56%	-4.71	6.51	5.36	<0.0001	0.810	0.791
	SARIMAX	0.136	0.033	0.183	6.64%	-146.02	-40.87	100.01	<0.0001	0.941	0.880
C4	SARIMA	0.099	0.024	0.154	5.48%	-139.70	-120.23	74.85	<0.0001	0.447	0.790
	SARIMAX	0.124	0.029	0.170	6.93%	-204.44	-107.08	127.22	<0.0001	0.941	0.839

3.9 Pronósticos

Con los modelos SARIMAX seleccionados en cada categoría, se elaboraron las predicciones para un horizonte h semanal, integrando las variables exógenas de promociones comerciales y los componentes de Fourier de orden previamente calibrado. Los intervalos de predicción al 80% y 95% reflejan tanto la incertidumbre inherente al proceso estocástico, así como la que deriva de la estimación de parámetros.

3.9.1 Categoría 1

En la Tabla 3.4 es donde se presentan los valores observados, los errores absolutos asociados y los intervalos de confianza al 80% y 95% junto con los pronósticos que fueron generados por el modelo SARIMAX para la **Categoría 1**. De igual forma, en la Figura 3.11 se visualiza la serie de ventas históricas, con los valores ajustados y los valores resultantes del pronóstico.

Los indicadores globales muestran un desempeño que es satisfactorio: el MAE se situó en 0.161, y la RMSE alcanzó 0.201, lo cual confirma que las discrepancias promedio se mantuvieron acotadas y sin desviaciones extremas. Con 0.040, el error cuadrático medio (MSE) refleja, por cierto,

variabilidad moderada en las diferencias, mientras que el error porcentual absoluto medio (MAPE), con 6.60%, evidencia un ajuste relativo adecuado respecto al nivel de ventas.

La Tabla 3.4 muestra que casi todos los valores reales se encuentran en los rangos pronosticados. Esto valida la consistencia en lo que respecta a las bandas de confianza estimadas. Los errores absolutos más elevados se concentran hacia las primeras semanas del horizonte. La baja aceptación de los clientes de la colección que ingresó a finales de año en este segmento de producto explica este desvío, impactando de forma negativa las ventas entre enero y junio. En ese período se registraron altos niveles de rebajas no vendidas y donaciones derivadas de dichas rebajas. Esto intensificó la diferencia con respecto al pronóstico y la realidad. No obstante, el modelo estadístico es válido pues los valores reales se conservaron en los intervalos de confianza.

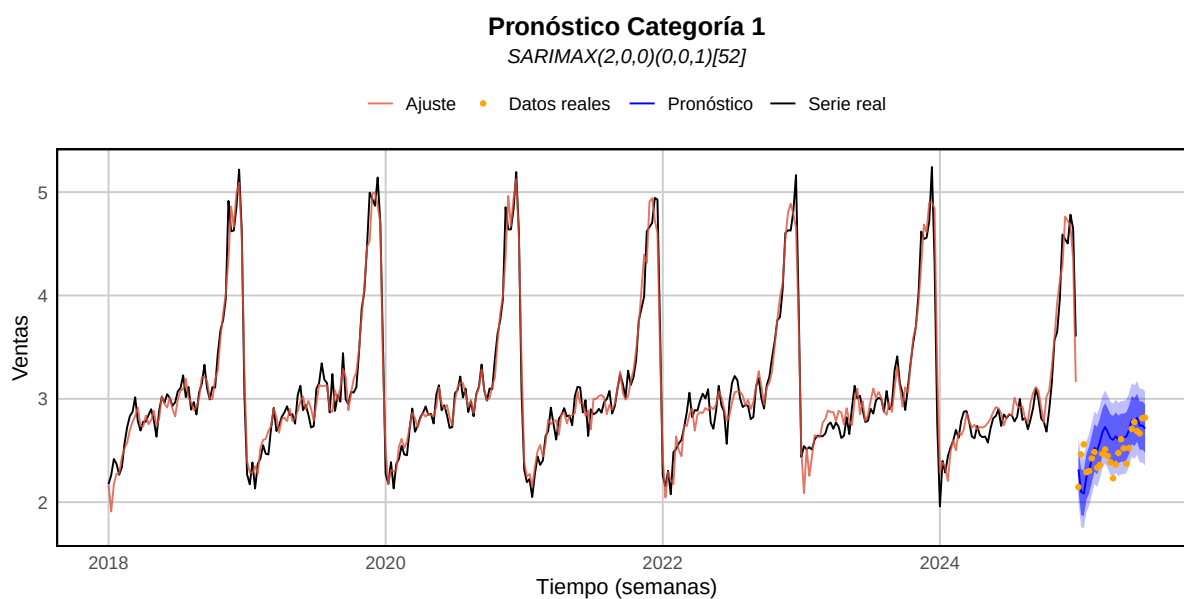


Figura 3.11.

Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 1.

Tabla 3.4.*Pronóstico de la Categoría 1 para las semanas 2025.001–2025.026.*

Semana	Pronóstico	LI 80%	LS 80%	LI 95%	LS 95%	Real	Error Absoluto
2025.001	2.318	2.118	2.517	2.012	2.623	2.146	0.172
2025.002	2.099	1.878	2.321	1.761	2.438	2.461	0.362
2025.003	2.084	1.863	2.305	1.746	2.422	2.559	0.475
2025.004	2.246	2.021	2.470	1.903	2.588	2.290	0.044
2025.005	2.311	2.084	2.538	1.964	2.658	2.301	0.010
2025.006	2.415	2.187	2.644	2.066	2.765	2.424	0.008
2025.007	2.526	2.297	2.754	2.176	2.876	2.486	0.040
2025.008	2.499	2.270	2.728	2.149	2.849	2.333	0.166
2025.009	2.576	2.346	2.807	2.224	2.928	2.359	0.217
2025.010	2.677	2.447	2.907	2.325	3.029	2.467	0.210
2025.011	2.728	2.497	2.958	2.375	3.080	2.511	0.217
2025.012	2.678	2.448	2.909	2.325	3.031	2.447	0.231
2025.013	2.619	2.388	2.850	2.265	2.972	2.385	0.233
2025.014	2.603	2.372	2.834	2.250	2.957	2.232	0.371
2025.015	2.637	2.406	2.868	2.283	2.991	2.364	0.273
2025.016	2.606	2.375	2.838	2.252	2.961	2.476	0.130
2025.017	2.625	2.393	2.857	2.270	2.980	2.609	0.015
2025.018	2.625	2.392	2.857	2.270	2.979	2.521	0.104
2025.019	2.638	2.406	2.870	2.283	2.993	2.373	0.265
2025.020	2.701	2.468	2.934	2.345	3.057	2.524	0.177
2025.021	2.795	2.562	3.028	2.439	3.151	2.709	0.085
2025.022	2.776	2.543	3.009	2.419	3.132	2.776	0.001
2025.023	2.816	2.583	3.049	2.459	3.173	2.687	0.129
2025.024	2.738	2.505	2.972	2.381	3.095	2.666	0.072
2025.025	2.742	2.509	2.976	2.385	3.100	2.813	0.071
2025.026	2.712	2.478	2.946	2.354	3.069	2.818	0.106

3.9.2 Categoría 2

La Tabla 3.5 presenta los valores observados, además de los errores absolutos y los pronósticos de la **Categoría 2** obtenidos con el modelo SARIMAX, junto con intervalos de confianza al 80% y 95%. La Figura 3.12 complementa este resumen mostrando la serie histórica más los valores ajustados. Además, presenta el horizonte de predicción.

El modelo exhibe un desempeño razonable para el horizonte considerado en términos agregados; es decir, el *MAE* fue de 0.155, el *RMSE* de 0.188 y el *MSE* de 0.035. El *MAPE* de 22.56% indica que, si bien el error relativo resulta moderado, este se mantiene en umbrales aceptables para un escenario semanal con señal de baja magnitud y también variabilidad estacional pronunciada. Se apoya la calibración de la incertidumbre del modelo porque la mayoría de los

valores observados está dentro de las bandas de predicción.

El patrón de errores a lo largo del horizonte sugiere dos comportamientos que están diferenciados. Los desvíos son pequeños durante las primeras seis semanas (errores absolutos típicamente por debajo de 0.06), lo cual es consistente con la inercia de la dinámica reciente capturada por el componente ARIMA. A partir de la semana 7 se ven episodios con errores más grandes (por ejemplo, en las semanas 7–9 y 11, con desvíos absolutos entre 0.16 y 0.27) y, otra vez, entre las semanas 15 y 21 (con picos entre 0.23 y 0.34 en 16–17 y 20–21). Tales discrepancias sí son compatibles con shocks promocionales o con diferencias en la intensidad efectiva de las campañas respecto a los patrones promedio que son representados por los términos de Fourier. Asimismo, son compatibles con posibles cambios al ser una categoría cuyo comportamiento está marcado por las temporadas. A pesar de lo mencionado, los valores registrados siguen contenidos por los intervalos de predicción. Esto indica que la medición de la incertidumbre es correcta aun cuando aumenta la distancia del pronóstico.

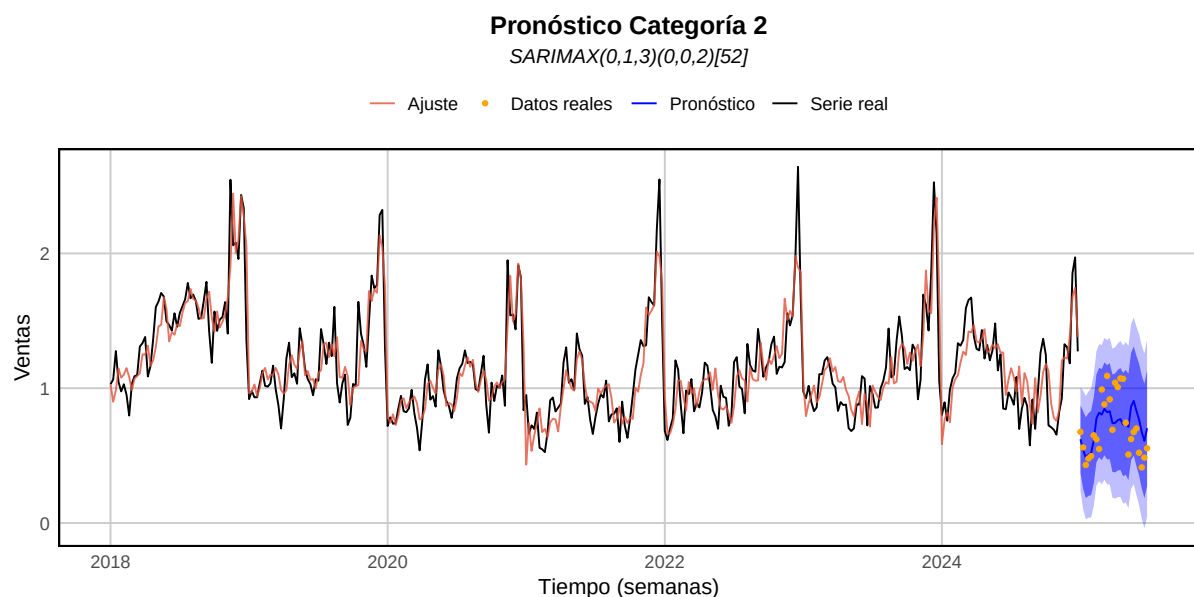


Figura 3.12.

Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 2.

Tabla 3.5.*Pronóstico de la Categoría 2 para las semanas 2025.001–2025.026.*

Semana	Pronóstico	LI 80%	LS 80%	LI 95%	LS 95%	Real	Error Absoluto
2025.001	0.621	0.370	0.872	0.237	1.005	0.676	0.055
2025.002	0.543	0.255	0.830	0.103	0.983	0.561	0.018
2025.003	0.485	0.186	0.785	0.027	0.943	0.431	0.054
2025.004	0.507	0.201	0.814	0.039	0.976	0.479	0.029
2025.005	0.520	0.207	0.832	0.041	0.998	0.496	0.023
2025.006	0.614	0.294	0.933	0.125	1.102	0.650	0.037
2025.007	0.777	0.451	1.103	0.279	1.275	0.621	0.156
2025.008	0.818	0.486	1.150	0.311	1.326	0.549	0.269
2025.009	0.804	0.466	1.142	0.287	1.321	0.990	0.186
2025.010	0.847	0.503	1.191	0.321	1.373	0.880	0.033
2025.011	0.822	0.472	1.172	0.286	1.357	1.091	0.269
2025.012	0.829	0.473	1.185	0.285	1.373	0.917	0.088
2025.013	0.740	0.379	1.102	0.187	1.293	0.692	0.049
2025.014	0.743	0.375	1.110	0.181	1.304	1.040	0.298
2025.015	0.762	0.389	1.135	0.192	1.332	1.008	0.246
2025.016	0.771	0.393	1.150	0.193	1.350	1.073	0.301
2025.017	0.731	0.347	1.114	0.144	1.317	1.069	0.339
2025.018	0.747	0.358	1.136	0.152	1.342	0.744	0.003
2025.019	0.713	0.319	1.107	0.110	1.316	0.508	0.205
2025.020	0.865	0.465	1.264	0.254	1.476	0.623	0.242
2025.021	0.905	0.501	1.310	0.287	1.524	0.675	0.230
2025.022	0.832	0.422	1.241	0.205	1.458	0.701	0.130
2025.023	0.767	0.353	1.182	0.133	1.401	0.521	0.246
2025.024	0.675	0.255	1.094	0.033	1.316	0.412	0.262
2025.025	0.609	0.185	1.034	-0.040	1.258	0.485	0.124
2025.026	0.703	0.274	1.132	0.047	1.359	0.555	0.149

3.10 Categoría 3

La Tabla 3.6 muestra los pronósticos de la **Categoría 3** que presenta la Figura 3.13, logrados con el modelo SARIMAX, abarcando valores observados, errores absolutos e intervalos al 80% y 95%, también la serie histórica, los valores ajustados y los pronósticos. El desempeño general es sólido, con $MAE = 0.111$, $RMSE = 0.149$, $MSE = 0.022$ e $MAPE = 5.53\%$. Dichos niveles sugieren variabilidad del error que está contenida así como sesgo reducido, lo cual es coherente con la capacidad del modelo de poder capturar la dependencia temporal y la estacionalidad.

El hecho de que la mayor parte de las observaciones se sitúen dentro de las bandas de predicción a lo largo del horizonte apoya el ajuste adecuado de la incertidumbre. Los desvíos más

notorios se concentran en las semanas específicas de: 2–3 ($\approx 0.31–0.33$), 15 (≈ 0.26), 17 (≈ 0.24) y 26 (≈ 0.26). En contraste con aquello, en ciertos puntos el error es casi nulo o bastante bajo. Esto se puede observar, por ejemplo, en las semanas 6, 9, 12, 19 y 24, todas ellas con $|e_t| \leq 0.02$.

De acuerdo con estos resultados, el SARIMAX de la Categoría 3 presenta un *trade-off* favorable en parsimonia y ajuste, con bandas predictivas informativas y errores absolutos acotados. Es recomendable el uso de los intervalos de confianza como una guía de cobertura en esas semanas cercanas a los cambios de fase estacional para las aplicaciones operativas, además de aprovechar la buena precisión promedio ($\text{MAPE} \approx 5.5\%$) para la planificación táctica del inventario y el suministro, ya que los desvíos tienden a incrementarse de forma leve.

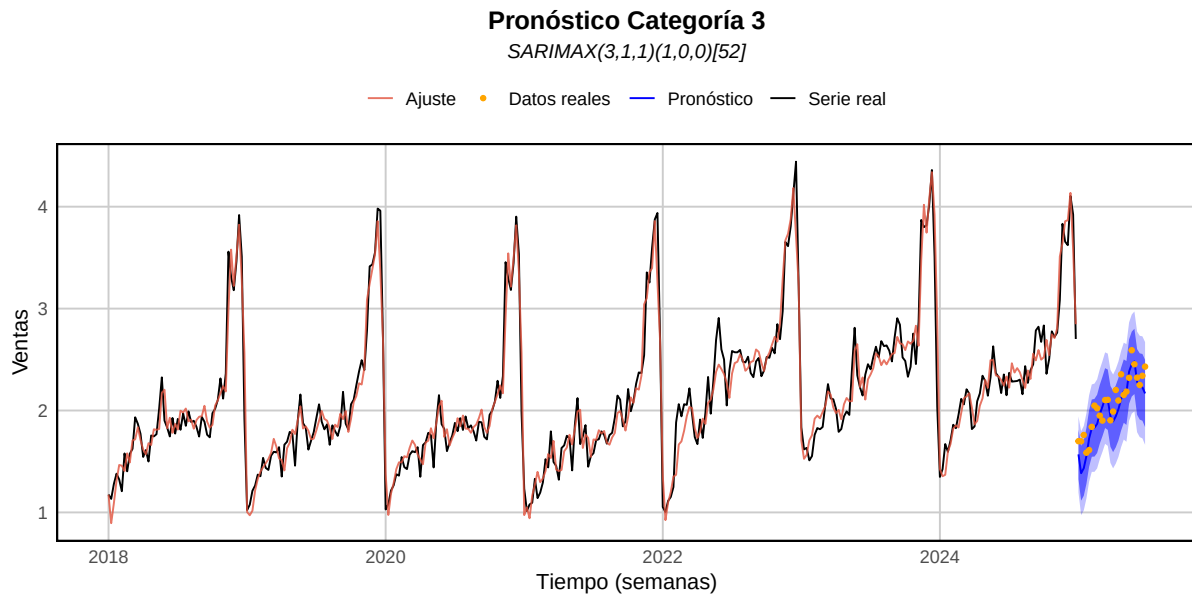


Figura 3.13.

Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 3.

Tabla 3.6.*Pronóstico de la Categoría 3 para las semanas 2025.001–2025.026.*

Semana	Pronóstico	LI 80%	LS 80%	LI 95%	LS 95%	Real	Error Absoluto
2025.001	1.570	1.326	1.813	1.197	1.942	1.699	0.129
2025.002	1.385	1.116	1.654	0.973	1.796	1.698	0.313
2025.003	1.434	1.164	1.705	1.020	1.848	1.760	0.326
2025.004	1.556	1.282	1.829	1.137	1.974	1.586	0.030
2025.005	1.726	1.448	2.003	1.301	2.150	1.612	0.114
2025.006	1.827	1.547	2.108	1.398	2.256	1.840	0.012
2025.007	1.831	1.548	2.114	1.398	2.264	2.048	0.217
2025.008	1.869	1.583	2.154	1.431	2.306	2.022	0.154
2025.009	1.963	1.675	2.252	1.522	2.404	1.950	0.013
2025.010	2.039	1.748	2.330	1.594	2.484	1.901	0.138
2025.011	2.122	1.828	2.415	1.673	2.571	2.103	0.019
2025.012	2.106	1.810	2.402	1.653	2.559	2.105	0.001
2025.013	1.942	1.644	2.241	1.486	2.399	1.904	0.039
2025.014	1.893	1.592	2.194	1.432	2.353	1.989	0.097
2025.015	1.942	1.639	2.246	1.478	2.407	2.201	0.259
2025.016	2.013	1.707	2.319	1.545	2.481	2.098	0.085
2025.017	2.112	1.803	2.421	1.640	2.584	2.355	0.243
2025.018	2.190	1.879	2.501	1.715	2.666	2.153	0.037
2025.019	2.172	1.859	2.486	1.693	2.652	2.185	0.013
2025.020	2.382	2.066	2.698	1.899	2.865	2.321	0.061
2025.021	2.456	2.137	2.774	1.969	2.942	2.591	0.136
2025.022	2.482	2.161	2.802	1.992	2.972	2.454	0.028
2025.023	2.281	1.958	2.603	1.787	2.774	2.323	0.042
2025.024	2.236	1.911	2.561	1.739	2.733	2.250	0.014
2025.025	2.227	1.900	2.555	1.726	2.728	2.341	0.113
2025.026	2.167	1.838	2.497	1.663	2.672	2.431	0.263

3.10.1 Categoría 4

En la Tabla 3.7 se reportan los pronósticos de la **Categoría 4**; además, la Figura 3.14 muestra la serie histórica, los valores ajustados, al igual que los valores que resultaron del pronóstico, incluyendo intervalos al 80% y 95% y los errores absolutos.

El desempeño general es adecuado según el periodo evaluado: $MAE = 0.147$, $RMSE = 0.192$, $MSE = 0.0369$ y $MAPE = 10.59\%$. Estos valores reflejan que las desviaciones medias permanecen acotadas. Se evidencia un ajuste adecuado de la incertidumbre, pues la mayoría de las observaciones se encuentran dentro de las bandas de confianza.

Las semanas 3 con 8 (con máximos en 4: $|e_t| \approx 0.46$, 6 e incluso 7: $|e_t| \approx 0.31–0.33$), así como las semanas 15 junto a 17 ($|e_t| \approx 0.29$ y $|e_t| \approx 0.32$, respectivamente) conforman al patrón

dentro del horizonte de pronósticos con discrepancias elevadas que se concentran al inicio y después. Por el contrario, pronósticos muy cercanos también se observan en las semanas 9, 11 y 19, con $|e_t| \leq 0.02$). Este patrón sugiere la coexistencia con una estacionalidad que está siendo capturada de manera efectiva. También sugiere la coexistencia de shocks específicos como intensidad promocional efectiva, disponibilidad de surtido que no coinciden con los promedios históricos. Aun así, lo que indica que la incertidumbre del pronóstico está razonablemente cuantificada es que los valores observados permanecen en su mayoría dentro de los intervalos del 80% y 95%.

Los resultados para esta categoría muestran, en general, un equilibrio entre el ajuste y la parsimonia ofrecida por el modelo, con errores absolutos moderados y bandas predictivas informativas.

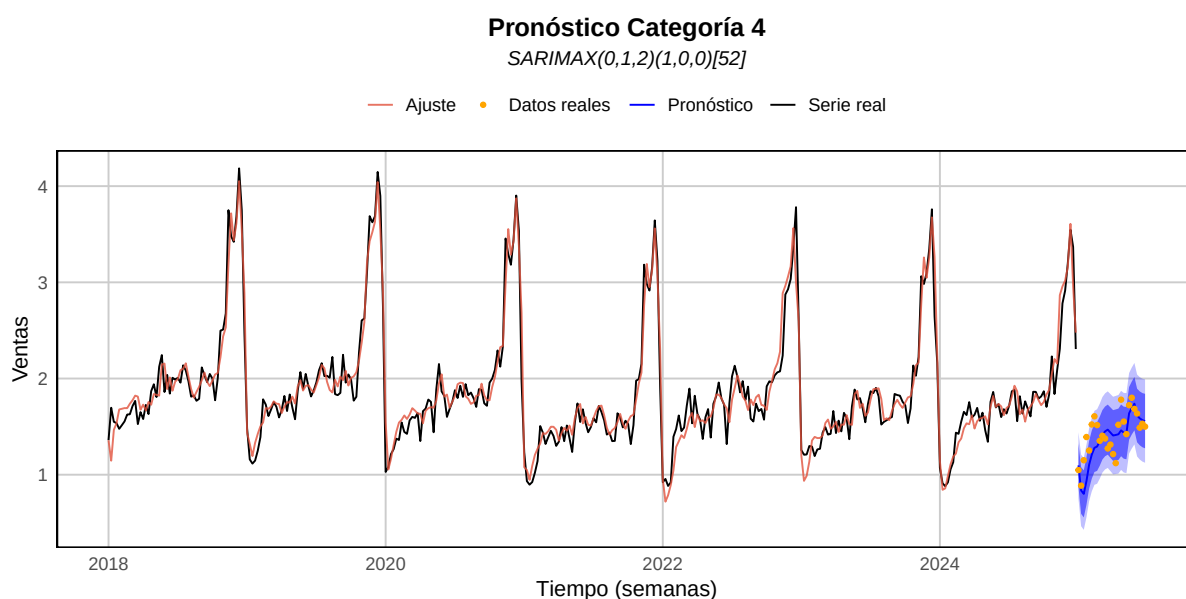


Figura 3.14.

Pronósticos e intervalos de predicción (80% y 95%) correspondientes a la categoría 4.

Tabla 3.7.*Pronóstico de la Categoría 4 para las semanas 2025.001–2025.026.*

Semana	Pronóstico	LI 80%	LS 80%	LI 95%	LS 95%	Real	Error Absoluto
2025.001	1.097	0.872	1.322	0.752	1.442	1.048	0.049
2025.002	0.836	0.593	1.079	0.465	1.207	0.887	0.051
2025.003	0.801	0.556	1.045	0.427	1.175	1.151	0.350
2025.004	0.933	0.687	1.179	0.557	1.310	1.392	0.459
2025.005	1.104	0.856	1.352	0.725	1.483	1.252	0.148
2025.006	1.209	0.959	1.458	0.827	1.591	1.522	0.314
2025.007	1.280	1.028	1.531	0.895	1.665	1.607	0.327
2025.008	1.293	1.040	1.547	0.906	1.681	1.516	0.222
2025.009	1.349	1.094	1.604	0.959	1.739	1.353	0.004
2025.010	1.415	1.158	1.672	1.022	1.808	1.406	0.010
2025.011	1.451	1.192	1.709	1.055	1.846	1.376	0.074
2025.012	1.469	1.208	1.729	1.070	1.867	1.273	0.196
2025.013	1.438	1.176	1.700	1.037	1.839	1.313	0.125
2025.014	1.407	1.144	1.671	1.004	1.811	1.214	0.193
2025.015	1.413	1.148	1.679	1.008	1.819	1.120	0.293
2025.016	1.421	1.153	1.688	1.012	1.829	1.519	0.099
2025.017	1.461	1.192	1.730	1.050	1.872	1.780	0.319
2025.018	1.437	1.166	1.707	1.023	1.850	1.552	0.116
2025.019	1.433	1.160	1.705	1.016	1.849	1.422	0.011
2025.020	1.638	1.365	1.912	1.220	2.057	1.723	0.085
2025.021	1.693	1.418	1.968	1.272	2.114	1.800	0.107
2025.022	1.739	1.462	2.016	1.315	2.162	1.684	0.054
2025.023	1.616	1.337	1.894	1.190	2.042	1.636	0.020
2025.024	1.586	1.306	1.866	1.158	2.015	1.489	0.097
2025.025	1.567	1.285	1.849	1.136	1.998	1.530	0.037
2025.026	1.558	1.274	1.841	1.124	1.991	1.500	0.058

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

Esta investigación aborda la evaluación de la influencia de la integración de los eventos comerciales como variables exógenas en la mejora de la precisión de modelos de pronósticos de demanda, en dólares, del sector retail. Mediante la comparación entre modelos de pronósticos tradicionales de la familia ARIMA estacional y SARIMAX, con el objetivo de demostrar cómo los factores externos permiten una estimación más precisa, un ajuste más parsimonioso y obteniendo una mejor gestión a nivel comercial al incluir factores relevantes para el negocio. A diferencia de enfoques tradicionales que se limitan a comportamientos históricos, el presente estudio evidencia el aporte de la consideración de factores externos y permite analizar el comportamiento de las ventas desagregadas, a nivel de categorías.

4.1 Conclusiones

Posterior a la aplicación de las metodologías planteadas empleadas en la presente investigación y tras examinar los resultados, se presentan a continuación las conclusiones más relevantes del estudio:

- Los datos obtenidos verifican la hipótesis de participaciones constantes que no se sostienen en el tiempo y evidencian la necesidad de gestionar la oferta comercial de manera diferenciada por categoría.
- El análisis de las frecuencias de cada categoría muestra que las ventas no solo responden a

la estacionalidad anual, sino también a subciclos de menor duración ligados a campañas o rotación de productos.

- La categoría 4, en términos de gestión, necesita una planificación continua, trimestral o semestral, además de la anual, porque los subciclos también impactan fuertemente las ventas.
- El análisis realizado permitió contrastar la capacidad predictiva y explicativa de modelos SARIMAX así como SARIMA en el estudio de series temporales con frecuencia semanal. La marcada estacionalidad que se evidenció a partir de la descomposición inicial de las series, junto con los componentes aleatorios y fluctuaciones de tendencia, justificó la incorporación de funciones de Fourier con el fin de capturar patrones complejos de periodicidad.
- Los pronósticos obtenidos mediante los modelos seleccionados presentaron en general un buen grado de ajuste frente a los valores reales. Los errores en la predicción fueron moderados y también dentro de los intervalos de confianza que se esperaban. En particular, la Categoría 1 reflejó mayores desviaciones durante algunos periodos. Esto se atribuye a factores del negocio como el mal desempeño de ciertas colecciones y el efecto de rebajas no vendidas, destacando así la importancia de combinar el análisis estadístico con el contexto del negocio.
- Las Campañas 4 y 5, destacaron al constituirse en los principales determinantes exógenos de la dinámica de ventas, ejerciendo un efecto inmediato como también persistente en todas las categorías de acuerdo con el test de causalidad. Por su parte, las Campañas 1 y 2, mostraron efectos tardíos, explicados por el incremento de la demanda al final de estas. La Campaña 3 no exhibió significancia estadística pese a que mantiene relevancia estratégica desde la perspectiva del negocio.

4.2 Recomendaciones

Tras haber cumplido los objetivos definidos en esta investigación, se identificaron oportunidades de mejora orientadas a la ampliación del alcance de los resultados obtenidos y que sea fructífera. En función de los hallazgos obtenidos, se presentan las recomendaciones principales para el fortalecimiento de la práctica empresarial como a la proyección de futuras líneas de investigación.

- Extender la metodología a otras líneas de productos que cuenten con suficiente data histórica, evaluando su aplicabilidad y beneficios en diferentes categorías.
- En escenarios con mayor volumen de datos, se recomienda explorar enfoques más avanzados, como técnicas de aprendizaje automático, que permitan complementar la interpretabilidad con mayor poder predictivo.
- Integrar los hallazgos en la planificación de inventarios y campañas de marketing para maximizar la efectividad comercial.
- Identificar las categorías con frecuencia trimestral o semestral con el objetivo de realizar un seguimiento diferenciado.
- El costo computacional puede aumentar con modelos de diversos regresores y con múltiples armónicos de Fourier. Se recomienda la implementación de métodos de reducción de dimensionalidad o selección de variables, así como también técnicas de selección automática de armónicos, que posibiliten balancear la precisión y la eficiencia.

BIBLIOGRAFÍA

- Alharbi, F. R. and Csala, D. (2022). A seasonal autoregressive integrated moving average with exogenous factors (sarimax) forecasting model-based time series approach. *Inventions*.
- Bobbitt, Z. (2020). Ljung-box test: Definition + example. *Statology*.
- Box, G. E. P, Jenkins, G. M., Reinsel, G., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley, 5th edition.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer, 2nd edition.
- Chiba, G. I. and Luna, M. M. M. (2020). Modelos de séries temporais: A análise da acurácia das previsões da demanda de uma linha de produtos em empresa do setor do vestuário. *Revista Gestão da Produção Operações e Sistemas*, 15.
- Chopra, S. y Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation*. Pearson.
- Esri (2025). Cómo funciona análisis de agrupamiento.
- Fildes, R., Ma, S., and Kolassa, S. (2019). Retail forecasting: Research and practice. *MPRA Paper*.
- Ghareeb, S., Mahyoub, M., and Mustafina, J. (2023). A comparative time series analysis of the different categories of items based on holidays and other events. In *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*.

- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.
- Haque, M. S., Amin, M. S., and Miah, J. (2023). Retail demand forecasting: A comparative study for multivariate time series. <https://arxiv.org/abs/2308.11939>. arXiv preprint arXiv:2308.11939.
- Hasan, M. R., Kabir, M. A., Shuvro, R. A., and Das, P. (2022). A comparative study on forecasting of retail sales.
- Holm, S. (2024). Arima explained: qué es y cómo funciona el modelo arima. <https://www.morpher.com/es/blog/arima-explained#:~:text=Esencialmente%2C%20especifica%20que%20el%20valor%20actual%20de,p%20observaciones%20pasadas%20para%20predecir%20valores%20futuros>.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne, Australia.
- Ibe, O. C. (2014). Chapter 11 - linear systems with random inputs. In Ibe, O. C., editor, *Fundamentals of Applied Probability and Random Processes*. Academic Press.
- Khalid, O. (2024). Short-term and long-term product demand forecasting with time series models. *Journal of Trends in Financial and Economics*, 1(3).
- Khan, S. I. and Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. *Journal of Big Data*.
- Khandelwal, V., Patel, D., and Malani, D. V. (2023). Process of sales prediction using arima and sarimax to forecast. *IJSDR*.

Kotz, S. and Johnson, N. L. (2000). *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Hoboken, NJ.

Lee, C.-F. (2020). Time-series analysis: Components, models, and forecasting. In *Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning*. World Scientific Publishing Co.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, pages 129–137.

Meca, I. and Belmonte, C. (2018). Ruido blanco. R Pubs (RStudio). Disponible en: <https://rpubs.com/Meca/376836>.

Morpher (2024). Arima explicado: Cómo el modelo de promedio móvil integrado autorregresivo predice las tendencias del mercado. <https://www.morpher.com/es/blog/arima-explained>.

Rao, A. and Meer, P. (2016). Simpson's paradox: Examples. *PMC*.

Rico, V. A. (2020). Estacionariedad. definición y test en r.

Rodó, P. (2019). Modelo arma. Economipedia Diccionario económico.

Rojas, K. (2022). Análisis de series de tiempo para ciencias naturales. capítulo 8: Componentes de las series de tiempo. https://bookdown.org/keilor_rojas/CienciaDatos/an%C3%A1lisis-de-series-de-tiempo.html.

Smith, T. (2024). Autocorrelation: What it is, how it works, tests.

Vellaisamy, P. (2014). Simpsons paradox and collapsibility. *arXiv*.

Waples, J. (2024). La trama q-q: qué significa y cómo interpretarlo. DataCamp Tutorial (en español).

Özden Gür Ali, Sayn, S., van Woensel, T., and Fransoo, J. (2009). Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10):12340–12348.

APÉNDICES

APÉNDICE A

A.1 Variaciones porcentuales interanuales de venta original

Tabla A.1.

Variaciones porcentuales interanuales de ventas -- Categoría 1 (sin imputar)

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-2,9%	4,7%	-6,3%	-10,7%	27,8%	-9,3%
02	-0,9%	5,7%	-8,6%	7,3%	2,3%	0,5%
03	-4,8%	-54,6%	121,8%	5,6%	-5,8%	3,1%
04	1,5%	-98,9%	7372,1%	28,0%	-9,7%	-1,2%
05	3,9%	-72,9%	222,9%	11,1%	0,6%	-6,8%
06	-5,2%	-27,8%	42,1%	-0,2%	0,1%	-2,5%
07	2,9%	-24,8%	19,6%	9,4%	-6,4%	-4,6%
08	2,7%	-13,6%	13,4%	-2,6%	2,2%	-4,5%
09	-3,4%	-2,8%	5,0%	-2,4%	5,1%	-7,6%
10	-9,1%	-1,2%	8,6%	2,4%	-3,0%	-10,8%
11	0,6%	-15,8%	11,0%	-0,1%	4,1%	-6,7%
12	3,5%	-5,5%	4,7%	1,0%	-6,6%	2,2%
Total general	-0,8%	-24,4%	27,9%	3,3%	-0,3%	-3,8%

Tabla A.2.*Variaciones porcentuales interanuales de ventas -- Categoría 2 (sin imputar)*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-14,8%	-33,7%	21,6%	-8,4%	37,3%	-9,7%
02	11,8%	-14,6%	-31,7%	58,7%	0,2%	24,1%
03	-18,7%	-63,9%	143,7%	12,6%	19,1%	36,2%
04	-8,7%	-100,0%	0,0%	80,2%	-16,7%	50,0%
05	-26,0%	-97,4%	2115,1%	18,2%	-11,3%	78,2%
06	-29,0%	-69,1%	176,2%	-2,0%	8,7%	5,6%
07	-24,4%	-66,4%	111,8%	24,5%	-17,2%	-2,2%
08	-20,1%	-53,0%	39,8%	29,8%	6,4%	-31,9%
09	-40,1%	1,2%	-19,3%	64,1%	7,2%	-20,1%
10	-22,7%	-27,1%	21,0%	21,9%	-2,3%	-31,2%
11	-24,1%	-32,5%	44,4%	-10,6%	4,6%	-29,6%
12	-7,4%	-26,2%	36,3%	-1,9%	-2,9%	-15,9%
Total general	-20,0%	-47,6%	51,3%	17,0%	1,8%	-0,8%

Tabla A.3.*Variaciones porcentuales interanuales de ventas -- Categoría 3 (sin imputar)*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-7,9%	7,8%	-10,4%	-6,6%	59,8%	-8,5%
02	3,3%	1,2%	-14,0%	44,0%	-2,5%	5,7%
03	-11,1%	-56,5%	127,2%	30,2%	3,5%	-0,2%
04	1,4%	-100,0%	0,0%	70,5%	3,8%	11,6%
05	-7,3%	-91,4%	686,4%	88,9%	-1,3%	2,4%
06	-4,0%	-36,8%	48,6%	50,5%	-7,4%	-1,3%
07	1,9%	-28,8%	24,3%	51,7%	-2,4%	-11,3%
08	-5,7%	-9,3%	7,1%	41,2%	7,0%	-13,0%
09	4,4%	-8,8%	14,9%	21,5%	11,7%	-2,5%
10	-1,4%	-10,3%	15,2%	19,4%	-2,1%	4,7%
11	-2,3%	-15,9%	14,1%	14,5%	6,7%	-4,6%
12	10,3%	-11,9%	11,7%	10,6%	-7,6%	1,5%
Total general	-0,6%	-28,2%	29,5%	30,9%	2,3%	-1,5%

Tabla A.4.*Variaciones porcentuales interanuales de ventas -- Categoría 4 (sin imputar)*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-20,2%	25,3%	-23,0%	-22,2%	35,1%	-31,6%
02	-0,3%	12,0%	-27,7%	19,1%	-15,5%	5,2%
03	0,7%	-55,3%	100,4%	13,1%	-12,1%	9,9%
04	2,7%	-99,8%	30087,5%	28,3%	-0,5%	6,2%
05	-11,1%	-81,9%	339,5%	15,4%	-14,4%	13,4%
06	-1,9%	-36,5%	22,5%	15,1%	-20,0%	22,3%
07	0,6%	-37,3%	4,8%	48,6%	-39,7%	44,9%
08	9,8%	-25,3%	-3,7%	24,4%	-14,1%	10,8%
09	-4,4%	6,2%	-23,4%	6,2%	7,8%	-11,4%
10	-12,0%	17,0%	-20,6%	16,6%	-10,1%	-14,8%
11	-3,2%	-18,2%	-10,4%	5,1%	11,3%	-12,5%
12	3,3%	-15,9%	2,1%	1,1%	-10,4%	8,9%
Total general	-2,4%	-25,7%	7,9%	12,3%	-8,7%	3,3%

A.2 Variaciones porcentuales interanuales de ventas imputadas

Tabla A.5.

Variaciones porcentuales interanuales de ventas -- Categoría 1

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-2,9%	0,2%	-2,1%	0,1%	14,0%	-9,3%
02	-0,9%	-0,2%	-3,2%	7,3%	2,3%	0,5%
03	-4,8%	-0,2%	0,9%	5,6%	-5,8%	3,1%
04	1,5%	-0,1%	-0,6%	4,6%	-9,7%	-1,2%
05	3,9%	0,1%	-0,6%	-2,3%	0,6%	-6,8%
06	-5,2%	-0,2%	2,8%	-0,2%	0,1%	-2,5%
07	2,9%	-3,2%	-7,1%	9,4%	-6,4%	-3,9%
08	2,7%	-3,0%	1,1%	-2,6%	2,2%	-4,5%
09	-1,5%	1,1%	-0,9%	-2,4%	5,1%	-7,6%
10	-4,7%	4,1%	-1,7%	2,4%	-3,0%	-10,8%
11	0,6%	-0,9%	-5,7%	-0,1%	4,1%	-6,7%
12	3,5%	-3,4%	2,3%	1,0%	-6,6%	2,2%
Total general	-0,3%	-0,7%	-1,0%	1,6%	-0,9%	-3,7%

Tabla A.6.*Variaciones porcentuales interanuales de ventas -- Categoría 2*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-14,8%	-20,8%	1,8%	-8,4%	37,3%	-9,7%
02	11,8%	-17,9%	-28,9%	58,7%	0,2%	24,1%
03	-18,7%	-18,2%	7,5%	12,6%	19,1%	36,2%
04	-8,7%	-14,1%	12,9%	-5,4%	-16,7%	50,0%
05	-26,0%	-13,5%	11,8%	-29,5%	-11,3%	78,2%
06	-29,0%	-15,7%	1,2%	-2,0%	8,7%	5,6%
07	-24,4%	-6,3%	-24,0%	24,5%	-17,2%	-2,2%
08	-20,1%	-12,8%	-24,7%	29,8%	6,4%	-31,9%
09	-40,1%	7,5%	-24,1%	64,1%	7,2%	-20,1%
10	-15,6%	-22,1%	3,6%	21,9%	-2,3%	-31,2%
11	-24,1%	-6,1%	3,8%	-10,6%	4,6%	-29,6%
12	-7,4%	-19,7%	25,3%	-1,9%	-2,9%	-15,9%
Total general	-19,4%	-13,6%	-1,4%	8,2%	1,8%	-0,8%

Tabla A.7.*Variaciones porcentuales interanuales de ventas -- Categoría 3*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-7,9%	0,3%	-3,7%	-2,2%	52,6%	-8,5%
02	3,3%	0,5%	-13,4%	44,0%	-2,5%	5,7%
03	-11,1%	0,3%	-1,4%	30,2%	3,5%	-0,2%
04	1,4%	-0,3%	-1,8%	14,6%	3,8%	11,6%
05	-7,3%	-0,7%	-1,6%	30,4%	-1,3%	2,4%
06	-4,0%	-1,0%	-5,2%	50,5%	-7,4%	-1,3%
07	1,9%	-3,7%	-8,2%	51,7%	-2,4%	-8,4%
08	-5,7%	3,3%	-6,0%	41,2%	7,0%	-13,0%
09	4,4%	-5,6%	11,0%	21,5%	11,7%	-2,5%
10	4,6%	-5,2%	2,6%	19,4%	-2,1%	4,7%
11	-2,3%	1,1%	-5,1%	14,5%	6,7%	-4,6%
12	10,3%	-9,0%	8,2%	10,6%	-7,6%	1,5%
Total general	-0,1%	-2,5%	-0,8%	25,5%	2,2%	-1,3%

Tabla A.8.*Variaciones porcentuales interanuales de ventas -- Categoría 4*

MES	2019/2018	2020/2019	2021/2020	2022/2021	2023/2022	2024/2023
01	-20,2%	-6,4%	-13,1%	-7,7%	35,1%	-23,9%
02	-0,3%	-7,0%	-10,8%	16,3%	-15,5%	0,6%
03	0,7%	-6,3%	-12,2%	23,1%	-12,1%	9,9%
04	2,7%	-7,3%	-12,1%	8,3%	-1,3%	7,1%
05	-11,1%	0,3%	-18,5%	12,2%	-1,0%	-2,0%
06	-1,9%	-8,6%	-11,7%	10,9%	0,1%	-2,4%
07	0,6%	-10,0%	-12,3%	23,6%	-7,5%	-5,5%
08	9,8%	-8,4%	-22,2%	25,6%	-14,1%	10,8%
09	-0,2%	-9,9%	-14,5%	7,5%	7,8%	1,0%
10	-5,2%	6,3%	-24,0%	24,3%	-10,1%	7,8%
11	-3,2%	-6,7%	-8,0%	-10,3%	11,3%	-12,5%
12	3,3%	-11,3%	-7,4%	5,7%	-4,5%	2,1%
Total general	-1,5%	-6,9%	-13,3%	10,5%	-2,3%	-0,4%