



UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA

Maestría Estadística Aplicada

Tema: Predicción de ventas y optimización de presupuestos digitales en comercio electrónico mediante modelos Prophet: una aplicación en el mercado chileno

Autora: Lic. Anabel Sonia Chiappello
Directora: Mag. Fernanda Méndez
Año: 2026

Agradecimientos

Este camino transitado para llegar hoy tiene 20 años desde el inicio del cursado. ¡Sí, pasaron 20 años! En este tiempo transité por 7 trabajos en relación de dependencia, muchas experiencias como profesional independiente y más de 10 años como docente en diferentes espacios.

También, me casé, tuve dos hijos y a este equipo es al primero que quiero agradecer, por sostenernos los 4 como familia en tiempo de mucha dedicación para alcanzar los objetivos profesionales y personales con los que me fui desafiando. ¡Gracias Guille, Lucio y Victoria!

Quiero sumar a mis padres (Raquel y Rubén), que me dieron la posibilidad de estudiar la carrera de grado que luego me trajo hasta aquí. Junto con ellos quiero agradecer a mi hermana Betsabé que siempre estuvo y está presente a lo largo de mi vida.

Quiero agradecer también, a esas amigas que siempre creyeron en que lo iba a lograr y me alentaron cada vez que volvía con ellas a charlar del tema: Caro A., Luciana, Gisela, Leticia, Caro C., Paula y Florentina.

Quiero resaltar y destacar a dos personas con las que armamos un gran equipo de estudio y apoyo, con las que logré vencer las barreras necesarias para hoy poder hasta acá: Caro Costa y Nanda Mendez. Sin ellas este camino no lo hubiese transitado.

Por último, quiero agradecerme a mí por creer que siempre vale la pena volver intentar, por ser paciente, perseverante, y por estar convencida que cerrar etapas ayuda a la evolución personal. ¡Gracias!

Resumen

En esta tesis se desarrolla un marco analítico para la predicción de ventas y la optimización del presupuesto en campañas de marketing digital, aplicado a una empresa de comercio electrónico que opera en Chile. El objetivo principal es estimar las ventas diarias a partir de variables relevantes (Identificación de Evento especial, Identificación de *Cyber*, Número de visitantes al sitio web, Número de visitas totales al sitio, Compras realizadas en el sitio web, Número total de unidades compradas, Presupuesto o inversión ingresada plataforma de *Google Ads*, Costo facturado en *Google Ads*, Número de *clicks* realizados sobre el anuncio o campaña, Cantidad de veces que se muestra el anuncio) de campañas publicitarias en *Google Ads*, y utilizar dichas predicciones para asignar de forma eficiente el presupuesto disponible entre múltiples campañas.

Se emplea el modelo Prophet, desarrollado por *Facebook*, por su capacidad para capturar estacionalidades múltiples y su robustez ante valores atípicos, eventos especiales y cambios estructurales en las series. Existen estudios previos que abordan la predicción de ventas con modelos clásicos y de aprendizaje automático, este trabajo amplía el enfoque al incorporar variables adicionales, tratar las series en forma individualizada por campaña y vincular directamente la predicción con una función objetivo de optimización presupuestaria.

El modelo se ajusta utilizando datos reales, con validación temporal (*train-test*) y métricas específicas para series de tiempo. Los resultados muestran un buen desempeño predictivo y evidencian que el enfoque propuesto permite mejorar la eficiencia en la distribución del presupuesto publicitario. Esta investigación demuestra la aplicabilidad de herramientas de ciencia de datos en la toma de decisiones estratégicas de marketing digital, combinando modelos de predicción y optimización en un entorno real de negocios.

Palabras clave: Predicción de ventas, Prophet, comercio electrónico, marketing digital, optimización presupuestaria.

Tabla de contenidos

1.Introducción	5
2.Objetivos.....	7
3. Metodología y Materiales	8
3.1. Modelo <i>Prophet</i>	8
3.1.1. Definición.....	11
3.1.2. Ajuste de modelos. Estrategia de separación de datos: <i>train-test</i>	17
3.1.3. Evaluación del desempeño predictivo	19
3.2. Modelo de asignación de presupuesto: Función Objetivo	22
3.2.1. Definición de la Función Objetivo	22
3.2.2. Variables de Decisión	23
3.2.3. Restricciones	23
3.3. Materiales	24
4. Aplicación práctica.....	27
4.1. Análisis descriptivo.....	27
4.2. Modelo <i>Prophet</i> : ajuste.....	46
4.3. Modelo de asignación: optimización Función Objetivo.....	65
5. Discusión	70
7. Anexo.....	76
7.1. Medidas descriptivas por campaña	76
7.2. Medidas de desempeño predictivo del modelo por campaña.....	93
7.3. Resultados optimización de presupuesto	98
7.4. Código <i>Python</i> utilizado	98

1.Introducción

La era digital ha transformado profundamente el panorama empresarial, en especial el del comercio electrónico. La expansión de plataformas digitales, redes sociales y dispositivos móviles ha generado un volumen masivo de datos sobre el comportamiento de los usuarios, modificando radicalmente el enfoque del marketing. Las empresas cuentan hoy con acceso en tiempo real a información detallada sobre las interacciones de los clientes y su respuesta a diversas estrategias publicitarias y de adquisición. Esta posibilidad de anticipar las necesidades y deseos de los clientes, y de entregar productos o servicios de manera satisfactoria, se ha vuelto posible gracias al análisis de estos grandes volúmenes de datos de la interacción digital. Para cualquier negocio inmerso en el marketing digital, el aprovechamiento adecuado de estos volúmenes masivos de datos es ahora una necesidad imperativa.

Para abordar esta necesidad, *Big Data*¹ y automatización se consolidan como aliados clave, potenciados por técnicas analíticas avanzadas —como los modelos de series de tiempo— y por capacidades informáticas de gran escala. La automatización de procesos analíticos y predictivos permite adaptarse con agilidad a los cambios en el comportamiento del consumidor, facilitando respuestas rápidas a las demandas del mercado. Asimismo, la implementación de técnicas de Aprendizaje Automático sobre conjuntos de datos históricos posibilita la automatización y mejora continua de los modelos predictivos utilizados en campañas digitales.

La estimación de ventas constituye uno de los pilares fundamentales del comercio electrónico. La capacidad de prever con precisión las ventas futuras permite no solo anticipar tendencias, personalizar estrategias de marketing y mejorar la planificación, sino también aumentar la eficiencia operativa y la rentabilidad del negocio. La integración de técnicas de series temporales con herramientas de *Big Data* y automatización ofrece una oportunidad única para obtener información valiosa que puede impulsar el crecimiento y la sostenibilidad empresarial en el actual entorno digital.

Esta tesis se desarrolló con el objetivo de profundizar en la predicción de ventas y optimización del presupuesto en campañas digitales. Existen antecedentes en esta temática que propusieron una solución a esta necesidad a través de un enfoque automatizado para la predicción de ventas en comercio electrónico mediante la evaluación de campañas de

¹ Término que se refiere a conjuntos de datos de gran volumen, complejos y registrados a alta velocidad que las herramientas y métodos de procesamiento de datos tradicionales no son capaces de capturar, almacenar, gestionar y analizar de manera eficiente para extraer valor y conocimiento significativo de estos datos para la toma de decisiones

marketing (Costa, 2024). Se abordó mediante la formulación, implementación y comparación de modelos predictivos avanzados, específicamente SARIMAX, VARX y XGBoost, con el fin de identificar y seleccionar el modelo que ofrezca las estimaciones más precisas y confiables de las ventas futuras en el contexto del comercio electrónico. En esta tesis se propone extender los avances sobre esta temática a través de la búsqueda de un modelo que incorpore múltiples estacionalidades para estimar las ventas, utilizando enfoques automatizados capaces de operar sobre cientos de series históricas vinculadas a campañas y sus distintos atributos. Otro aspecto para abordar es poder **optimizar el presupuesto** asignado a **campañas de marketing digital**, tomando como base las estimaciones de **ventas neta**.

Este documento se estructura en siete capítulos, cada uno enfocado en aspectos específicos de la investigación, con el propósito de brindar una comprensión integral del tema. Luego de esta introducción, el capítulo 2 expone objetivos principales y secundarios que guían el estudio, definiendo el alcance y las metas específicas del trabajo. El capítulo 3 comprende la metodología empleada para el desarrollo del modelo predictivo junto con una propuesta de la optimización del presupuesto de marketing. Además, se presentan los datos que se utilizarán y se expondrán en el siguiente capítulo. El capítulo 4 presenta un caso de estudio práctico, en el que se aplica el sistema de pronóstico desarrollado a un escenario real en el comercio electrónico chileno. Este capítulo ilustra la eficacia del modelo en la predicción de ventas netas y su capacidad para optimizar la asignación de recursos en campañas de marketing digital. En el capítulo 5, se presentan la discusión y consideraciones finales, incluyendo una comparación con los resultados obtenidos en otros estudios. El capítulo 6 contiene la lista de referencias bibliográficas, facilitando la profundización en los temas tratados. Por último, el capítulo 7 incluye anexos que complementan y profundizan lo presentado en el trabajo con información adicional, como gráficos y tablas de datos que respaldan los análisis realizados.

2.Objetivos

2.1. Objetivo general

El objetivo principal de esta tesis consiste en desarrollar y evaluar un marco analítico para la predicción de ventas diarias en un comercio electrónico mediante el uso del modelo *Prophet*, incorporando variables relevantes de campañas publicitarias en *Google Ads*². A partir de dichas predicciones, se busca formular un modelo de optimización presupuestaria que permita asignar de manera eficiente los recursos disponibles entre múltiples campañas.

2.2. Objetivos específicos

- Implementar un modelo *Prophet* para predecir las ventas diarias por campaña, considerando múltiples estacionalidades y covariables relevantes, con un horizonte de predicción de hasta 90 días.
- Evaluar la capacidad predictiva del modelo y analizar su robustez ante valores atípicos y eventos especiales (*cyberdays*³, feriados).
- Diseñar una función objetivo para la asignación óptima del presupuesto publicitario, basada en las predicciones obtenidas mediante el Modelo *Prophet*, minimizando el costo de inversión en publicidad.
- Comparar los resultados obtenidos con aquellos alcanzados en estudios previos, donde el modelo *XGBoost*⁴ fue el que mejor desempeño mostró para la mayoría de las campañas; para así, discutir sus implicancias para la toma de decisiones en marketing digital.

² Programa de publicidad online de Google que permite a empresas crear anuncios para llegar a audiencias específicas en el momento exacto en que están buscando productos y servicios. Es una plataforma de publicidad de pago.

³ Eventos de compras en línea donde numerosas tiendas ofrecen descuentos y promociones especiales en sus productos y servicios.

⁴ XGBoost (Extreme Gradient Boosting): Algoritmo de aprendizaje supervisado basado en un conjunto de árboles de decisión que se entrenan de forma secuencial. Destaca por su alta eficiencia y precisión al corregir iterativamente los errores de modelos anteriores mediante técnicas de optimización y regularización.

3. Metodología y Materiales

3.1. Modelo *Prophet*

*Prophet*⁵ fue desarrollado internamente por Sean J. Taylor y Ben Letham en *Facebook* en 2017, con el propósito de facilitar y democratizar la creación de pronósticos. Su diseño busca superar dos barreras principales del análisis de datos tradicional: por un lado, la poca flexibilidad técnica de las herramientas automáticas previas que por su rigidez no permitían que los analistas expertos añadieran su conocimiento sobre el negocio (como eventos especiales o cambios de tendencia); por otro, los métodos más precisos solían ser tan complejos que solo analistas expertos podían usarlos. *Prophet* simplifica este proceso, permitiendo obtener resultados profesionales sin necesidad de un dominio profundo en estadística avanzada.

El diseño de *Prophet* se centra en potenciar la capacidad de los analistas, permitiéndoles integrar su conocimiento del sector para generar predicciones de series temporales que sean tanto precisas como aplicables a la realidad operativa empresarial.

Principales ventajas de *Prophet*

- **Facilidad de uso y automatización:** fue diseñado pensando en la facilidad de uso para analistas y no necesariamente para expertos en series de tiempo. Permite generar pronósticos de manera rápida y con una configuración mínima, automatizando gran parte del proceso de modelado. Esto lo hace ideal para generar cientos de pronósticos para múltiples series temporales de forma masiva y simultánea, pronósticos en escala. Sin embargo, se debe tener presente que, al confiar en parámetros por defecto, se pueden omitir dinámicas particulares de series específicas (como rupturas estructurales, valores atípicos o cambios drásticos de tendencia). También puede haber una menor precisión técnica dado que los modelos ajustados manualmente (como *ARIMA*⁶ o modelos econométricos dedicados) suelen superar a las configuraciones automáticas en series de alta complejidad o comportamiento no lineal. Por otra parte, la automatización detecta patrones estocásticos, pero no interpreta eventos del entorno (campañas de marketing, cambios regulatorios o interrupciones de stock) a menos que un analista los incorpore explícitamente como covariables.

⁵ Taylor, S. J., & Letham, B. (2017). Forecasting at scale.

⁶ *ARIMA* (Autoregressive Integrated Moving Average): Modelo estadístico que predice valores futuros de una serie de tiempo analizando tres factores: sus valores pasados (autorregresivo), la eliminación de tendencias (integrado) y los errores de pronósticos previos (media móvil)

- **Robustez frente a datos atípicos:** Es robusto a la existencia de datos faltantes y a valores atípicos, lo cual es una ventaja significativa frente a los modelos estadísticos clásicos, los cuales exigen una secuencia temporal estricta, uniforme y sin discontinuidades. Mientras que *ARIMA/SARIMA*⁷ requieren técnicas previas de imputación de datos o diferenciación para no desestabilizar las matrices de covarianza, *Prophet* aborda la predicción como un problema de regresión ajustado en función del tiempo, manteniendo la fiabilidad de las predicciones sin alterar la estructura temporal subyacente.
- **Manejo de la estacionalidad y tendencias:** sobresale en la identificación y modelado de patrones estacionales complejos (diarios, semanales, anuales) y cambios en la tendencia. Descompone la serie temporal en componentes de tendencia, estacionalidad y efectos de días festivos, lo que facilita la interpretación de los resultados y permite entender por qué se observan ciertos patrones en días específicos. A diferencia de los métodos de suavizamiento o *SARIMA* que típicamente abordan una sola frecuencia estacional, *Prophet* utiliza series de Fourier para modelar de forma simultánea múltiples capas de estacionalidad (diaria, semanal y anual) de manera flexible.
- **Incorporación de días festivos y eventos especiales:** permite incorporar explícitamente el impacto de días festivos, promociones o eventos especiales (tanto pasados como futuros) que pueden influir significativamente en el comportamiento de la serie. Esto es crucial en el marketing digital y el comercio electrónico, donde las campañas y fechas especiales tienen un impacto directo en las ventas. Facilita la parametrización del efecto de promociones o días festivos especificando ventanas de influencia alrededor de cada fecha, integrándolas directamente como componentes del modelo. En modelos clásicos como *ARIMA* o *VARMAX*⁸, la inclusión de múltiples variables exógenas *dummy* suele sobredimensionar la matriz de parámetros y causar problemas de sobreajuste. *Prophet* facilita la parametrización del efecto de promociones o festivos especificando ventanas de influencia ajustables ($\pm n$ días alrededor del evento), integrándolas de forma nativa y estable en el modelo.
- **Flexibilidad en la modelización de la tendencia:** la tendencia se modela mediante una función continua a trozos que puede cambiar de pendiente en puntos específicos. Estos puntos de cambio pueden ser detectados automáticamente por el algoritmo o definidos

⁷ ARIMA (Seasonal Auto-Regressive Integrated Moving Average): Modelo estadístico de series temporales que extiende el modelo ARIMA incorporando componentes estacionales para capturar patrones periódicos y tendencias en los datos

⁸ VARMAX (Vector Autoregressive Moving Average with Exogenous Variables): Modelo estadístico multivariante de series temporales que extiende el modelo VARMA al incluir variables explicativas o externas (exógenas) para predecir múltiples variables interrelacionadas simultáneamente.

manualmente, ofreciendo gran flexibilidad para adaptarse a cambios estructurales en los datos. No asume una tendencia fija o global como los modelos *ARIMA*.

- **Generación de intervalos de confianza:** Proporciona intervalos de predicción calculados mediante simulación Monte Carlo (o muestreo Bayesiano MCMC⁹), lo que permite cuantificar la incertidumbre asociada al pronóstico. El modelo proyecta cientos de trayectorias futuras considerando la varianza histórica observada en los cambios de tendencia (*changepoints*) y el error residual, obteniendo los límites del intervalo a partir de los percentiles empíricos de las simulaciones.
- **Velocidad computacional:** comparado con otros métodos de pronóstico de series temporales multivariados o iterativos clásicos (*VARMAX* o *ARIMA*) que sufren de alta complejidad computacional al aumentar las dimensiones, *Prophet* tiende a ser computacionalmente eficiente, lo que lo hace adecuado para escenarios que requieren predicciones en tiempo real o el procesamiento de un gran volumen de series.
- **Interpretación intuitiva:** al descomponer la serie en sus componentes fundamentales, los resultados de *Prophet* son relativamente fáciles de interpretar en comparación con modelos *ARIMA*, incluso para usuarios sin un profundo conocimiento estadístico. Esto facilita la comunicación de los hallazgos a las partes interesadas del negocio.

Prophet es un modelo que fusiona principios de la estadística con enfoques de *Machine Learning*¹⁰. Es por eso que algunos expertos lo clasifican como un modelo híbrido o como un modelo estadístico que utiliza técnicas de aprendizaje automático para su ajuste y robustez.

Los puntos en común con *Machine Learning*, son los siguientes:

- **Aprendizaje a partir de datos:** en su esencia, el *Machine Learning* implica que un algoritmo "aprende" patrones y relaciones a partir de los datos para hacer predicciones o tomar decisiones. *Prophet* cumple con esto dado que ajusta sus componentes (tendencia, estacionalidad, días festivos) basándose en los datos históricos.
- **Optimización de parámetros:** aunque gran parte de la configuración es automática, *Prophet* involucra la optimización de parámetros subyacentes para lograr el mejor ajuste a los datos. Esto es una característica común en los algoritmos de *Machine Learning*.
- **Flexibilidad y no-linealidad:** a diferencia de muchos modelos estadísticos clásicos para series de tiempo (como *ARIMA*), que a menudo asumen linealidad o estacionalidad fija,

⁹ MCMC (Markov Chain Monte Carlo): Algoritmo de simulación que combina muestreo aleatorio (Monte Carlo) con secuencias donde cada estado depende del anterior (Cadenas de Markov). Se utiliza en inferencia Bayesiana para aproximar distribuciones de probabilidad complejas mediante la generación de miles de muestras representativas de los parámetros del modelo.

¹⁰ Machine Learning o Aprendizaje Automático es una rama de la Inteligencia Artificial (IA) que dota a los sistemas informáticos de la capacidad de aprender y mejorar a partir de la experiencia, sin ser programados explícitamente para cada tarea.

Prophet puede modelar tendencias no lineales y estacionalidades complejas de forma flexible, lo que es una característica de muchos modelos de *Machine Learning*.

- **Uso de componentes de regresión:** *Prophet* se basa en un modelo aditivo que se puede ver como una forma de regresión donde las características de entrada son el tiempo y otros factores (como los indicadores de días festivos). La regresión, históricamente originada en la estadística clásica, es una de las aplicaciones fundamentales del *Machine Learning*, dado que permite modelar la variable de respuesta como una función paramétrica continuada en el tiempo.
- **Diseñado para la automatización y la escala:** *Prophet* fue creado para manejar una gran cantidad de series de tiempo de forma automática y robusta, lo que lo alinea con el espíritu de muchas soluciones de *Machine Learning* para problemas a gran escala.

Los argumentos que lo acercan a los modelos estadísticos tradicionales son:

- **Fundamentos estadísticos claros:** *Prophet* se basa en un modelo aditivo generalizado, que tiene una sólida base en la estadística. Descompone la serie de tiempo en componentes interpretables (tendencia, estacionalidad, días festivos y errores), lo cual es una práctica común en el análisis de series de tiempo.
- **Interpretabilidad:** los componentes de *Prophet* son interpretables, se puede entender cómo cada factor contribuye a la predicción final.
- **Enfoque paramétrico:** aunque flexible, *Prophet* utiliza funciones matemáticas definidas (como la regresión logística o lineal a trozos para la tendencia, y series de Fourier para la estacionalidad) para modelar los componentes, lo que lo diferencia de algoritmos de *Machine Learning* no paramétricos.

3.1.1. Definición

Prophet es un modelo aditivo generalizado¹¹, descompone la serie temporal en diversos componentes que representan patrones diferenciados en los datos. La predicción final resulta de la suma de estos componentes.

El modelo general puede representarse como

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (3.1.1.1)$$

donde,

$g(t)$ es la componente de tendencia. Representa la tendencia no lineal de la serie temporal.

¹¹ Rafferty, G. (2021).

$s(t)$ es la componente estacional. Captura los patrones periódicos que se repiten en la serie temporal.

$h(t)$ es la componente de Días Festivos y Eventos (*Holiday Component*). Modela el impacto de eventos especiales e irregulares (como días festivos nacionales, promociones específicas, lanzamientos de productos, etc.) que no se ajustan a patrones estacionales regulares.

$e(t)$ es el término de error.

Es importante mencionar que puede ser un modelo tanto aditivo como multiplicativo. Por defecto, es un modelo aditivo generalizado. Vale resaltar que el modo elegido afecta a todos los componentes estacionales y de días festivos, mientras que la tendencia siempre es aditiva (lineal o logística).

A continuación, se describen en profundidad cada una de las componentes del modelo.

Componente de tendencia $g(t)$

Prophet ofrece dos funciones de tendencia:

- **Crecimiento logístico (*Logistic Growth*)**: se utiliza cuando la serie temporal exhibe un crecimiento que tiende a saturarse en un límite superior (capacidad). Es útil para modelar fenómenos que tienen una cota superior que no es posible traspasar, como la adopción de un producto que no puede superar una cierta cuota de mercado. En este caso, la tendencia se modela con una curva logística, que se aproxima a una asíntota superior.

De aquí que la tendencia se modela mediante la función logística que se define como:

$$g(t) = \frac{C}{1+e^{-k(t-m)}} \quad (3.1.1.2)$$

donde:

- C es la capacidad o valor máximo de la curva. *Prophet* puede inferir este parámetro a partir de los datos o bien, el usuario puede especificarlo explícitamente.
- k es la constante de crecimiento, refiere a la inclinación de la curva.
- m es un parámetro de *offset*, es decir, valor constante que ajusta la posición vertical de la curva de tendencia en el eje del tiempo.

Prophet permite que los parámetros k (tasa de crecimiento) y m (punto de *offset*) cambien en el tiempo a través de puntos de cambio (*changepoints*). Entonces, si hay S puntos de cambio en los tiempos t_j (donde $j = 1, \dots, S$), la función logística se vuelve más compleja para reflejar

estos cambios de pendiente. La tasa de crecimiento k se ajusta en cada punto de cambio t_j por un valor δ_j . Así, la tasa de crecimiento en cualquier momento t se puede expresar como:

$$k(t) = k_0 + \sum_1^S \delta_j * I(t > t_j) \quad (3.1.1.3)$$

donde

k_0 es la tasa de crecimiento inicial

δ_j es el ajuste de la tasa de crecimiento en los puntos de cambio

$I(\cdot)$ es la función indicadora (1 si la condición es verdadera, 0 en caso contrario)

Uno de los modelos más conocidos en marketing y economía es el Modelo de Difusión de Bass, publicado por Frank Bass en 1969, se utiliza para predecir cómo se adopta un nuevo producto o tecnología en el mercado a lo largo del tiempo. El gran aporte de Bass fue explicar por qué las ventas de productos nuevos suelen seguir una gráfica con forma de "S": al principio crecen despacio, luego explotan en ventas y finalmente se estabilizan cuando el mercado se satura. Este es el punto de encuentro con la estimación de la tendencia como una función de crecimiento logístico en modelos *Prophet*, ambos comparten la misma esencia conceptualmente hablando. En Modelo de Difusión de Bass, las ventas no pueden superar la capacidad m (el total de personas en el mercado) y en *Prophet* la curva de tendencia no puede superar la capacidad $C(t)$ fija cuando se utilizar el crecimiento logístico para la tendencia. Ambos modelos sostienen que ningún producto puede crecer infinitamente al mismo ritmo, sino que en momento del tiempo se alcanza el “techo” del mercado.

- **Crecimiento lineal a trozos (*Piecewise Linear Growth*)**: es la opción más utilizada por defecto. Este modelo de tendencia se aplica cuando no existe un límite superior definido para el crecimiento de la serie temporal. Modela la tendencia como una serie de segmentos lineales conectados, donde la pendiente puede cambiar en puntos específicos, llamados puntos de cambio (*changepoints*). *Prophet* puede detectar estos puntos de cambio automáticamente a partir de los datos, o bien, el usuario puede especificarlos.

En términos matemáticos, modelar la tendencia como una función lineal cuya pendiente puede cambiar en puntos de cambio discretos, se define como:

$$g(t) = (k_0 + \mathbf{a}(t)^T \boldsymbol{\delta})t + (m_0 + \mathbf{a}(t)^T \boldsymbol{\gamma}) \quad (3.1.1.4)$$

donde:

- k_0 es la tasa de crecimiento inicial

- $\delta = (\delta_1, \dots, \delta_S)^T$ es un vector (transpuesto) de ajustes de la tasa de crecimiento en los puntos de cambio
- $\mathbf{a}(t) = (I(t > t_1), \dots, I(t > t_S))^T$ es un vector (transpuesto) que indica qué puntos de cambio han ocurrido hasta el tiempo t .
- m_0 es un *offset inicial*, un valor constante que ajusta la posición vertical de la curva de tendencia en el eje del tiempo, especialmente al inicio de la serie temporal.
- γ es un vector de ajustes de *offset* para mantener la continuidad de la función en los puntos de cambio.

La flexibilidad en la tendencia es una de las grandes ventajas de *Prophet*, ya que permite capturar cambios estructurales en los datos de forma robusta.

Componente estacional $s(t)$

Captura los patrones periódicos que se repiten en la serie temporal. *Prophet* puede modelar múltiples estacionalidades:

- Estacionalidad anual: patrones que se repiten cada año (por ejemplo, ventas que aumentan en Navidad).
- Estacionalidad semanal: patrones que se repiten cada semana (por ejemplo, más tráfico web los fines de semana).
- Estacionalidad diaria: patrones que se repiten cada día (por ejemplo, más ventas *online* en la noche).

Prophet ajusta una serie de términos de Fourier a los datos con el fin de modelar los ciclos estacionales. El orden de Fourier —es decir, el número de términos utilizados— determina la flexibilidad del modelo para capturar patrones estacionales más complejos.

La serie de Fourier dice que cualquier función periódica se puede descomponer y representar como una suma de ondas sinusoidales (senos y cosenos) simples, con diferentes frecuencias, amplitudes y fases.

Una serie de Fourier general para una función periódica $f(t)$ con un período P se expresa de la siguiente manera:

$$f(t) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right] \quad (3.1.1.5)$$

donde:

a_0 es el término constante. Representa el valor promedio de la función durante un período completo.

$\cos(2\pi nt)$ y $\sin(2\pi nt)$ son componentes sinusoidales. Son las ondas base (senos y cosenos).

n es un número de armónico u "orden de Fourier" que determina la frecuencia de cada onda sinusoidal.

Cuando $n = 1$, es el primer armónico o componente fundamental. Su período es igual al período original de la función P .

Cuando $n = 2$, es el segundo armónico, con una frecuencia doble y un período de $P/2$.

Cuanto mayor es n , más rápidamente oscila la onda, lo que permite capturar detalles más finos en el patrón periódico.

$\frac{2\pi}{P}$: es la frecuencia angular fundamental.

t : es la variable de tiempo.

a_n y b_n son los coeficientes de Fourier.

En términos del modelo *Prophet*, la estacionalidad queda expresada como:

$$s(t) = \sum_1^N \left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (3.1.1.6)$$

donde N es el orden Fourier que se especifica.

a_n y b_n representan las amplitudes de cada componente de coseno y seno, respectivamente. Estos coeficientes determinan cuánto contribuye cada onda sinusoidal a la forma total de la función periódica. Se calculan mediante integrales de la función original, lo que esencialmente mide la correlación entre la función original y cada armónico.

En general, se usa el valor por defecto orden 10 para la estacionalidad anual y orden 3 para la estacionalidad semanal. El orden 10 genera una combinación de 20 variables regresoras (10 senos y 10 cosenos) en la serie de Fourier, esto le otorga al modelo la flexibilidad necesaria para trazar curvas complejas y capturar patrones detallados anuales sin caer en un sobreajuste excesivo. Mientras que, un orden 3 genera 6 variables regresoras (3 senos y 3 cosenos) y es suficiente para aproximar curvas suaves de pocos días.

Componente de días festivos y eventos (*Holiday Component*) $h(t)$

Modela el impacto de eventos especiales e irregulares (como días festivos nacionales, promociones específicas, lanzamientos de productos, etc.) que no se ajustan a patrones estacionales regulares.

Estos eventos se incorporan al modelo mediante variables indicadoras (*dummies*). *Prophet* estima su efecto de forma aditiva o multiplicativa sobre la serie, según la configuración establecida.

La contribución de los días festivos al modelo puede expresarse matemáticamente como:

$$h(t) = \sum_{j=1}^J \beta_j \cdot D_j(t) \quad (3.1.1.7)$$

donde:

J : es el número total de festivos/eventos definidos.

β_j : es el parámetro estimado para el efecto del día festivo j (estimado a partir de los datos).

$D_j(t)$: es una función indicadora (*dummy*) que vale 1 si el día t cae en el rango de influencia del festivo j , y 0 en otro caso.

Componente Residuos $e(t)$

Es el término de error. Representa el ruido residual o las variaciones aleatorias que el modelo no puede explicar con las otras componentes. Se asume que este término sigue una distribución normal con media cero y varianza constante, es decir,

$$e_t \sim N(0, \sigma^2) \quad (3.1.1.8)$$

Los modelos *Prophet* también contemplan la opción de incorporar regresores o covariables adicionales. En este caso, el modelo presentado en 3.1.1.1. se expresa como:

$$y(t) = g(t) + s(t) + h(t) + \sum_{i=1}^k \beta_i x_i(t) + e(t) \quad (3.1.1.9)$$

donde:

$x_i(t)$: valor del i -ésimo regresor en el tiempo t

β_i : coeficiente que representa el impacto del regresor x_i constante para todo t

También es importante definir el modelo *Prophet* multiplicativo, dado que suele ser el que se utiliza para series de tiempo del sector de comercios (*retail*), donde los picos de ventas en ciertos momentos del año (como Navidad) son mucho más altos que el resto de la serie.

Para definir el modelo *Prophet* multiplicativo, es fundamental resaltar que, a diferencia del modelo aditivo donde los componentes se suman, aquí los efectos de la estacionalidad y los eventos externos actúan como porcentajes que amplifican o reducen la tendencia base. El modelo definido en 3.1.1.1 queda representado por la siguiente ecuación:

$$y(t) = g(t) * (1 + s(t) + h(t)) + e(t) \quad (3.1.1.10)$$

donde,

$g(t)$ es la componente de tendencia.

$s(t)$ es la componente estacional.

$h(t)$ es la componente de Días Festivos y Eventos (*Holiday Component*).

$e(t)$ es el término de error.

La incorporación de covariables en el modelo multiplicativo tiene un efecto similar a la estacionalidad y los días festivos o eventos externos. El modelo definido en 3.1.1.9, en su versión multiplicativa queda expresado como:

$$y(t) = g(t) * (1 + s(t) + h(t) + \sum_{i=1}^k \beta_i x_i(t)) + e(t) \quad (3.1.1.11)$$

donde,

$x_i(t)$: valor del i-ésimo regresor en el tiempo t

β_i : coeficiente que representa el impacto del regresor x_i

Si bien *Prophet* utiliza la filosofía compositiva de modelos aditivos generalizados (*GAM*). Desde la perspectiva del pronóstico de series temporales, *Prophet* adopta una noción más amplia que la de los *GAM*. En contexto de series de tiempo, la finalidad de un *GAM* radica en su aptitud para separar la variación atribuible a patrones temporales intrínsecos (tendencia y ciclos) de la variación causada por covariables externas con comportamientos no lineales, permitiendo interpretar el impacto marginal de cada predictor de manera independiente y transparente. El tratamiento de la componente de tiempo se basa en funciones no paramétricas o semi-paramétricas (como *smoothing splines*). *Prophet* sustituye los *splines* de suavizado no paramétricos por un enfoque de ajuste de curvas (*curve-fitting*) estrictamente paramétrico, donde las componentes temporales se modelan mediante fórmulas matemáticas

cerradas: la tendencia se define mediante regresiones lineales o logísticas por tramos (*piecewise*), mientras que la estacionalidad y los ciclos se aproximan mediante series finitas de Fourier. Esta simplificación paramétrica le permite abordar el tiempo no como una covariable más dentro de un espacio estocástico complejo, sino como el eje estructurante del modelo. De este modo, *Prophet* adopta una noción más amplia e instrumental de la aditividad: sacrifica la flexibilidad para modelar relaciones no lineales complejas con covariables exógenas continuas, fortaleza central del *GAM*, a cambio de obtener una elevada robustez operativa frente a datos faltantes, cambios abruptos en la tendencia (*change points*) y múltiples patrones estacionales superpuestos. En consecuencia, mientras que el *GAM* utiliza la aditividad como un vehículo de inferencia y análisis causal no lineal, *Prophet* la emplea como un marco de descomposición modular paramétrica diseñado para maximizar la precisión, interpretabilidad y automatización de las proyecciones futuras.

3.1.2. Ajuste de modelos. Estrategia de separación de datos: *train-test*

El modelado de datos tiene como objetivo identificar patrones que permitan predecir el comportamiento de nuevos casos. Para evaluar la capacidad predictiva de un modelo, es necesario contrastar su desempeño en observaciones que no fueron utilizadas durante su ajuste. En el ámbito del aprendizaje automático, es común dividir el conjunto de datos disponible en dos subconjuntos: entrenamiento (*train*) y prueba (*test*).

Los datos de entrenamiento son los que se utilizan para entrenar los diferentes modelos propuestos, mientras que los datos de prueba o *test*, se reservan para evaluar la capacidad predictiva del modelo en datos no vistos previamente. Una división típica del conjunto de datos es 80/20, aunque esta depende del horizonte necesario de pronóstico y tamaño de serie disponible.

En el caso particular de las series temporales, es crucial respetar el orden cronológico de los datos durante la separación. A diferencia de otros tipos de datos, una partición aleatoria no es adecuada.

Por este motivo, la serie temporal debe ordenarse cronológicamente y establecerse un punto de corte definido. Los datos previos a este corte conforman el conjunto de entrenamiento, mientras que los posteriores se destinan al conjunto de prueba. Este procedimiento permite simular un escenario realista, en el cual los modelos se entrenan con datos históricos y luego se aplican para predecir valores futuros, reproduciendo así el proceso típico de pronóstico en contextos empresariales.

3.1.3. Evaluación del desempeño predictivo

La evaluación del ajuste de un modelo constituye un paso fundamental en el análisis, ya que garantiza que las predicciones generadas sean confiables y pertinentes para apoyar la toma de decisiones. En el caso particular de *Prophet*, dada su aplicación sobre series temporales, una de las estrategias más robustas para evaluar el desempeño del modelo es la validación cruzada temporal (*Time Series Cross-Validation*). A diferencia de la validación cruzada tradicional, no es posible mezclar los datos aleatoriamente, ya que el orden secuencial contiene información esencial sobre la dinámica temporal.

El procedimiento implementado por *Prophet* para la validación cruzada aborda los siguientes puntos:

- Puntos de corte (*Cutoff points*): el historial de datos se divide en múltiples puntos de corte.
- Entrenamiento y pronóstico: para cada punto de corte, el modelo se entrena solo con los datos anteriores a ese punto y luego se genera un pronóstico para un horizonte futuro específico.
- Comparación: las predicciones de cada horizonte se comparan con los valores reales observados en ese período.

Esto simula cómo se habría comportado el modelo si se hubiera entrenado y utilizado en diferentes momentos del pasado, proporcionando una evaluación más realista de su rendimiento futuro.

Los parámetros clave para la validación cruzada son:

- Inicio (*initial*): el período de tiempo mínimo para el entrenamiento inicial del modelo. Debe ser lo suficientemente largo para capturar todas las estacionalidades (al menos un año si hay estacionalidad anual).
- Período (*period*): la frecuencia con la que se realizan los puntos de corte (ej. cada mes, cada 3 meses).
- Horizonte (*horizon*): la duración del período de pronóstico para cada punto de corte (ej. 30 días, 90 días).

En esta estrategia de evaluación por ventanas temporales móviles o acumulativas (*backtesting*). Se divide la serie en múltiples cortes temporales ($k = 1, \dots, K$), entrenando hasta un tiempo t_k y evaluando la métrica de error elegida sobre un horizonte de pronóstico h

$$Error\ Promedio = \frac{1}{K} \sum_{k=1}^K M(y_{t_k+1:t_k+h}, \hat{y}_{t_k+1:t_k+h}) \quad (3.1.3.1)$$

Donde $M(\cdot)$ representa la función de error seleccionada.

Ventajas: Respeta el orden temporal sin fuga de datos. Refleja el rendimiento real del modelo a lo largo de variadas estacionalidades. Muestra cómo varía el error según el horizonte de predicción.

Desventajas: Requiere un mayor costo y tiempo computacional. Depende de la elección correcta de la ventana inicial, el paso de corte y el horizonte.

A continuación, se presentan estas funciones de error que permiten evaluar el desempeño del modelo. Para ello, es necesario aclarar que:

y_i : refiere al valor observado en la posición i

$\hat{y}_i$: refiere al valor predicho por el modelo para la misma posición i

n : es el número total de predicciones del conjunto de datos.

Error Cuadrático Medio (MSE):

$$MSE = \frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2 \quad (3.1.3.2)$$

Ventajas: penaliza fuertemente los errores grandes debido al cuadrado. Útil para optimización.

Desventajas: la unidad de medida es el cuadrado de la variable original, lo que lo hace menos interpretable. Es sensible a los valores atípicos.

Raíz de Error Cuadrático Medio (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2} \quad (3.1.3.3)$$

Ventajas: está en las mismas unidades que la variable original, lo que lo hace más interpretable que el MSE.

Desventajas: también es sensible a los valores atípicos.

Error Absoluto Medio (MAE):

$$MAE = \frac{1}{n} \sum_1^n |y_i - \hat{y}_i| \quad (3.1.3.4)$$

Ventajas: fácil de interpretar ("el error promedio es de X unidades"). Menos sensible a los valores atípicos que el MSE y RMSE.

Desventajas: no penaliza los errores grandes tan fuertemente.

Error Porcentual Absoluto Medio (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (3.1.3.5)$$

Ventajas: expresado como un porcentaje, lo que facilita la comparación del rendimiento del modelo entre diferentes series temporales con diferentes escalas. Es muy intuitivo ("el modelo se equivoca en un X% en promedio").

Desventajas: se vuelve infinito o indefinido si los valores reales (y_i) son cero o muy cercanos a cero. Esto es una limitación importante para series con muchos ceros.

Error Porcentual Absoluto Medio Simétrico (SMAPE)

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \times 100 \quad (3.1.3.6)$$

Ventajas: intenta abordar la limitación del MAPE cuando los valores reales son cero o pequeños, ya que el denominador promedia el valor real y el pronóstico. También se expresa en porcentaje.

Desventajas: puede tener problemas si tanto y_i como $\hat{y}_i$ son cero. No es completamente simétrico en todas las situaciones.

Error Porcentual Absoluto Mediano (MDAPE)

Una alternativa al MAPE que usa la mediana para ser más robusta a valores extremos.

$$MDAPE = mediana\left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100 \quad (3.1.3.7)$$

Cobertura (Coverage)

Mide el porcentaje de puntos reales que caen dentro de los intervalos de incertidumbre de la predicción del modelo.

Mide la proporción de observaciones reales (y_i) que quedan comprendidas dentro del intervalo de confianza o incertidumbre predicho por el modelo para un nivel determinado (por ejemplo, 80% o 95%).

$$Coverage = \frac{1}{n} \sum_{i=1}^n I(\hat{y}_{lower,i} \leq y_i \leq \hat{y}_{upper,i}) \times 100 \quad (3.1.3.8)$$

Donde $I(\cdot)$ es la función indicadora (vale 1 si la condición se cumple y 0 si no), y $\hat{y}_{lower}, \hat{y}_{upper}$ son los límites del intervalo predicho.

Ventajas: Evalúa la calibración de la incertidumbre del modelo. Muy intuitiva de interpretar frente al nivel de confianza fijado. Esencial para análisis de riesgo e inventarios, planificación presupuestaria y financiera.

Desventajas: No mide el ancho del intervalo (un intervalo infundadamente ancho da 100% de cobertura). Debe complementarse con métricas de amplitud del intervalo.

Un valor de cobertura del 80% con un intervalo de confianza del 80% indica un buen calibrado. Si la cobertura es mucho menor que el coeficiente de confianza, los intervalos son demasiado estrechos; si es mucho mayor, son demasiado amplios.

3.2. Modelo de asignación de presupuesto: Función Objetivo

El punto central de un problema de optimización se aborda con una función objetivo con restricciones. En términos generales, se trata de un planteo matemático o computacional cuyo propósito es identificar los valores óptimos de un conjunto de variables de decisión, maximizando o minimizando un criterio predefinido, respetando al mismo tiempo ciertas condiciones o restricciones.

En el ámbito de la publicidad, el problema de asignación de presupuesto es un desafío importante. En este contexto, el objetivo consiste en determinar la cantidad de recursos financieros que se deben asignar a cada categoría de productos que ofrece una empresa, a fin de maximizar la efectividad de su campaña publicitaria.

En el presente trabajo se utiliza este enfoque para formular el problema de asignación de presupuesto publicitario en una empresa con múltiples categorías de productos. El objetivo es maximizar la efectividad publicitaria, para lo cual se busca asignar un presupuesto determinado a cada categoría de manera que se alcance el mayor volumen de ventas posible con un gasto mínimo.

3.2.1. Definición de la Función Objetivo

La resolución de un problema mediante una función objetivo implica encontrar los valores específicos de $(x_1, x_2, \dots, x_n)$ que maximicen o minimicen la función elegida, sin violar ninguna de las condiciones de presupuesto o proporción.

Dicha resolución viene dada por una expresión matemática que cuantifica lo que se pretende optimizar:

- Minimizar esa función: si representa algo "malo" o indeseable, como costos, errores, riesgos, tiempo, o consumo de energía. Por ejemplo, minimizar el presupuesto asignado en publicidad.
- Maximizar esa función: si representa algo "bueno" o deseable, como ganancias, ingresos, eficiencia, rendimiento, o satisfacción del cliente. Por ejemplo, maximizar las ventas o el beneficio neto.

Matemáticamente, la función objetivo se denota como

$$f(x_1, x_2, \dots, x_n) = \max \sum_{i=1}^n c_i x_i \text{ o } \min \sum_{i=1}^n c_i x_i \quad (3.2.1.1)$$

donde

x_i es el valor asignado a la variable de decisión i .

n es el número total de variables de decisión

c_i es el coeficiente de contribución de la variable

3.2.2. Variables de Decisión

Las variables de decisión son los elementos sobre los cuales se tiene control y cuyos valores se deben determinar para alcanzar el objetivo, ya sea maximizar o minimizar la función. Son los "qué" del problema. Por ejemplo, en un problema de asignación de presupuesto publicitario, las variables de decisión podrían ser la cantidad de dinero a invertir en cada canal (redes sociales, televisión, etc.).

3.2.3. Restricciones

Las restricciones son un conjunto de condiciones o limitaciones que las variables de decisión deben satisfacer. Estas restricciones reflejan las limitaciones del mundo real, los recursos disponibles, las políticas internas, las regulaciones o las capacidades técnicas. Sin restricciones, muchos problemas de optimización tendrían soluciones triviales o infinitas.

Las restricciones pueden ser de varios tipos:

- Restricciones de Igualdad (o Balance): se utilizan cuando se quiere que la suma de ciertas variables cumpla una cuota exacta. Esto es, las variables de decisión deben satisfacer una ecuación exacta como la siguiente:

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_i x_i = C \quad (3.2.3.1)$$

- Restricciones de Desigualdad: se utilizan para representar recursos limitados, no sobre pasar un "techo", o bien, para representar requisitos mínimos, "pisos" que deben

cumplirse obligatoriamente. Esto es, las variables de decisión deben ser menores o iguales, o mayores o iguales que un cierto valor:

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_i x_i \leq K \quad (3.2.3.2)$$

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_i x_i \geq L \quad (3.2.3.3)$$

Como caso particular de esta restricción se encuentra la restricción de recursos globales (o presupuesto) que pone un límite superior o máximo como es el capital disponible, reflejado en la expresión 3.2.3.2.

- Restricciones de No Negatividad: muchas variables de decisión, como cantidades o precios, no pueden ser negativas. Esto es,

$$x_i \geq 0 \text{ para todas las } i \quad (3.2.3.4)$$

- Restricciones de Proporción (o Frontera): se utilizan para definir límites específicos para cada variable, límites inferior y superior. Esto es,

$$L_i \leq x_i \leq U_i \text{ para todas las } i \quad (3.2.3.5)$$

- Restricciones Enteras/Binarias: las variables deben tomar valores enteros (ej. número de *clicks*) o binarios (0 o 1, para decisiones de sí/no).

3.3. Materiales

Los datos utilizados en este estudio provienen de una empresa de comercio electrónico que opera en el mercado chileno. La empresa centra su estrategia en campañas que forman parte del Marketing de Pago, específicamente en las campañas publicitarias denominadas *Performance Max (PMAx)* de Google, representan aproximadamente el 70% de las ventas totales del sitio. Estas campañas en particular se llaman *always on*, lo que significa que no han sido creadas o encendidas únicamente para épocas de eventos (*black out*¹² y *hot sales*¹³) o *cybers*, sino que se encuentran permanentemente activas.

Las ventajas de centrarse en la estrategia de Marketing de Pago es que se orienta principalmente a la conversión, con el propósito de atraer, alcanzar y convertir audiencias de manera ágil. Esta estrategia permite llegar a personas que, de otro modo, no conocerían la existencia del sitio o de sus productos.

¹² Es un período en el que no se pueden realizar promociones, descuentos o campañas publicitarias.

¹³ Evento de comercio electrónico con grandes descuentos y ofertas especiales organizado durante unos días específicos del año para impulsar las ventas online.

Los datos que se utilizan en esta tesis corresponden exclusivamente a campañas publicitarias *PMAX* de *Google* para 12 categorías de productos. En la Tabla 3.3.1. se presentan las 12 campañas correspondientes a cada categoría de producto en estudio.

Tabla 3.3.1.: Campañas publicitarias de *PMAX* de *Google Ads*

Número	Campaña
1	Automotriz
2	Belleza y Bienestar
3	Climatización
4	Grandes electrodomésticos - Cocina
5	Grandes electrodomésticos - Lavado
6	Grandes electrodomésticos - Refrigeración
7	Pequeños electrodomésticos
8	Telefonía
9	Mascotas
10	Jardinería
11	Menaje
12	Decoración y Blanco

Los datos disponibles para este proyecto se recolectaron en forma diaria y corresponden al período comprendido entre el 17 de enero de 2022 y el 22 de agosto de 2023.

Figura 3.3.1. Período bajo estudio. Recopilación de datos.



En la Tabla 3.3.2. se presentan las variables recopiladas y disponibles para su análisis y modelado.

Tabla 3.3.2.: Variables recopiladas y disponibles para analizar.

Variable	Descripción	Tipo
date	Fecha	Fecha
campaign	Campaña o anuncio <i>PMAX</i>	Categorico
is_event	Identificación de Evento especial (1 día de evento; 0 día sin evento)	Categorico
is_cyber	Identificación de <i>Cyber</i> (1 día de <i>cyber</i> ; 0 día sin <i>cyber</i>)	Categorico
mnv_USD (y)	Venta neta diaria en dólares (<i>Net Media Value</i>)	Continua
visitors	Número de visitantes al sitio web	Discreta
visits	Número de visitas totales al sitio web	Discreta
orders	Compras realizadas en el sitio web	Discreta
total_units_net	Número total de unidades compradas	Discreta
budget_USD	Presupuesto o inversión ingresada plataforma de <i>Google Ads</i>	Continua
cost_USD	Costo facturado en <i>Google Ads</i>	Continua
clicks	Número de <i>clicks</i> realizados sobre el anuncio o campaña	Discreta
impressions	Cantidad de veces que se muestra el anuncio	Discreta

Su estructura puede verse en la Tabla 3.3.3. a continuación, donde cada fila representa toda la información de una campaña específica para un día determinado. Por ejemplo, el día 17 de enero del 2022 la campaña C1 tuvo una inversión de 200 dólares, recibió 10.000 *clicks* y generó al comercio electrónico una venta neta de 1.000 dólares.

Tabla 3.3.3.: Estructura de los datos

date	campaign	mnv_USD (y)	visits	...	budget_USD	clicks
17/1/2022	C1	1.000	9.500	...	200	10.000
17/1/2022	C2	900	6.789	...	200	7.000
...	...	...	...	...	...	...
22/8/2023	C12	750	4.795	...	100	5.000

4. Aplicación práctica

4.1. Análisis descriptivo

El análisis descriptivo de un conjunto de datos es una etapa fundamental y estratégica en todo proyecto debido a su capacidad para caracterizar y comprender a profundidad el tema bajo estudio. Antes de cualquier análisis inferencial o explicativo más complejo, es crucial tener una visión clara de qué está sucediendo. Este apartado permite la comprensión y entendimiento del conjunto de datos disponible.

A modo de referencia se recuerda que los datos recopilados corresponden a 12 campañas de marketing de una empresa de comercio electrónico chilena. Se cuenta con un total de 13 variables registradas diariamente entre el 17 de enero de 2021 hasta el 22 de agosto de 2023. Se trata de datos completos para todas las campañas, salvo para Automotriz, Climatización, Telefonía y Pequeños Electrodomésticos que cuentan con un registro menos que corresponde a un día de *cyber* en el cual no fue registrado el presupuesto asignado, tal como se muestra en la Figura 4.1.1. A excepción de ese día particular, no observan datos perdidos para ninguna de las variables en estudio.

Figura 4.1.1. Fechas sin registro de información por campaña.

```
Campaign: Automotriz  
Missing Dates: [Timestamp('2022-05-29 00:00:00')]  
Campaign: Climatizacion  
Missing Dates: [Timestamp('2022-05-29 00:00:00')]  
Campaign: Pequeños_Electrodomesticos  
Missing Dates: [Timestamp('2022-01-17 00:00:00')]  
Campaign: Telefonia  
Missing Dates: [Timestamp('2022-05-29 00:00:00')]
```

Fuente: elaboración propia

Antes de comenzar con el análisis descriptivo es necesario destacar que para la comprensión general del tema de cara al futuro ajuste de modelos, hay dos variables en específico que son las de principal interés:

- **Venta neta diaria en dólares (*Net Media Value*)** (mnv_USD o y): es la variable respuesta que se pretende estimar.
- **Presupuesto o inversión ingresada plataforma de *Google Ads*** (budget_USD): es la variable que se pretende optimizar, a partir de las estimaciones halladas para la venta neta diaria en dólares.

Para tomar dimensión del volumen del negocio se dispone en la Tabla 4.1.1., que sigue a continuación, un conjunto de medidas descriptivas de cada variable para las 12 campañas juntas, sin discriminarlas.

Tabla 4.1.1. Medidas descriptivas.

Métricas	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto/ inversión en Google Ads	Costo facturado en Google Ads
Recuento	6.992	6.992	6.992	6.992	6.992	6.992	6.992	6.992	6.992
Promedio	23.992	14.111	1.032.426	11.712	15.596	306	579	987	1.039
Desvío estándar	42.621	15.376	1.233.059	12.479	17.396	549	1.102	1.775	2.311
Mínimo	44	3	182	42	44	1	1	0	0
25%	7.077	3.954	305.759	3.737	4.819	75	135	302	266
50%	14.267	8.46	609.066	7.072	9.182	149	283	598	612
75%	28.838	19.541	1.335.617	15.721	20.827	368	673	1.326	1.205
Máximo	981.317	203.857	22.517.597	146.420	238.770	14.772	32.941	68.612	69.710

Fuente: elaboración propia

Durante los meses de medición se encontró un volumen máximo de ventas netas (en dólares) que supera los USD 980.000 dólares. Mientras que, el presupuesto máximo asignado en *Google Ads* ronda los USD 68.000 dólares, que representa el 7% del máximo de ventas netas.

El nivel de interés de un anuncio se suele medir por la cantidad *clicks* vs. las impresiones o veces que se muestra el anuncio. En promedio se observan 14.111 *clicks* entre el millón (1.032.426) de veces que se mostró el aviso, esto implica que la tasa de *clicks* (CTR) ronda el 1,4%; que, mirado desde la perspectiva del negocio, refiere a un nivel de interés moderado¹⁴ del público objetivo respecto del anuncio.

Al observar las métricas referidas a tráfico web, se puede ver un promedio de 11.712 visitantes únicos que realizan 15.596 visitas en total, lo que implica que cada persona que visita el sitio

¹⁴ En los canales de generación de demanda o tipo *push* (como *Google Display* o publicidad en redes sociales), el consumidor no manifiesta una intención de compra activa previa, por lo que la atención se capta mediante interrupción pasiva (Chaffey & Ellis-Chadwick, 2019). Por este motivo, la tasa de clics (CTR) promedio de la industria para anuncios gráficos y de descubrimiento oscila históricamente entre el 0.46% y el 1.5% (LocaliQ, 2024; Varos, 2024)

web 1,33 veces en promedio; indicando un nivel moderado¹⁵ de atracción (*engagement*) hacia el contenido o las ofertas propuestas. Las conversiones (ventas) muestran un promedio de 306 órdenes (compras) en el período, con 579 unidades vendidas, esto es, cada orden contiene aproximadamente 1,96 unidades.

El rendimiento financiero de las campañas es notable, dado que con un presupuesto promedio de USD 987 se alcanza una venta neta promedio de USD 23.992. Mostrando un retorno de la inversión (*ROI*¹⁶) aproximado de 23,3, un número atractivo que sugiere una eficacia importante en la conversión de inversión en ingresos.

Por otra parte, el rendimiento del negocio mirado desde la perspectiva de la inversión publicitaria tiene varias aristas a considerar. En este caso, vale destacar que, dependiendo de los objetivos trazados para cada empresa, un valor de *AdCos*¹⁷ de 0,30 puede considerarse como aceptable para algunas empresas, mientras que, otras pueden considerarlo demasiado alto o bajo¹⁸. En este caso, el valor hallado es razonable.

Para este trabajo el valor de *AdCost* hallado es razonable. Sin embargo, cuando se observa el presupuesto en *Google Ads* contra el costo facturado en *Google Ads*, se detecta un ligero exceso del costo sobre el presupuesto, situación que podría indicar que existe una oportunidad para el equipo de marketing en lo que respecta a presupuestos establecidos y su optimización.

En el Anexo 7.1 se presentan las medidas descriptivas por campaña, junto con *CTR*, *ROI* y *AdCost* calculado para cada una. Es importante remarcar aquí que en términos de rendimientos de *CTR*, *ROI* y *AdCost*, las campañas de Belleza-Bienestar y Mascotas son las mejores. Algunas otras campañas tienen mayores ventas netas (en dólares) pero la relación con el presupuesto es mejor para estas dos campañas. Respecto de la tasa de *clicks*, la mejor relación también se observa para estas dos campañas. Estos indicadores no son los únicos para monitorear el rendimiento de una campaña, existen otros más como Costo por *Click* (*CPC*¹⁹) y Tasa de Conversión (*CR*²⁰) que no se expondrán en este trabajo.

¹⁵ En la analítica de comercio electrónico, valores entre 1,20 y 1,50 representan un patrón típico de consideración o "fase de descubrimiento", donde el usuario realiza una primera visita de exploración y retorna de forma acotada para reevaluar la oferta antes de tomar una decisión de compra (Kaushik, 2010; Chaffey & Ellis-Chadwick, 2019).

¹⁶ $ROI \text{ (Return on Investment)} = (\text{Ventas Netas} - \text{Presupuesto o Inversión publicitaria}) / \text{Presupuesto o Inversión publicitaria} * 100$

¹⁷ $AdCost = \text{Presupuesto o Inversión publicitaria} / \text{Ventas Netas}$

¹⁸ Reibstein, D., Pfeifer, P., Bendle, N., Farris, P. (2015). *Marketing Metrics: The Manager's Guide to Measuring Marketing Performance*. Reino Unido: Pearson Education.

¹⁹ Costo por Click (*CPC*) = $\text{Presupuesto o Inversión Publicitaria} / \text{Cantidad de clicks}$.

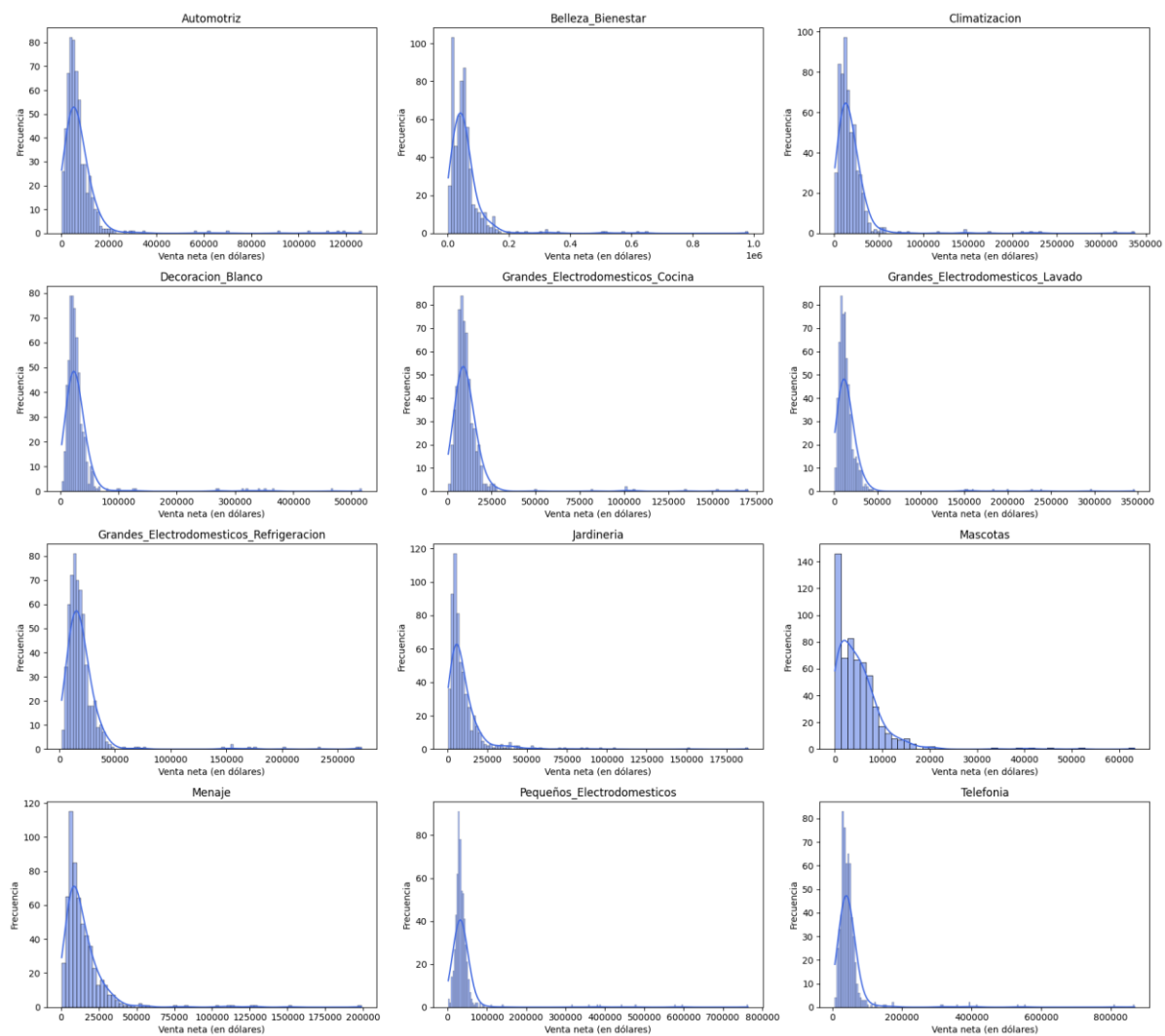
²⁰ Tasa de Conversión (*CR*) = $\text{Compras u Ordenes} / \text{Visitas} * 100$.

Por otra parte, a partir de las medidas descriptivas, se desprende que las distribuciones de estas variables presentan dispersión y asimetría, dado que el valor del desvío estándar es grande con respecto a la media, sumado a que media y mediana están muy alejadas entre sí.

Para ilustrar esta situación, a continuación, se presenta y analiza la distribución de la venta neta (en dólares) por campaña y el presupuesto o inversión en *Google Ads* (dólares). Tal como se mencionó anteriormente, se toman como referencia estas dos variables dado que la primera responde al objetivo de estimación de la venta neta mediante la aplicación de modelos Prophet con el fin de utilizar estas estimaciones para luego optimizar el presupuesto diario y poder distribuirlo entre campañas.

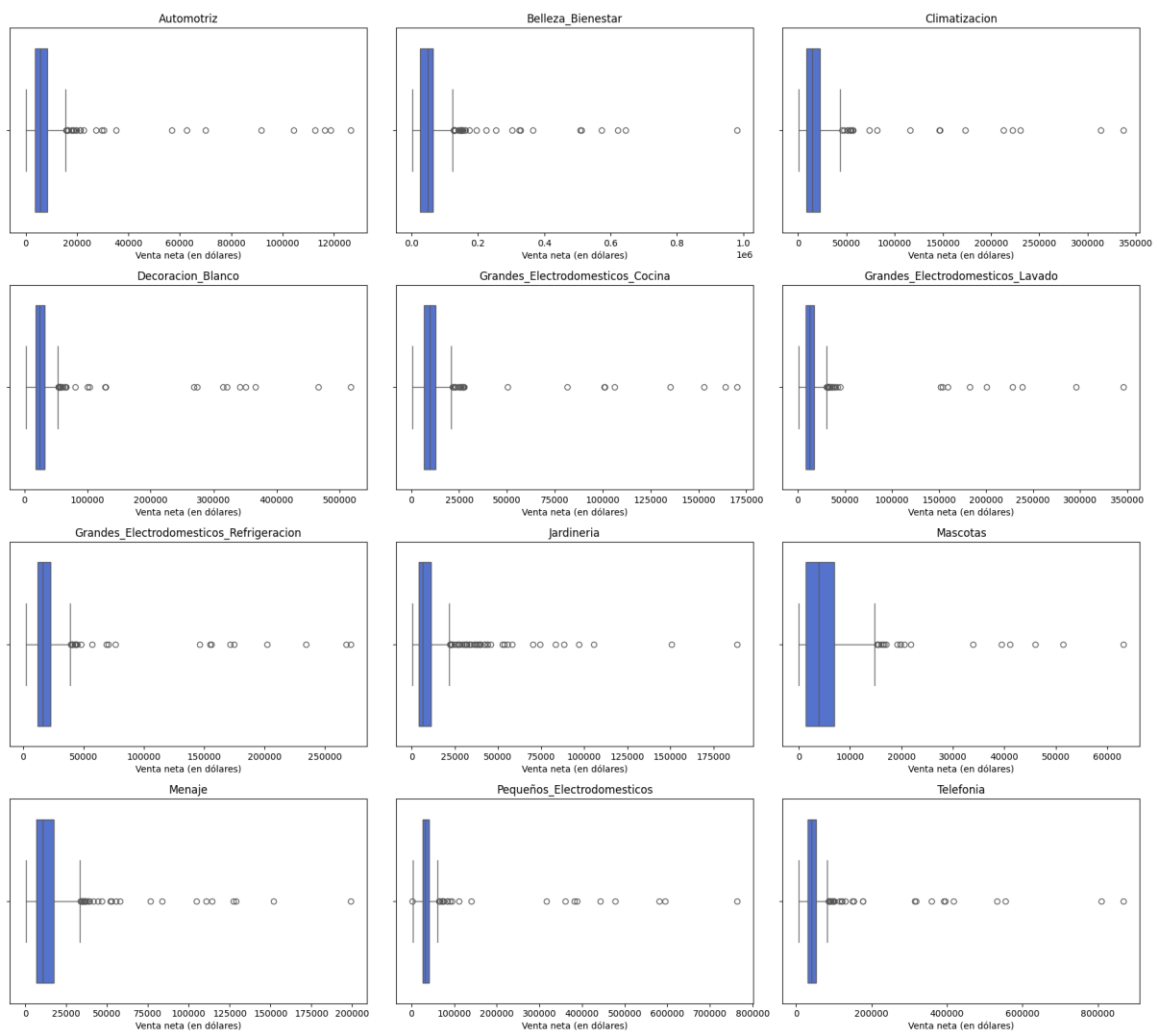
En la Figura 4.1.2. se pueden ver los histogramas de frecuencias de las ventas netas (en dólares) por campaña. En la Figura 4.1.3. se exponen los gráficos *box-plot* por campaña. En ambas figuras se evidencia la asimetría de los datos y la presencia de valores extremos por campaña.

Figura 4.1.2. Distribución de la venta neta (en dólares) según campaña.



Fuente: elaboración propia

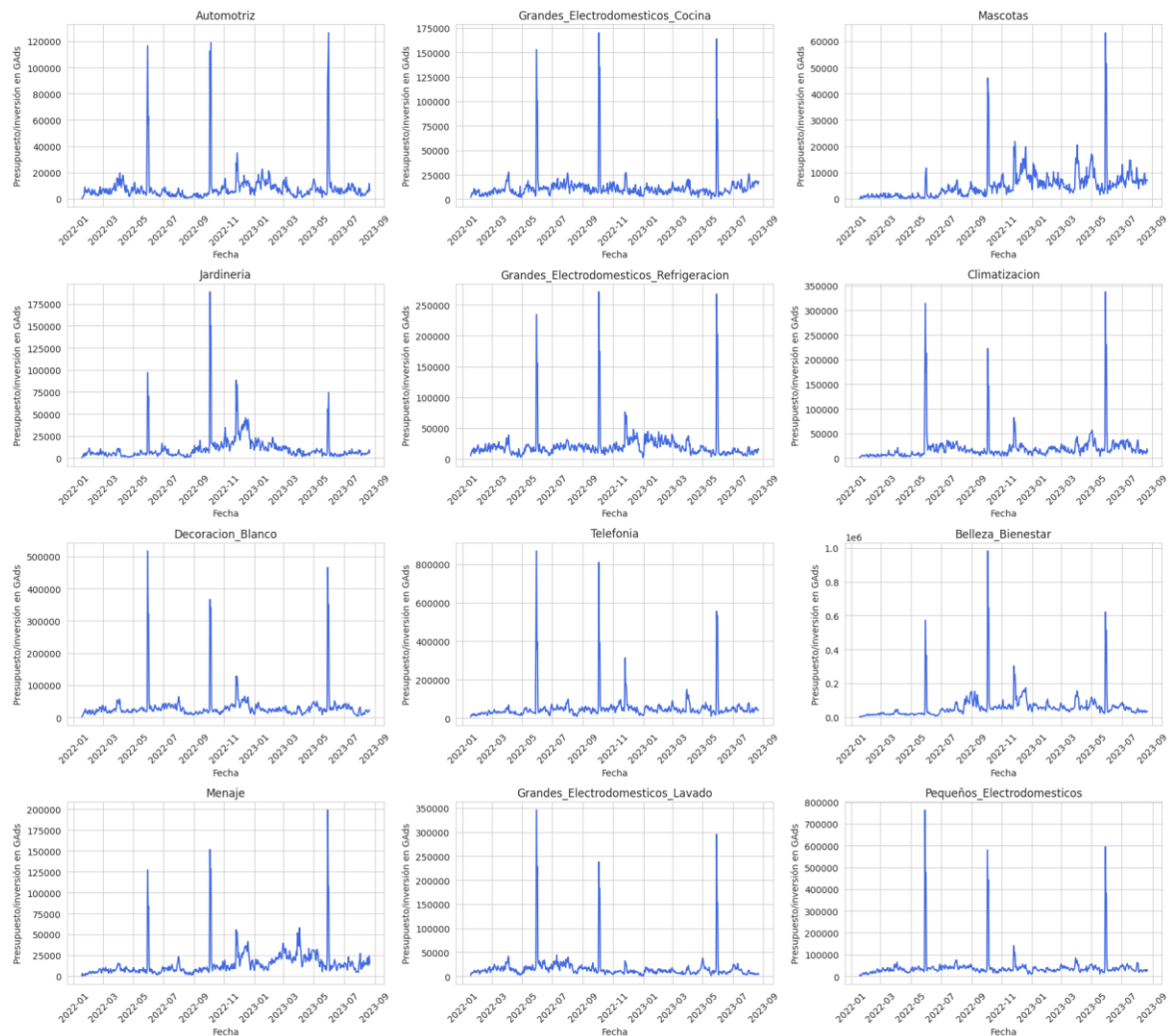
Figura 4.1.3. Distribución de la venta neta (en dólares) según campaña.



Fuente: elaboración propia

En la Figura 4.1.4, que se presenta a continuación, se muestra la evolución de la venta neta (en dólares) por campaña. Como se puede observar se trata de serie con ciertos picos altos, diferentes entre campaña. También parecerían tener diferentes tendencias.

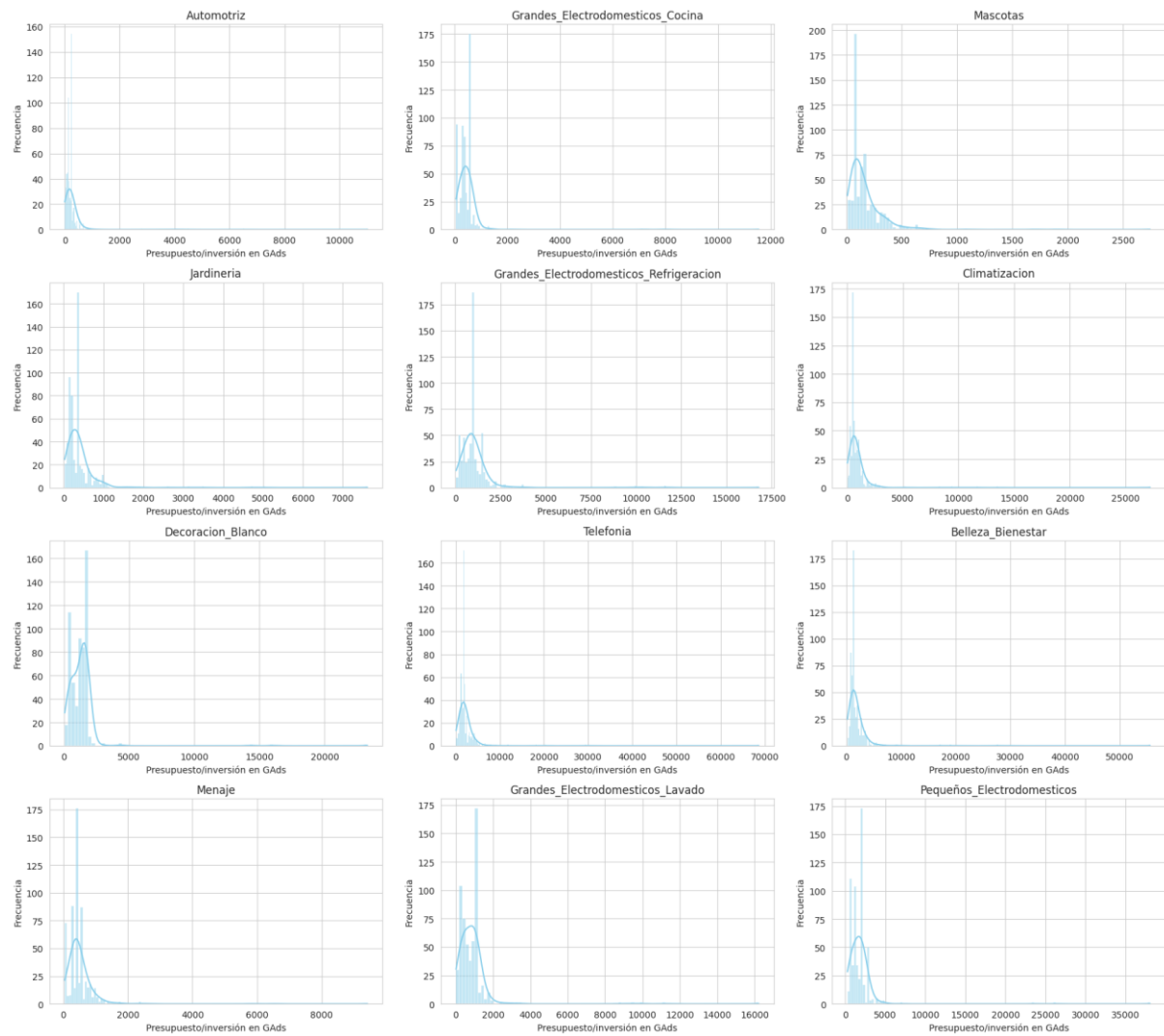
Figura 4.1.4. Distribución de la venta neta (en dólares) según campaña.



Fuente: elaboración propia

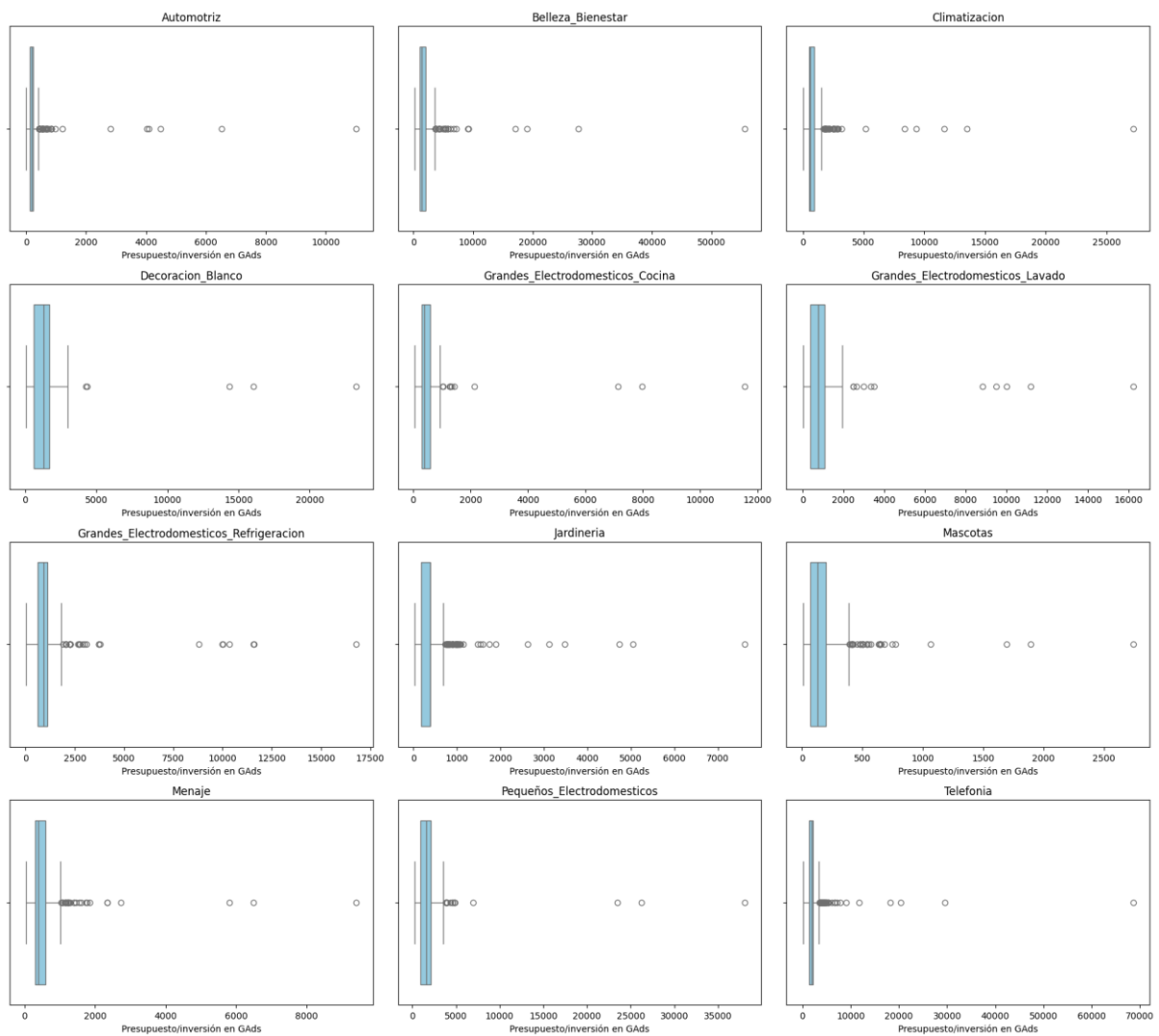
Una situación similar se observa para el presupuesto o inversión diaria en *Google Ads*. Se visualizan distribuciones asimétricas y presencia de valores extremos (Figura 4.1.5., Figura 4.1.6 y Figura 4.1.7).

Figura 4.1.5. Distribución del presupuesto o inversión diaria en *Google Ads* (en dólares) según campaña



Fuente: elaboración propia

Figura 4.1.6. Distribución del presupuesto o inversión diaria en *Google Ads* (en dólares) según campaña.



Fuente: elaboración propia

Figura 4.1.7. Distribución del presupuesto o inversión diaria en *Google Ads* (en dólares) según campaña.

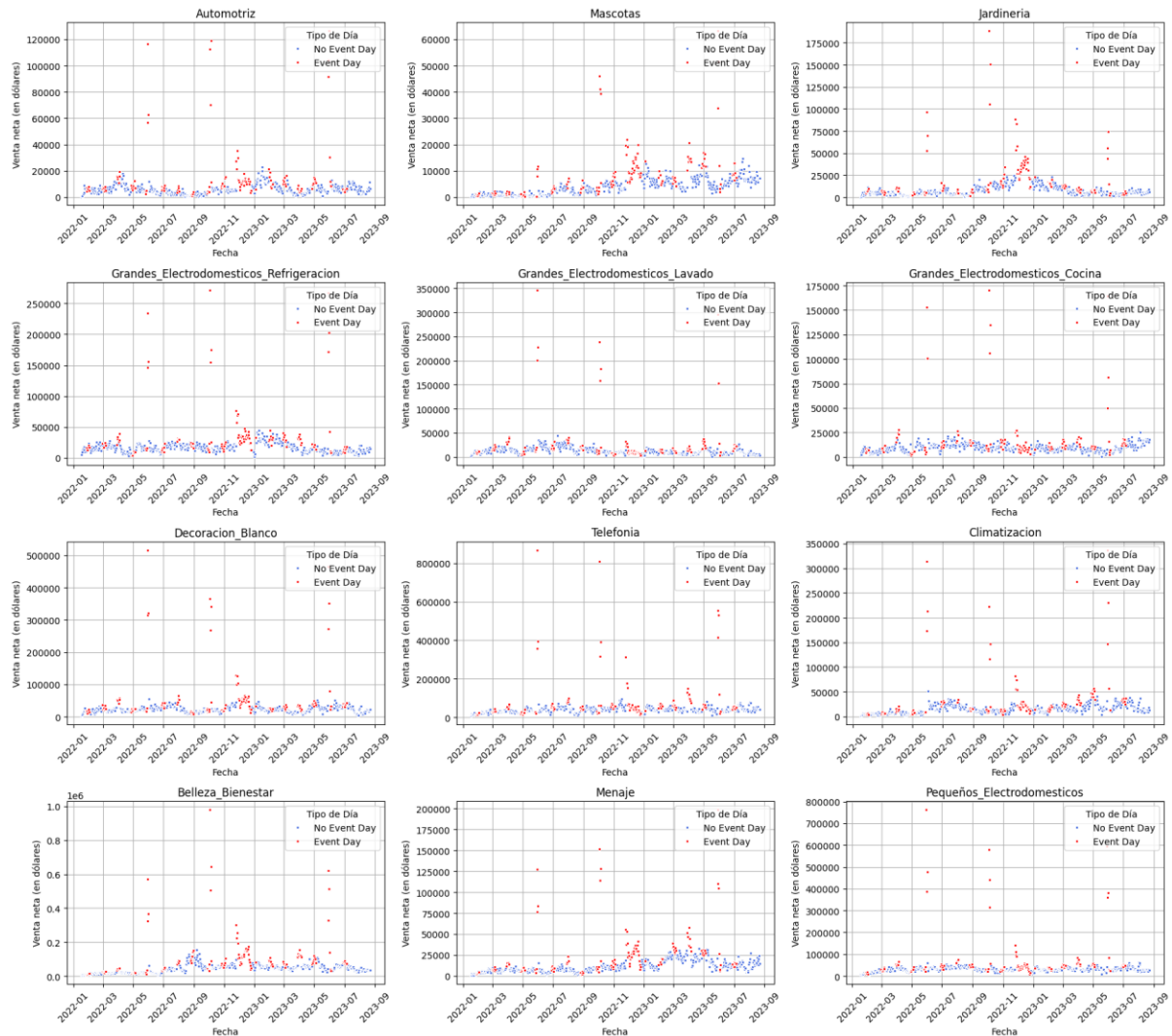


Fuente: elaboración propia.

Dadas las características de los datos expuestas que manifiestan la presencia de valores extremos, distribuciones asimétricas, algunos datos faltantes, diferentes comportamientos por campaña, es una gran oportunidad trabajar con los datos crudos tal cual están en su base de datos original puesto que una de las ventajas de los modelos *Prophet* es que son modelos robustos ante este tipo de situaciones. Dado que *Prophet* contempla la componente *Holiday* como parte del modelo para tratar este tipo de sucesos. Si bien con estas características en los datos, cualquier modelo podría adaptarse, sería necesario aplicar transformaciones o imputaciones porque los supuestos de un modelo tipo *ARIMA* o *SARIMA* se ven afectados por valores extremos, datos faltantes y asimetría extrema.

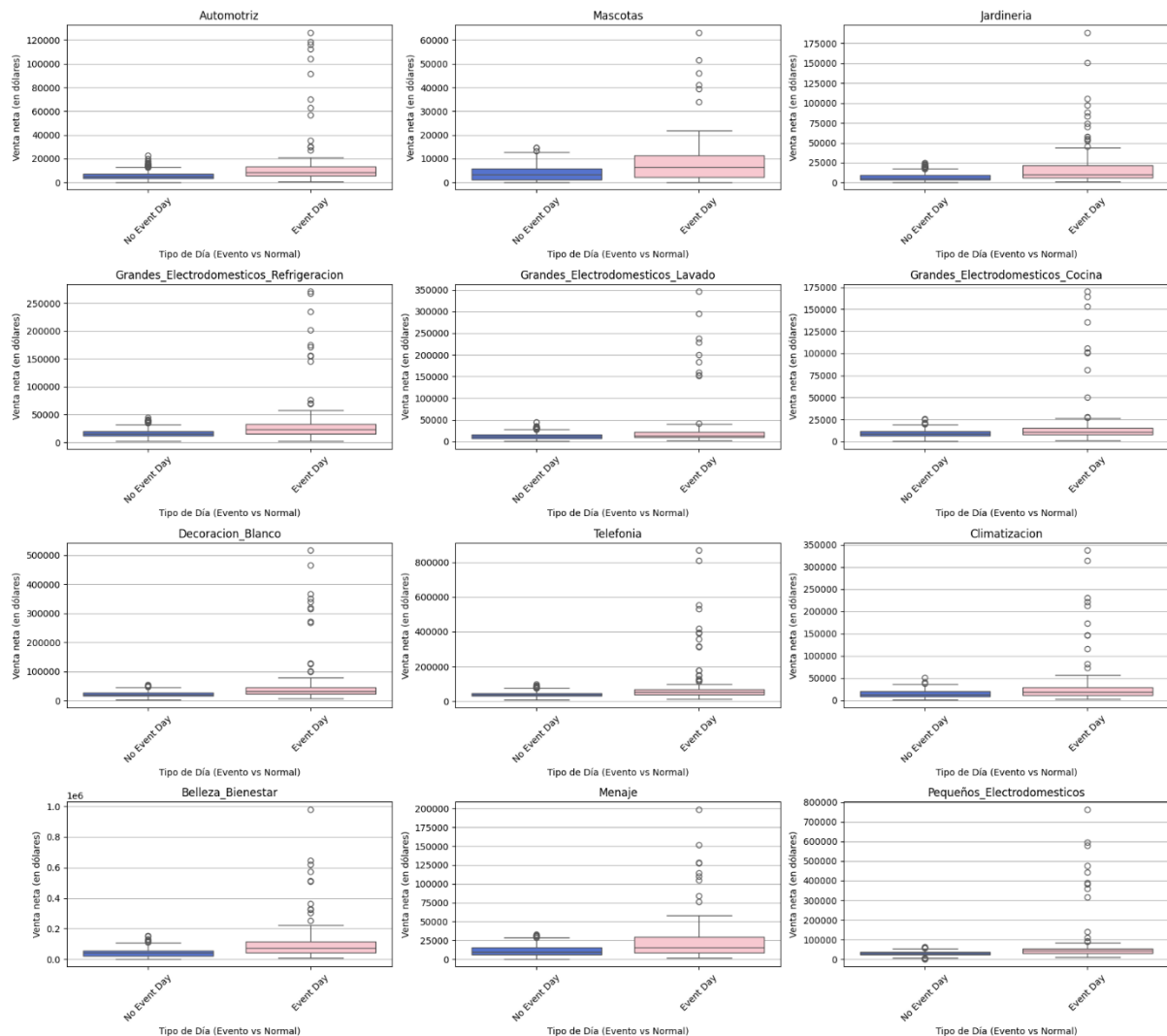
Al revisar la venta neta (en dólares) en el tiempo y su distribución para las distintas campañas, pudo verse que los valores correspondientes a días de eventos o de cyber se desprenden del resto de los valores, tal como se observa en los ejemplos a continuación.

Figura 4.1.8. Distribución de la Venta Neta (en dólares) según Días con Evento vs No Eventos por campaña



Fuente: elaboración propia

Figura 4.1.9. Distribución de la Venta Neta (en dólares) según Días con Evento vs No Eventos por campaña



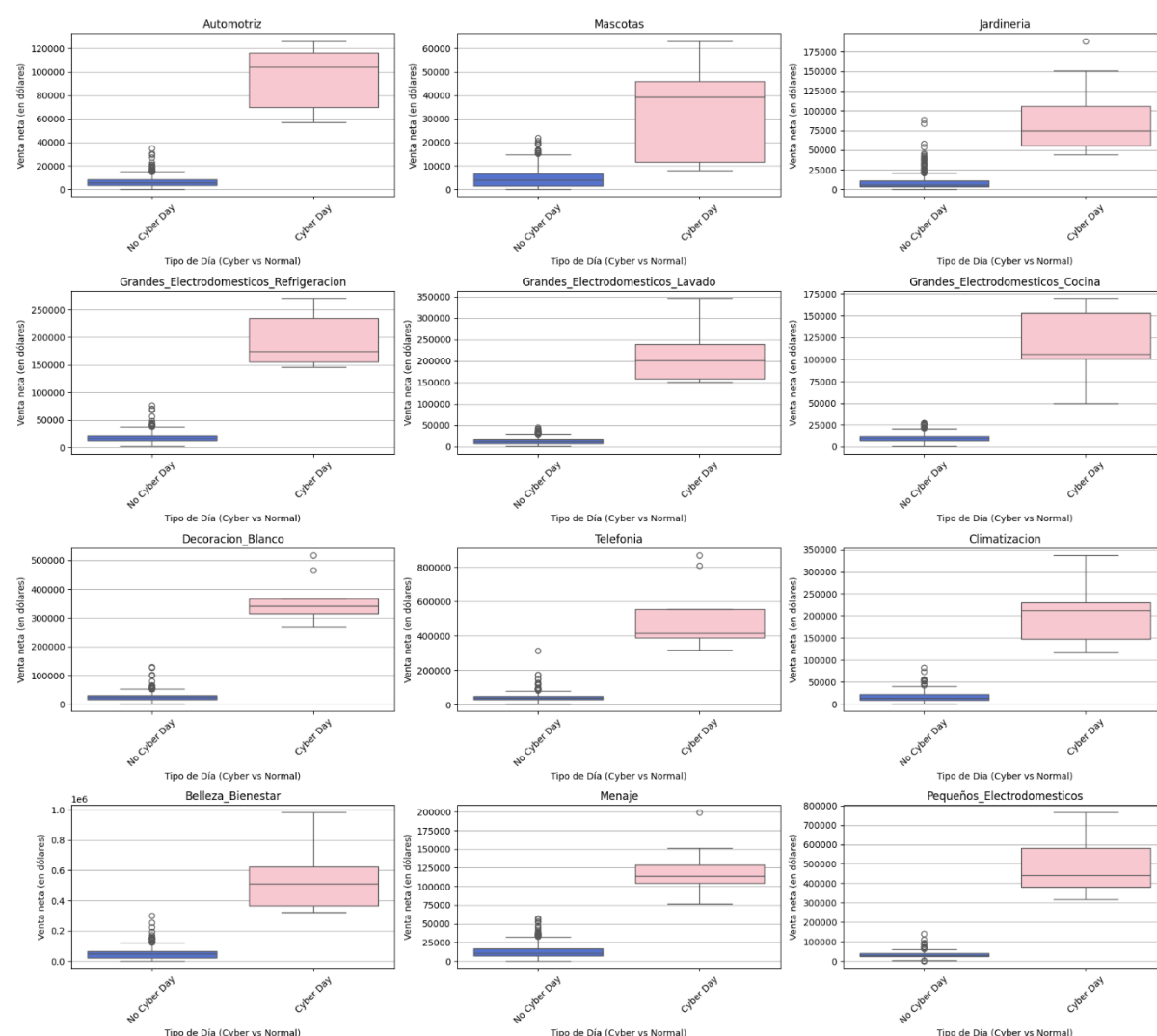
Fuente: elaboración propia

Figura 4.1.10. Distribución de la Venta Neta (en dólares) según Días con *Cyber* vs *No Cyber* por campaña



Fuente: elaboración propia

Figura 4.1.11. Distribución de la Venta Neta (en dólares) según Días con *Cyber* vs *No Cyber* por campaña



Fuente: elaboración propia

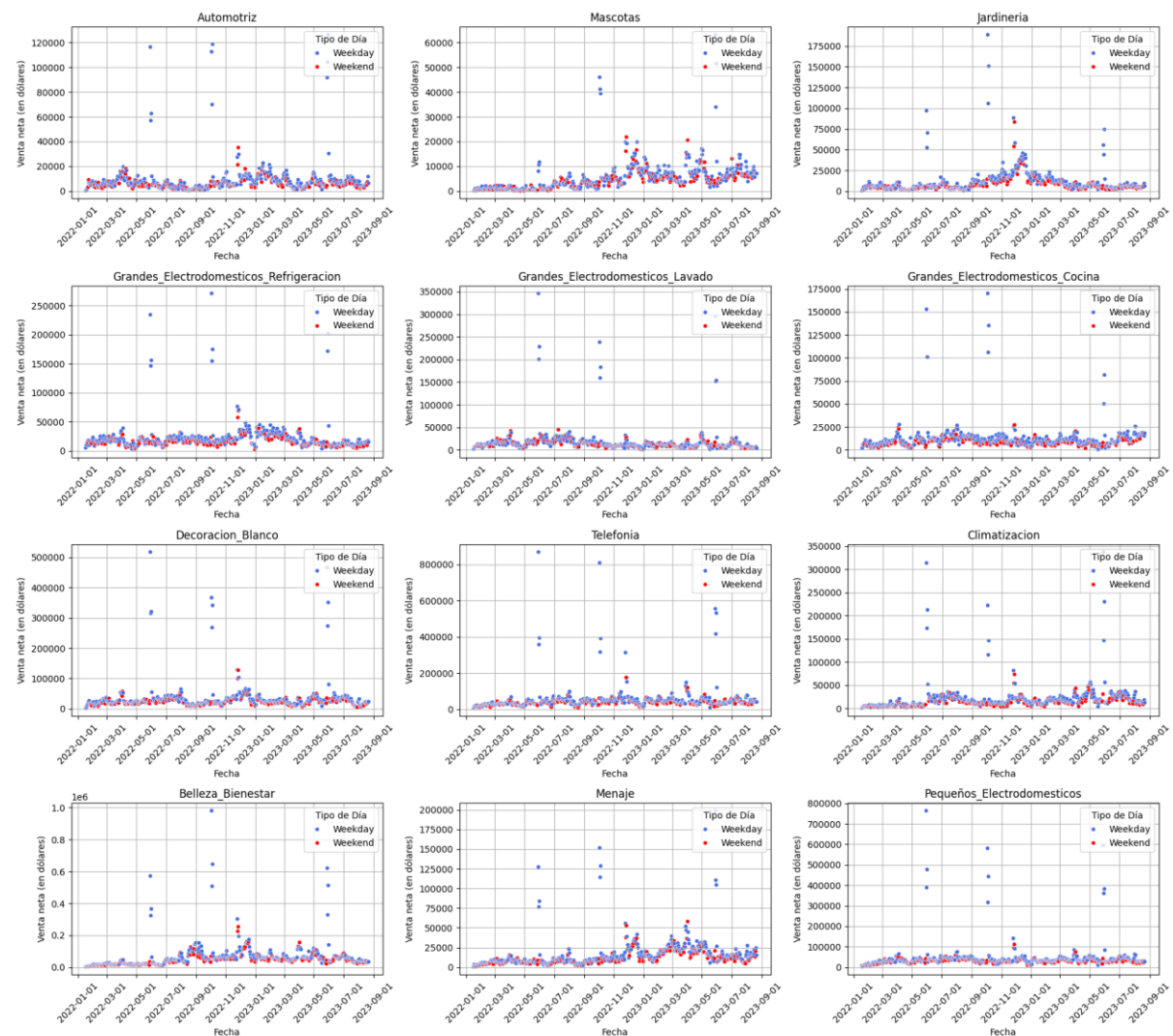
Claramente los días especiales, tanto de *cyber* como de eventos, se observan valores de venta neta (en dólares) superiores a los días regulares o normales.

En el Anexo 7.1 se presentan las Tabla 7.1.14 y Tabla 7.1.15 que muestran las medidas de *CTR*, *ROI* y *AdCost* para cada campaña según días especiales (*cyber* o eventos) versus días regulares. Allí se puede apreciar que la tasa de conversión es mayor en días especiales; sin embargo, el retorno de la inversión es menor en relación con los días regulares. Respecto del costo invertido en publicidad por venta neta es menor en días especiales.

Con respecto a los fines de semana, se realiza un análisis similar al anterior comparando las distribuciones en el tiempo para las distintas campañas, así también como las principales

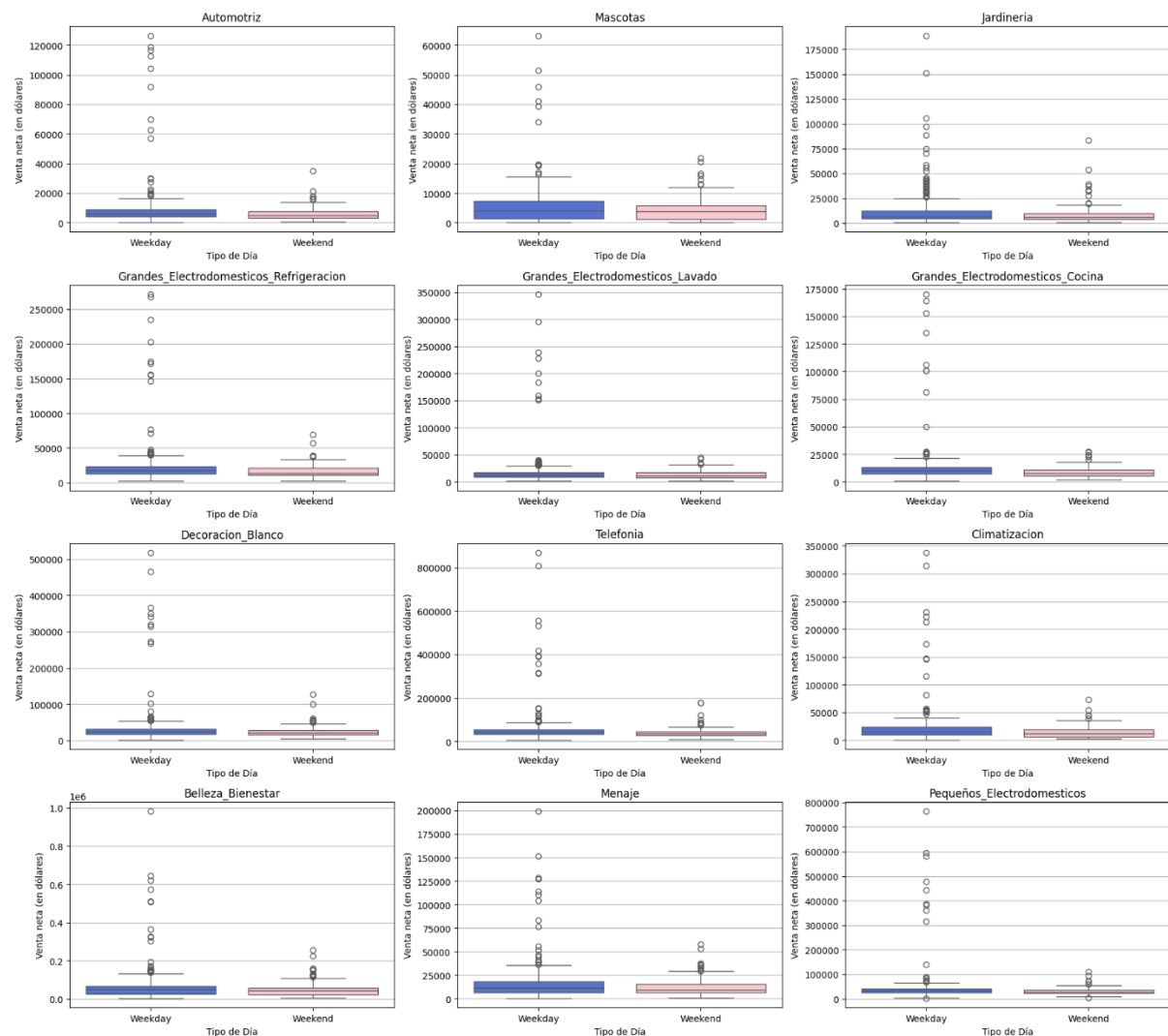
métricas para el negocio. Según se puede ver en la Figura 4.1.12, no pareciera que el comportamiento de los usuarios difiere entre días de semana y fines de semana. Sumado a esto, las diferentes métricas de negocio también muestran un comportamiento similar, a excepción del ROI, que es ligeramente más alto los días de semana (Anexo 7.1). Respecto del AdCost, los días de fin de semana son ligeramente mayores.

Figura 4.1.12. Distribución de la Venta Neta (en dólares) según Días de semana vs fines de semana por campaña



Fuente: elaboración propia

Figura 4.1.13. Distribución de la Venta Neta (en dólares) según Días de semana vs fines de semana por campaña



Fuente: elaboración propia

Para finalizar con la descripción de los datos, en el Anexo 7.1 se presentan las distribuciones por campaña para las demás variables medidas en este estudio: Número de visitantes al sitio web, Número de visitas totales al sitio web, Compras realizadas en el sitio web, Número total de unidades compradas, Costo facturado en *Google Ads*, Número de *clicks* realizados sobre el anuncio o campaña y Cantidad de veces que se muestra el anuncio. Además de las estadísticas descriptivas para todas las variables por días de semana, *cyber* y eventos.

Tal como se mencionó anteriormente, en esta oportunidad, se trabajará con los datos crudos, tal cual están en su base de datos original puesto que una de las ventajas de los modelos *Prophet*. En lugar de trabajar con datos imputados tal como lo hizo Costa, C. (2024). Si es importante remarcar que se necesita la identificación de todos los datos anómalos para

modelarlos. De aquí que, se utiliza la misma técnica de identificación de valores extremos que referencia Costa, C. (2024), en su estudio.

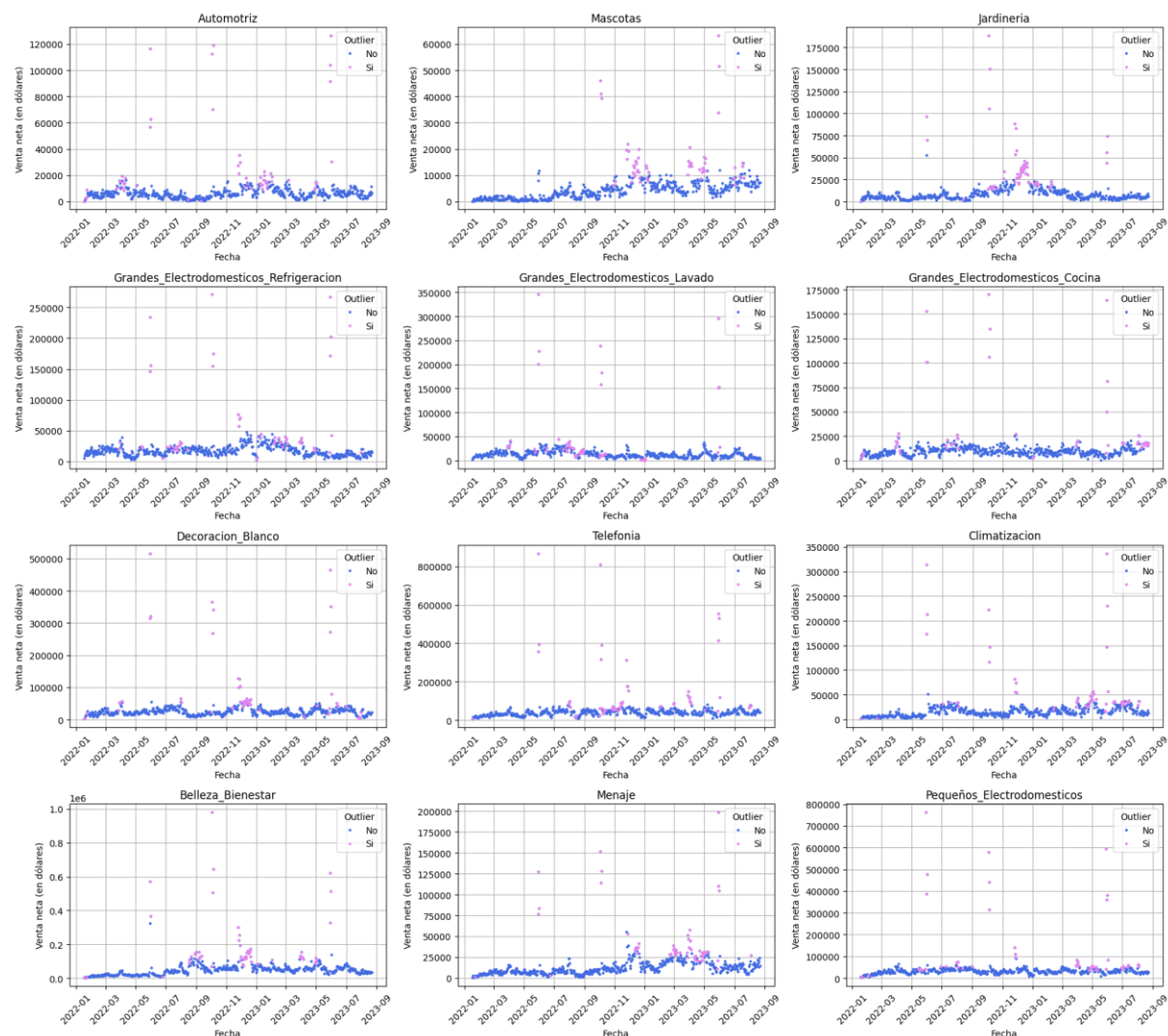
El procedimiento utilizado para su identificación comienza con la aplicación de la descomposición STL²¹, permite descomponer la serie temporal en sus componentes de tendencia, estacionalidad y residuos, facilitando la identificación de valores anómalos. Es importante resaltar que esta técnica es particularmente útil cuando las series de tiempo exhiben patrones estacionales o tendencias que no siguen un modelo lineal, lo cual es común en muchos contextos del mundo real. Aislado las componentes de estacionalidad y tendencia, se logra identificar con mayor confiabilidad aquellas observaciones que se desvían de manera significativa del comportamiento regular de la serie, convirtiendo este proceso en una opción versátil y eficaz para la detección de anomalías en la serie original. En esta oportunidad se trabaja con la librería de *Python PyOD (Python Outlier Detection)* diseñada especialmente con herramientas y funcionalidades que facilitan esta tarea. Se utiliza esta librería por ser la más escalable para detectar anomalías en datos multivariados y debido a que incluye diversos métodos estadísticos, métodos basados en distancia y otros basados en modelos, que permite una elección más apropiada tomando en cuenta los datos con los que se trabaja. Adicionalmente, *PyOD* tiene una buena integración con *scikit-learn*, una popular librería de aprendizaje automático en Python, que permite sistematizar el proceso de búsqueda de valores atípicos con otras funcionalidades como el pre-procesamiento de los datos y la evaluación de modelos.

Como referencia de importancia, Costa, C. (2024) concluye que las fluctuaciones en el comportamiento de los usuarios sugieren una diferencia en la navegación (*clicks* realizados, visitas hechas, etc.) y métricas en general durante los días de *cyber*, coincidiendo con lo analizado en este trabajo. Sin embargo, para obtener un mejor rendimiento de los modelos propuestos en su trabajo (*VARX*, *SARIMAX* y *XGBoost*), toma la decisión de eliminar y reimputar estos registros de los días de *cyber* junto con los valores extremos que identificó con la metodología descrita en el párrafo anterior; de modo tal que excluye esos días para que los modelos ajustados sean más representativos del comportamiento habitual de los consumidores, mejorando la precisión y relevancia de las estrategias basadas en estos datos.

En la Figura 4.1.14. se muestra gráficamente la identificación de observaciones extremas por campaña para la venta neta.

²¹ *STL: Seasonal and Trend decomposition using Loess*, Cleveland et al., 1990.

Figura 4.1.14. Identificación de observaciones atípicas por campaña para la venta neta (en dólares).



Fuente: elaboración propia

Este proceso de identificación de *outliers*, detectó 708 observaciones en el total del conjunto de datos, 59 valores atípicos por campaña. Tal como se muestra en la siguiente figura (Figura 4.1.15.). Esta situación de hallazgo de 59 valores atípicos en cada campaña es producto de sostener la comparabilidad con lo propuesto en el estudio de Costa, C. (2024). En esa oportunidad, con el propósito de identificar las perturbaciones estructurales en el comportamiento agregado del canal de comercio electrónico, el procedimiento de detección de valores atípicos (*outliers*) se implementó mediante un enfoque multivariado conjunto utilizando el algoritmo COPOD (*Copula-Based Outlier Detection*) que evalúa la función de distribución acumulada empírica basada en cópulas, lo que permite estimar la probabilidad conjunta de anomalía sin asumir normalidad multidimensional ni independencia entre campañas. De esta forma, el algoritmo identifica como anomalías (59

observaciones en la muestra) únicamente aquellos días donde el vector de ventas global presentó una desviación conjunta estadísticamente significativa. Esta identificación permitió garantizar que la posterior incorporación de variables *dummy* de intervención (o el tratamiento de imputación/limpieza de *outliers*) mantuviera la alineación temporal exacta a lo largo de todos los modelos individuales o multivariados (*XGBoost*, *VARX*), evitando distorsiones por falsos positivos locales.

La detección de observaciones extremas para evaluar el comportamiento individual de cada campaña, identificando un modelo independiente por campaña, no está dentro del alcance de este trabajo.

Figura 4.1.15. Frecuencia de *outliers* por campaña

Campaña	Outliers	No Outliers
Automotriz	59	523
Belleza_Bienestar	59	524
Climatizacion	59	523
Decoracion_Blanco	59	524
Grandes_Electrodomesticos_Cocina	59	524
Grandes_Electrodomesticos_Lavado	59	524
Grandes_Electrodomesticos_Refrigeracion	59	524
Jardineria	59	524
Mascotas	59	524
Menaje	59	524
Pequeños_Electrodomesticos	59	523
Telefonia	59	523

Fuente: elaboración propia

Los modelos *Prophet* se destacan por su robustez ante observaciones extremas y valores atípicos, habiendo sido diseñados específicamente para procesar series temporales largas con fuertes componentes estacionales. Para maximizar su precisión, resulta fundamental identificar e incorporar al modelo las anomalías, tales como registros faltantes, el impacto de eventos comerciales masivos (como los *cybers*) y los *outliers*, tal como se presentó en este análisis descriptivo.

En el siguiente apartado se presenta la aplicación de esta metodología para poder estimar las ventas netas (en dólares) tal como se planteó entre los objetivos de este trabajo.

4.2. Modelo *Prophet*: ajuste.

Para comenzar el abordaje de los modelos *Prophet*, lo primero que se debe mencionar es que el equipo de *Data Science* de *Facebook* (ahora *Meta*) diseñó, no sólo la metodología, sino también una biblioteca de código abierto en *Python*, que se llama *Prophet*. Fue creada, como ya se mencionó, para trabajar series de tiempo con fuertes efectos estacionales y varios períodos de datos históricos, lo que la hace muy útil para problemas de pronóstico empresarial facilitando a los científicos de datos y analistas la creación de pronósticos de alta calidad, incluso para aquellos con poca experiencia en modelos de series de tiempo complejos. También se destaca por la posibilidad de trabajar en simultáneo con múltiples series tiempo, cada una con su estacionalidad y tendencia.

Tal como se expresó en los primeros párrafos del análisis descriptivo, la variable a estimar o predecir es la **Venta neta diaria en dólares (*Net Media Value*)** (*mnv_USD* o *y*).

De aquí que como primera estrategia se realiza un análisis de correlaciones entre todas las variables. Esta técnica cuantitativa permite estimar y examinar la interdependencia entre dos o más variables. Específicamente en el contexto de las series de tiempo, el análisis de correlaciones es un instrumento utilizado para identificar y abordar la multicolinealidad. Al eliminar aquellas variables que muestran una alta correlación entre sí, se busca reducir la redundancia en el conjunto de datos y mejorar la estabilidad del modelo. Además, al eliminar información redundante, se puede mejorar la eficiencia computacional del proceso de modelado sin sacrificar significativamente el poder predictivo. Por lo tanto, este procedimiento no solo simplifica la metodología de modelado, sino que además contribuye a la interpretación y robustez del modelo predictivo.

En la Figura 4.2.1 que sigue a continuación, se presenta la matriz de correlación de las variables numéricas disponibles para cada campaña.

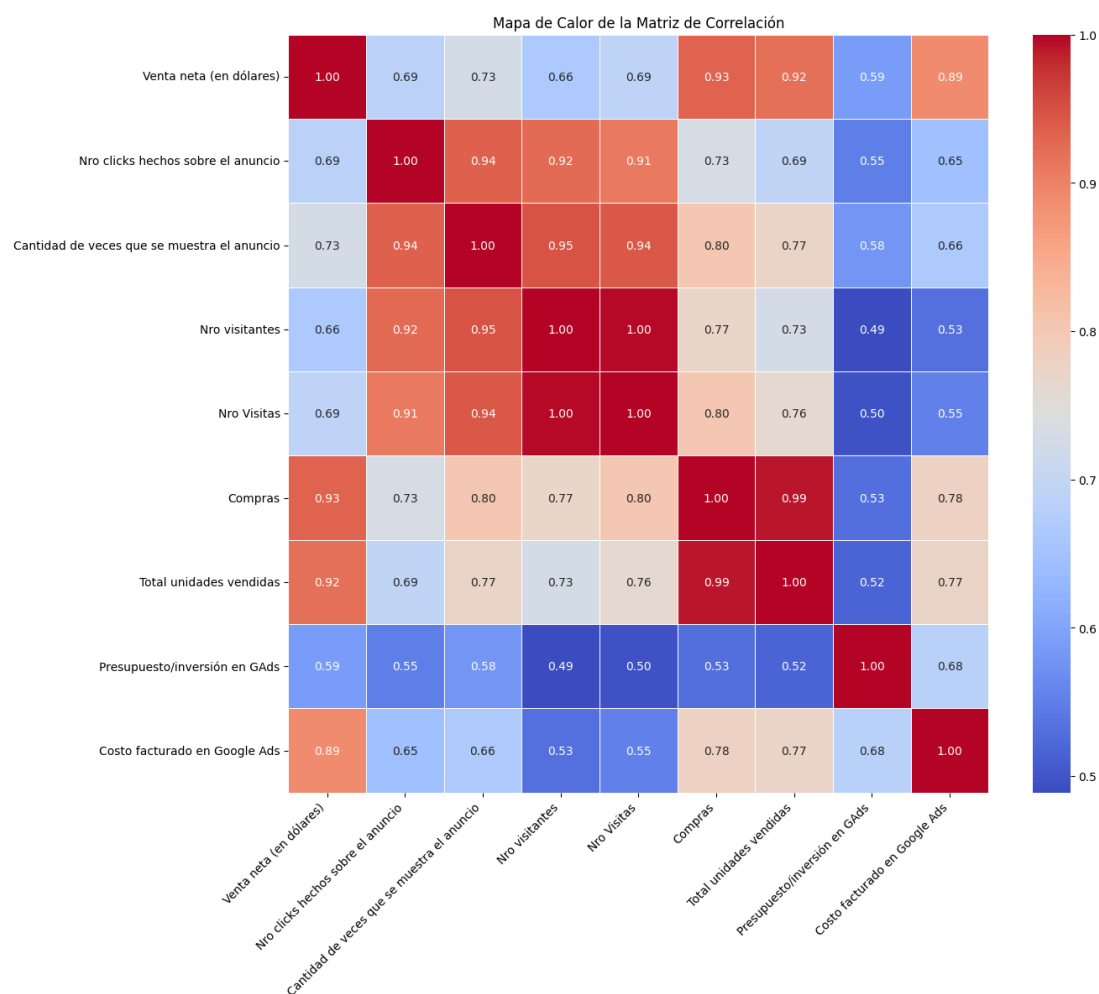
Se puede observar que todas las variables tienen un cierto grado de correlación con la venta neta (en dólares). Se diferencian dos grupos: las de correlación superior a 0,8 como compras, total de unidades vendidas y el costo facturado en *Google Ads*; y, las de correlación inferior a 0,8 como *clicks* sobre el anuncio, las impresiones o cantidad de veces que se muestra el anuncio, las visitas, los visitantes y el presupuesto diario invertido en *Google Ads*.

Por otro lado, el número de visitas y el número de visitantes tienen entre sí una alta correlación que podría deberse a que los visitantes tienen un promedio de visitas cercano a 1; es decir,

visitan el sitio una vez. Las impresiones y los *clicks* sobre el anuncio, también muestran un valor alto de correlación. Al igual que el total de unidades vendidas y las órdenes o compras.

Para la formulación y ajuste del modelo según lo conversado con el equipo de marketing, todas las covariables son importantes. De modo tal que el modelo propuesto debe contemplar estas estas variables cuantitativas junto con las variables indicadoras de eventos, *cyber* y *outliers* para cada campaña.

Figura 4.2.1. Matriz de correlación de variables registradas por campaña



Fuente: elaboración propia.

En el capítulo 3 “Metodología y Materiales” se presentó un modelo *Prophet* aditivo y otro multiplicativo. Para este estudio, el modelo multiplicativo es la opción adecuada para estimar las ventas netas (en dólares) por días. El modelo teórico en términos del problema en estudio se representa como:

$$y(t) = g(t) \cdot (1 + s(t) + h(t) + \sum_{j=1}^8 \beta_j x_j(t)) + \varepsilon_t$$

- $y(t)$: venta neta (en dólares)
- $g(t)$: tendencia lineal
- $s(t)$: estacionalidad. Se utilizó un ajuste automático de la estacionalidad, el resultado fue un ajuste semanal en este caso.
- $h(t)$ representa los efectos de días festivos o eventos especiales. Se estableció como componente *holiday* a los días de *cyber*, los días de eventos y los días que corresponden a valores *outliers* o son datos faltantes.
- $x_j(t)$: valor del j -ésimo regresor en el tiempo t . Los regresores o covariables expresados en vector tienen la siguiente definición:

$$x(t) = [\text{clicks}(t), \text{impressions}(t), \text{visitors}(t), \text{visits}(t), \text{orders}(t), \text{total_units_net}(t), \text{budget_USD}(t), \text{cost_USD}(t)]^T$$

- β_j : coeficiente que representa el impacto del regresor x_j
- ε_t : error con distribución Log-Normal $(0, \sigma^2)$ o ruido multiplicativo con media 1

A continuación, se presenta el modelado en términos de *Python* mediante el uso de la librería “*prophet*”:

```
model = Prophet(
    growth='linear',
    yearly_seasonality='auto',
    weekly_seasonality='auto',
    daily_seasonality='auto',
    seasonality_mode='multiplicative',
    seasonality_prior_scale=6.0,
    holidays=holidays_df,
    holidays_prior_scale=8.0,
    holidays_mode='multiplicative',
    n_changepoints=100,
    changepoint_range=0.6,
    changepoint_prior_scale=0.9,
    interval_width=0.95)

for cov in covariables:
    model.add_regressor(cov)
```

Componente de tendencia: `growth='linear'`

Se propuso lineal dado las ventas netas (en dólares) no tienen un valor de saturación o cota como para pensar un crecimiento logístico. *Prophet* modela la tendencia como una serie de segmentos lineales y los puntos de cambio (*changepoints*) son los lugares donde una pendiente puede cambiar, la fracción del historial (*changepoint_range*) de datos es la fracción en que el modelo puede ubicar puntos de cambio (*changepoints*) en la tendencia y un parámetro (*changepoints_prior_scale*) que controla cuán flexible es el modelo para adaptarse a cambios en la tendencia en los puntos de cambio (*changepoints*).

```
n_changepoints=100,  
changepoint_range=0.6,  
changepoint_prior_scale=0.9,
```

Según Taylor y Letham, la cantidad de puntos de cambio depende de la cantidad de datos que tiene la serie en estudio. Las sugerencias existentes son:

- si la serie de datos diaria es menor de un año, los puntos de cambio deben ser entre 10 y 25;
- si la serie de datos diaria tiene entre 1 o 2 años, los puntos de cambio deben ser entre 25 y 50
- si la serie de datos tiene más de 2 años, los puntos de cambio deben ser entre 50 y 100.

Puesto que las ventas en *e-commerce* suelen tener múltiples factores que generan cambios de tendencia como ser promociones, temporadas, lanzamientos, *Black Friday*, etc., se sugieren usar 100 puntos o más. En este caso, se fijaron 100 puntos.

Para la fracción de datos históricos (*changepoints_range*) las referencias son para cambios estructurales recientes se sugiere entre 0.5 y 0.7, para una distribución homogénea de los cambios se propone utilizar 0.8 que es el valor por defecto en cualquier ajuste. En este caso, de usó 0.6 lo que implica que se utiliza el 60% inicial de los datos históricos para encontrar los 100 puntos de cambio establecidos.

Respecto de la flexibilidad (*changepoints_prior_scale*), Manandhar, et. al. (2024) sugieren un valor de 0.01 a 0.1 cuando la tendencia tiene pocos cambios, es más bien estable; entre 0.2 y 0.5 cuando existen cambios estructurales frecuentes; entre 0.5 y 1.0 cuando la tendencia es altamente variable.

En temas de ventas por *e-commerce*, se recomienda utilizar hasta 0.3 si no existen promociones, temporadas, lanzamientos, *Black Friday*, etc., si no, debe ser mayor. En este caso se utilizó 0.9 dada la multiplicidad de campañas que se trabajan y cada una tiene su propia distribución.

Componente de estacionalidad: `seasonality_mode='multiplicative'`

Prophet puede detectar automáticamente la estacionalidad presente en los datos, puede trabajar con estacionalidad diaria, semanal y anual si los datos tienen la frecuencia y duración suficientes. En ese proyecto se trabajó con la detección automática de estacionalidad dado que son 12 series de tiempo con particularidades diferentes en cada una de ellas. Esto se estableció mediante los siguientes parámetros:

```
yearly_seasonality='auto',  
weekly_seasonality='auto',  
daily_seasonality='auto',  
seasonality_prior_scale=6.0,
```

Los primeros 3 parámetros marcan la identificación automática de estacionalidad. Mientras que el parámetro `seasonality_prior_scale` actúa como un parámetro de regularización que manifiesta qué tan fuerte o pronunciado debe ser el efecto de estas estacionalidades. Este parámetro puede asumir cualquier valor mayor que 0. La bibliografía²² propone

- Valores muy pequeños (por ejemplo: 0.01, 0.1): hacen que la estacionalidad sea muy suave y restringen fuertemente su capacidad para ajustarse a las fluctuaciones en los datos. Esto puede ser útil cuando no se necesita que el modelo capture el ruido como estacionalidad.
- Valores intermedios (por ejemplo: 1.0, 5.0, 10.0): son los más comunes y, en general, se usan como punto de partida. Permiten un ajuste razonable a los patrones estacionales sin un sobreajuste excesivo.
- Valores grandes (por ejemplo: 20.0, 50.0, 100.0 o más): permiten que la estacionalidad sea muy flexible y se ajuste estrechamente a las fluctuaciones de los datos. Esto puede ser útil cuando los datos tienen estacionalidad muy pronunciada y compleja, pero aumenta el riesgo de sobreajuste y puede generar pronósticos con intervalos de incertidumbre más amplios.

²² Taylor S. y Letham B. (2017)

Los valores por defecto que tiene *Prophet* son:

- para estacionalidad global: 10.0;
- para estacionalidad semanal: por defecto es 3.0;
- para estacionalidad anual: por defecto es 10.0;
- para la estacionalidad diaria: por defecto es 10.0.

La mejor manera de determinar el valor óptimo para `seasonality_prior_scale` es a través de la optimización de hiperparámetros y la validación cruzada. Esto es,

- definir un rango de valores a probar.
- Entrenar y evaluar el modelo para cada combinación de parámetros utilizando la validación cruzada de *Prophet*.

Luego, seleccionar el valor que minimiza una métrica de error (como RMSE o MAPE) en el conjunto de datos de prueba.

Es fundamental considerar que el parámetro `seasonality_prior_scale` no opera de forma aislada, sino que interactúa estrechamente con variables como `changepoint_prior_scale` (que rige la flexibilidad de la tendencia) y el orden de Fourier (`fourier_order`) de cada ciclo (que determina la complejidad geométrica de la estacionalidad). Tras realizar un proceso exhaustivo de experimentación y análisis conjunto de estos hiperparámetros, se determinó que un valor de 6.0 es el más adecuado para optimizar la estimación de la venta neta en dólares por campaña.

Componente *Holiday*: `holidays_mode='multiplicative'`

Optar por el modo multiplicativo para los días festivos es ideal cuando su impacto no es constante, sino proporcional al comportamiento del negocio. Esto significa que, si el nivel general de la serie de tiempo sube, el efecto de los días festivos también se vuelve más pronunciado, adaptándose mejor a la realidad de los datos. Por ejemplo,

- Ventas: una promoción o un día festivo importante (como *Black Friday* o Navidad) podría aumentar las ventas en un 50%. Si las ventas base son 1.000, el aumento es 500. El impacto absoluto crece con el nivel de las ventas.
- Tráfico web: un evento noticioso importante puede duplicar el tráfico a un sitio.

Para los datos bajo estudio, como la estacionalidad tiene un efecto multiplicativo, la definición de *holiday* también. Lo primero que se hizo fue evaluar si los días festivos chilenos tenían impacto en las ventas netas. Como se suponía, los días festivos no tienen efecto en las ventas

netas. Con lo cual se utilizó esta componente para modelar los momentos que si se sabe que tienen impacto. Estos son, días de eventos, días de *cyber* y días con valores extremos.

```
holidays=holidays_df,  
holidays_prior_scale=8.0,  
holidays_mode='multiplicative',
```

Por último, `holidays_prior_scale` es un parámetro de regularización que se encarga de controlar qué tan fuerte o pronunciado puede ser el impacto estimado de un día festivo en el pronóstico. Como regla general, valores altos para `holidays_prior_scale` suponen que las desviaciones causadas por los días festivos pueden ser grandes y significativas. En este caso, el modelo tendrá más libertad para asignar un efecto más fuerte a cada día festivo, ajustándose más a las fluctuaciones históricas observadas en esos días. Permite que el impacto de un día festivo sea más "*picudo*" o "*agudo*". Mientras que, valores bajos marcan una restricción de la magnitud del efecto que un día festivo puede tener. El modelo es más conservador, asumiendo que los días festivos tienen un impacto más suave o menos pronunciado.

Los valores prácticos para este parámetro son siempre mayores que cero y no tienen un límite superior. En la práctica, se suelen utilizar los siguientes rangos de referencia:

- 0.1 a 1.0: permite una flexibilidad baja a moderada.
- 1.0 a 10.0: donde se encuentra el valor por defecto.
- 10.0 a 50.0 (o incluso 100.0): permite una flexibilidad alta, haciendo que los días festivos tengan un impacto muy pronunciado.

Por defecto, *Prophet* propone un valor de 10.0, en este caso se utilizó 8.0. La selección del valor fue empírica, observando los valores de métricas como *MAPE* o *RMSE*.

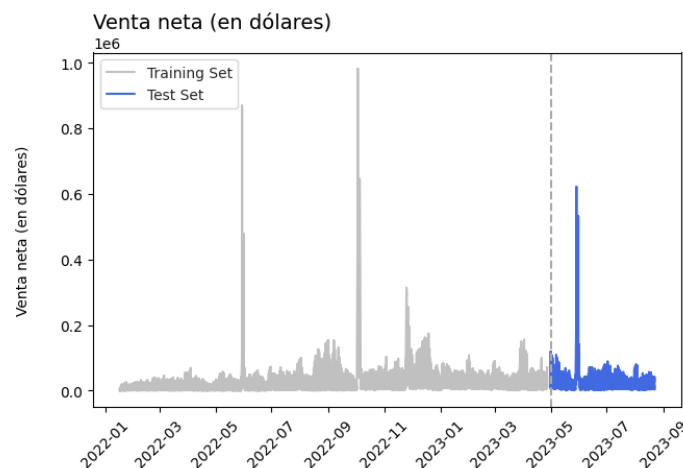
Covariables: *Prophet* permite incorporar covariables o regresores y se hace mediante la siguiente sentencia

```
for cov in covariables:  
    model.add_regressor(cov)
```

Una vez detallada cada componente desde el punto de vista teórico, para aplicar el ajuste del modelo, es necesario separar los datos en dos conjuntos, datos de entrenamiento (*train*) y datos de prueba (*test*), bajo la regla del 80/20. Tal como se presenta en la Figura 4.2.2 a modo

de ejemplo con la venta neta (en dólares), los datos de entrenamiento son aquellos comprendidos entre el 17 de enero de 2022 y el 27 de abril de 2023. Mientras que los datos de prueba corresponden al período comprendido entre el 28 de abril de 2023 y el 22 de agosto de 2023.

Figura 4.2.2.: Separación del conjunto de datos en entrenamiento y prueba para la variable respuesta venta neta en dólares.



Fuente: elaboración propia

A continuación, se presentan los resultados de cada componente del modelo por campaña para el conjunto de datos de entrenamiento. En la Figura 4.2.3. se muestran las tendencias, en la Figura 4.2.4. se presenta la componente estacional que se detectó automáticamente (semanal/*weekly*); en la Figura 4.2.5. se observa la componente *holiday* y, por último, la Figura 4.2.6. contiene los residuos que, deben seguir una distribución log-normal con media 1.

Según las Figuras mencionadas en el párrafo anterior, se puede ver que existen campañas con cierta tendencia descendente y una posterior estabilización a un nivel inferior como Menaje, Automotriz, Decoracion_Blanco, Pequeños Electrodomesticos, Jardineria. También, existen campañas que exhiben un patrón de fluctuante, pero con una dirección general que tiende a estabilizarse o a mostrar un cambio menos drástico a largo plazo, estas son: Telefonía, Belleza Bienestar, Grandes Electrodomesticos_Lavado, Grandes Electrodomesticos Refrigeracion, Grandes Electrodomesticos_Cocina, Por otra parte, las campañas Climatización y Mascotas tienen otras características en cuanto a tendencia.

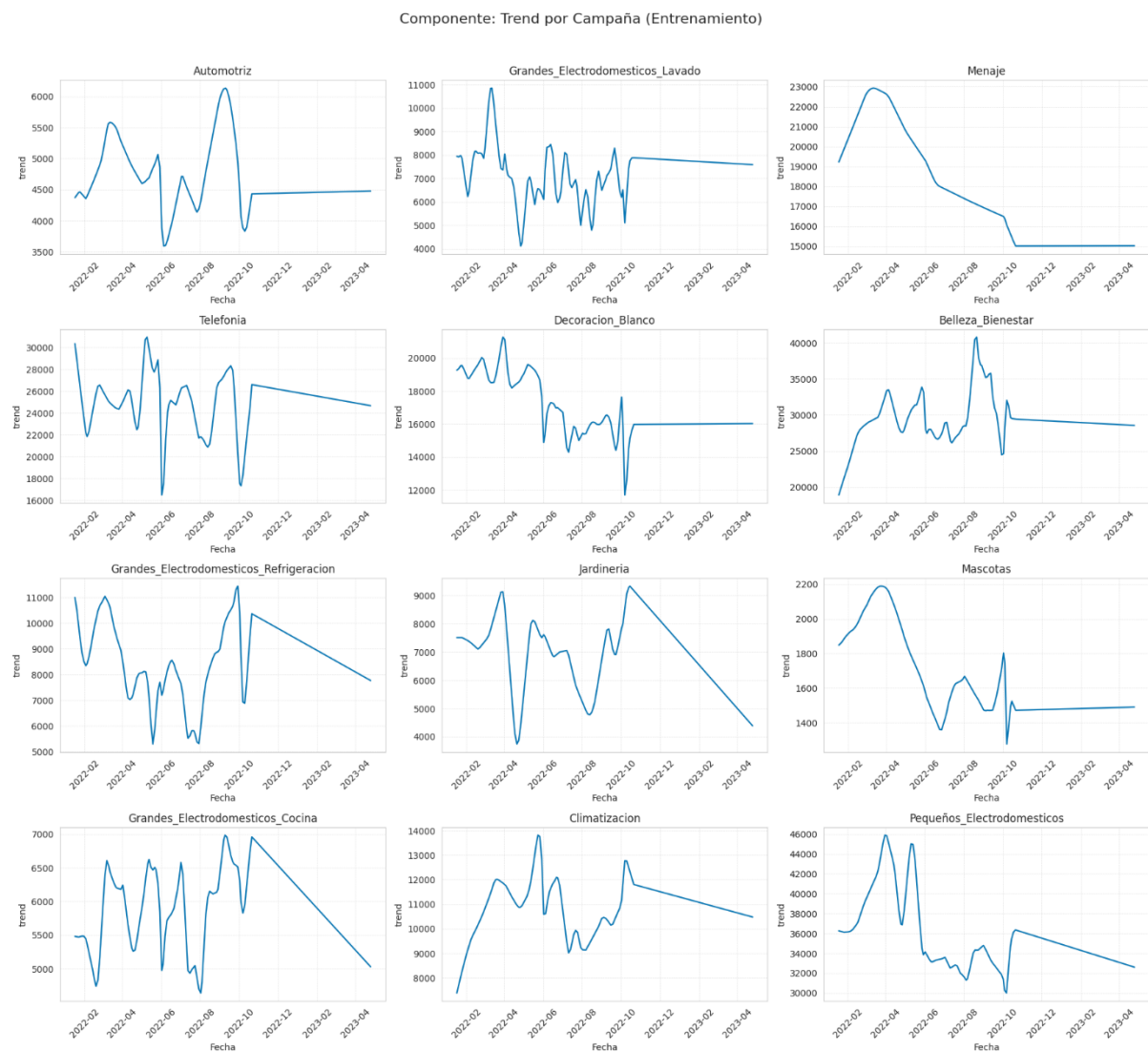
La Figura 4.2.3 presenta el componente de tendencia estimado para cada campaña durante el período de entrenamiento. En términos generales, se observan comportamientos heterogéneos entre campañas, aunque pueden identificarse algunos patrones comunes.

Por un lado, las campañas Menaje, Automotriz, Decoración_Blanco y Pequeños Electrodomésticos evidencian una tendencia descendente a lo largo del período analizado, seguida de una estabilización en niveles inferiores hacia el final de la serie. Este comportamiento sugiere una reducción progresiva del nivel estructural de búsquedas o interacciones asociado a dichas categorías.

En contraste, campañas como Telefonía, Belleza_Bienestar, Grandes_Electrodomésticos_Lavado, Grandes_Electrodomésticos_Refrigeración, Grandes_Electrodomésticos_Cocina y Jardinería presentan una mayor variabilidad temporal, con fluctuaciones más pronunciadas en la estimación de la tendencia. No obstante, estas oscilaciones no alteran el comportamiento general de largo plazo y convergen hacia trayectorias más estables en los períodos finales.

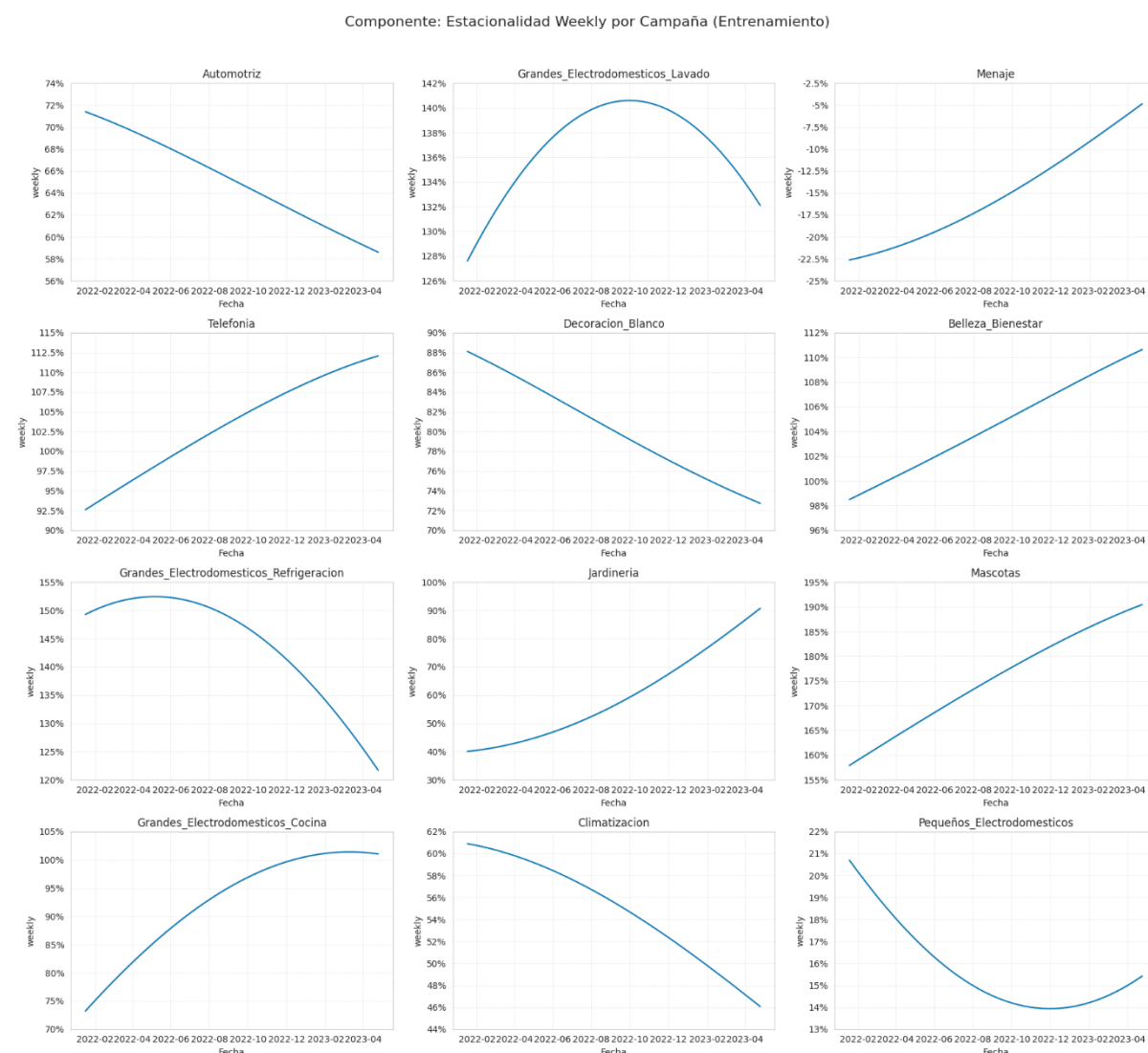
Los casos de Climatización y Mascotas no muestran un patrón sustancialmente diferente al resto de las campañas. Si bien exhiben niveles y amplitudes de variación particulares, la evolución de su tendencia también se encuentra dentro de los comportamientos observados en las demás categorías. En ambas campañas puede apreciarse que, a partir de aproximadamente octubre de 2022, la componente de tendencia adopta una trayectoria prácticamente lineal. Este comportamiento no debe interpretarse como una dinámica intrínseca del negocio, sino como una consecuencia de la metodología de estimación utilizada por *Prophet*. Dado que el modelo representa la tendencia mediante segmentos lineales unidos por puntos de cambio (*changepoints*), cuando no identifica evidencias suficientes de cambios estructurales adicionales en el tramo final de la serie, la tendencia estimada continúa siguiendo la pendiente más reciente observada. Como resultado, se obtienen trayectorias lineales suavizadas que reflejan la extrapolación del comportamiento estructural dominante en los datos históricos disponibles.

Figura 4.2.3. Componente de tendencia por campaña.



Fuente: elaboración propia

Figura 4.2.4. Componente estacional que se detectó automáticamente (semanal/weekly) por campaña



Fuente: elaboración propia

A partir de la Figura 4.2.4., se puede comprender cómo fluctúa la estacionalidad dentro de una semana típica para cada campaña. El eje vertical (*weekly*) representa el porcentaje de estacionalidad semanal para cada campaña o categoría de producto. Un valor del 100% (o una línea de base similar, como 1.0) significa que la métrica está en su promedio para esa semana. Se puede ver que para las campañas Menaje, Jardineria Mascotas, Belleza_Bienestar, Telefonía, Grandes_Electrodomesticos_Cocina el valor semanal fue aumentando a medida que avanza el tiempo. Esto significa que la estacionalidad para esa categoría sería más fuerte (o más alta) hacia el final del período observado. Mientras que, para las campañas Automotriz, Climatizacion y Decoracion_Blanco se observa que el valor

semanal va disminuyendo a medida que avanza el tiempo; significando que la estacionalidad para esas categorías sería más débil (o más baja) hacia el final del período observado.

Grandes_Electrodomesticos_Refrigeracion y Pequeños_Electrodomesticos muestran una estacionalidad descendente, pero con formas muy diferentes entre sí y con el primer grupo descrito anteriormente. Finalmente, Grandes_Electrodomesticos_Lavado presenta una forma de “U” invertida. Lo que significa que el punto máximo de 140% representa un volumen de actividad un 40% superior al promedio.

Figura 4.2.5. Componente *holiday* por campaña

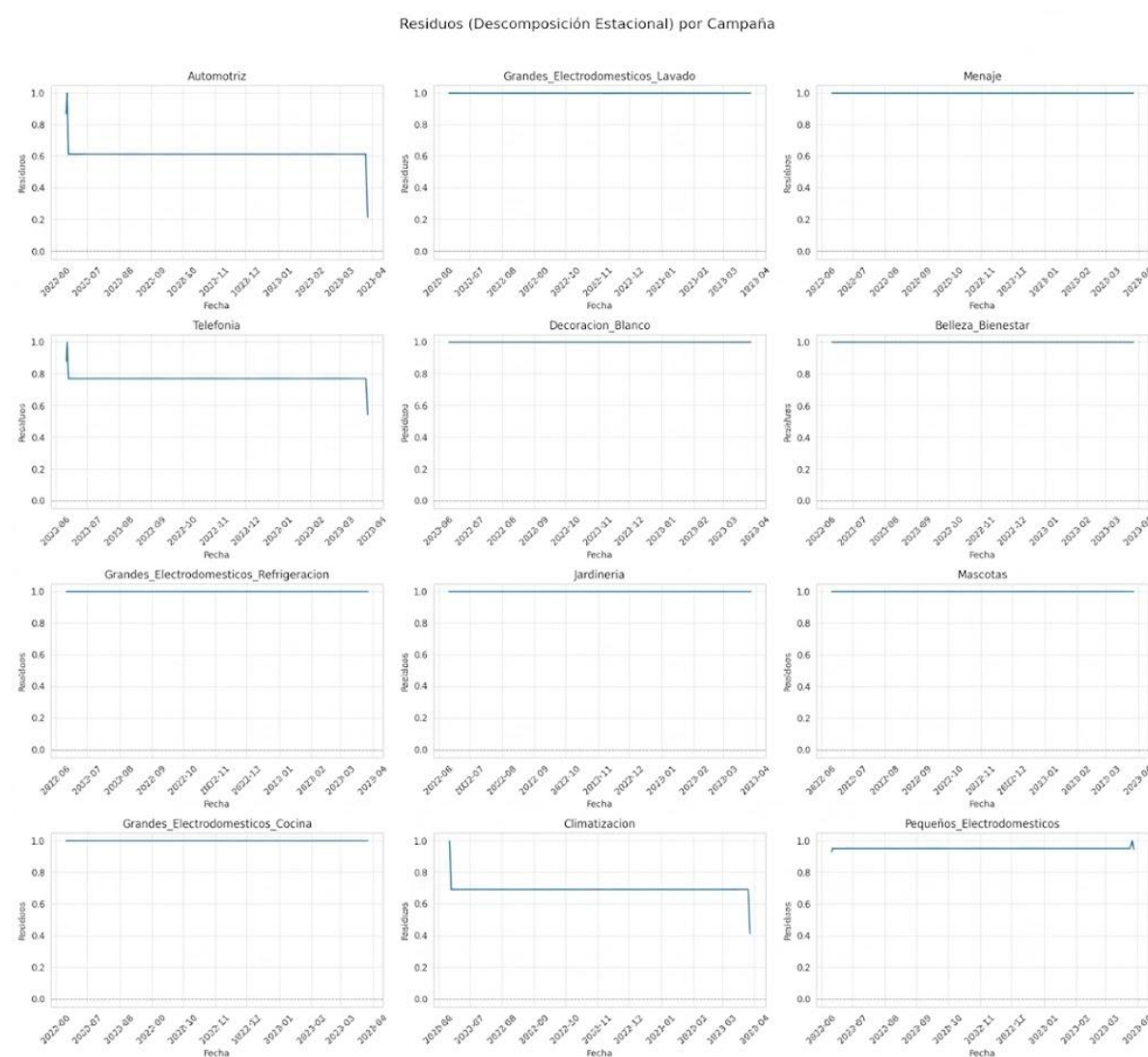


Fuente: elaboración propia

La componente de días festivos (*holiday*) que se presenta en la figura anterior, muestra la influencia de los días feriados o eventos específicos en el comportamiento de las ventas para cada categoría/campaña. Allí se observa, si tienden a aumentar o disminuir la actividad en

esos días en particular, una vez aislados los efectos de la tendencia y la estacionalidad regular. Por ejemplo, para Automotriz el efecto de los festivos es variable, hay días con un impacto positivo significativo (alrededor del +2.5%) y otros con un impacto negativo considerable (alrededor del -0.5%). Esto sugiere que algunos festivos pueden impulsar la actividad automotriz, mientras que otros la detienen. Situación similar se observa en Menaje, Grandes_Electrodomesticos_Cocina, Grandes_Electrodomesticos_Lavado, Climatizacion y Decoracion_Blanco. Existe otro grupo de campañas con campañas con impacto de días festivos predominantemente negativo: Belleza_Bienestar, Jardinería, Pequeños Electrodomesticos, Telefonía. En contraposición están las campañas de Grandes_Electrodomesticos_Refrigeracion y Mascotas donde el impacto de días festivos es predominantemente positivo. Esta situación de que algunos festivos tienen efectos positivos y otros negativos radica principalmente en el tipo de producto o campaña. Por ejemplo, cuando se trata de categorías ligadas a la rutina laboral o al consumo cotidiano, durante los días festivos, la rutina se interrumpe y la necesidad percibida de adquirir estos artículos cae. Mientras que cuando se trata de necesidades esenciales y de urgencia, como la categoría refrigeración, en días festivos, con la familia reunida en el hogar o preparando eventos, la exigencia sobre estos productos aumenta, disparando la búsqueda y compra urgente si se presenta una falla o si se aprovecha un descuento de fecha clave.

Figura 4.2.6. Residuos por campaña.



Fuente: elaboración propia

En la Figura 4.2.6 puede observarse que las campañas Automotriz, Telefonía y Climatización exhiben algunas desviaciones respecto de la media, representada por el valor 1, particularmente al inicio y al final de las series. Sin embargo, para el resto de las campañas, los residuos oscilan aleatoriamente alrededor de dicho valor sin presentar tendencias ni patrones persistentes, validando el supuesto correspondiente a los residuos de un modelo *Prophet*.

Para terminar de completar el desarrollo del modelo propuesto, se presentan los coeficientes de cada variable regresora o independiente por campaña en el conjunto de entrenamiento.

Tabla 4.2.1.: Coeficientes estimados para cada variable independiente por campaña.

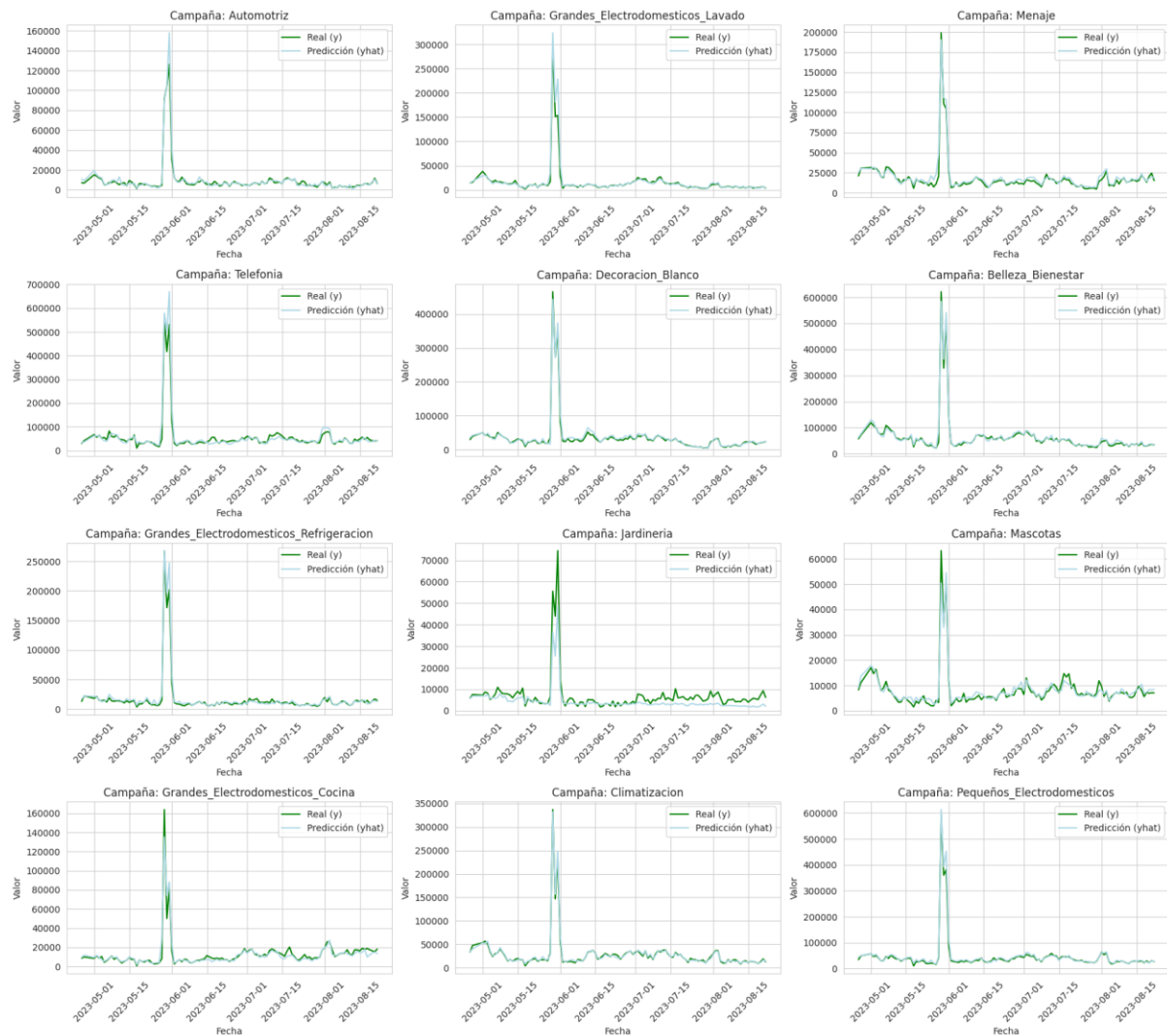
Variable regresora / Campaña	Presupuesto/inversión en Google Ads (budget_USD)	Nro clicks hechos sobre el anuncio (clicks)	Costo facturado en Google Ads (cost_USD)	Cantidad de veces que se muestra el anuncio (impressions)	Compras (orders)	Total unidades vendidas (total_units_net)	Nro visitantes (visitors)	Nro visitas (visits)
Automotriz	0,000122	0,000010	0,000327	0,000000	0,014188	0,003643	-0,000378	0,000200
Belleza_Bienestar	0,000058	0,000001	0,000078	0,000000	0,001266	0,000384	-0,000003	-0,000002
Climatizacion	-0,000089	-0,000003	0,000132	0,000000	0,009071	0,000114	-0,000106	0,000082
Decoracion_Blanco	-0,000050	0,000005	0,000158	0,000000	0,003452	0,000322	-0,000030	0,000027
Grandes_Electrodomesticos_Cocina	0,000436	0,000008	0,000184	0,000000	0,021298	0,001038	-0,000070	0,000014
Grandes_Electrodomesticos_Lavado	0,000132	-0,000009	0,000413	0,000000	0,018406	0,000886	0,000042	-0,000068
Grandes_Electrodomesticos_Refrigeracion	0,000047	-0,000002	0,000033	0,000000	0,018134	0,000395	0,000051	-0,000025
Jardineria	-0,000059	0,000015	0,001101	0,000000	0,009825	0,000933	-0,000070	-0,000076
Mascotas	0,000244	-0,000027	0,001101	0,000000	0,028565	0,003824	-0,000580	0,000373
Menaje	0,000189	-0,000015	0,000280	0,000000	0,002052	0,000675	-0,000072	0,000041
Pequeños_Electrodomesticos	0,000000	0,000003	0,000102	0,000000	0,001851	0,000195	-0,000034	0,000026
Telefonia	0,000115	0,000017	0,000151	0,000000	0,004325	0,000356	-0,000156	0,000085

Fuente: elaboración propia

A partir de la tabla anterior se puede observar que la venta neta (en dólares) por campaña está compuesta por diferentes signos y magnitudes de los coeficientes de cada variable regresora. En el Anexo 7.2., se presenta una representación gráfica de estos coeficientes para cada campaña (Figura 7.2.1.)

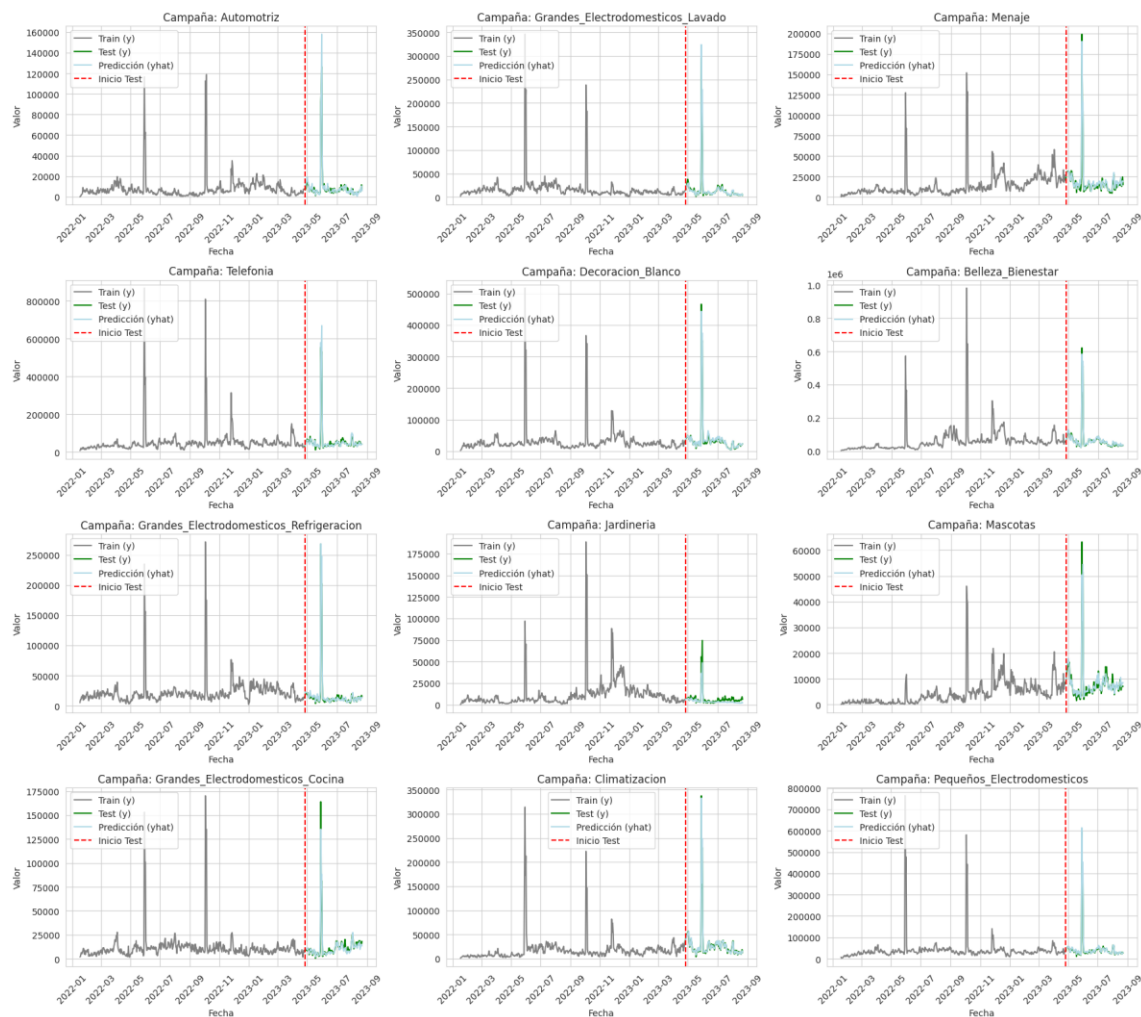
Una vez entrenado el modelo, se ajusta sobre los datos de prueba para determinar qué tan bien predice los valores de la variable dependiente. En este caso, la venta neta (en dólares). Los resultados hallados se pueden ver en la Figura 4.2.7. donde se presenta la venta neta (en dólares) estimada junto con la venta neta (en dólares) del conjunto de datos de prueba por campaña. En la Figura 4.2.8. se presenta esta misma información integrando toda la serie de datos, marcando la separación de datos de entrenamiento y su ajuste sobre el conjunto de prueba.

Figura 4.2.7.: Venta neta estimada (en dólares) vs venta neta (en dólares) por campaña para el conjunto de datos de prueba.



Fuente: elaboración propia

Figura 4.2.8.: Venta neta estimada (en dólares) vs venta neta (en dólares) por campaña para todo el conjunto de datos.

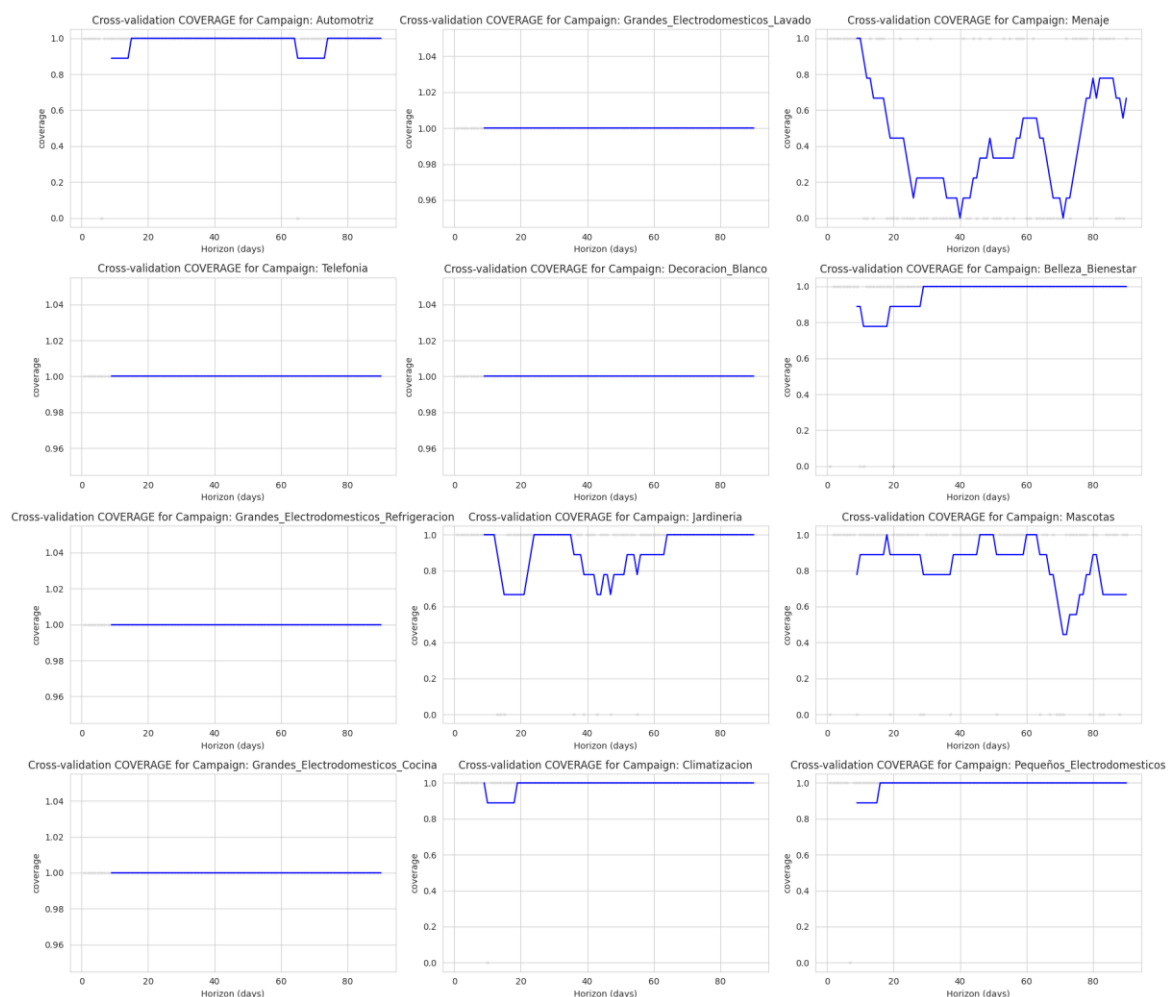


Fuente: elaboración propia

A partir de las gráficas anteriores se puede apreciar un buen ajuste y poder predictivo del modelo propuesto para cada campaña, tanto en los días regulares como los días de cyber. Esta misma observación se puede validar con las medidas de validación del modelo y capacidad de predicción planteadas en el apartado 3.1.2. Inicialmente, se presenta la Figura 4.2.9. con una medida específica de los modelos *Prophet* para evaluar su poder predictivo, *Coverage* que mide el porcentaje de puntos reales que caen dentro de los intervalos de incertidumbre (límite inferior de “*y estimado*” y límite superior de “*y estimado*”) del modelo para diferentes horizontes de tiempo (validación cruzada). En 5 campañas, el % de cobertura hasta un horizonte de 90 días, es de 1, lo que implica un buen pronóstico de la venta neta (en dólares) en las campañas de Grandes Electrodomésticos Lavado, Telefonía, Decoración Blanco, Grandes Electrodomésticos Refrigeración y Grandes Electrodomésticos Cocina. En

las campañas de Climatización y Pequeños Electrodomésticos, la cobertura o certidumbre del pronóstico hasta los 20 días fluctúa entre 0.8 y 1, de ahí en adelante es 1. La campaña Menaje es la que mayor variación tiene en cuanto al % de cobertura, situaciones similares ocurren con las campañas de Mascotas y Jardinería. Un aspecto relevante de esta métrica es que permiten evaluar de manera anticipada la capacidad predictiva del modelo para un momento determinado en el tiempo, así como cuantificar el nivel de confianza asociado a las predicciones generadas. En la Tabla 4.2.2. se exponen los valores de *MAPE*, *MAE* y *RMSE* para el conjunto de entrenamiento (*train*) y para el conjunto de prueba (*test*). Se eligen estas medidas a fin de poder establecer una comparación de con lo presentado por Costa, C. (2024) en su trabajo final, se toman las medidas generales; sin embargo, en el Anexo 7.2. se presenta la opción de validación cruzada para cada una de ellas (Figura 7.2.1., Figura 7.2.2. y Figura 7.2.3.). En la Tabla 4.2.3. se presentan los valores de *MAPE* para todos los modelos, las medidas restantes se presentan en el Anexo 7.2. (Tabla 7.2.1 y Tabla 7.2.2).

Figura 4.2.9.: Validación cruzada para Cobertura (*coverage*) por campaña.



Fuente: elaboración propia

Tabla 4.2.2.: valores de *MAPE*, *MAE* y *RMSE* para el conjunto de entrenamiento (*train*) y para el conjunto de prueba (*test*)

Campaña / Medidas	Entrenamiento (Train)			Prueba (Test)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Automotriz	1.012,12	1.346,85	20,51	1.571,41	3.533,72	21,45
Belleza_Bienestar	3.185,82	4.901,59	7,20	5.206,34	8.465,48	9,41
Climatizacion	1.871,14	2.609,97	15,27	2.314,10	3.286,81	10,63
Decoracion_Blanco	2.156,36	3.030,37	8,85	3.681,21	5.858,61	12,33
Grandes_Electrodomesticos_Cocina	1.285,43	1.703,39	14,66	2.273,98	4.656,87	20,02
Grandes_Electrodomesticos_Lavado	1.597,07	2.150,50	13,35	2.863,85	8.652,89	17,05
Grandes_Electrodomesticos_Refrigeracion	2.019,23	2.615,22	11,98	2.789,95	5.717,91	19,97
Jardineria	1.390,47	2.012,97	20,62	2.589,64	4.288,11	36,51
Mascotas	659,22	959,69	28,32	1.238,20	1.830,16	19,60
Menaje	1.386,07	2.010,09	13,90	2.834,44	4.259,45	20,96
Pequeños_Electrodomesticos	2.584,42	3.458,55	8,27	4.605,02	9.002,51	11,18
Telefonia	4.994,60	7.510,95	11,63	9.769,94	19.537,21	17,56

Fuente: elaboración propia

Tabla 4.2.3.: comparación de valores de *MAPE* del modelo *Prophet* con modelos propuestos en el trabajo final de Costa, C. (2024)

Campaña	Modelo	Mejor Modelo (2024)	
	Prophet	XGBoost	VARMAX
Automotriz	21,45	18,13	
Belleza_Bienestar	9,41	15,42	
Climatizacion	10,63	15,49	
Decoracion_Blanco	12,33	13,74	
Grandes_Electrodomesticos_Cocina	20,02		20,18
Grandes_Electrodomesticos_Lavado	17,05	359,86	
Grandes_Electrodomesticos_Refrigeracion	19,97	33,83	
Jardineria	36,51		28,44
Mascotas	19,60	18,76	
Menaje	20,96	20,25	
Pequeños_Electrodomesticos	11,18	12,05	
Telefonia	17,56		15,14

Fuente: elaboración propia

Como se puede observar a partir de la tabla anterior, *Prophet* tiene un mejor desempeño que los modelos *XGBoost* y *VARMAX* en 7 campañas: Belleza_Bienestar, Climatización, Decoración Blanco, Grandes Electrodomésticos Cocina+Lavado+Refrigeración y Pequeños

Electromésticos. Para Menaje y Mascotas, alcanzan valores similares; mientras que, en Jardinería, Automotriz y Telefonía, *Prophet* tiene un desempeño inferior. Con respecto a MAE y RMSE (Tabla 7.2.1 y Tabla 7.2.2 del Anexo 7.2.) *Prophet* tiene mejor ajuste en Automotriz, de modo tal que solo para Jardinería y Telefonía en las tres medidas evaluadas tiene valores que marcan un ajuste menor. A pesar de esto y a partir de esta evidencia, se puede considerar que los modelos *Prophet* son una buena opción para la estimar la venta neta (en dólares) para cada campaña dado que, en comparación con los otros modelos, el costo de análisis y procesamiento previo de datos es mucho menor. No existe la necesidad de imputación de valores extremos, faltantes o días de *cyber*, tampoco fue necesario identificar y evaluar la influencia de días previos y días posteriores a los eventos y/o *cyber* para incorporarlos al modelo como sí se realizó en el ajuste de los modelos *VARMAX* y *XGBoost*. *Prophet* lo resuelve mediante los parámetros de pesos o ponderación de tendencia y estacionalidad, además de *holiday*.

Dados los resultados hallados en el ajuste de los modelos *Prophet* para estimar la venta neta (en dólares) por campaña, se procede a evaluar el siguiente objetivo a cumplir en este trabajo, la optimización del presupuesto o inversión diario por campaña. Se utilizará el resultado de este modelo, como dato de entrada para generar un modelo de asignación de presupuesto óptimo que se presenta en el apartado siguiente.

4.3. Modelo de asignación: optimización Función Objetivo

El problema de asignación de presupuesto publicitario es un desafío importante en el ámbito empresarial, especialmente en el área de marketing y ventas. Es por lo que se intenta determinar la cantidad de recursos financieros que se deben asignar a cada categoría de productos o campaña con el fin de maximizar la efectividad de la campaña de *Google Ads*.

Se busca asignar un presupuesto determinado a cada categoría de manera que se alcance el mayor volumen de ventas netas posible con un gasto mínimo. Esto se aborda desde la programación lineal, con un modelo de optimización lineal que busca encontrar la asignación de presupuesto óptima que permita cumplir con una serie de restricciones. El modelo está compuesto por una función objetivo y un conjunto de restricciones.

La función objetivo para este caso en estudio queda definida del siguiente modo:

$$\text{función objetivo: } \min \sum_{i=1}^{12} \sum_{d=1}^n \frac{X_{id}}{C_{id}} = \min \sum_{i=1}^{12} \sum_{d=1}^n AdCost_{id} \quad (4.3.1)$$

donde:

X_{id} representa la cantidad²³ de dólares a invertir en la categoría i para el día d

C_{id} representa la venta neta (en dólares) correspondientes a la categoría i en el día d

En particular para este estudio, C_{id} representa la venta neta estimada que se obtuvo mediante el ajuste del modelo *Prophet* presentado en la sección anterior. Esta función objetivo busca minimizar la *ratio* total de inversión (X_{id}) respecto a la venta neta correspondiente a las conversiones (C_{id}), lo que implica asignar de manera eficiente el presupuesto para maximizar el rendimiento de las campañas publicitarias en cada categoría y cada día; en términos la función propuesta para este trabajo, lo que se pretende es minimizar el valor AdCost, presentado en el apartado de descripción del conjunto de datos. Este enfoque permite asignar recursos de manera estratégica y priorizar aquellas categorías que generan un mayor impacto con una menor inversión. Para lograr este objetivo, es necesario establecer una serie de restricciones sobre la función objetivo:

- Restricción de inversión máxima diaria o presupuesto diario no podrá superar el valor máximo asignado para ese día para alguna de las campañas.
- Restricción de presupuesto diario no negativo por campaña, inversión mayor a 0 dólares.
- Restricción de presupuesto mínimo debe ser el valor correspondiente a la mínima inversión asignada para ese día para alguna de las campañas.
- Restricción de inversión máxima diaria o presupuesto diario en categorías menos importantes: 'Automotriz', 'Jardinería', 'Mascotas'. Entre las 3 campañas no debe superar el 15% del presupuesto diario.
- Restricción de inversión máxima diaria o presupuesto diario en categorías más importantes: deben acumular el 75% del presupuesto diario. Junto con una restricción de inversión mínimo que indica que al menos debe tener 5% del presupuesto diario en cada campaña de grupo de categorías más importantes.
- Restricción de función objetivo (AdCost): debe ser siempre positivo (>0).

La optimización de la función se realizó mediante la librería `scipy.optimize(minimize)` de *Python* y se utilizó el método de optimización “Programación Secuencial de Mínimos Cuadrado ('SLSQP'²⁴)” que es un algoritmo adecuado para problemas con restricciones y cotas tal como se expresó en las restricciones del párrafo anterior. Observando los resultados por campaña para todo el conjunto de datos de prueba para los cuales se dispone de los valores de venta neta (en dólares) estimada, se puede ver que esta función de optimización

²³ budget_USD: Presupuesto o inversión ingresada plataforma de Google Ads

²⁴ SLSQP: Sequential Least Squares Programming

mejora los valores de presupuesto o inversión publicitaria a partir de la minimización de los de la función AdCost, optimizando el presupuesto en USD 414.517,47, logrando pasar del valor USD 1.441.276 a USD 1.026.759. Los resultados se presentan en la siguiente tabla (Tabla 4.3.1.). Es importante resaltar que, para este análisis, la venta neta (en dólares) estimada utilizada para el cálculo de optimización de AdCost se consideró como una variable exógena fija, sin recálculo de diario a partir de la optimización diaria del presupuesto o inversión publicitaria. Lo que implica que el modelo optimiza el costo asumiendo que la demanda no cambia y la minimización del AdCost tenderá a reducir el presupuesto hasta los mínimos dados por las restricciones sin modelar cómo esa reducción afectaría a las ventas. Sería propicio, en un futuro, conectar *Prophet* con el algoritmo de optimización para que las ventas reaccionen en tiempo real. Así, si el modelo decide recortar presupuesto, se podrá saber con precisión cuántas ventas se perderían y cuál es el punto exacto donde conviene dejar de invertir.

Tabla 4.3.1.: comparación de presupuesto optimizado por campaña con los valores de presupuesto original por campaña para alcanzar los valores de venta neta estimada (en dólares) hallados con el modelo *Prophet*.

Campaña	Venta neta estimada (en dólares) (Yhat)	Presupuesto o inversión ingresada plataforma de Google Ads (Budget_USD)	Presupuesto optimizado (Budget_optimizado)	AdCost (Budget_USD/ Yhat)	AdCost optimizado (por la función objetivo)
Automotriz	\$ 1.096.088	\$ 30.904	\$ 4.636	0,0282	0,0042
Belleza_Bienestar	\$ 7.554.565	\$ 199.005	\$ 149.254	0,0249	0,0125
Climatizacion	\$ 3.274.068	\$ 131.956	\$ 120.091	0,0456	0,0228
Decoracion_Blanco	\$ 4.285.898	\$ 160.122	\$ 80.060	0,0475	0,0238
Grandes_Electrodomesticos_Cocina	\$ 1.397.849	\$ 79.867	\$ 59.900	0,0543	0,0271
Grandes_Electrodomesticos_Lavado	\$ 1.979.783	\$ 95.460	\$ 71.595	0,0523	0,0261
Grandes_Electrodomesticos_Refrigeracion	\$ 2.056.719	\$ 109.138	\$ 81.854	0,0595	0,0298
Jardineria	\$ 507.754	\$ 34.159	\$ 5.124	0,0732	0,0366
Mascotas	\$ 932.402	\$ 25.268	\$ 3.790	0,0278	0,0139
Menaje	\$ 2.247.056	\$ 79.349	\$ 59.512	0,0329	0,0164
Pequeños_Electrodomesticos	\$ 5.554.301	\$ 261.350	\$ 196.013	0,0487	0,0243
Telefonia	\$ 6.751.362	\$ 234.698	\$ 176.023	0,0361	0,018
Total	\$ 37.637.845	\$ 1.441.276	\$ 1.026.759	0,0383	0,0273

Fuente: elaboración propia.

A partir de la tabla anterior se puede ver que para alcanzar el mismo valor de venta neta estimada con el presupuesto o inversión en *Google Ads* originales se necesita mucho más dinero que luego de la optimización. Por ejemplo, para la campaña Automotriz para el período comprendido entre 1 de mayo de 2023 y el 22 de agosto de 2023 (datos de prueba) se

necesitó una inversión o presupuesto de 0.0282 centavos de dólar por cada dólar que ingresó por venta neta estimada; mientras que, esta relación luego de la optimización se reduce a 0.0042 centavos de dólar en *Google Ads* por cada dólar que se estimó por venta neta para esa campaña. Aplicando la Prueba de *Wilcoxon* para las medianas de AdCost y Adcost optimizado, se puede concluir que la optimización ha logrado reducir significativamente el AdCost ($p=0.000489$). La prueba se presenta en el Anexo Figura 7.3.1.

Otra ventaja de este modelo de optimización es que se puede distribuir un valor fijo de presupuesto para un día específico. Por ejemplo, para el día 28 de junio de 2023 se dispone de USD 10. para alcanzar las ventas netas estimadas para ese día optimizando la inversión publicitaria con la minimización de AdCost con las restricciones antes previstas, la distribución por campaña hubiese sido de USD 7.499,99 tal como se observa en la Tabla 4.3.2.

Tabla 4.3.2.: presupuesto optimizado por campaña para el día 28 de junio de 2023 a partir de un valor fijo de USD 10.000

Campaña	Venta neta estimada (en dólares) (Yhat)	Presupuesto o inversión ingresada plataforma de Google Ads (Budget_USD)	AdCost (Budget_USD/ Yhat)	Presupuesto optimizado (Budget_optimizado)	AdCost optimizado (por la función objetivo)
Automotriz	6.137,72	92,69	0,0151	10,00	0,0016
Belleza_Bienestar	81.821,79	2.418,85	0,0296	1.912,21	0,0234
Climatizacion	32.651,16	1.690,83	0,0518	1.184,19	0,0363
Decoracion_Blanco	38.298,48	1.282,74	0,0335	776,10	0,0203
Grandes_Electrodomesticos_Cocina	12.756,81	522,69	0,0410	16,05	0,0013
Grandes_Electrodomesticos_Lavado	13.334,06	650,37	0,0488	143,73	0,0108
Grandes_Electrodomesticos_Refrigeracion	9.023,08	741,03	0,0821	234,39	0,0260
Jardineria	2.943,14	125,39	0,0426	10,00	0,0034
Mascotas	8.391,90	258,71	0,0308	10,00	0,0012
Menaje	15.858,49	496,08	0,0313	10,00	0,0006
Pequeños_Electrodomesticos	47.695,68	2.355,89	0,0494	1.849,25	0,0388
Telefonia	39.696,73	1.850,70	0,0466	1.344,06	0,0339
TOTAL	10.000			7.499,99	

Fuente: elaboración propia.

Claramente se puede ver que la optimización del presupuesto a partir de esta función permite un ahorro de un 25% del presupuesto para un día determinado.

A continuación, se presentan dos ejemplos de optimización para dos días específicos: para el día 29 de mayo de 2023 es un día de *cyber*, y otro, para un día regular como lo fue el 22 de agosto de 2023. En esta oportunidad, se fijan los valores de presupuesto diario asignados para

esos días específicos (USD 123.590,42 y USD 7.229,12, respectivamente) a fin de evaluar cuál hubiese sido el resultado si se hubiese tenido este trabajo desarrollado como *input* de asignación. Tal como se presenta en la Tabla 4.3.3. y Tabla 4.3.4., tanto para el día de *cyber* como el día regular, se logró un valor menor de presupuesto asignado luego de utilizar la función de optimización. Al igual que en el caso anterior, el presupuesto se optimiza en un 25%.

Tabla 4.3.3.: Optimización de presupuesto para el día 29 de mayo de 2023 (día *cyber*)

Campaña	Venta neta estimada (en dólares) (Yhat)	Presupuesto o inversión ingresada plataforma de Google Ads (Budget_USD)	Presupuesto optimizado (Budget_optimizado)
Automotriz	90.343,10	4.018,39	602,76
Belleza_Bienestar	584.102,54	17.078,41	12.808,81
Climatizacion	332.600,07	8.352,85	6.264,64
Decoracion_Blanco	443.446,76	14.363,23	10.772,42
Grandes_Electrodomesticos_Cocina	135.279,62	7.127,40	5.345,55
Grandes_Electrodomesticos_Lavado	324.025,05	10.008,14	7.506,11
Grandes_Electrodomesticos_Refrigeracion	268.134,30	10.360,40	7.770,30
Jardineria	36.490,91	3.114,17	467,13
Mascotas	50.446,76	1.692,07	253,81
Menaje	190.274,70	5.807,09	4.355,32
Pequeños_Electrodomesticos	613.747,23	23.481,96	17.611,47
Telefonia	579.525,89	18.186,31	13.639,73

Presupuesto total REAL 123.590,42 USD

Presupuesto total OPTIMIZADO 87.398,04 USD

Diferencia -36.192,38 USD

Fuente: elaboración propia

Tabla 4.3.4.: Optimización de presupuesto para el día 22 de agosto de 2023 (día regular)

Campaña	Venta neta estimada (en dólares) (Yhat)	Presupuesto o inversión ingresada plataforma de Google Ads (Budget_USD)	Presupuesto optimizado (Budget_optimizado)
Automotriz	5787,97	132,49	19,87
Belleza_Bienestar	35443,16	835,18	626,38
Climatizacion	13491,95	514,59	385,94
Decoracion_Blanco	24040,29	598,58	448,93
Grandes_Electrodomesticos_Cocina	13512,25	826,90	620,17
Grandes_Electrodomesticos_Lavado	4362,15	220,03	165,02
Grandes_Electrodomesticos_Refrigeracion	13063,38	831,63	623,72
Jardineria	2056,77	212,93	31,93
Mascotas	8500,91	176,76	26,43
Menaje	18891,18	587,94	440,95
Pequeños_Electrodomesticos	28942,06	778,39	583,79
Telefonia	40741,25	1514,20	1135,65

Presupuesto total REAL 7229,12 USD

Presupuesto total OPTIMIZADO 5108,83 USD

Diferencia -2120,29 USD

Fuente: elaboración propia

Como síntesis de este apartado se puede ver que utilizar una función de objetivo para distribuir el presupuesto diario de inversión publicitaria, tuvo sus beneficios. Si bien fue una opción simple de asignación, es claro que hay un camino intenso por recorrer en este sentido. En términos de negocios, no siempre responde a una estimación lineal la optimización del presupuesto, en esta oportunidad se propuso una posible solución de asignación entre las infinitas disponibles. Este ensayo deja como valor agregado que existe la posibilidad de trabajo conjunto entre marketing, datos y negocio para lograr un mejor rendimiento de las campañas y del negocio en sí. En el apartado siguiente se presentan algunas líneas de reflexión sobre el trabajo realizado tanto en aplicación de *Prophet* como en la optimización de la función objetivo.

5. Discusión

En el desarrollo del proyecto se llevaron a cabo una serie de revisiones para abordar de manera integral el planteamiento del problema inicial. La presentación y utilización de modelos *Prophet* para estimar las ventas netas por campañas para la empresa de comercio electrónico chilena ha mostrado resultados interesantes, no sólo por la bondad del ajuste del modelo si no por simpleza y practicidad que tienen al momento de su aplicación. Desde la perspectiva

del negocio, poder trabajar con una serie de tiempo en su estado original alcanzando resultados viables para la toma de decisiones es punto para resaltar. En comparación con los resultados hallados entre los antecedentes en esta temática, que propusieron una solución a esta necesidad a través de un enfoque automatizado para la predicción de ventas en comercio electrónico mediante la evaluación de campañas de marketing (Costa, 2024), este modelo tuvo un buen desempeño y es de destacar el tiempo que se puede ahorrar en un equipo de trabajo durante el proceso de limpieza de datos si el logro de los resultados es muy similar. La utilización de estas estimaciones permitió dar un marco de claridad sobre la asignación eficiente de recursos publicitarios, especialmente cuando se gestionan múltiples campañas publicitarias en forma simultánea como en este trabajo. Mediante una programación lineal, con un modelo de optimización lineal a través de una función objetivo y un conjunto de restricciones, se logró desarrollar un plan de inversión diaria para las 12 campañas de la empresa de comercio electrónico chileno con el objetivo de maximizar la efectividad publicitaria, asignando un presupuesto determinado a cada categoría de manera que se alcance el mayor volumen de ventas netas posible con un gasto mínimo. Como reflexión final, alcanzar estos objetivos tuvo como propósito mostrar la sinergia entre negocio y ciencia, demostrando que pueden ayudarse a optimizar la toma de decisiones estratégicas en marketing digital, incrementar la agilidad empresarial y mejorar la competitividad en un entorno altamente dinámico con buenos resultados y a un tiempo adecuado para el ritmo del negocio.

Es importante resaltar que en materia de futuro sería propicio considerar un modelo con métricas de rendimiento del negocio como *AdCost*, *ROI*, *CTR*, etc. en lugar de las variables originales correlacionadas. El modelado con estas métricas podría ser una alternativa para superar el punto de la correlación entre las variables predictoras. También, sería interesante ajustar un modelo con otras opciones de parámetros para encontrar el equilibrio ante una posible situación de sobreajuste. Por último, en lo que respecta a la asignación de presupuesto o inversión publicitaria hay un largo camino por recorrer, agregando nuevos escenarios de mínimos y máximos, de restricciones, o bien, la utilización de otras metodologías de optimización.

Limitaciones del estudio y recomendaciones para trabajos futuros.

Si bien los resultados obtenidos en esta tesis son alentadores, es importante reconocer ciertas limitaciones que abren camino a nuevas líneas de trabajo y mejora.

Una primera limitación radica en la naturaleza de los datos utilizados. El estudio se basó en un conjunto específico de campañas publicitarias de una única empresa de comercio electrónico chilena, lo que restringe la generalización de los resultados a otros contextos o industrias. Asimismo, algunas variables clave del negocio, como métricas de rentabilidad o “*engagement*” (por ejemplo, *AdCost*, *ROI*, *CTR*), no estaban disponibles o no fueron incorporadas en esta primera versión del modelo, limitando el alcance interpretativo de los resultados desde una perspectiva de negocio integral.

Otra limitación se relaciona con el uso de modelos *Prophet* individuales por campaña. Si bien esta estrategia permitió un ajuste rápido y eficiente, no se exploraron alternativas multivariadas o jerárquicas que pudieran capturar patrones comunes entre campañas o beneficiarse del aprendizaje conjunto. Tampoco se consideraron enfoques automatizados de ajuste de hiperparámetros, lo que podría mejorar el rendimiento del modelo y reducir el riesgo de sobreajuste en determinadas campañas.

En cuanto a la etapa de optimización presupuestaria, el modelo propuesto se formuló como un problema lineal con restricciones simples, centrado en maximizar las ventas netas previstas bajo un presupuesto total fijo. No se evaluaron escenarios más complejos, como presupuestos dinámicos, múltiples objetivos (por ejemplo, maximizar conversiones o minimizar el costo por adquisición), ni restricciones por canal, categoría de producto o calendario promocional.

Recomendaciones para trabajos futuros:

- Incorporar métricas de negocio más completas y sintéticas como *ROI*, *CPA* o *AdCost*, lo que permitiría evaluar no solo la magnitud de las ventas, sino también su eficiencia económica.
- Experimentar con modelos jerárquicos, multivariados o de series correlacionadas que permitan aprovechar similitudes entre campañas y mejorar la estabilidad de las predicciones.
- Implementar esquemas automatizados de validación cruzada y selección de hiperparámetros para mejorar la generalización del modelo y reducir el esfuerzo manual.
- Explorar algoritmos de optimización no lineales o heurísticos para asignación presupuestaria bajo restricciones más realistas y flexibles.
- Replicar el enfoque propuesto en otros sectores o geografías para validar su aplicabilidad y robustez en diferentes contextos.

- Plantear una estrategia adaptativa para la determinación de puntos de cambio potenciales puesto que la especificación nativa de *Prophet* asigna de manera determinística y uniforme una malla de puntos de cambio candidatos en el tramo inicial de la serie histórica delegando la selección final a una distribución a priori Laplace, y podría resultar potencialmente ineficiente ante discontinuidades no uniformes. Podría considerarse el uso de algoritmos de detección de quiebres estructurales basados en minimización de costos penalizados (como PELT²⁵ o GeDS²⁶) para alimentar exógenamente la posición de los *changepoints* en *Prophet*, garantizando que el modelo reaccione únicamente a quiebres estadísticamente significativos.

Estas líneas futuras no solo ampliarían la robustez técnica del enfoque, sino que también permitirían un mayor impacto práctico en la toma de decisiones del marketing digital en entornos reales de negocio.

²⁵ PELT (*Pruned Exact Linear Time*): Tiempo Lineal Exacto con Poda, es uno de los algoritmos más populares y eficientes en la estadística moderna para la detección de puntos de cambio (*changepoints*) en series temporales y conjuntos de datos secuenciales.

²⁶ GeDS (Generalized Spline Regression: Regresión mediante Splines Generalizados, metodología estadística desarrollada para ajustar curvas y tendencias de forma flexible e inteligente, sin necesidad de que el analista defina manualmente cuántos nudos o puntos de cambio usar ni dónde colocarlos.

6. Referencias bibliográficas

1. Baron Jaramillo, D. (2024). Modelo predictivo de ventas usando aprendizaje automático (Machine Learning) para pronosticar las ventas diarias de una empresa. Disponible en <https://repository.unad.edu.co/handle/10596/62872>
2. Bustos, F. (2021). Haciendo click al éxito: Predicción de Ventas y Optimización del Gasto en Marketing para Empresas Ecommerce. Repositorio Digital Universidad Torcuato Di Tella <https://repositorio.utdt.edu/handle/20.500.13098/12024>
3. Centro Economía Digital CCS: Sobre la Medición de la Economía Digital en Chile. Disponible en https://static.cepchile.cl/uploads/cepchile/2024/05/29-195436_mlao_pdr695_quijada.pdf
4. Chaffey, D., & Smith, P. (2017). Digital Marketing Excellence: Planning, optimizing and integrating online marketing. <https://www.taylorfrancis.com/books/mono/10.4324/9781315640341/digital-marketing-excellencedave-chaffey-pr-smith>
5. Chaffey, D., & Ellis-Chadwick, F. (2019). Digital Marketing: Strategy, Implementation and Practice (7th ed.). Pearson.
6. Christiansen Agar, M. (2020). Desarrollo de un modelo de predicción de tráfico de clientes para optimizar la dotación de personal en tiendas físicas del Retail. Disponible en <https://repositorio.uchile.cl/handle/2250/177936>
7. Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics.
8. Costa, C. (2024). Un enfoque automatizado para la predicción de ventas en comercio electrónico mediante la evaluación de campañas de marketing. Universidad Nacional de Rosario.
9. Kaushik, A. (2010). Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity. Wiley.
10. LocaliQ. (2024). Search Advertising Benchmarks for Your Industry. <https://localiq.com/blog/search-advertising-benchmarks/>
11. Manandhar, P., Rafiq, H., Rodriguez-Ubinas, E., Palpanas, T (2024). New Forecasting Metrics Evaluated in Prophet, Random Forest, and Long Short-Term Memory Models for Load Forecasting. Energies 2024, 17, 6131. <https://doi.org/10.3390/en17236131>
12. Rafferty, G. (2021). Forecasting Time Series Data with Facebook Prophet. Reino Unido. Packt Publishing Ltd.

13. Ramón Tadghighi, P. (2024). Innovación en la gestión de campañas comerciales: desarrollo de un modelo predictivo para la eficiencia comercial desde la predicción global hasta la precisión local en Inditex. Universitat Politècnica de València. <https://riunet.upv.es/handle/10251/207208>
14. Taylor, S., Letham, B. (2017). Forecasting at Scale. Disponible en https://facebook.github.io/prophet/static/prophet_paper_20170113.pdf
15. Yang, Y., & Zhai, P. (2022). Click-through rate Prediction in Online Advertising: A Literature review. Information Processing and Management, 59(2), 102853. <https://doi.org/10.1016/j.ipm.2021.102853>
16. Zunic, E. a. (2020). *Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data*. https://www.researchgate.net/publication/341071241_Application_of_Facebook's_Prophet_Algorithm_for_Successful_Sales_Forecasting_Based_on_Real-world_Data

7. Anexo

7.1. Medidas descriptivas por campaña

Tabla 7.1.1. Medidas descriptivas para campaña Automotriz

Automotriz									
Métricas	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	582	582	582	582	582	582	582	582	582
mean	7.895	5.28	369.669	4.215	5.409	99	184	256	313
std	12.223	6.031	344.132	2.883	3.696	131	264	618	570
min	87	16	494	42	44	3	6	0	1
0,25	3.671	2.312	181.732	2.401	3.089	50	87	117	93
0,5	5.656	3.641	301.116	3.693	4.736	75	136	180	167
0,75	8.453	5.954	443.034	5.272	6.837	112	204	236	328
max	126.437	52.486	2.982.496	24.564	33.264	1.647	3.057	11.02	7.084

Fuente: elaboración propia.

Tabla 7.1.2. Medidas descriptivas para campaña Belleza y Bienestar

Belleza_Bienestar									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	59.341	35.572	3.152.858	36.417	50.914	1.055	1.967	1.898	1.788
std	73.983	23.719	2.254.238	20.588	29.365	1.154	2.42	2.867	3.33
min	2.579	746	44.903	2.881	3.127	57	81	202	10
0,25	24.976	16.016	1.423.625	16.44	22.861	426	778	1.069	709
0,5	48.29	33.882	2.834.934	38.248	53.856	916	1.622	1.393	1.088
0,75	65.102	48.462	4.219.742	50.71	71.283	1.229	2.215	2.084	1.969
max	981.317	203.857	22.517.597	146.42	238.77	14.772	32.941	55.6	58.838

Fuente: elaboración propia.

Tabla 7.1.3. Medidas descriptivas para campaña Climatización

Climatizacion									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	582	582	582	582	582	582	582	582	582
mean	19.362	10.201	707.542	7.881	10.169	189	341	875	908
std	27.671	7.553	508.479	4.655	6.304	252	469	1.473	1.678
min	754	245	37.667	681	720	7	13	12	5
0,25	8.816	5.784	439.024	4.726	6.033	79	153	501	409
0,5	14.249	8.582	607.256	7.149	9.032	144	262	595	650
0,75	22.912	12.502	887.257	9.802	12.431	224	393	956	956
max	337.363	64.31	5.640.520	44.112	67.104	3.202	5.521	27.208	29.194

Fuente: elaboración propia.

Tabla 7.1.4. Medidas descriptivas para campaña Decoración “Blanco”

Decoracion_Blanco									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	30.988	19.805	1.488.620	15.889	21.535	481	1.031	1.294	1.430
std	44.232	12.068	1.010.050	7.272	10.524	597	1.346	1.353	2.696
min	2.044	757	56.972	3.072	3.756	38	89	71	6
0,25	17.338	12.874	898.368	11.292	15.188	259	540	629	690
0,5	23.64	18.165	1.346.222	15.311	20.526	366	792	1.293	1.093
0,75	31.936	24.232	1.853.048	19.27	25.853	526	1.124	1.712	1.48
max	517.063	132.74	11.751.415	79.878	118.705	6.35	13.835	23.282	39.227

Fuente: elaboración propia.

Tabla 7.1.5. Medidas descriptivas para campaña Grandes Electrodomésticos “Cocina”

Grandes_Electrodomesticos_Cocina									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	11.641	5.54	386.73	4.238	5.396	90	168	460	587
std	14.859	4.303	250.037	2.277	2.922	106	223	660	884
min	516	238	17.961	573	753	8	8	52	7
0,25	6.708	3.125	234.29	2.769	3.524	54	91	294	297
0,5	9.455	4.35	332.586	3.622	4.605	69	119	389	434
0,75	12.561	6.58	487.04	4.928	6.273	95	183	586	667
max	170.167	34.331	2.297.249	17.676	25.176	1.121	2.461	11.553	11.563

Fuente: elaboración propia.

Tabla 7.1.6. Medidas descriptivas para campaña Grandes Electrodomésticos “Lavado”

Grandes_Electrodomesticos_Lavado									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	16.562	9.379	457.843	5.456	6.875	110	190	864	933
std	27.379	11.312	425.214	4.001	5.113	186	348	1.087	1.965
min	1.028	210	10.351	471	697	8	8	19	1
0,25	8.261	3.172	197.53	2.774	3.532	55	89	372	350
0,5	12.147	5.365	320.66	4.147	5.231	76	127	770	614
0,75	17.175	10.868	579.558	6.937	8.814	114	188	1.072	1.001
max	346.226	122.358	3.865.279	29.781	41.142	2.341	4.485	16.222	24.748

Fuente: elaboración propia.

Tabla 7.1.7. Medidas descriptivas para campaña Grandes Electrodomésticos “Refrigeración”

Grandes_Electrodomesticos_Refrigeracion									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	20.834	11.939	660.702	7.117	8.970	136	241	1.053	1.087
std	24.701	11.261	673.108	4.608	5.972	170	326	1.271	1.494
min	2.135	287	27.947	1.201	1.313	18	24	34	6
0,25	11.733	5.112	363.184	4.412	5.596	78	129	604	539
0,5	16.416	8.449	533.03	6.193	7.876	105	179	913	865
0,75	22.653	15.438	819.926	8.422	10.673	148	258	1.125	1.238
max	271.289	126.43	9.696.511	63.528	84.814	1.915	3.64	16.793	22.181

Fuente: elaboración propia.

Tabla 7.1.8. Medidas descriptivas para campaña Jardinería

Jardineria									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	10.434	6.363	533.364	5.367	6.930	136	310	398	384
std	14.886	4.661	348.945	3.397	4.417	161	370	511	463
min	393	19	1.293	182	196	4	8	28	0
0,25	4.101	3.236	272.055	3.146	4.126	58	121	170	158
0,5	6.305	4.891	463.311	4.266	5.411	90	198	382	265
0,75	11.323	8.036	682.622	6.804	8.682	148	372	382	444
max	188.691	26.466	2.233.466	19.042	26.803	1.871	4.873	7.616	5.134

Fuente: elaboración propia.

Tabla 7.1.9. Medidas descriptivas para campaña Mascotas

Mascotas									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	5.030	3.093	250.228	3.311	4.495	93	177	169	163
std	5.758	2.615	213.18	2.506	3.419	97	192	196	240
min	44	3	182	80	84	1	1	7	0
0,25	1.363	811	72.092	850	1.126	20	41	69	50
0,5	3.859	2.482	203.396	2.927	3.962	75	142	124	108
0,75	6.869	4.656	367.718	5.094	7.014	129	236	196	191
max	63.222	19.403	1.896.905	14.127	20.815	964	1.818	2.743	3.791

Fuente: elaboración propia.

Tabla 7.1.10. Medidas descriptivas para Menaje

Menaje									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	583	583	583	583	583	583	583	583	583
mean	14.630	8.942	778.5	8.616	11.655	265	577	515	541
std	16.7	6.362	549.823	5.573	7.629	253	570	589	777
min	334	23	3.903	733	972	15	23	46	1
0,25	6.74	3.975	376.782	4.151	5.574	111	231	308	242
0,5	10.575	6.985	621.111	6.966	9.283	193	434	399	359
0,75	17.386	13.664	1.079.866	12.902	17.383	350	764	599	627
max	198.959	46.899	4.700.001	35.098	52.651	2.63	5.791	9.413	11.536

Fuente: elaboración propia.

Tabla 7.1.11. Medidas descriptivas para campaña Pequeños Electromésticos

Pequeños_Electrodomesticos									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	582	582	582	582	582	582	582	582	582
mean	40.467	24.802	1.547.636	19.829	25.930	552	977	1.808	2.096
std	59.014	15.431	908.221	8.6	12.159	702	1.335	2.201	3.805
min	1.839	826	65.77	3.123	3.794	21	38	278	20
0,25	25.741	16.576	1.076.378	15.356	19.885	355	606	975	1.007
0,5	32.047	22.128	1.419.759	19.596	25.72	450	772	1.611	1.512
0,75	40.809	28.708	1.870.436	23.451	30.543	589	1.014	2.115	2.206
max	763.63	181.599	12.227.608	105.984	155.994	8.394	15.523	38.063	48.027

Fuente: elaboración propia.

Tabla 7.1.12. Medidas descriptivas para campaña Telefonía

Telefonia									
	Venta neta (en dólares)	Nro clicks hechos sobre el anuncio	Cantidad de veces que se muestra el anuncio	Nro visitantes	Nro Visitas	Compras	Total unidades vendidas	Presupuesto /inversión en Google Ads	Costo facturado en Google Ads
count	582	582	582	582	582	582	582	582	582
mean	50.764	28.44	2.056.371	22.219	28.885	468	780	2.257	2.241
std	66.888	15.582	1.188.938	9.273	12.709	506	930	3.352	3.871
min	7.169	914	82.818	2.603	2.802	65	79	107	30
0,25	30.735	19.253	1.352.008	17.06	21.728	306	485	1.305	1.179
0,5	40.92	26.79	1.845.852	21.994	28.754	391	626	1.942	1.681
0,75	52.257	34.302	2.465.262	26.492	34.263	491	800	2.165	2.357
max	868.503	135.874	13.848.318	84.813	129.413	5.803	11.237	68.612	69.71

Fuente: elaboración propia.

Tabla 7.1.13. CTR, ROI y AdCost por campaña

Campaña	CTR	ROI	AdCost
Automotriz	1,28	25,22	0,03
Belleza_Bienestar	1,40	33,19	0,03
Climatizacion	1,29	21,32	0,05
Decoracion_Blanco	1,36	21,67	0,04
Grandes_Electrodomesticos_Cocina	1,27	19,83	0,04
Grandes_Electrodomesticos_Lavado	1,26	17,75	0,05
Grandes_Electrodomesticos_Refrigeracion	1,26	19,17	0,05
Jardineria	1,29	27,17	0,04
Mascotas	1,36	30,86	0,03
Menaje	1,35	27,04	0,04
Pequeños_Electrodomesticos	1,31	19,31	0,04
Telefonia	1,30	22,65	0,04

Fuente: elaboración propia.

Tabla 7.1.14. CTR, ROI y AdCost por campaña para Días con Eventos vs. Día sin Eventos

event_status campaign	CRT		ROI		AdCost	
	Event Day	No Event Day	Event Day	No Event Day	Event Day	No Event Day
Automotriz	1.305	1.273	27.454	23.471	0.032	0.033
Belleza_Bienestar	1.408	1.393	30.254	36.190	0.030	0.034
Climatizacion	1.305	1.284	20.412	22.068	0.041	0.048
Decoracion_Blanco	1.381	1.344	21.275	22.021	0.028	0.053
Grandes_Electrodomesticos_Cocina	1.289	1.267	18.931	20.449	0.030	0.046
Grandes_Electrodomesticos_Lavado	1.291	1.247	16.618	18.740	0.041	0.061
Grandes_Electrodomesticos_Refrigeracion	1.280	1.251	18.865	19.401	0.046	0.054
Jardineria	1.292	1.290	27.597	26.798	0.034	0.042
Mascotas	1.350	1.362	27.099	34.615	0.033	0.034
Menaje	1.365	1.347	25.562	28.350	0.025	0.043
Pequeños_Electrodomesticos	1.329	1.299	18.584	19.905	0.031	0.055
Telefonia	1.320	1.292	22.064	23.103	0.039	0.049
Total	1.349	1.324	22.295	23.763	0.034	0.047

Fuente: elaboración propia.

Tabla 7.1.15. CTR, ROI y AdCost por campaña para Días con Cyber vs. Día sin Cyber

cyber_status campaign	CRT		ROI		AdCost	
	Cyber	No Cyber	Cyber	No Cyber	Cyber	No Cyber
Automotriz	1.426	1.276	25.917	25.047	0.029	0.033
Belleza_Bienestar	1.604	1.391	25.397	34.939	0.024	0.033
Climatizacion	1.456	1.283	19.315	21.783	0.035	0.047
Decoracion_Blanco	1.503	1.348	17.653	22.802	0.013	0.048
Grandes_Electrodomesticos_Cocina	1.437	1.267	17.444	20.332	0.018	0.043
Grandes_Electrodomesticos_Lavado	1.447	1.249	15.260	18.527	0.025	0.059
Grandes_Electrodomesticos_Refrigeracion	1.394	1.253	17.772	19.429	0.033	0.054
Jardineria	1.444	1.286	30.018	26.782	0.018	0.041
Mascotas	1.473	1.354	25.081	31.719	0.021	0.035
Menaje	1.501	1.348	22.516	27.896	0.014	0.038
Pequeños_Electrodomesticos	1.487	1.299	16.741	19.990	0.015	0.051
Telefonia	1.486	1.293	20.029	23.216	0.030	0.047
Total	1.500	1.324	19.549	23.916	0.023	0.045

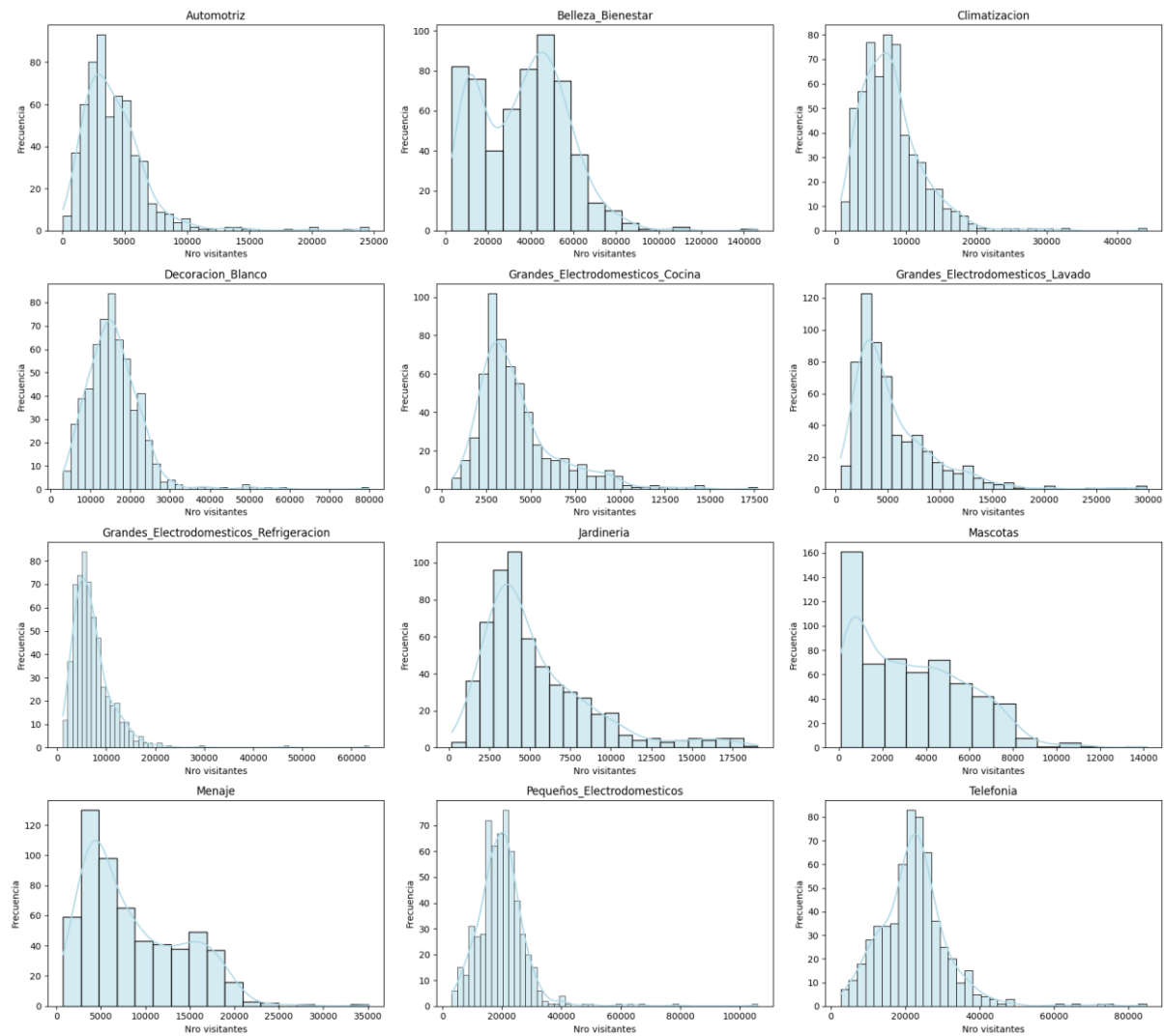
Fuente: elaboración propia.

Tabla 7.1.16. CTR, ROI y AdCost para cada campaña para Días de semana vs. Fines de semana

day_type campaign	CRT		ROI		AdCost	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Automotriz	1.295	1.253	26.616	21.107	0.030	0.042
Belleza_Bienestar	1.410	1.367	33.783	31.356	0.030	0.038
Climatizacion	1.299	1.267	22.398	18.086	0.043	0.055
Decoracion_Blanco	1.368	1.322	22.062	20.424	0.038	0.054
Grandes_Electrodomesticos_Cocina	1.284	1.246	20.629	17.367	0.036	0.052
Grandes_Electrodomesticos_Lavado	1.272	1.229	17.967	17.086	0.049	0.062
Grandes_Electrodomesticos_Refrigeracion	1.269	1.237	19.843	17.180	0.047	0.062
Jardineria	1.301	1.266	29.073	22.181	0.036	0.046
Mascotas	1.369	1.328	31.626	28.736	0.031	0.040
Menaje	1.364	1.324	27.685	25.213	0.033	0.043
Pequeños_Electrodomesticos	1.320	1.277	19.805	17.721	0.041	0.058
Telefonia	1.310	1.274	23.470	20.186	0.042	0.054
Total	1.343	1.302	23.757	21.054	0.038	0.051

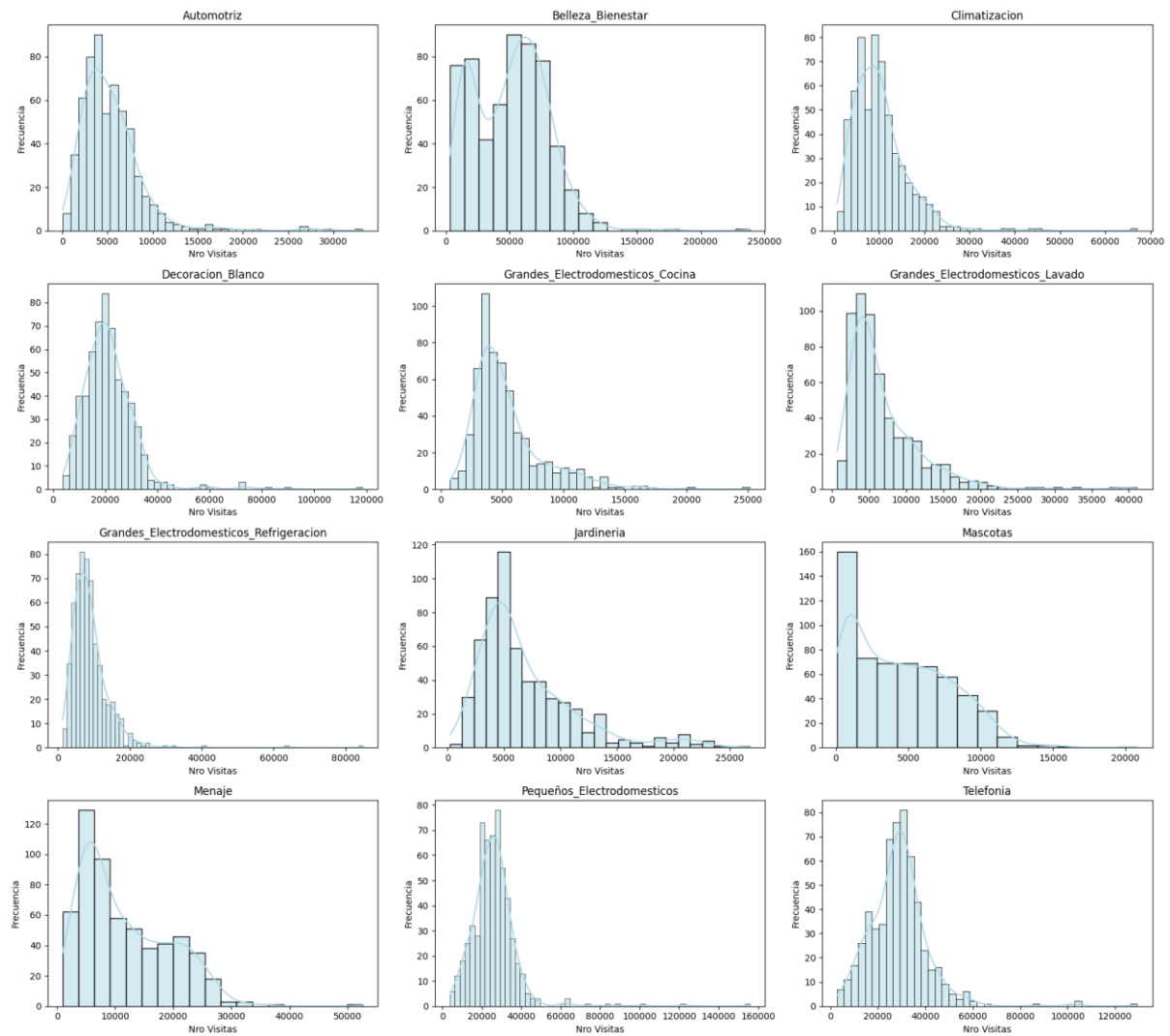
Fuente: elaboración propia.

Figura 7.1.1. Distribución del Número de visitantes por campaña



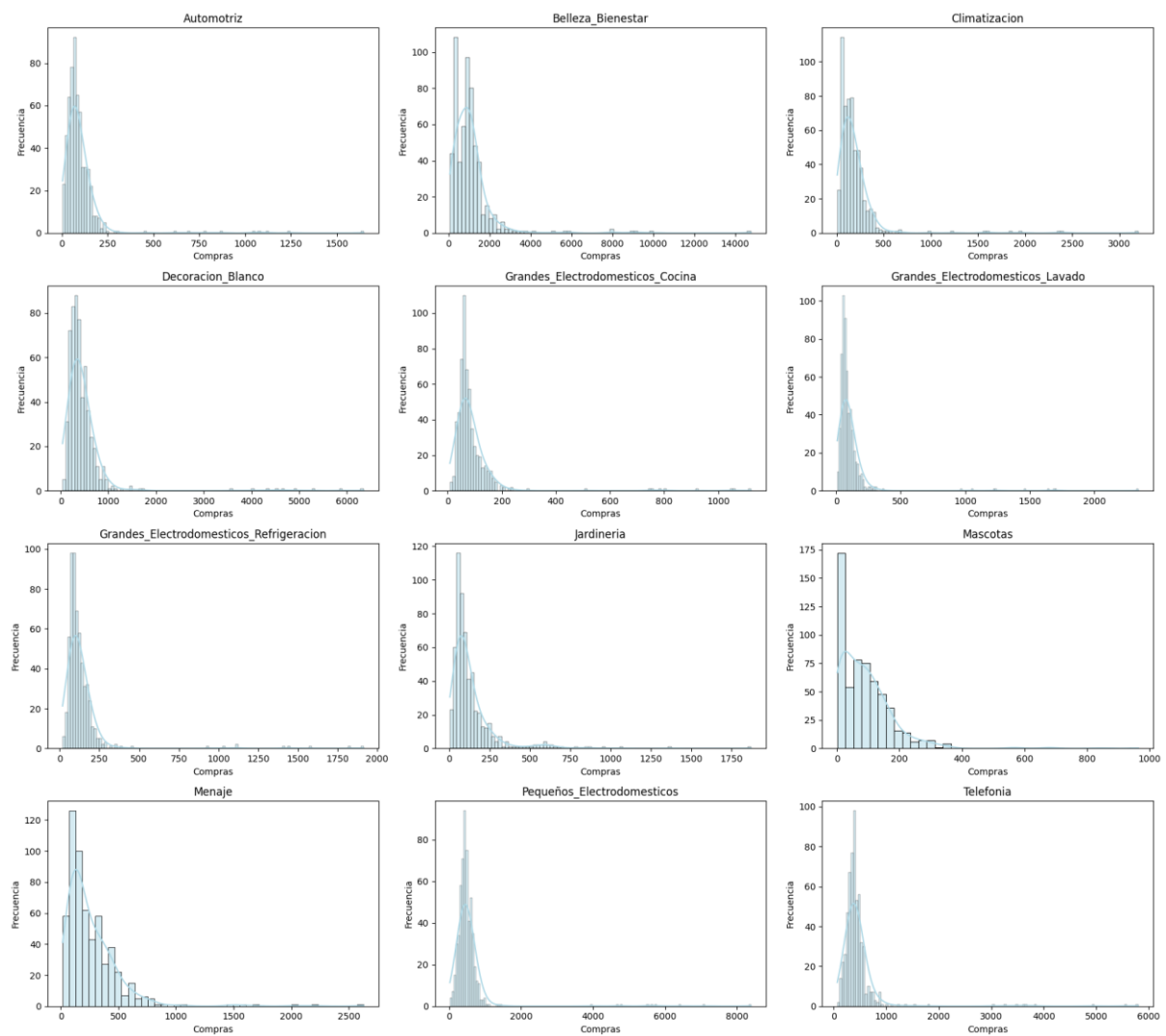
Fuente: elaboración propia.

Figura 7.1.2. Distribución del Número de visitas por campaña



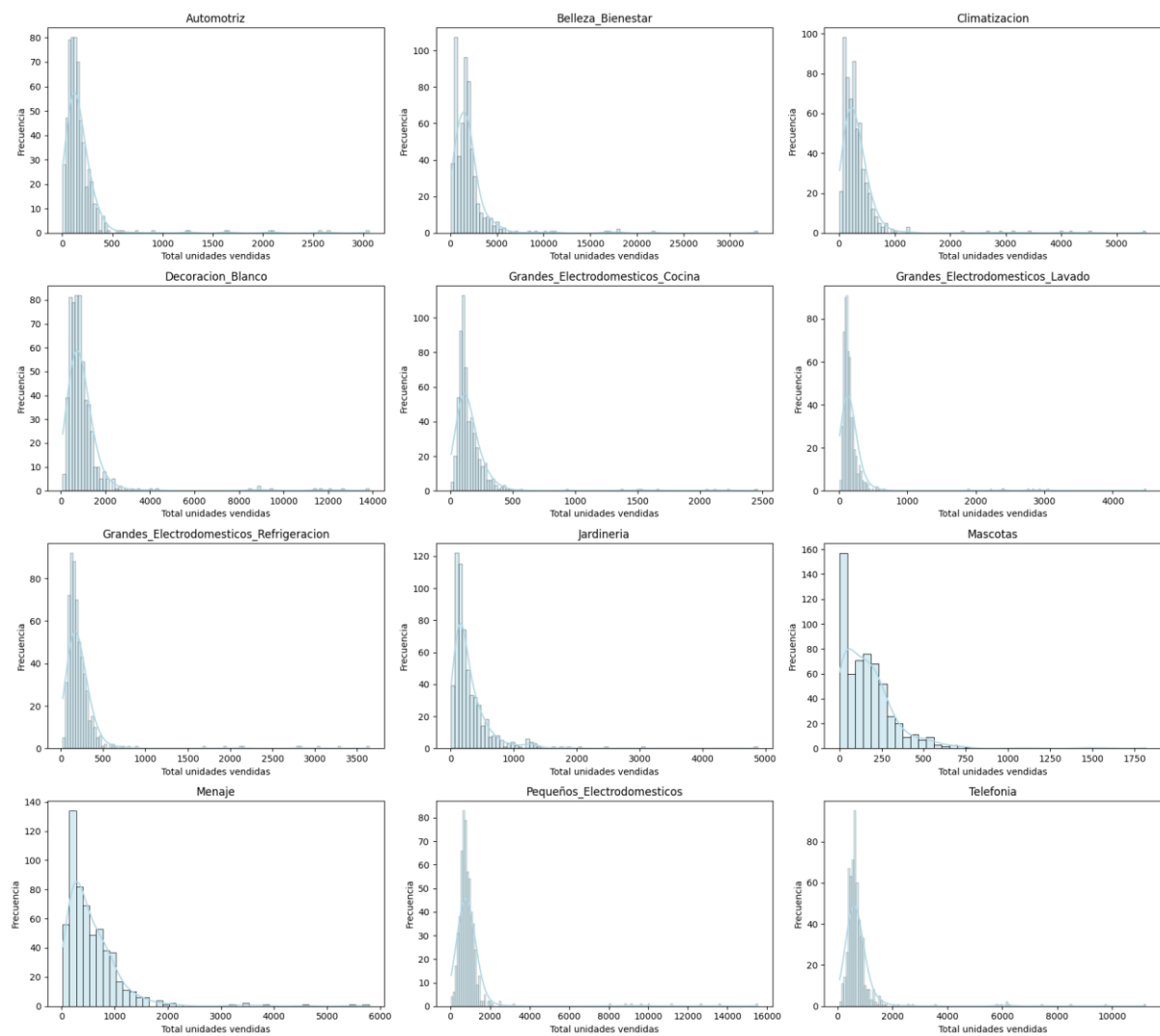
Fuente: elaboración propia.

Figura 7.1.3. Distribución de la cantidad de comprar (ordenes) por campaña



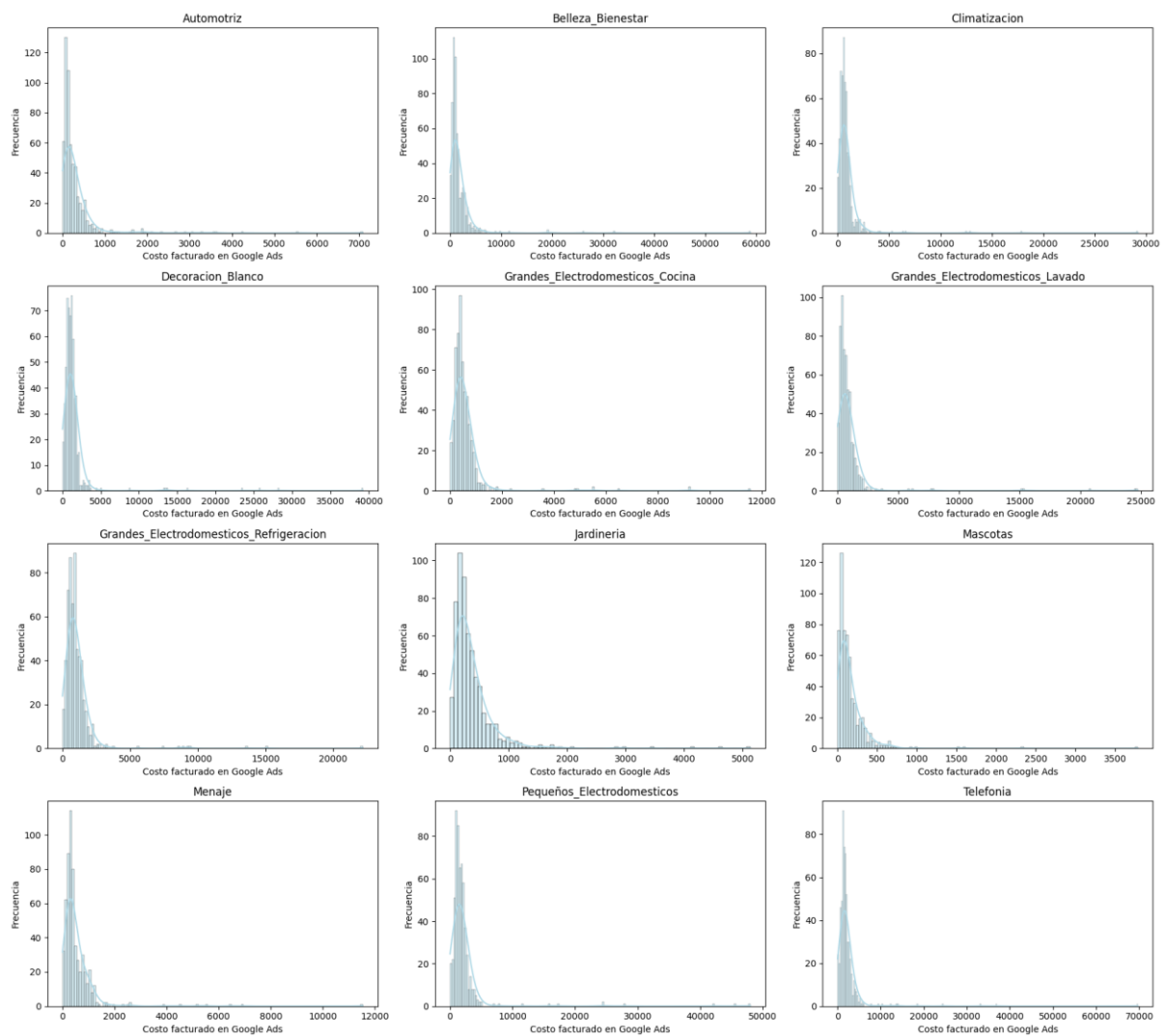
Fuente: elaboración propia.

Figura 7.1.3. Distribución del total de unidades vendidas por campaña



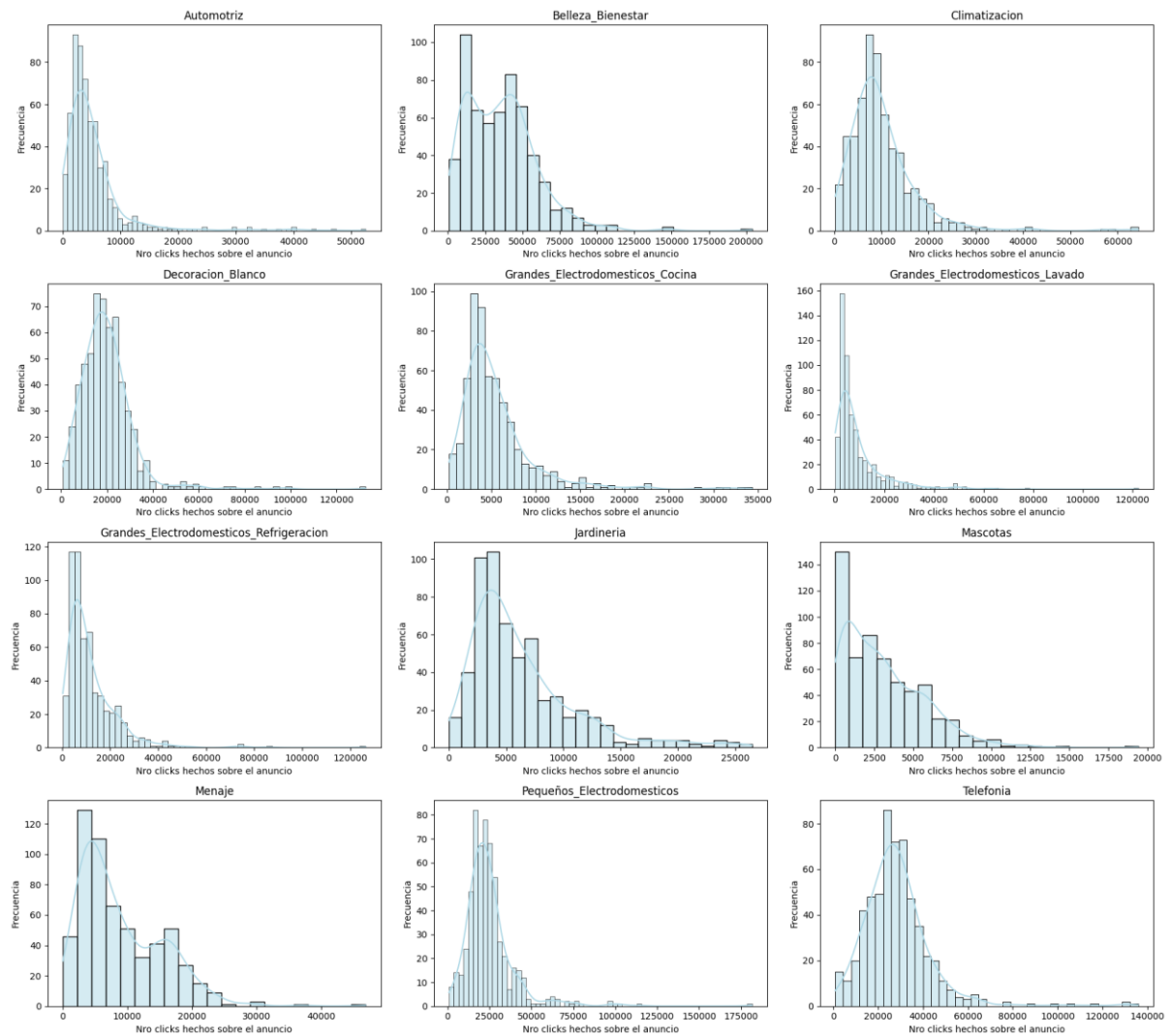
Fuente: elaboración propia.

Figura 7.1.4. Distribución del costo facturado en *Google Ads* por campaña



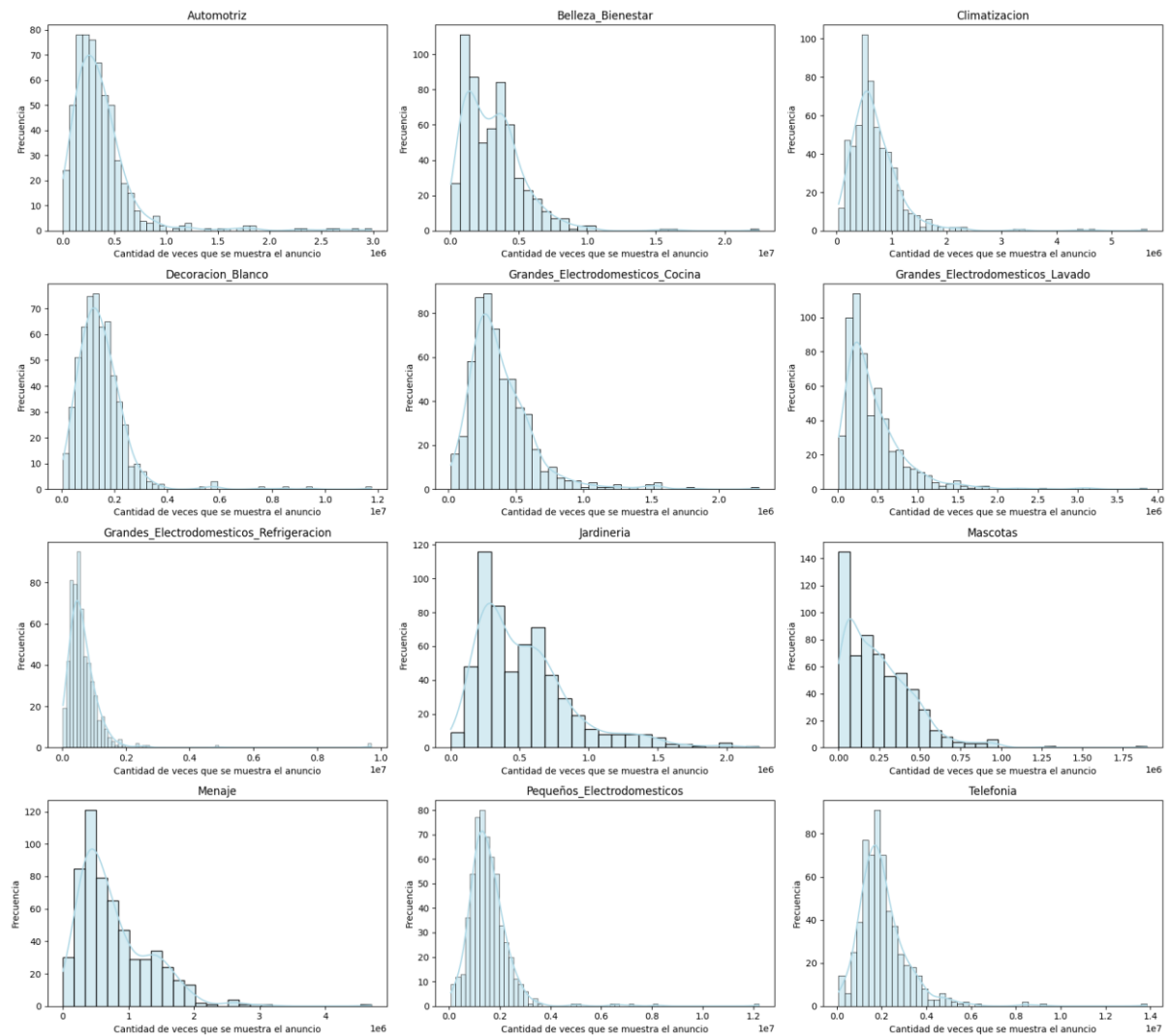
Fuente: elaboración propia.

Figura 7.1.5. Distribución del número de *clicks* por campaña



Fuente: elaboración propia.

Figura 7.1.6. Distribución del número de veces que se muestra el anuncio por campaña



Fuente: elaboración propia.

Tabla 7.1.16. Estadísticas descriptivas por campaña para las variables cuantitativas

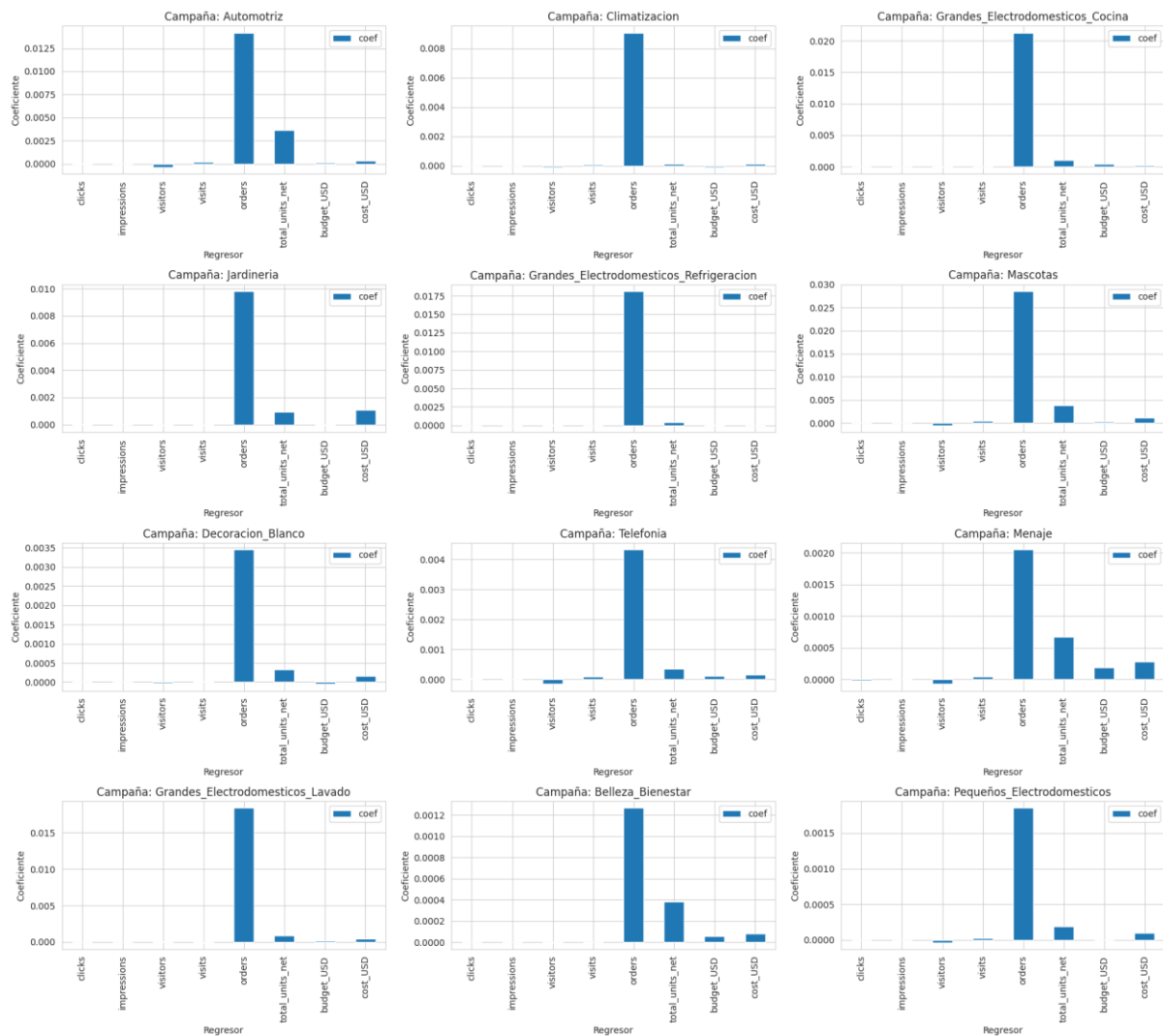
Variable	Estadística	Día de semana	Fin de semana	No Cyber	Cyber	No Evento	Evento
Venta neta (en dólares)	count	5.003	1.989	6.884	108	5.171	1.821
	mean	25.995	18.954	20.472	248.374	18.104	40.713
	std	48.618	20.129	20.786	198.186	16.211	76.505
	min	44	95	44	7.972	44	117
	25%	7.497	6.245	6.985	105.398	6.407	9.578
	50%	15.036	12.371	14.016	173.909	12.786	19.634
	75%	29.947	25.884	27.839	347.374	25.220	40.821
	max	981.317	254.041	313.674	981.317	153.209	981.317
Presupuesto/inversión en Google Ads	count	5.003	1.989	6.884	108	5.171	1.821
	mean	998	960	913	5.725	848	1.382
	std	1.801	1.708	1.156	9.840	1.169	2.830
	min	0	12	0	69	12	0
	25%	302	287	302	553	248	382
	50%	599	586	586	1.917	586	833
	75%	1.342	1.293	1.293	7.907	1.21	1.611
	max	68.612	38.063	38.063	68.612	38.063	68.612
Nro clicks hechos sobre el anuncio	count	5.003	1.989	6.884	108	5.171	1.821
	mean	14.378	13.440	13.609	46.132	12.822	17.773
	std	15.932	13.860	13.979	42.721	12.975	20.292
	min	3	27	3	1.779	3	27
	25%	3.906	4.069	3.904	15.040	3.742	5.157
	50%	8.506	8.387	8.297	29.456	7.708	10.762
	75%	20.126	18.175	19.062	62.772	18.276	22.565
	max	203.857	115.199	115.199	203.857	115.199	203.857
Cantidad de veces que se muestra el anuncio	count	5.003	1.989	6.884	108	5.171	1.821
	mean	1.045.556	999.402	988.733	3.817.482	917.666	1.358.305
	std	1.280.465	1.104.472	1.089.418	3.878.462	970.952	1.737.413
	min	182	2.459	182	163.746	182	2.459
	25%	303.494	307.634	301.959	1.272.953	287.846	404.842
	50%	606.944	611.625	598.342	2.407.366	563.355	766.864
	75%	1.349.218	1.297.644	1.301.640	5.255.476	1.229.164	1.634.879
	max	22.517.597	9.430.149	10.210.938	22.517.597	9.367.409	22.517.597
Nro visitantes	count	5.003	1.989	6.884	108	5.171	1.821
	mean	11.876	11.299	11.403	31.382	10.886	14.057
	std	12.822	11.564	11.833	27.759	11.030	15.656
	min	42	96	42	1.606	42	169
	25%	3.732	3.784	3.708	10.987	3.545	4.403
	50%	7.137	6.897	6.916	19.022	6.607	8.231
	75%	16.008	15.385	15.532	44.619	15.150	17.698
	max	146.420	80.728	91.049	146.420	83.493	146.420
Nro Visitas	count	5.003	1.989	6.884	108	5.171	1.821
	mean	15.947	14.711	15.102	47.058	14.408	18.968
	std	18.065	15.556	16.207	43.123	15.086	22.373
	min	44	122	44	2.382	44	241
	25%	4.843	4.745	4.784	16.635	4.574	5.652

Compras	50%	9.33	8.791	8.986	26.963	8.592	10.756
	75%	21.15	19.836	20.555	66.701	19.872	23.271
	max	238.770	104.150	122.455	238.770	108.002	238.770
	count	5.003	1.989	6.884	108	5.171	1.821
	mean	330	247	268	2.748	238	498
	std	617	309	326	2.59	264	953
	min	1	2	1	136	1	2
	25%	78	68	74	987	70	96
	50%	156	133	146	1.598	134	215
	75%	384	334	356	4.088	335	538
	max	14.772	3.837	4.252	14.772	2.347	14.772
	count	5.003	1.989	6.884	108	5.171	1.821
	mean	628	454	501	5.549	439	974
	std	1.245	593	623	5.39	488	1.943
	min	1	5	1	267	1	10
	25%	142	120	133	2.062	126	182
	50%	298	243	276	3.051	248	410
	75%	714	589	650	8.596	608	988
	max	32.941	8.488	9.093	32.941	4.832	32.941
	count	5.003	1.989	6.884	108	5.171	1.821
Costo facturado en Google Ads	mean	1.094	900	856	12.705	762	1.826
	std	2.662	962	861	12.714	735	4.259
	min	0	1	0	235	0	1
	25%	262	278	262	4.146	231	376
	50%	613	608	600	8.743	546	881
	75%	1.212	1.194	1.171	16.571	1.061	1.742
	max	69.710	10.417	10.417	69.710	7.96	69.710

Fuente: elaboración propia.

7.2. Medidas de desempeño predictivo del modelo por campaña

Figura 7.2.1.: coeficientes de cada variable regresora del modelo *Prophet* ajustado.



Fuente: elaboración propia.

Tabla 7.2.1.: comparación de valores de *MAE* del modelo *Prophet* con modelos propuestos en el trabajo final de Costa, C. (2024)

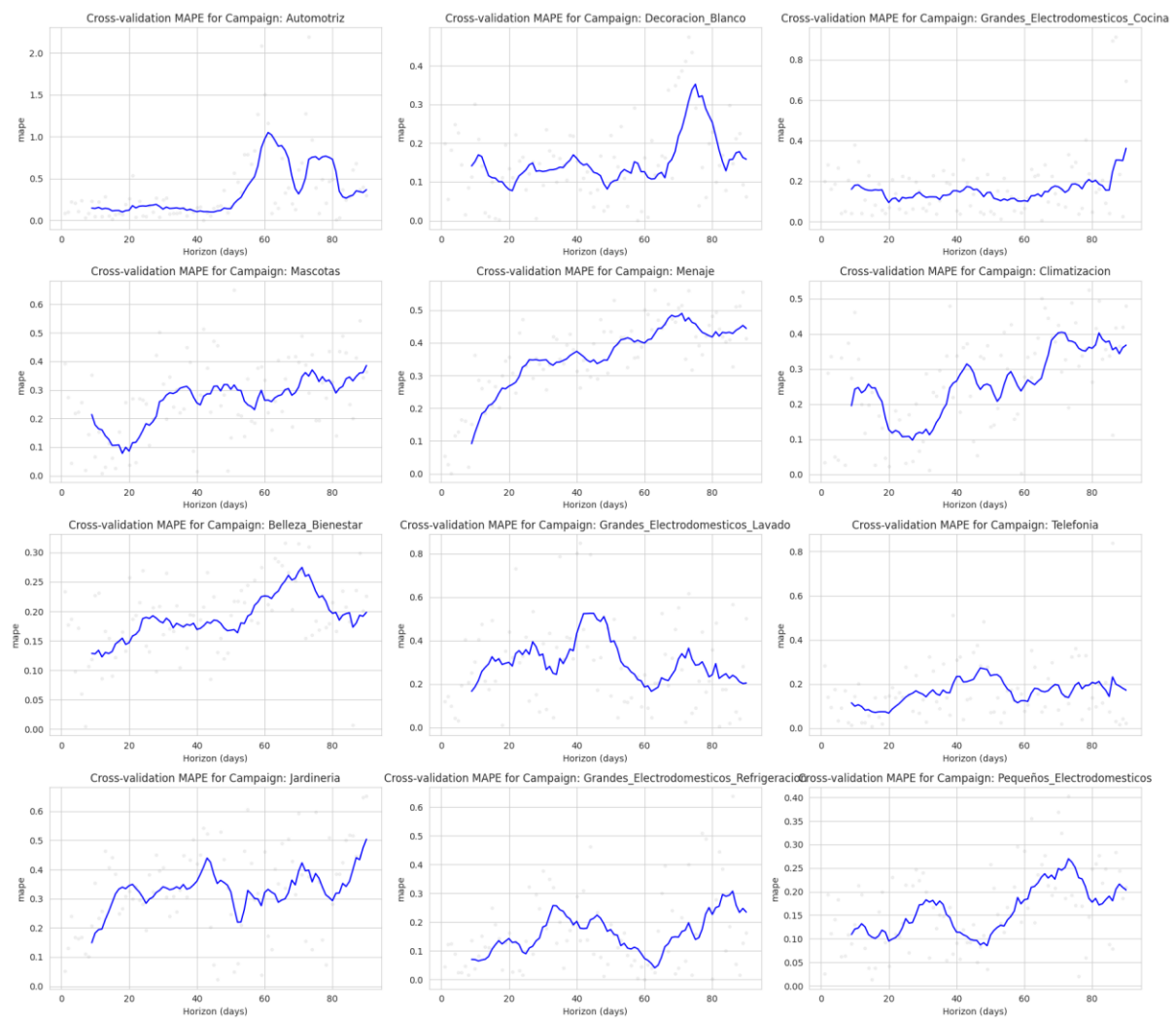
	Modelo	Mejor Modelo (2024)	
Campaña	Prophet	XGBoost	VARMAX
Automotriz	1.571,41	2.875,11	
Belleza_Bienestar	5.206,34	7.062,91	
Climatizacion	2.314,10	3.156,82	
Decoracion_Blanco	3.681,21	3.378,03	
Grandes_Electrodomesticos_Cocina	2.273,98		1.801,02
Grandes_Electrodomesticos_Lavado	2.863,85	2.084,79	
Grandes_Electrodomesticos_Refrigeracion	2.789,95	3.859,28	
Jardineria	2.589,64		1.520,35
Mascotas	1.238,20	1.065,43	
Menaje	2.834,44	2.503,58	
Pequeños_Electrodomesticos	4.605,02	4.846,19	
Telefonia	9.769,94		8.611,03

Tabla 7.2.2.: comparación de valores de *RMSE* del modelo *Prophet* con modelos propuestos en el trabajo final de Costa, C. (2024)

	Modelo	Mejor Modelo (2024)	
Campaña	Prophet	XGBoost	VARMAX
Automotriz	3.533,72	8.171,15	
Belleza_Bienestar	8.465,48	13.749,75	
Climatizacion	3.286,81	4.415,07	
Decoracion_Blanco	5.858,61	5.033,63	
Grandes_Electrodomesticos_Cocina	4.656,87		2.267,13
Grandes_Electrodomesticos_Lavado	8.652,89	2.906,16	
Grandes_Electrodomesticos_Refrigeracion	5.717,91	5.738,26	
Jardineria	4.288,11		2.216,40
Mascotas	1.830,16	1.366,89	
Menaje	4.259,45	3.123,98	
Pequeños_Electrodomesticos	9.002,51	9.262,02	
Telefonia	19.537,21		14.597,10

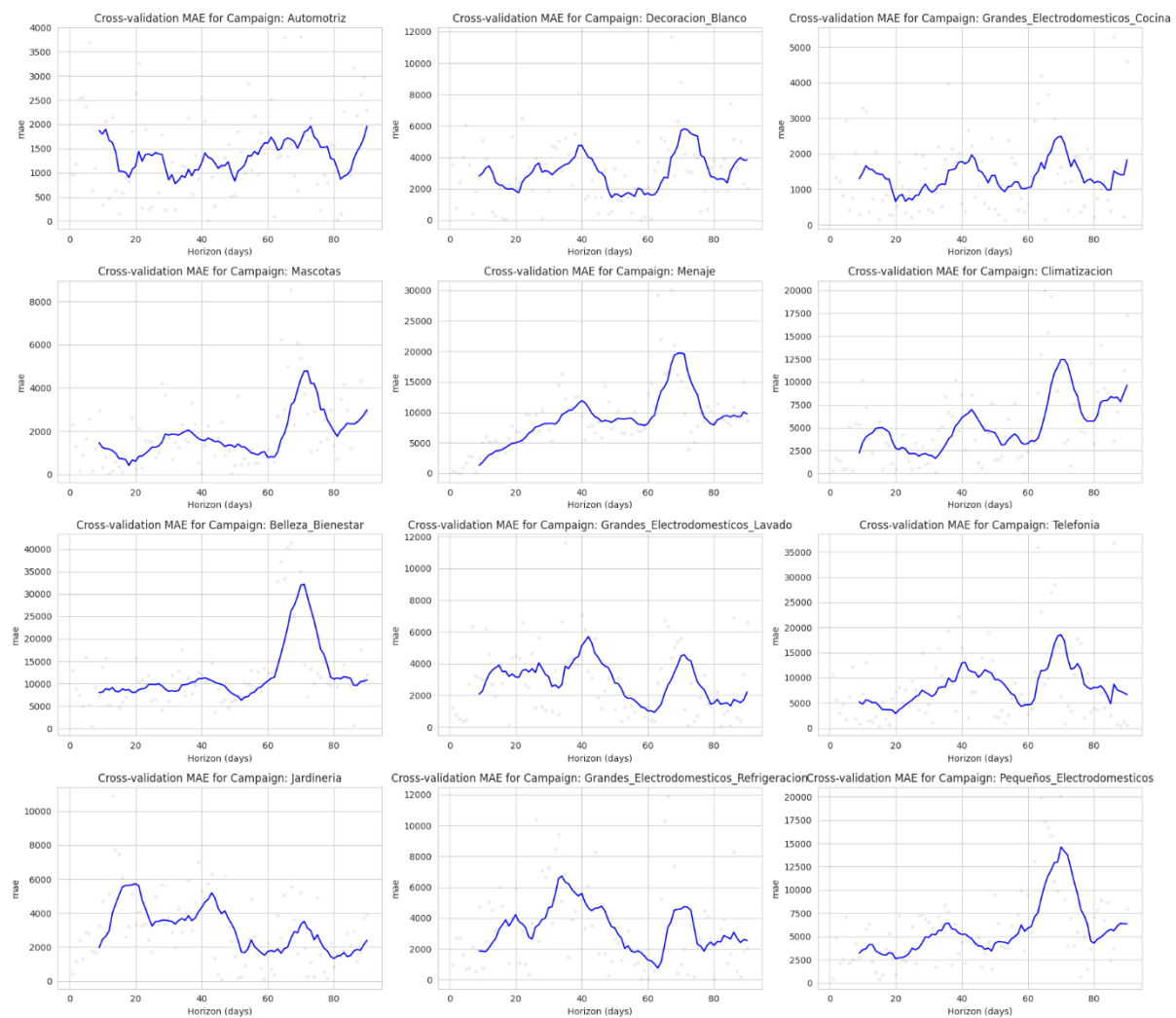
Fuente: elaboración propia.

Figura 7.2.1.: Validación cruzada para *MAPE* por campaña.



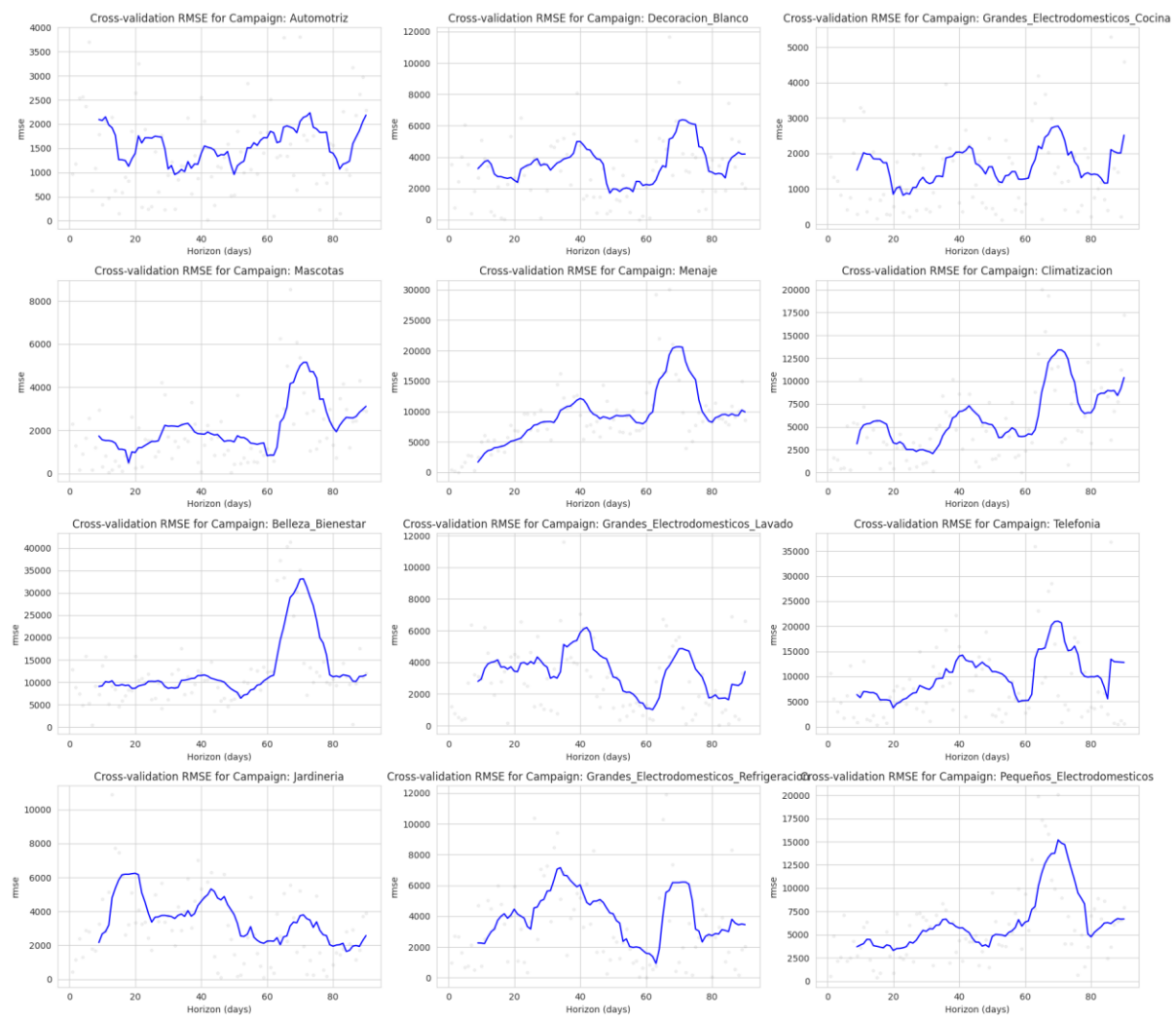
Fuente: elaboración propia.

Figura 7.2.2.: Validación cruzada para MAE por campaña.



Fuente: elaboración propia.

Figura 7.2.3.: Validación cruzada para *RMSE* por campaña.



Fuente: elaboración propia.

7.3. Resultados optimización de presupuesto

Figura 7.3.1. Prueba de Wilcoxon

campaign	yhat	budget_USD_real	budget_USD_optimizado	delta_budget	adcost_real	adcost_optimizado
Automotriz	1.09609e+06	30904.5	4635.67	-26268.8	0.0282	0.0042
Belleza_Bienestar	7.55456e+06	199005	149254	-49751.4	0.0263	0.0198
Climatizacion	3.27407e+06	131956	98967.2	-32989.1	0.0403	0.0302
Decoracion_Blanco	4.2859e+06	160122	120091	-40030.5	0.0374	0.028
Grandes_Electrodomesticos_Cocina	1.39785e+06	79866.8	59900.1	-19966.7	0.0571	0.0429
Grandes_Electrodomesticos_Lavado	1.97978e+06	95460	71595	-23865	0.0482	0.0362
Grandes_Electrodomesticos_Refrigeracion	2.05672e+06	109138	81853.6	-27284.5	0.0531	0.0398
Jardineria	507754	34158.6	5123.79	-29034.8	0.0673	0.0101
Mascotas	932402	25267.5	3790.13	-21477.4	0.0271	0.0041
Menaje	2.24706e+06	79349.4	59512.1	-19837.3	0.0353	0.0265
Pequeños_Electrodomesticos	5.5543e+06	261350	196013	-65337.5	0.0471	0.0353
Telefonia	6.75136e+06	234698	176023	-58674.5	0.0348	0.0261

Realizando el Test de Rango con Signo de Wilcoxon para AdCost (Original vs Optimizado):
Estadístico de Wilcoxon: 0.0
Valor P: 0.00048828125

Fuente: elaboración propia.

7.4. Código Python utilizado

