

Tipo de documento: Tesis de maestría

Master in Management & Analytics

Haciendo click al éxito: Predicción de Ventas y Optimización del Gasto en Marketing para Empresas Ecommerce

Autoría: Bustos, Florencia

Fecha de defensa de la tesis: 2023

¿Cómo citar este trabajo?

Bustos, F. (2021) "Haciendo click al éxito: Predicción de Ventas y Optimización del Gasto en Marketing para Empresas Ecommerce". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella
<https://repositorio.utdt.edu/handle/20.500.13098/12024>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 2.5 Argentina (CC BY-NC-SA 2.5 AR)
Dirección: <https://repositorio.utdt.edu>



Haciendo click al éxito: Predicción de Ventas y Optimización del Gasto en Marketing para Empresas Ecommerce

Alumna: Florencia Bustos

Tutor: Pablo Roccatagliata

Abstract

En esta tesis se aborda el desafío de mejorar la eficiencia y la toma de decisiones en el ámbito de la publicidad digital, específicamente en el contexto de los anuncios en Google. El objetivo principal es desarrollar un algoritmo que utilice los datos proporcionados por los anuncios de Google entre los años 2020 y 2022 para predecir el volumen de conversiones, es decir, las ventas que se pueden lograr al ajustar los gastos en las diferentes categorías de una empresa ecommerce. Esta predicción permitirá pronosticar y planificar los gastos semanales, buscando una distribución óptima de la inversión a lo largo de los días de la semana en cada categoría.

Para lograr este objetivo, se propone un enfoque basado en dos etapas principales. En primer lugar, se utiliza un modelo estadístico conocido como "Modelo SUR" para estimar las ventas. A partir de esta estimación de ventas, se avanza a la segunda etapa, que consiste en desarrollar un modelo de asignación de gastos publicitarios que determine cuánto invertir en cada categoría por día para una semana dada, teniendo en cuenta un límite de presupuesto establecido. El modelo de asignación se basará en técnicas de optimización, específicamente en programación lineal entera, con el objetivo de encontrar la distribución óptima de los gastos publicitarios en cada día de la semana, maximizando la eficiencia de la empresa.

Los resultados obtenidos en este análisis, demuestran que el modelo propuesto para la asignación de gastos publicitarios ofrece una forma potencialmente más eficiente de distribuir el presupuesto, en comparación con el enfoque actualmente utilizado por la empresa analizada. Esto representa una mejora potencial del 23% en la eficiencia, lo cual afecta positivamente la toma de decisiones brindando a la empresa la capacidad de obtener mejores resultados y maximizar el retorno de la inversión en sus campañas publicitarias.



Clicking to Success: Sales Forecasting and Marketing Spend Optimization for Ecommerce Companies

Student: Florencia Bustos

Thesis Advisor: Pablo Roccatagliata

Abstract

This thesis addresses the challenge of improving efficiency and decision making in digital advertising, specifically in the context of Google Ads. The main objective is to develop an algorithm that uses the data provided by Google ads between the years 2020 and 2022 to predict the volume of conversions, i.e. the sales that can be achieved by adjusting the expenses in the different categories of an ecommerce company. This prediction will allow forecasting and planning weekly expenses, looking for an optimal distribution of the investment along the days of the week in each category.

To achieve this objective, an approach based on two main steps is proposed. First, a statistical model known as the "SUR Model" is used to estimate sales. From this sales estimate, we proceed to the second stage, which consists of developing an advertising expenditure allocation model that determines how much to invest in each category each day for a given week, taking into account a set budget limit. The allocation model will be based on optimization techniques, specifically integer linear programming, with the objective of finding the optimal distribution of advertising expenditures on each day of the week, maximizing the company's efficiency.

The results obtained in this analysis will demonstrate that the proposed model for the allocation of advertising expenses offers a potentially more efficient way of distributing the budget, compared to the approach currently used by the analyzed company. This represents an improvement in strategic decision making in digital advertising with a 23% potential improvement in efficiency and, providing the company with the ability to obtain better results and maximize the return on investment in their advertising campaigns.

Índice

Abstract	1
Abstract	2
1.Introducción del Problema	5
1.1. Contexto	5
1.1.2 Presupuestos Específicos por Categoría	5
1.1.3 Presupuestos Específicos por Día de la Semana	6
1.2. Metodología actual y Problema	6
1.3. Objetivo	7
1.4. Output Esperado del Proyecto	8
1.4.1 Relación Conversiones Predichas y Asignación de Gasto del Modelo de Optimización	9
2.Marco Teórico	9
2.1 Optimización en base a Google Ads	9
2.2 Predicción de Ventas	10
2.3 Sur	11
2.3.1 Ejemplos de aplicación en SUR	12
2.3.2 SUR en Nuestro Contexto	13
2.4. Problema de Asignación	13
2.4.1 Solver ZIMPL	14
2.5 Medida de Eficiencia	15
2.6 Antecedentes	16
3. Datos	17
3.1. Diccionario de datos	17
3.2 Descripción de los datos	18
3.3 Análisis y Visualizaciones de Datos Crudos (con outliers)	20
3.3.1 Análisis de outliers	20
3.3.1.a Exclusión Fechas Especiales de Industria	22
3.3.2 Análisis Descriptivo del Periodo	23
3.3.2.a Análisis de la distribución de datos	23
3.3.2.b Remover Outliers de Week_Num	28
3.3.3 Análisis Descriptivo de la Base de Datos	28
3.3.3.a Importancia de las Categorías para la Empresa	28
3.3.3.b Análisis de Distribución y Comportamiento de los Días de la Semana	31
3.3.3.c Análisis de la Brecha del Tipo de Cambio y las Ventas	34
3.4 Preprocesamiento de las Variables	35
3.4.1 Ajuste por Inflación de las Variables Nominales	36
3.4.1.a Metodología utilizada	36
3.4.2 Estacionariedad de las Series de Tiempo	37
3.4.2.a Diferenciación Para Variable no Estacionarias	39
3.4.3 Multicolinealidad entre Variables	40
3.4.3.a Resultados VIF	40
3.4.3.a Abordaje Multicolinealidad	41

4. Predicción de Ventas y Optimización de Presupuesto	42
4.1. SUR	42
4.1.2 Variables del Modelo de Predicción	42
4.1.2.a Creación de Variables Dummy utilizando One-Hot Encoding	44
4.1.3 Entrenamiento del Modelo	44
4.1.3.a Codificación de referencia completa para la interpretación directa de variables	46
4.1.3.b Ajuste del modelo	47
4.1.4 Predicción	50
4.1.4.a RECM	51
4.2 Descripción Modelo de Asignación	52
4.2.1 Función Objetivo	52
4.2.2 Restricciones	53
4.2.3 Modelo de Optimización Final	58
4.2.3.a Consideraciones del Modelo de Optimización de Presupuesto	59
4.2.3.b Aplicación en ZIMPL	60
4.2.3.d Determinación Solución	60
4.3 Incorporación de predicciones del modelo SUR en el método ZIMPL	61
4.3.1 Datos del Modelo Sur	61
4.3.1.a Obtención de las predicciones del modelo SUR	61
4.3.2 Ventanas Temporales de las Variables	62
4.3.2.a Definición Ventana Temporal para las Variables propias de Google Ads	62
4.3.2.b Definición Periodo de Tiempo Anterior para la Brecha del tipo de cambio	64
4.3.3 Incorporación de las predicciones del modelo SUR en ZIMPL	65
4.4 Restricción Adicional a nuestro modelo de Asignación	66
4.4.1 Incorporación Nuevas restricciones	67
4.4.2 Validación Nuevas Restricciones en nuestro modelo	68
5. Validación de Resultados	69
5.1 Validación Cruzada con Semanas Excluidas	70
5.2 Evaluación del modelo de asignación de presupuesto mediante el Test de Giacomini y White	70
5.2.2 Comparación Gráfica de los Modelos	72
5.3 Benchmark	73
5.3.1 Impacto de la Predicción de Ventas	77
6. Conclusiones	79
7. Referencias	80
8. Anexo	82
8.1 Código	82
8.2 Data	82

1.Introducción del Problema

1.1. Contexto

Las empresas más importantes del mundo utilizan el servicio de Google Ads como herramienta de Marketing Digital. Este servicio es un programa accesible a cualquier tipo de empresa, pero muy amplio y complejo, tanto en el entendimiento como en su aplicación. Esto último, hace necesario que los usuarios de esta herramienta se vean obligados a indagar en las métricas particulares de cada una de las empresas, perfeccionando y ajustando el rendimiento del servicio, con el objetivo de sobresalir, y en paralelo ser eficientes al momento de publicitar su compañía. Las empresas más importantes del mundo usan Google Ads como herramienta de marketing.¹

En este proyecto nos centraremos en una de las e-commerce más importantes de Latino América que cuenta con una plataforma de comercio electrónico extremadamente amplia que abarca una gran variedad de productos y servicios en línea. Con el objetivo de facilitar la navegación y la búsqueda de productos específicos, el sitio de esta e-commerce se divide en categorías que agrupan los productos según su tipo o función. Estas categorías son una herramienta fundamental para los usuarios de la plataforma, ya que les permiten filtrar y ordenar los productos que están buscando de manera eficiente y rápida. Las categorías cubren una amplia variedad de productos, desde electrónica y hogar hasta moda y belleza, y son una forma efectiva de organizar y presentar la gran cantidad de productos que se ofrecen en el sitio. Si bien esta e-commerce opera en casi todo latinoamérica, para este análisis solo nos enfocaremos en los datos de Argentina.

1.1.2 Presupuestos Específicos por Categoría

Proporcionar categorías claras y bien organizadas mejora la experiencia de compra en línea, permitiendo a los usuarios encontrar fácilmente lo que están buscando. Al mismo tiempo, esto ayuda a los vendedores a publicar sus productos de manera efectiva en el sitio, al mostrarlos en la categoría correcta para alcanzar a su público objetivo de manera más eficiente. Para lograr esto, es necesario asignar presupuestos específicos de marketing a cada categoría. Al asignar presupuestos específicos para cada categoría, se evitan gastos excesivos en áreas que no generan un retorno de inversión significativo. De esta manera, se puede concentrar el presupuesto en las categorías que generan mayores ingresos, maximizando así el retorno de inversión.

Cada categoría tiene su propia demanda y pertenece a un mercado diferente, por lo que es importante analizar la manera más eficiente de asignarles un presupuesto individual. De esta forma, podremos enfocarnos en los objetivos particulares de cada categoría y medir los resultados de manera precisa, facilitando así la evaluación del desempeño de cada una y permitiendo realizar ajustes y cambios en la estrategia de marketing para maximizar el éxito.

Con esta idea en mente, se busca crear un prototipo de asignación de presupuesto para estas categorías en particular cada semana.

¹ Marketing Metrics: The definitive guide to measuring marketing performance, 2010

1.1.3 Presupuestos Específicos por Día de la Semana

Los consumidores suelen tener comportamientos de compra distintos a lo largo de la semana. Los días laborables, como de lunes a viernes, no suelen tener igual comportamiento que los fines de semana, la intención de compra varía según qué se quiera comprar y qué día de la semana. Además, los patrones de compra también pueden estar influenciados por factores estacionales o por eventos especiales, como días festivos o promociones especiales. Es importante tener en cuenta estas variables al asignar un presupuesto de marketing en función del día de la semana para maximizar la eficacia de la inversión y llegar al público objetivo en el momento adecuado.

La correcta asignación de un presupuesto de marketing en base al día de la semana puede ser crucial para optimizar la efectividad de una campaña publicitaria. Esto se debe a que el comportamiento del consumidor varía según el día de la semana, lo que puede afectar el rendimiento de los anuncios y, por lo tanto, el retorno de inversión (ROI) de la campaña.² En general, la asignación de un presupuesto de marketing en base al día de la semana puede ayudar a las empresas a optimizar su estrategia publicitaria y maximizar el ROI, al adaptar su presupuesto a los patrones de comportamiento de los consumidores. Sin embargo, este enfoque también puede plantear desafíos en términos de planificación y gestión del presupuesto, y requerir un análisis cuidadoso de los datos de comportamiento del consumidor y de las tendencias del mercado para determinar la mejor asignación de inversión.³

Teniendo todo esto en cuenta, se hace necesario asignar un presupuesto de marketing de manera estratégica, considerando tanto la categoría de productos como el día de la semana. Al hacerlo, se puede optimizar el gasto y maximizar el retorno de inversión. Por ejemplo, los datos pueden mostrar que los consumidores compran más productos electrónicos los fines de semana, mientras que los artículos de moda se venden mejor los días laborables de la semana. Asignar un presupuesto más alto a la categoría de productos que tiene un mayor volumen de ventas durante la semana, junto con un aumento en la inversión en días específicos, puede ayudar a impulsar el éxito de la estrategia de marketing y aumentar las ganancias de la e-commerce.

1.2. Metodología actual y Problema

Haciendo foco en nuestro caso puntual, hoy en día la optimización del presupuesto para cada día de la semana y para cada categoría se hace de forma manual. Se ve una oportunidad en la mejora de esta asignación incorporando un modelo que tenga en cuenta data pasada para poder predecir de manera eficiente las conversiones e incorporarlo a la asignación de presupuesto semanal. Actualmente, las decisiones de inversión no se hacen en forma granular ni se tiene en mente la conversión que puede llegar a tener cada categoría específica cada día de la semana.

Asignar un presupuesto de marketing de forma manual y sin tener en cuenta datos pasados o modelos de predicción puede llevar a riesgos en términos de resultados de marketing. En primer

² Dave Bussière (2011)

³ Marketing Metrics: The definitive guide to measuring marketing performance, 2010

lugar, sin un análisis exhaustivo de los datos de comportamiento de los consumidores, puede haber una asignación desequilibrada del presupuesto, lo que significa que se pueden asignar demasiados recursos a ciertas categorías o días de la semana que no necesariamente ofrecen un retorno de inversión importante. Además, la falta de una función objetivo clara en la asignación de presupuesto puede resultar en una falta de enfoque y objetivos poco claros, lo que puede derivar a una estrategia de marketing ineficaz. En última instancia, la asignación de presupuesto de marketing debe basarse en datos y modelos de predicción robustos para garantizar la optimización del retorno de inversión y una estrategia de marketing enfocada.

Es por eso que, con el fin de mejorar la eficiencia y la toma de decisiones, se busca generar un algoritmo con los datos brindados por los anuncios de Google con el objetivo de predecir las conversiones (ventas en pesos) que se pueden lograr al ajustar los gastos de las diferentes categorías. Logrando así, poder pronosticar y planear de la mejor manera los gastos a utilizar semanalmente.

1.3. Objetivo

El objetivo principal de este proyecto es desarrollar un modelo de predicción para estimar la cantidad de conversiones de cada categoría de productos para cada día de la semana en una de las e-commerce más importantes de Latino America para luego optimizar la asignación de presupuesto semanal. Se buscará generar un modelo que prediga las ventas diarias para cada categoría, para luego utilizarlas en un modelo de asignación de inversión en base a una función de utilidad. La finalidad de esta predicción es optimizar el presupuesto de marketing asignado, con el fin de mejorar la medida de eficiencia más importante para la empresa, que es el Adcost.

El Adcost es la inversión que se realiza en publicidad por cada venta generada. Es decir, por cada peso ganado en ventas, se calcula cuánto se ha invertido en publicidad. El objetivo es minimizar el valor del Adcost para reducir la cantidad de inversión necesaria para obtener una cantidad de conversiones constante. El AdCost se puede expresar matemáticamente como:

$$AdCost = \frac{Inversión}{Ventas}$$

Es importante optimizar el presupuesto de marketing en función del AdCost (inversión/ventas) porque es una medida clara de la eficiencia de una campaña de marketing. Si el AdCost es bajo, significa que se está invirtiendo una cantidad menor de dinero para obtener una mayor cantidad de ventas, lo que indica una campaña, en términos de marketing, eficiente. Por otro lado, si el AdCost es alto, significa que se está invirtiendo una gran cantidad de dinero para obtener pocas ventas, lo que indica que la campaña de marketing no está funcionando de manera efectiva y se podrían estar desperdiciando recursos.⁴

⁴ Manuals y MicroStrategy (2013)

Al optimizar el presupuesto de marketing en función del AdCost, se puede mejorar la eficiencia de la campaña y maximizar el retorno de inversión. Esto se logra al enfocarse en las categorías de productos y los días de la semana que tienen una mayor probabilidad de generar conversiones y minimizar el gasto en aquellos que no lo hacen. De esta manera, se logra una mejor asignación del presupuesto de marketing, lo que se traduce en una campaña de marketing más rentable y eficiente en términos de inversión.

1.4. Output Esperado del Proyecto

En principio, se busca presentar una metodología más exhaustiva en comparación con la práctica actualmente empleada. El output esperado de la predicción de las ventas y asignación de presupuesto en base al día de la semana y categoría de producto es obtener una estrategia de marketing más eficiente en términos de Adcost, que permita maximizar las ventas con la menor inversión posible. Para ello, se utilizarán técnicas de modelado predictivo y aprendizaje automático para predecir el valor de las conversiones de cada categoría para cada día de la semana y, a partir de estas predicciones, asignar el presupuesto de marketing de manera óptima.

El propósito de este estudio es contrastar la eficiencia del modelo propuesto con la eficiencia derivada de las decisiones tomadas en la realidad sin la intervención del modelo. Se utilizará el AdCost (inversión/ventas) como métrica para evaluar el impacto de la predicción de ventas en la eficiencia del presupuesto asignado. El objetivo final es obtener una estrategia de marketing más completa y rentable para la empresa.

Además de comparar el rendimiento del modelo de asignación de presupuesto con la realidad, también se analizará la performance predictiva de las ventas en base al día de la semana y categoría de producto. Esto nos permitirá evaluar cómo influye la exactitud de la predicción de ventas en la eficiencia del modelo de asignación. Si la predicción de ventas es altamente precisa, podemos esperar que el modelo de asignación de presupuesto también lo sea. En cambio, si la predicción de ventas no tiene una performance predictiva considerablemente buena, es probable que el modelo de asignación de presupuesto no sea tan confiable. Por lo tanto, la evaluación de la estimación de las futuras ventas será un factor crítico para determinar la eficacia general del modelo propuesto de asignación de presupuesto en base al día de la semana y categoría de producto.

1.4.1 Relación Conversiones Predichas y Asignación de Gasto del Modelo de Optimización

En el modelo de optimización, se utilizan las conversiones predichas como una guía para asignar el gasto en publicidad de manera eficiente. Sin embargo, esto no significa que las conversiones estén fijas de antemano y que el gasto en publicidad no tenga impacto en la cantidad de ventas. El objetivo del modelo de optimización es encontrar la asignación de gasto óptima que maximice las conversiones esperadas. Esto implica ajustar la asignación de gasto en función de la predicción de conversiones, pero también considerando otras variables y restricciones del problema, como los presupuestos disponibles y las metas comerciales. Si el optimizador asigna un gasto más bajo en publicidad, esto no implica necesariamente una reducción en las conversiones predichas. El modelo

de asignación busca encontrar un equilibrio entre el gasto y las conversiones, considerando la eficiencia y la rentabilidad de las inversiones en publicidad.

En la sección de resultados podremos ver una comparación con los resultados y eficiencia reales, en este caso, el modelo asume que los gastos son similares a los registrados en las últimas 3 semanas para basar las predicciones de conversiones. Sin embargo, esto no implica que cualquier reducción en el gasto planteadas por el optimizador se traduzca automáticamente en una reducción proporcional en las conversiones predichas. Es importante tener en cuenta que el modelo de optimización tiene en cuenta múltiples factores y busca maximizar las conversiones dadas las restricciones y las metas establecidas. La relación entre el gasto en publicidad y las conversiones puede ser compleja y depende de varios elementos, como la efectividad de las estrategias publicitarias, la competencia en el mercado y otros factores externos.

En resumen, el modelo de optimización busca asignar el gasto de manera eficiente para maximizar las conversiones predichas, pero esto no significa que las conversiones estén fijas de antemano ni que una reducción en el gasto siempre implique una reducción proporcional en las conversiones. El modelo considera múltiples variables y busca encontrar un equilibrio óptimo entre el gasto y las conversiones dadas las condiciones y restricciones del problema.

2. Marco Teórico

En esta sección, abordaremos la fundamentación teórica de las técnicas y algoritmos que serán implementados en la presente propuesta. Nos basaremos en la revisión de papers y bibliografía que respaldan y explican de manera concisa su aplicación práctica. Asimismo, examinaremos los problemas y desafíos en el campo del marketing que subyacen en nuestros modelos y análisis, proporcionando un contexto para la comprensión de la investigación.

2.1 Optimización en base a Google Ads

Cuando se habla de optimización en Google Ads nos referimos al proceso de mejorar el rendimiento de una campaña de Google Ads. Entre los pasos clave de la optimización de los anuncios de Google se incluyen la búsqueda y selección de palabras clave, la redacción y prueba de anuncios, la optimización de la segmentación y la supervisión y el análisis continuos para medir y mejorar el rendimiento.⁵ En este último punto es donde nos centraremos para hacer nuestros análisis.

Para predecir conversiones y optimizar el gasto en marketing en una empresa de ecommerce, se puede hacer foco en distintos puntos:

1. **Análisis de datos:** Recopilar datos sobre el comportamiento del cliente, incluyendo información demográfica, patrones de compra, canales de marketing más efectivos, etc.

⁵ Del Valle (2016)

2. **Modelos de predicción de conversiones**: Utilizar herramientas analíticas como machine learning para desarrollar modelos predictivos que proporcionen una estimación precisa de la tasa de conversión para diferentes grupos de clientes.
3. **Optimización de las estrategias de marketing**: Basado en las predicciones de conversión, optimizar estrategias de marketing para maximizar la eficacia y reducir el costo por conversión. Esto incluye la selección de los canales de marketing adecuados, la segmentación de audiencia, la optimización de anuncios y el ajuste de la estrategia de pujas.
4. **Seguimiento y análisis**: Monitorear continuamente las métricas clave para evaluar el éxito de las estrategias de marketing y realizar ajustes en consecuencia.

Al implementar estas prácticas, una empresa de e-commerce puede mejorar la eficiencia de sus gastos en marketing, aumentar las conversiones y obtener un retorno más alto de la inversión. En el análisis de este proyecto se buscará proponer un modelo que abarque todos estos puntos en un conjunto para poder asignar de la mejor forma posible el presupuesto en marketing para cada día de la semana y para cada campaña. Cada categoría, en nuestro caso puntual, representa una campaña distinta en Google Ads.

2.2 Predicción de Ventas

Para predecir las ventas de una campaña de marketing en Google Ads para un e-commerce, es necesario tener en cuenta una serie de factores importantes como el presupuesto asignado a la campaña, el tipo de producto que se está vendiendo en esa campaña, la época del año, las tendencias del mercado y el comportamiento de los consumidores. Además, también es necesario tener en cuenta las métricas y estadísticas que se pueden obtener de la propia plataforma de Google Ads, como el costo por clic (CPC), el número de impresiones y los clics en los anuncios de cada campaña.⁶ Predecir las ventas de una campaña de marketing en Google Ads para una e-commerce requiere una cuidadosa consideración de múltiples factores y el uso de herramientas y métodos adecuados. Al tomar en cuenta las métricas de Google Ads, las tendencias del mercado y el comportamiento de los consumidores, así como las herramientas de análisis y predicción, se pueden obtener predicciones de las ventas futuras y tomar decisiones informadas sobre las estrategias publicitarias para maximizar el retorno de inversión.

En el caso que abordaremos en este proyecto, resulta importante para la empresa predecir las ventas de las campañas de marketing en Google Ads porque permite anticiparse a la demanda y optimizar su presupuesto de publicidad con el objetivo de tomar decisiones estratégicas. Existen diferentes modelos de Machine Learning que pueden utilizarse para predecir ventas. En este caso, el método Seemingly Unrelated Regression (SUR) para predecir las ventas de las campañas en Google Ads nos parece adecuado porque permite modelizar varias series de tiempo simultáneamente, lo que es útil en el caso de una empresa que tiene múltiples productos y servicios que pueden tener distintos patrones de ventas. Además, el método SUR permite tener en cuenta las correlaciones entre los

⁶ Marqués (2014)

errores de las diferentes series de tiempo, lo que ayuda a mejorar las predicciones. En la siguiente sección abordaremos más en profundidad este modelo.

2.3 Sur

La naturaleza básica del modelo de regresión múltiple es que describe el comportamiento de una determinada variable de estudio a partir de un conjunto de variables explicativas. Cuando el objetivo es explicar todo el sistema, puede haber más de una ecuación de regresión múltiple. Por ejemplo, en un conjunto de ecuaciones de regresión múltiple lineales individuales, cada ecuación puede explicar algún fenómeno económico. Un enfoque para tratar un conjunto de ecuaciones de este tipo consiste en considerar la configuración de un modelo de ecuaciones simultáneas en el que una o varias de las variables explicativas de una o varias ecuaciones sean a su vez la variable dependiente (endógena) asociada a otra ecuación del sistema completo.

Por otra parte, supongamos que ninguna de las variables del sistema es a la vez explicativa y dependiente. Aún puede haber interacciones entre las distintas ecuaciones si los componentes de error aleatorio asociados con al menos algunas de las distintas ecuaciones están correlacionados entre sí. Esto significa que las ecuaciones pueden estar vinculadas estadísticamente, aunque no estructuralmente, a través de la distribución conjunta de los términos de error. Este comportamiento se refleja en el modelo de ecuaciones de regresión aparentemente no relacionadas (SUR), en el que las ecuaciones individuales están de hecho relacionadas entre sí, aunque superficialmente no lo parezcan.⁷

En resumen, el modelo SUR (Seemingly Unrelated Regression) es una técnica de análisis multivariante que se utiliza para modelar y predecir múltiples variables dependientes que pueden estar correlacionadas entre sí debido a posibles correlaciones en los términos de error. En el contexto de la predicción de ventas en un e-commerce, el modelo SUR puede ser utilizado para modelar y predecir las ventas de diferentes productos que se venden en el sitio. Cuando se tienen múltiples categorías de productos en un sitio de comercio electrónico, como en este caso, es probable que las ventas de un producto estén relacionadas con las ventas de otros productos.⁸ Por ejemplo, si un sitio vende productos electrónicos, es probable que las ventas de teléfonos móviles estén correlacionadas con las ventas de accesorios para teléfonos móviles debido a posibles correlaciones en los términos de error de las distintas ecuaciones.

En lugar de estimar cada variable dependiente por separado, el modelo SUR modela las variables simultáneamente, teniendo en cuenta la correlación entre sus errores generando un modelo para cada variable. De esta manera, el modelo puede mejorar las predicciones al considerar las interacciones entre las variables. Las métricas que se utilizan para evaluar el modelo SUR son similares a las que se utilizan en el modelo de regresión lineal múltiple, como el R cuadrado, el error cuadrático medio (MSE) y el error absoluto medio (MAE). En ese caso nos enfocamos en el R cuadrado ajustado (R^2 Ajustado) que indica qué tan bien se ajustan los datos a un modelo de regresión. Va de 0 a 1, donde 1 significa un ajuste perfecto. Esta métrica representa la proporción de

⁷ Hayashi (2011)

⁸ Zucchini et al. (2017)

la variabilidad de la variable dependiente que puede ser explicada por las variables independientes en el modelo. Un valor cercano a 0 indica un mal ajuste, mientras que un valor cercano a 1 indica un buen ajuste.⁹

2.3.1 Ejemplos de aplicación en SUR

Un ejemplo destacado de aplicación del modelo SUR (Seemingly Unrelated Regression) es un estudio titulado "A Seemingly Unrelated Regression Analysis of Online Sales"¹⁰. En este trabajo, los investigadores emplearon el modelo SUR para analizar y predecir las ventas de diversos productos en un sitio web de comercio electrónico. Los resultados obtenidos demostraron que el modelo SUR mejoró significativamente en términos estadísticos las predicciones en comparación con el modelo de regresión lineal múltiple tradicional. Este hallazgo resalta la utilidad del modelo SUR en el ámbito de la predicción de ventas en e-commerce.

Otro artículo relevante que aborda este tema es "Modeling and Forecasting Multivariate Time Series of Counts with Random Effects"²¹. En este estudio, los autores emplearon modelos SUR para modelar y predecir las ventas de diversos productos en una tienda minorista. La utilización de modelos SUR permitió capturar las interrelaciones entre las diferentes variables y mejorar las predicciones en comparación con enfoques más tradicionales. Este trabajo destaca la aplicabilidad del modelo SUR en el análisis de ventas y pronóstico de series de tiempo multivariadas.

Estos dos estudios ejemplifican cómo el modelo SUR puede ser utilizado con éxito en diferentes contextos, como la predicción de ventas en un e-commerce y el análisis de series de tiempo multivariadas en una tienda minorista. Estos casos demuestran las ventajas y la efectividad del modelo SUR en la mejora de la capacidad predictiva de las predicciones y en el análisis de relaciones complejas entre variables.

2.3.2 SUR en Nuestro Contexto

En particular, cuando se trata de un comercio electrónico que ofrece productos en múltiples categorías, es fundamental comprender cómo las ventas varían en cada una de ellas y cómo se ven afectadas por diferentes factores. En este sentido, surge la necesidad de utilizar el Modelo de Regresión Seemingly Unrelated (SUR) para analizar y predecir las ventas en cada categoría individualmente. El modelo SUR permite considerar las correlaciones entre los errores de las variables dependientes, lo que proporciona una estimación más precisa y robusta de las relaciones entre las variables independientes y dependientes.

El procedimiento consiste en organizar los datos en dos dataframes: uno para las variables dependientes (Y) y otro para las variables independientes (X). En nuestro caso, las variables dependientes son las ventas ajustadas de cada categoría en particular, mientras que las variables independientes incluyen información relevante como clics, costo ajustado, día de la semana, brechas

⁹ Martín (2022)

¹⁰ Srivastava y Giles (2020)

cambiarías, costo promedio por clic, número de semana, número de mes y relación costo/ventas del anuncio.

Con esto en mente, se crea una ecuación por cada categoría de productos, donde la variable dependiente es la venta ajustada de la categoría y las variables independientes son las mencionadas anteriormente. Esto nos permite ajustar múltiples regresiones para cada categoría, considerando las correlaciones entre ellas. Al ajustar el modelo SUR, buscamos encontrar la combinación óptima de variables predictoras que explique de manera precisa las variaciones en las ventas de cada categoría, teniendo en cuenta la relación entre las variables dependientes. Al considerar las correlaciones entre las categorías, el modelo SUR mejora las predicciones en comparación con un modelo de regresión lineal simple.

Una vez que se ha ajustado el modelo SUR para cada categoría, podemos utilizar las ecuaciones resultantes para predecir las ventas por categoría en función del día de la semana. Esta información se vuelve especialmente relevante para nuestro modelo de asignación, ya que nos permite tomar decisiones más informadas sobre la asignación de recursos y estrategias de marketing en cada categoría en función de las ventas esperadas para cada día de la semana.

2.4. Problema de Asignación

Los problemas de asignación en una ecommerce se refieren a cómo asignar los recursos limitados (como el presupuesto de publicidad) a las diferentes campañas de marketing de manera óptima para maximizar el retorno de la inversión. La asignación de recursos puede ser complicada porque hay múltiples campañas y canales de publicidad que compiten entre sí por el presupuesto limitado, y cada uno tiene diferentes costos y beneficios esperados. En este trabajo nos concentramos en el canal de google ads pero cabe aclarar que el problema también se puede expandir a asignación de presupuesto a diferentes canales tales como Facebook, TikTok, entre otras.

En un modelo de asignación, la función objetivo es una medida matemática que se optimiza dadas las limitaciones o restricciones determinadas, y con variables que necesitan ser minimizadas o maximizadas usando técnicas de programación lineal o no lineal. En el caso de un modelo de asignación para las campaña de marketing de una ecommerce, el objetivo es determinar cómo deben hacerse las n asignaciones de presupuesto para maximizar la eficiencia total. Para construir una función objetivo en un modelo de asignación, se deben tener en cuenta los objetivos específicos del problema, las variables involucradas y las restricciones del modelo. Es importante destacar que el proceso de construcción de una función objetivo puede variar según el tipo de modelo de asignación que se esté utilizando y la naturaleza específica del problema. Por lo tanto, es necesario tener en cuenta los detalles del problema y utilizar técnicas adecuadas para construir una función objetivo efectiva.¹¹

En un problema de asignación como el que se presentará en este trabajo donde se busca maximizar la eficiencia (Adcost) de una empresa asignando el presupuesto a las distintas campañas de

¹¹ Kolman y Beck (2014)

marketing, las restricciones para la función objetivo son importantes para poder definir bien el problema teniendo en cuenta todas las especificaciones para que el problema sea aplicable. Algunas posibles restricciones que se pueden considerar incluyen: limitaciones de presupuesto, ya que no se tiene un presupuesto ilimitado para las campañas, objetivos de retorno de inversión, limitaciones de recursos disponibles para cada campaña, entre otras.

2.4.1 Solver ZIMPL

El modelo de asignación puede resolverse utilizando diferentes métodos de optimización, y uno de ellos es ZIMP, un lenguaje de programación orientado a la formulación y resolución de problemas de programación matemática. Es ampliamente utilizado para maximizar o minimizar una función objetivo sujeta a un conjunto de restricciones lineales. En el contexto de los problemas de asignación en la publicidad, ZIMPL es especialmente adecuado debido a su capacidad para manejar grandes cantidades de datos y restricciones, que son comunes en este tipo de problemas. Por lo tanto, utilizaremos ZIMPL para resolver nuestro problema de asignación de inversión en campañas de marketing.¹²

Existen varios tipos de problemas de asignación que pueden ser resueltos mediante el método ZIMPL. A continuación, se detallan algunos de ellos:

Problemas de asignación de transporte: Estos problemas se utilizan para determinar la forma más eficiente de enviar bienes desde los proveedores hasta los destinos. En este tipo de problemas, se busca minimizar los costos de transporte, mientras se satisfacen las demandas y las ofertas.

Problemas de asignación de tareas: Estos problemas se utilizan para asignar trabajos o tareas a los trabajadores o máquinas de manera que se minimice el tiempo de producción o se maximice la eficiencia. Para resolver estos problemas, se utiliza el método ZIMPL junto con la técnica de asignación de menor costo.

Problemas de asignación de recursos: Estos problemas se utilizan para asignar recursos limitados, como materiales o máquinas, a diferentes proyectos de manera que se maximice la ganancia total o se minimice el costo total. Para resolver estos problemas, se utiliza el método ZIMPL junto con la técnica de asignación de costo mínimo.

En nuestro caso, al distribuir la inversión en campañas de marketing para cada día de la semana con el objetivo de maximizar la eficiencia, nos encontramos en el ámbito de los problemas de asignación de recursos. Estos problemas buscan asignar recursos limitados a diversas actividades de manera que se minimice una función objetivo, en este caso, la eficiencia medida como la relación entre la inversión y las ventas.

2.5 Medida de Eficiencia

Como se mencionó previamente, la medida de eficiencia en la que se basa el problema para esta tesis es el Adcost que viene dado por: inversión/ventas. El AdCost es una métrica utilizada en publicidad digital para medir la eficiencia de una campaña publicitaria. Esta métrica indica el

¹² Gorai (2022)

porcentaje de ingresos generados por una venta que se ha gastado en publicidad para obtenerla. Dicha métrica nos permite evaluar la rentabilidad de la campaña publicitaria y hacer ajustes en caso de que los costos publicitarios estén siendo demasiado altos en relación a los ingresos generados. Una buena medida de AdCost dependerá del margen de ganancia de cada producto y de los objetivos de la campaña, por lo que no existe un número fijo que se considere aceptable para todas las empresas. Algunas empresas pueden considerar un AdCost del 30% como aceptable, mientras que para otras puede ser demasiado alto o bajo.¹³ A modo de ejemplo, si se exige un adcost de 1.2, se puede leer como: por cada venta generada, se invirtió en promedio 1.2 unidades monetarias en publicidad digital.

Esto último, es un indicador clave para las empresas de comercio online ya que permite evaluar la rentabilidad de la inversión realizada en la plataforma. En términos generales, esta medida permite comparar el costo de las inversiones realizadas en la plataforma con los ingresos obtenidos a través de la misma. Es importante destacar que la eficiencia de inversión/ventas (AdCost) no solo nos indica la rentabilidad de la inversión realizada en la plataforma, sino que también nos da una idea clara sobre la salud financiera de nuestro negocio en línea.

En términos prácticos, la eficiencia de inversión/ventas nos aporta información valiosa para la toma de decisiones de negocio en línea. Por ejemplo, si esta métrica representa una eficiencia alta, se podría concluir que la inversión realizada en la plataforma de comercio electrónico está generando ingresos suficientes para cubrir los costos de la inversión y obtener ganancias significativas. En este caso, la empresa podría considerar aumentar la inversión en la plataforma para continuar mejorando su eficiencia. Por otro lado, si la eficiencia de inversión/ventas es baja, la empresa debería revisar los costos de la plataforma y evaluar la posibilidad de reducirlos para aumentar su rentabilidad. También podría ser necesario revisar la estrategia de marketing y ventas para mejorar la conversión de visitantes en clientes y aumentar los ingresos generados.

En conclusión, la medida de eficiencia de inversión/ventas es una métrica clave para las empresas de comercio electrónico. Nos permite evaluar la rentabilidad de la inversión realizada en la plataforma y nos da información valiosa para la toma de decisiones de negocio en línea. Por lo tanto, esta métrica parece ser una excelente medida de éxito para una empresa de comercio electrónico.

2.6 Antecedentes

En el campo de la predicción de ventas y optimización del gasto en marketing digital, se ha desarrollado una amplia gama de investigaciones y estrategias que buscan mejorar la eficiencia y precisión en estas áreas clave del negocio. Para contextualizar la importancia de los hallazgos encontrados en este trabajo y valorarlos adecuadamente, es esencial realizar una revisión de literatura exhaustiva que analice los enfoques y resultados previos.

¹³ Reibstein, D., Pfeifer, P., Bendle, N., Farris, P. (2015). Marketing Metrics: The Manager's Guide to Measuring Marketing Performance. Reino Unido: Pearson Education.

En la literatura académica y científica, se pueden encontrar diversos estudios y modelos que abordan la predicción de ventas y la optimización del gasto en marketing digital. Por ejemplo, en un artículo de Google Scholar titulado "E-commerce system for sale prediction using machine learning technique"¹⁴, los autores proponen el uso de algoritmos de aprendizaje automático, como Random Forest y Regresión Logística, para predecir las ventas en tiendas en línea. El estudio demuestra que estos enfoques pueden proporcionar estimaciones precisas de las ventas futuras, lo que resulta valioso para la planificación y toma de decisiones empresariales

Además, otro trabajo de investigación relevante encontrado titulado "Digital Marketing Excellence: Planning, Optimizing and Integrating Online Marketing"¹⁵ Presenta una estrategia de optimización de gasto en marketing digital basada en el análisis de datos. Los autores proponen el uso de técnicas de minería de datos y análisis de comportamiento del consumidor para identificar patrones y tendencias que guíen la asignación efectiva del presupuesto de marketing en diferentes canales digitales. Esta aproximación ha demostrado mejorar el retorno de la inversión en campañas de marketing y maximizar la eficiencia del gasto

Asimismo, investigaciones recientes han explorado el uso de técnicas más avanzadas, como el aprendizaje profundo (deep learning), para la predicción de ventas y la personalización de estrategias de marketing. Un estudio en particular, titulado "Deep Neural Networks for Sales Forecasting in Retail"¹⁶ muestra cómo las redes neuronales profundas pueden capturar relaciones complejas entre múltiples variables y mejorar la precisión de las estimaciones de ventas.

En general, se puede observar que existe un amplio conjunto de enfoques, modelos y herramientas disponibles en la literatura para abordar la predicción de ventas y la optimización del gasto en marketing digital. Esta contextualización es fundamental para destacar la relevancia y el impacto de los hallazgos encontrados en el presente trabajo, así como para proporcionar una perspectiva para futuras investigaciones en esta área.

3. Datos

Los datos para esta tesis fueron tomados de un informe hecho desde la plataforma de Google Ads. Google Ads proporciona estos informes en un formato relativamente fácil de usar donde los clientes pueden acceder en cualquier momento y desde cualquier lugar, lo que nos permite tomar decisiones informadas sobre las campañas publicitarias y mejorar el rendimiento de nuestros anuncios. Estos informes pueden contener filtros en base a fechas específicas con métricas específicas para cada campaña. Al revisar estos informes, los clientes podemos obtener una visión detallada del rendimiento de las campañas publicitarias en un período específico de tiempo, incluyendo el número de impresiones, clics, conversiones, el costo por conversión, entre otras métricas.

En el contexto de este proyecto, se decidió elegir métricas de mercado que pudieran proporcionar una idea de la performance de las ventas según el día de la semana y la categoría de los productos.

¹⁴ Singh et al. (2020)

¹⁵ Chaffey y Smith (2017)

¹⁶ Loureiro et al. (2018)

Estas métricas nos parecen importantes para tener una visión global del comportamiento del mercado y poder tomar decisiones informadas en la asignación del presupuesto de marketing a cada categoría en función de su rendimiento. Al considerar métricas de mercado específicas para cada categoría, es posible evaluar de forma precisa el rendimiento de cada una en relación con el mercado en general. De esta manera, podemos utilizar estas métricas para predecir la cantidad de conversiones de cada categoría para cada día de la semana y optimizar la asignación del presupuesto de marketing en consecuencia.

Con el objetivo de tener un análisis más completo, incluimos variables adicionales que serán utilizadas posteriormente en nuestro modelo de predicción y asignación de presupuesto. Estas variables abarcan tanto métricas relacionadas con el tiempo (como el número de día de la semana y el número de semana del año) como métricas que nos proporcionarán información valiosa sobre el contexto económico, como la brecha cambiaria. Al agregar estas variables, buscamos enriquecer nuestro análisis y obtener una visión más completa de los factores que influyen en nuestro modelo de predicción y asignación de presupuesto.

3.1. Diccionario de datos

La base de datos con la que se trabajó cuenta con un total de 35,256 filas y 16 columnas que contienen diferentes variables para nuestro modelo de predicción de conversiones por día por categoría. La tabla 1 muestra una breve descripción de cada variable.

Tabla 1: Descripción de la base de datos.

Variable	Descripción	Tipo	Predictora/De Respuesta
Campaign	Nombre de la campaña publicitaria. Cada una representa una categoría de productos.	Categórica	Predictora
Day	Fecha específica de cada observación (Día, Mes y Año)	Numérica discreta	Predictora
Day of the week	Día de la semana (lunes a domingo en texto)	Categórica	Predictora
N°Day	Día de la semana en formato numero (Lunes =1)	Numérica discreta	Predictora
Month	Mes del año en formato texto	Categórica	Predictora
Year	Año en que se registraron los datos	Numérica discreta	Predictora
Clicks	Número de clics en el anuncio (Cada vez que se hizo Click en el anuncio mostrado)	Numérica discreta	Predictora
Impr.	Número de impresiones del anuncio (Cada vez que se mostró el Anuncio)	Numérica discreta	Predictora
Cost	Costo total de la campaña publicitaria (Inversión)	Numérica continua	Predictora
Avg. CPC	Costo promedio por clic (Costo/Clicks)	Numérica continua	Predictora
Adcost	Métrica de eficiencia promedio de la campaña según la plataforma	Numérica continua	Predictora
ConvValue	Valor en Pesos Argentinos de las Conversiones (Ventas)	Numérica continua	De Respuesta
Brecha	Diferencia % entre la cotización oficial del dólar contra la cotización CCL	Numérica continua	Predictora
Week_Num	Representa el número de la semana del año	Numérica continua	Predictora
Month_Num	Indica el número del mes del año	Numérica continua	Predictora
Campaign_Num	Número que se utiliza a modo de código para representar cada campaña en el conjunto de datos	Numérica continua	Predictora

*Fuente: Elaboración propia

3.2 Descripción de los datos

La base de datos utilizada contiene información detallada sobre 23 campañas publicitarias, cada una correspondiente a una categoría de productos específica. Además de la información de la campaña, se registran datos relacionados con la fecha, el día de la semana, el mes y el año en que se llevó a cabo la publicidad (entre los años 2020 y 2023). También se incluyen datos sobre la moneda utilizada (en este caso, se consideran únicamente los Pesos Argentinos), el número de clics, las impresiones, el costo total, el costo promedio por clic (Avg. CPC), el valor de las conversiones (Ventas) y el Adcost.

Con el objetivo de enriquecer el análisis, se han agregado variables numéricas basadas en segmentos semanales, utilizando técnicas de feature engineering. Estas variables incluyen el número de semana del año (Week_Num), el número de mes (Month_Num), el número de día de la semana (N°Day), donde el lunes es representado por el número 1 y el domingo por el número 7, y un índice asignado a cada campaña (Campaign_Num). Estas variables adicionales, generadas a través del proceso de feature engineering, serán utilizadas en la sección 2.4 para visualizar los datos y en el procesamiento de los datos, brindando una mayor profundidad en la comprensión de los resultados obtenidos en la sección 4. El feature engineering permite extraer y crear características nuevas a partir de los datos existentes, mejorando así la capacidad del modelo para capturar patrones y relaciones relevantes en los datos.

Por último, se agrega una variable a la base de datos que hace referencia a la brecha cambiaria entre la cotización del dólar oficial y la cotización del dólar contado con liqui (CCL). Agregar la brecha cambiaria entre la cotización oficial y la cotización de dólar CCL se evaluará en este contexto ya que puede proporcionar información valiosa sobre el impacto de los cambios en la economía en el desempeño de las ventas, tanto generales como de campañas específicas en un país donde la inflación y la economía suelen impactar tanto en los precios como en el comportamiento de los consumidores. El funcionamiento del dólar contado con liqui consiste en la compra-venta de acciones o títulos de deuda en pesos que cotizan en el país y en otro mercado internacional. La brecha cambiaria puede indicar la falta de confianza en la economía, la inflación, el aumento de los precios, la devaluación de la moneda y hasta incentivar a comprar en categorías como por ejemplo electrónica antes de que se devalúe el tipo de cambio oficial.¹⁷

En términos de ventas, una brecha cambiaria más amplia puede afectar la economía en general y, en consecuencia, influir en el poder adquisitivo de los consumidores. Si la inflación aumenta y la moneda se devalúa, los precios pueden aumentar y los consumidores pueden reducir su gasto en bienes y servicios, lo que puede afectar negativamente las ventas de la empresa. Además, una brecha cambiaria mayor puede afectar a los costos de producción, la disponibilidad de materiales y otros factores que pueden afectar la rentabilidad de la empresa. Un ejemplo claro es lo que ocurre con los productos de electrónica, si la brecha cambiaria se percibe en aumento, los consumidores pueden anticipar que los precios de estos productos aumentarán en el futuro debido a la inflación y la devaluación del tipo de cambio. Como resultado, pueden decidir comprar electrónica en el presente para satisfacer una demanda futura y evitar pagar precios más altos en el futuro. Por lo tanto, parece importante considerar la brecha cambiaria como un factor que puede afectar el desempeño de las ventas y las decisiones empresariales.

En la sección 2.3 se proporcionará información sobre la importancia de cada categoría para la empresa y el por qué es importante asegurarnos un mayor presupuesto en algunas de ellas tanto por su performance como por la inversión y atención que la empresa brinda en productos puntuales que se muestran en cada categoría. Para dar mayor contexto sobre las campañas a las que hacemos referencia durante los análisis, en la sección 8.3 del anexo se podrán encontrar definiciones breves para cada una de ellas.

¹⁷ Viana (1988)

El período disponible en la base de datos va desde el 6 de enero del 2020 hasta el 6 de marzo del 2023, y para los primeros análisis y modelos se utilizará la información del 6 de enero del 2020 hasta el 31 de agosto del 2022 con el objetivo de separar data para futuros análisis y validación de datos. La data, luego de ser dividida para estos futuros análisis, cuenta con 21.844 filas con 17 columnas que son las mencionadas previamente. En la sección 8.1 del Anexo se puede observar un breve resumen de la base de datos y el tipo de dato que tiene cada variable.

3.2.1 Estrategia de Validación

La decisión de utilizar la información desde el 6 de enero de 2020 hasta el 31 de agosto de 2022, y excluir semanas puntuales, se basa en un análisis previo de la distribución de datos y un criterio de separación para asegurar representatividad y generalización en los resultados. El 6 de enero de 2020 se eligió como punto de inicio debido a su estabilidad antes de eventos significativos que pudieran influir en los datos. El 31 de agosto de 2022 se seleccionó como corte para obtener un modelo que nos permita comparaciones futuras con datos actualizados. Esta partición ayuda a que el modelo no se sobreajuste o subajuste a patrones temporales específicos y minimiza el impacto de eventos atípicos. Además, se eligieron estas fechas para evaluar el desempeño del modelo con datos más recientes. Incluir datos hasta el 31 de agosto de 2022 permite evaluar la capacidad predictiva del modelo en comparación con datos históricos más antiguos. Esto es esencial para verificar si el modelo puede generalizar y adaptarse a posibles cambios o tendencias emergentes. Dado que los patrones pueden variar con el tiempo, la evaluación con datos cercanos al presente en el conjunto de validación brinda una mejor comprensión de la precisión y eficacia del modelo en situaciones actuales.

En resumen, esta elección busca evitar sesgos y garantizar una buena generalización del modelo, permitiendo su prueba con datos más recientes para evaluar su capacidad predictiva y aplicabilidad en escenarios cercanos a la realidad presente.

3.3 Análisis y Visualizaciones de Datos Crudos (con outliers)

Como primer paso en nuestro análisis, se realizan visualizaciones detalladas con el objetivo de comprender mejor la estructura y características de la base de datos. Esta exploración inicial nos permite identificar posibles outliers, es decir, valores atípicos que pueden distorsionar nuestros análisis posteriores. Una vez detectados estos outliers, se procede a realizar una limpieza de los datos con el fin de mejorar la calidad y confiabilidad de nuestros análisis futuros. Esta limpieza puede implicar diferentes enfoques, como la eliminación de registros con valores extremos, la imputación de valores faltantes o la corrección de datos erróneos. Al llevar a cabo estas acciones de limpieza, buscamos minimizar el impacto negativo de los outliers en nuestros resultados y asegurarnos de que los análisis posteriores se basen en datos representativos.

3.3.1 Análisis de outliers

Se calcularon estadísticas descriptivas básicas, como la media, mediana y desviación estándar, para cada variable. Esto permitió obtener una comprensión cuantitativa de la dispersión de los datos. Si se

encontraron diferencias llamativas entre las variables en términos de sus estadísticas descriptivas, esto podría indicar un desequilibrio en la distribución de los datos.

Tabla 2 : Estadísticas Descriptivas de Variables Numéricas

	Clicks	Impr	Cost	AvgCPC	Adcost	Convvalue
count	21529,00	21529,00	21529,00	21529,00	21529,00	21529,00
mean	7178,08	73310,74	11731,56	1,40	0,67	2025037,46
std	8464,91	88325,14	27778,50	1,05	0,93	2286862,52
min	0,00	4,00	0,00	0,00	0,00	0,00
25%	1471,00	16339,00	1462,97	0,73	0,24	507296,16
50%	4636,00	50462,00	5243,01	1,27	0,60	1350326,06
75%	9893,00	101776,00	14122,46	1,85	0,93	2737398,23
max	118531,00	1603122,00	957968,83	17,80	84,84	36013371,75

*Fuente: Elaboración Propia

En base a la tabla 2 podemos hacer los siguientes comentarios:

- **Clicks:** El número de clics varía desde 0 hasta 118,531. Lo que indica una amplia variación en los datos, lo que sugiere una diversidad en la cantidad de clics recibidos por campaña.
- **Impr:** Las impresiones van desde 4 hasta 1,603,122. Al igual que los clics, existe una amplia variación en los datos, lo que indica que las campañas han sido vistas en diferentes magnitudes.
- **Cost:** El costo varía desde 0 hasta 957,968.83. Los datos indican que ha habido diferentes niveles de inversión en las campañas a lo largo de los años, con una amplia gama de costos registrados.
- **AvgCPC:** El costo promedio por clic oscila entre 0 y 17.80. Esto indica que el costo medio por cada clic varía según la campaña y el momento del año.
- **Adcost:** La medida de eficiencia de nuestras campañas tiene un rango desde 0 hasta 84.84. Se observa una variación importante, lo que sugiere que algunas campañas han tenido un peor retorno sobre la inversión que otras.
- **Convvalue:** El valor de las conversiones oscila desde 0 hasta 36,013,371.75. Los datos indican que ha habido diferentes niveles de ventas de las campañas.

Al ver estas métricas , se realiza un análisis descriptivo de la influencia de los valores extremos (outliers) para las distintas variables numéricas de la base de datos. En la tabla 3 se encuentran calculados los cuartiles para cada una de las variables. Con estos cuartiles se calcula el límite inferior y superior a partir del cual los datos se consideran outliers. El límite inferior se calcula haciendo $Q25\% - 1.5 * IQR$, y el límite superior se calcula haciendo $Q75\% + 1.5 * IQR$. Se entiende como IQR la diferencia entre $Q75\%$ y $Q25\%$. Por último, se encuentra el porcentaje de outliers dado ese criterio.

Tabla 3: Análisis de outliers para las variables numéricas.

Variable	Cuantiles					Outliers		
	Q0%	Q25%	Q50%	Q75%	Q100%	Límite Inferior	Límite Superior	% de Outliers
N°Day	1.0	2.0	4.0	6.0	7.0	0.0*	12.0	0.0
Clicks	0.0	1468.0	4592.0	9828.0	118531.0	0.0*	22368.0	5.4
Cost	0.0	1458.5	5204.6	14065.0	957968.8	0.0*	32974.7	6.5
Brecha	59.14	45.85	40.94	37.33	17.78	58.63	24.55	4.3
AvgCPC	0.0	0.7	1.3	1.9	17.8	0.0*	3.5	3.0
Week_Num	1.0	13.0	24.0	36.0	53.0	0.0*	70.5	0.0
Month_Num	1.0	3.0	6.0	8.0	12.0	0.0*	15.5	0.0
Impr	4.0	16282.0	50116.5	101510.3	1603122.0	0.0*	229352.6	3.8
Adcost	0.0	0.2	0.6	0.9	84.8	0.0*	2.0	2.3
Convvalue	0.0	498423.4	1339585.3	2723334.3	36013371.8	0.0*	6060700.6	5.2

*Se utiliza 0.00 como límite real de la variable, ya que no tiene sentido pensar en números negativos.

*Fuente: Elaboración Propia en base a Monegal Ferrán, M. (1999)

La tabla 3 presenta el resultado del análisis de outliers para diferentes variables numéricas en la base de datos. Cada fila en la tabla corresponde a una variable y proporciona información sobre los límites inferior y superior calculados, así como el porcentaje de outliers encontrado para cada variable. La presencia de outliers puede indicar valores extremos o inusuales en los datos, que pueden afectar el análisis y los resultados. El cálculo de los límites inferior y superior se basa en el concepto de rango intercuartílico (IQR) y utiliza un factor de escala de 1.5 para identificar valores atípicos.¹⁸

Al interpretar la tabla 3, podemos observar que algunas variables tienen un porcentaje de outliers relativamente alto, mientras que otras tienen un porcentaje más bajo o incluso cero outliers. Por ejemplo, la variable "Cost" tiene aproximadamente un 6.51% de outliers, lo que indica que hay valores inusuales en el costo de los anuncios. Por otro lado, las variables "N°Day" "Week_Num" y "Month_Num" no tienen outliers según el criterio utilizado. Hacemos breve foco en las variables que presentan un porcentaje de outliers mayor y vemos que pertenecen a días de eventos puntuales donde la inversión y las conversiones suelen tener picos muy altos y nuestros incentivos de marketing al igual que el comportamiento de los consumidores no suele ser el mismo que tenemos en la operación diaria.

En consideración a los resultados obtenidos, se ha tomado la decisión de excluir las semanas que presenten eventos específicos de mercado en los diferentes años evaluados (ya sea Hot Sale, Cyber, Black Friday). Específicamente, se han identificado como semanas outliers aquellas en las que los costos superan los 5 millones de unidades monetarias. Esta elección se basa en el análisis de los datos atípicos en la variable de costo que se pueden observar en la tabla 3, cuyos valores extremos se encuentran alrededor de 32,974 unidades monetarias. Al considerar los días de la semana y el número de campañas registradas en nuestra base de datos, se estima que el umbral de 5 millones de unidades monetarias proporciona una aproximación adecuada.

¹⁸ Ferrán (1999)

Al eliminar estas semanas, se realiza una corrección en relación a las variables que presentan desviaciones significativas, ya que estas desviaciones son en gran medida atribuibles a eventos especiales de industria. Además, se ha decidido descartar las semanas cuyo promedio de brecha supere el valor de 58% ya que dichas diferencias se consideran atípicas en la distribución de datos.

Esta estrategia de exclusión selectiva permite obtener una muestra más representativa y coherente con el análisis en curso. Al eliminar semanas con eventos especiales y brechas promedio excepcionales, se reduce el sesgo potencial generado por influencias externas y se focaliza el análisis en periodos más estables y comparables. En consecuencia, se espera que esta selección cuidadosa de semanas contribuya a una interpretación más precisa y robusta de los resultados, al minimizar la influencia de valores atípicos y eventos que distorsionen la comprensión general de los datos.

3.3.1.a Exclusión Fechas Especiales de Industria

Tras el análisis previo, hemos identificado que estas fechas especiales, como el Hot Sale o el Cyber Monday, generan valores atípicos o extremos dentro de los años de estudio. Estos valores atípicos pueden distorsionar el análisis y dificultar la obtención de resultados representativos y confiables.

Además, al revisar la cantidad de datos disponibles para estas fechas especiales, se ha evidenciado que no se cuenta con una cantidad suficiente de observaciones representativas para realizar un análisis abarcativo de estas fechas. Esto se debe a que estas fechas son eventos anuales y no se tienen registros históricos extensos para construir modelos predictivos aplicables.

Para abordar esta situación, se sugiere la utilización de variables dummies en el modelo. Estas variables dummies permitirán capturar y tener en cuenta la presencia de estos eventos especiales en el análisis, sin tener que incluir los datos de las fechas exactas. Las variables dummies se pueden construir asignando un valor binario (1 o 0) a cada fecha especial, indicando si el evento se produjo o no en una fecha determinada. La inclusión de las variables dummies en el modelo permitirá que este sea entrenado y ajustado para tener en cuenta la influencia potencial de estos eventos especiales en las ventas. Sin embargo, es importante tener en cuenta que, debido a la falta de datos representativos suficientes en nuestro caso, las predicciones específicas para estos eventos podrían ser limitadas en confiabilidad por lo que optamos excluirlas de nuestro análisis.

3.3.2 Análisis Descriptivo del Periodo

Se lleva a cabo un análisis para evaluar la distribución de nuestras observaciones en las variables "week_num", "year", "n°day", "month_num" y "campaign". El objetivo de este análisis es determinar si es necesario realizar algún ajuste a la base de datos antes de llevar a cabo un análisis más detallado. Consideramos esencial verificar la distribución de las variables en términos de datos antes de realizar un análisis más exhaustivo.

El análisis de la distribución de datos en las variables nos permite evaluar la calidad y representatividad de nuestros datos, lo que a su vez impacta en la fiabilidad de los resultados y conclusiones obtenidos. Al asegurarnos de contar con una distribución equilibrada de observaciones en estas variables, estaremos mejor preparados para realizar un análisis más detallado y confiable,

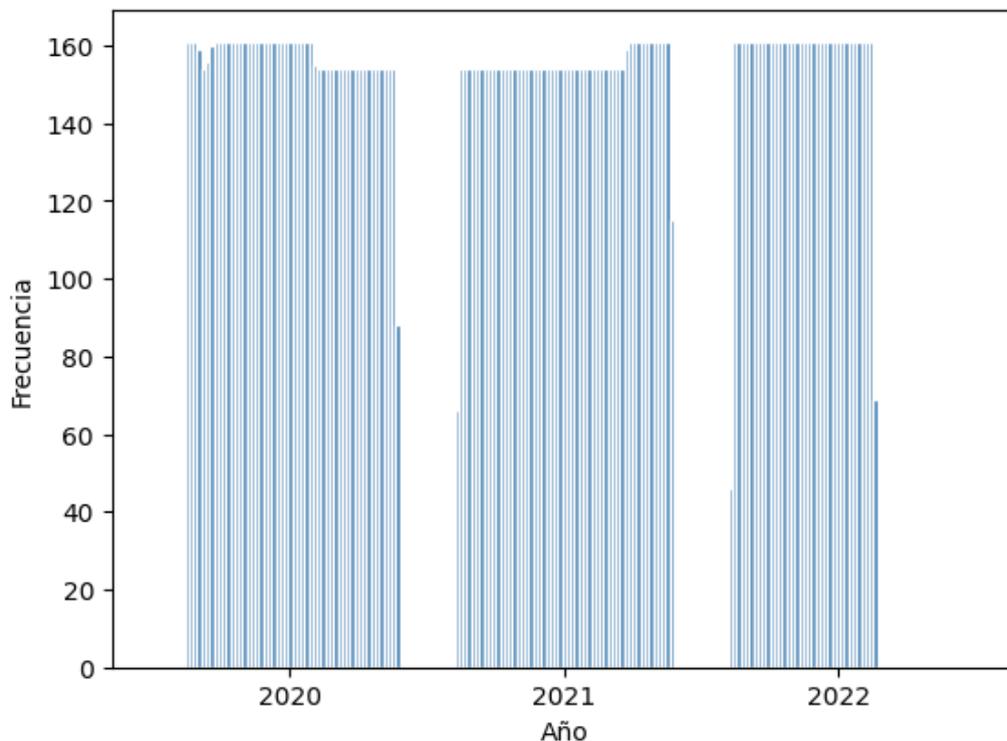
proporcionando una base sólida para la toma de decisiones informada en nuestro estudio. En resumen, la evaluación de la distribución de las variables en nuestra base de datos es un paso esencial antes de llevar a cabo un análisis más detallado. Al garantizar una distribución equilibrada, estamos fortaleciendo la confiabilidad y validez de nuestros resultados, evitando distorsiones potenciales.¹⁹

3.3.2.a Análisis de la distribución de datos

Se examina la distribución de observaciones en cada una de las variables categóricas mencionadas para comprender cómo están distribuidos los datos. Se utilizaron gráficos del tipo histogramas para visualizar la frecuencia o el conteo de valores en cada variable. Si se observó una distribución de datos desproporcional, con una concentración importante de valores en una categoría o rango particular, se consideró como una indicación de un posible desequilibrio en la cantidad de datos.

El gráfico 1 representa la distribución de observaciones a lo largo de las semanas del año y su evolución a través de los distintos años. Cada barra en el eje horizontal representa una semana, mientras que las columnas agrupadas dentro de cada barra indican la frecuencia de observaciones para cada año correspondiente. La utilización de múltiples columnas por semana permite visualizar la variación interanual de las observaciones y comprender si las diferencias en la cantidad de datos se deben a patrones sistemáticos inherentes a ciertas semanas del año o si, en cambio, se debe a una discrepancia en la cantidad de años considerados en el análisis.

Gráfico 1: Cantidad de Observaciones en la Variable Week_Num



*Fuente: Elaboración Propia

¹⁹ Roldán et al. (2019)

El análisis del Gráfico 1 revela que hay un desbalance en las observaciones de la variable "Week_Num" en las primeras y últimas semanas del año. Estas semanas muestran una cantidad menor de datos en comparación con las semanas intermedias. Este desbalance puede ser atribuido a varias razones, como la incompletitud de los años calendario, días festivos o vacaciones, patrones de actividad diferentes y el cierre o reinicio de operaciones comerciales. Es importante respaldar estas observaciones gráficas con medidas estadísticas, como la asimetría y la curtosis. En este caso, los valores obtenidos son:

Asimetría = 0.1895

Curtosis = -1.038.

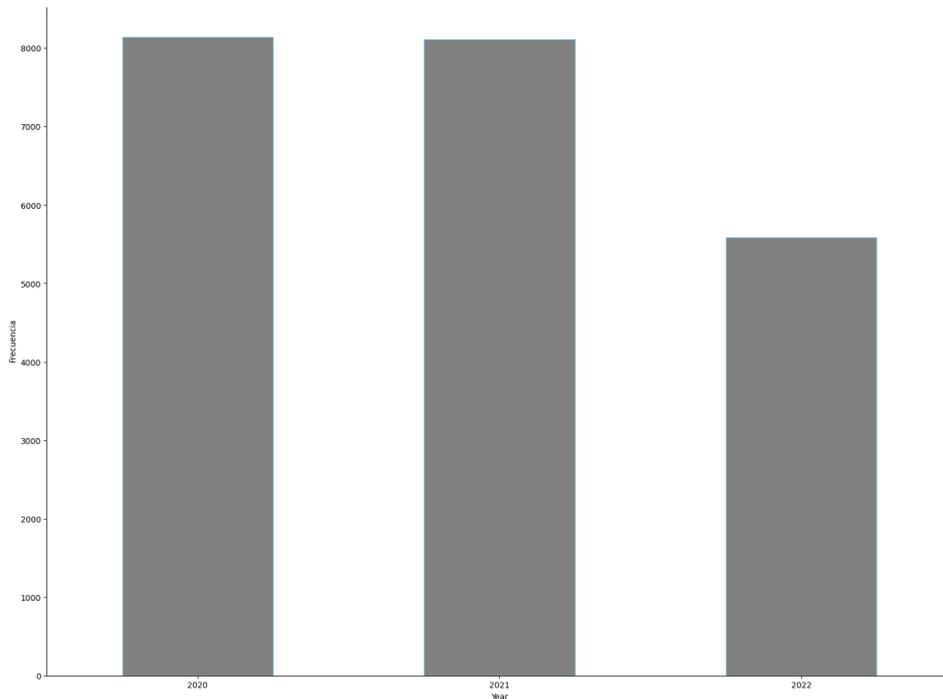
Estos resultados apoyan la interpretación visual del gráfico 1. La asimetría cercana a cero indica una distribución aproximadamente simétrica, lo cual respalda la observación de que las semanas intermedias del año tienen una cantidad similar de observaciones. Sin embargo, la asimetría ligeramente positiva sugiere una leve asimetría hacia la derecha, lo que podría indicar que las semanas finales del año tienen valores bajos en comparación con la media. La curtosis negativa de -1.038 respalda la observación visual de que las últimas semanas del año tienen una menor cantidad de datos. Indica que la distribución es ligeramente más achatada en comparación con una distribución normal, lo que podría deberse al menor número de observaciones en esas semanas.²⁰

En resumen, el análisis gráfico y las medidas estadísticas de asimetría y curtosis respaldan la observación de que hay un desbalance en las observaciones de la variable "Week_Num", particularmente en la primera y últimas semanas del año. Si consideramos que una semana completa tiene 7 días, es posible que la primera y última semana del año tengan menos días que las demás semanas. Esto puede generar un desequilibrio en la cantidad de observaciones en comparación con las semanas completas. A su vez, al no tomar el año 2022 completo por motivos de análisis posterior de la data, hace sentido que tengamos un desbalance en las últimas semanas del año.

Por otro lado, observamos cómo se distribuyen los datos para los distintos años en nuestra base de datos en el gráfico 2:

²⁰ Alamo (1952b)

Gráfico 2: Cantidad de Observaciones en la Variable Year

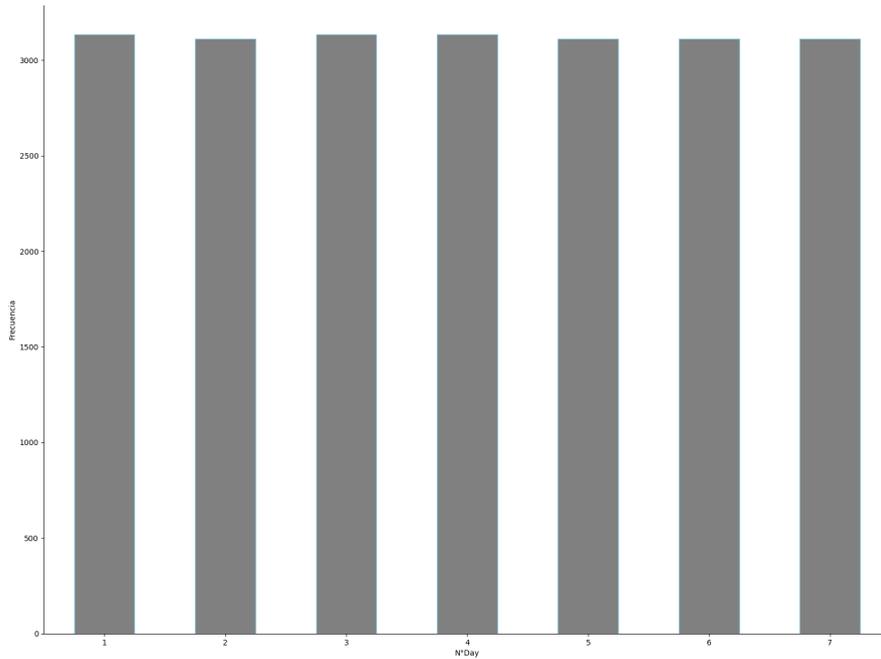


**Fuente: Elaboración Propia*

Podemos observar que se presentan menos datos para el año 2022 debido a una decisión previa de excluir la data más reciente en esta etapa inicial. Esta selección se realizó con el propósito de validar nuestros resultados en futuros análisis. Por lo tanto, se optó por omitir la data más reciente y utilizar únicamente datos disponibles hasta cierto punto en el tiempo, lo cual puede explicar la disminución en la cantidad de datos presentados para el año 2022. Esta estrategia nos permite asegurar la validez de nuestros análisis al trabajar con datos más completos y confiables en etapas posteriores de la investigación.

Realizamos este mismo análisis pero para la variable que tiene en cuenta los días de la semana siendo Lunes el primer día (1) y Domingo el último (7).

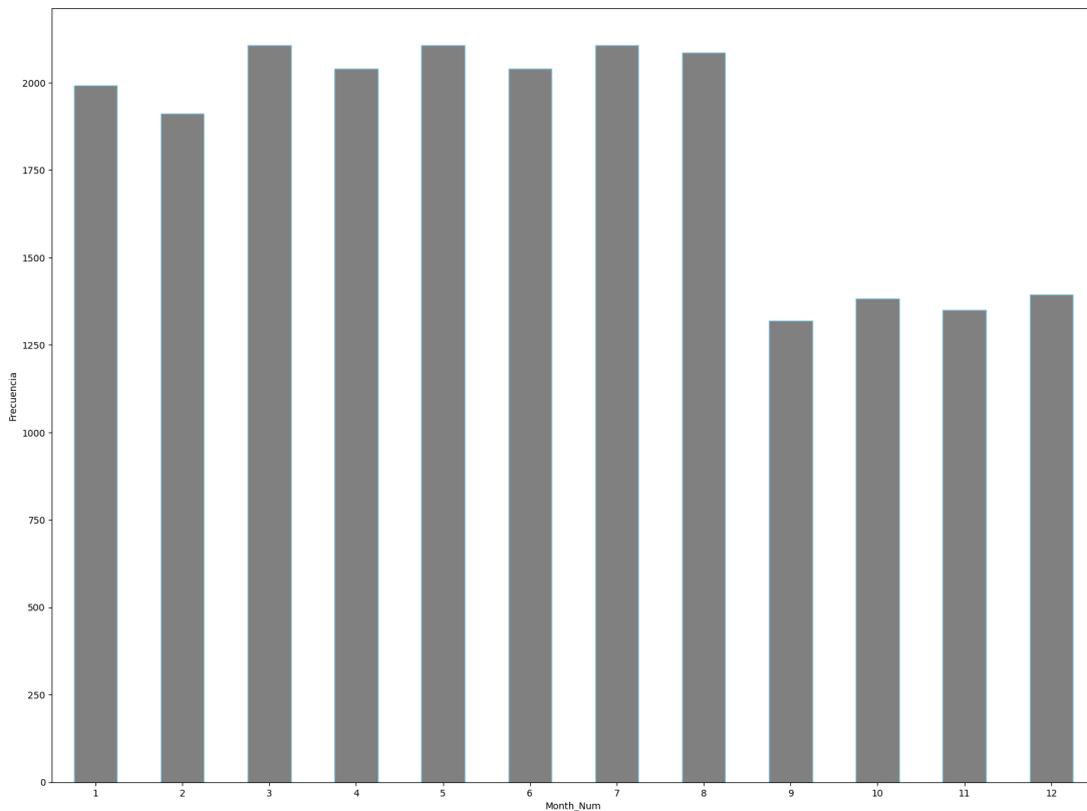
Gráfico 3: Cantidad de Observaciones en la Variable N°Day



*Fuente Elaboración Propia

Vemos que tenemos la misma cantidad de observaciones para cada día de la semana.

Gráfico 4: Cantidad de Observaciones en la Variable Month_Num



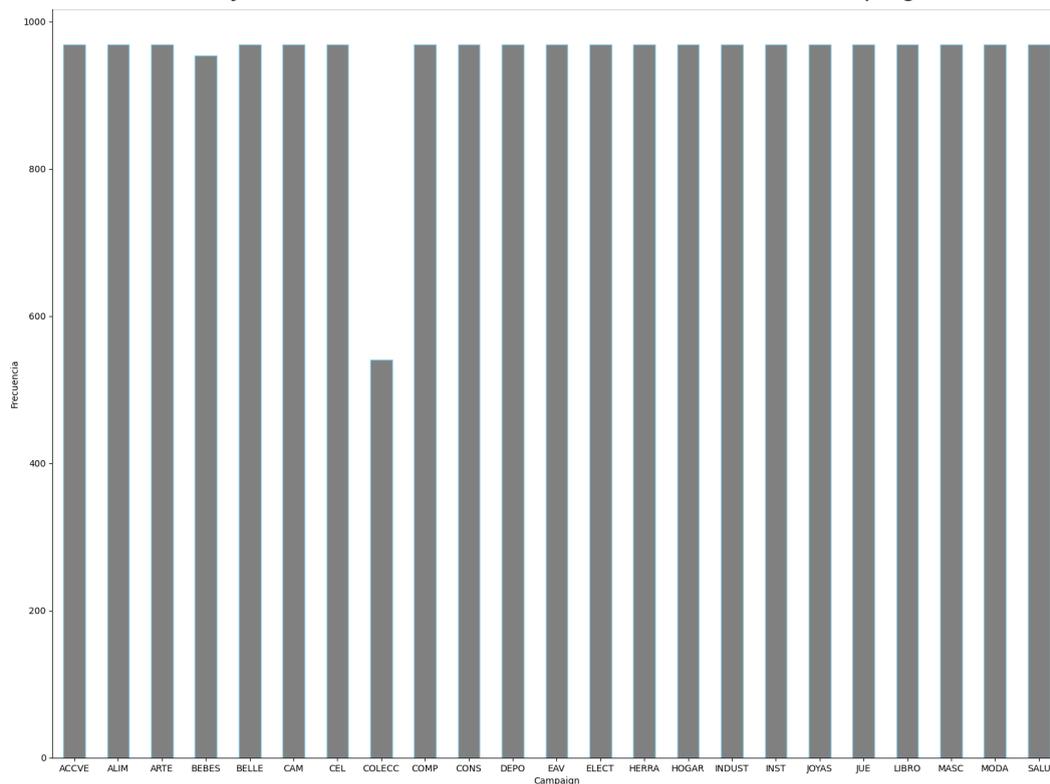
*Fuente: Elaboración Propia

En el gráfico 4 analizamos la cantidad de observaciones en la variable que hace referencia al Mes del año. En este caso, vemos también que los últimos meses del año tienen menos observaciones

debido al recorte de fechas del año 2022. A su vez notamos que no todos los meses están igual de representados, esto es puramente por la cantidad de día que contiene cada mes. Es cuestión de esperar que los meses con más días presenten más observaciones que los que tienen menor cantidad de días (Por ejemplo el mes 2 que hace ilusión a febrero lógicamente presentará menos observaciones).

Por último, visualizamos en el gráfico 5 la cantidad de observaciones que tenemos para cada Campaña:

Gráfico 5: Cantidad de Observaciones en la Variable Campaign



*Fuente: Elaboración Propia

Gracias a este gráfico podemos identificar que la variable que hace referencia a la categoría "Coleccionables y Hobbies" tiene una cantidad visiblemente menor de datos en comparación con otras campañas. Esta disparidad se debe a que en ciertos días o momentos del año, esta campaña tiende a no gastar dinero en publicidad, lo que resulta en la falta de datos para esos períodos. La razón detrás de esta discrepancia radica en que la campaña "Coleccionables y Hobbies" es considerada como una campaña de menor importancia y con poca prioridad dentro de la estrategia publicitaria de la empresa. Esta campaña no genera un retorno importante en términos de inversión, lo que lleva a la decisión de no destinar recursos financieros a ciertos días o momentos específicos del año.

Sin embargo, es importante destacar que aunque esta campaña no sea una de las más relevantes por el poco porcentaje que representa en nuestra operación diaria, aún tiene el potencial de generar cierto valor para la empresa. Es por esto que en este proyecto se propone la implementación de un modelo que evite que esta situación ocurra, asegurando que se destine una cantidad mínima de

dinero para obtener datos incluso en los días en los que tradicionalmente la campaña no genera gastos. La idea detrás de este enfoque es aprovechar todas las oportunidades posibles en la categoría de "Coleccionables y Hobbies", aunque no se tenga información detallada de ciertos días. Al mantener la campaña activa y asignar un presupuesto mínimo, podemos capturar datos valiosos que podrían ser utilizados para análisis posteriores y para tomar decisiones informadas en relación con esta categoría.

3.3.2.b Remover Outliers de Week_Num

En base a estos análisis, se ha observado que las variables "year", "n°day", "month_num" y "campaign" muestran una distribución de datos en la cual cada semana, día, mes y campaña respectivamente tienen una cantidad similar o parecida de observaciones.

Como nuestro objetivo es realizar un análisis de datos para predecir conversiones y asignar inversión de forma semanal, nos parece prudente considerar la comparabilidad de los datos y sacar de nuestra base de datos la primera y la última semana del año. Al excluir la primera y la última semana del año (Semana 1 y Semana 53), podemos mantener la consistencia en la duración de las semanas (Que todas tengan 7 días) y garantizar una comparación justa entre ellas. Esto nos permite obtener resultados más precisos y confiables al analizar la data, predecir conversiones y asignar inversión de forma semanal.

Este análisis ofrece una perspectiva inicial sobre la distribución de los datos y su relevancia para el estudio. Además, contribuye a la interpretación y posible generalización de los resultados en etapas posteriores del análisis de la base de datos. Al excluir estas dos semanas, nos quedamos con un total de 21529 observaciones, la cantidad de observaciones removida representaba solo un 1.44% del total de la data original.

3.3.3 Análisis Descriptivo de la Base de Datos

El objetivo de esta sección es representar gráficamente las relaciones entre las variables, utilizando técnicas como gráficos de dispersión o boxplot, entre otras. Esto nos ayuda tanto a identificar posibles patrones o relaciones entre las variables, como a evaluar si hay algún desequilibrio evidente en esos patrones y ver cómo se relacionan con nuestra variable a predecir (Ventas).

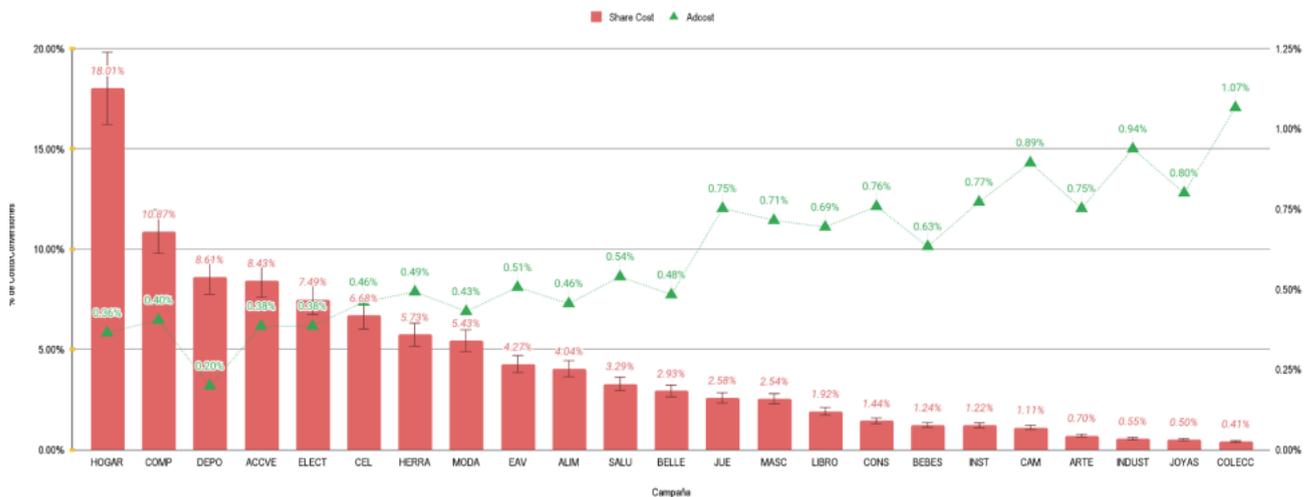
3.3.3.a Importancia de las Categorías para la Empresa

En la siguiente sección, graficamos las diferentes campañas que refieren a las categorías dentro de la empresa e-commerce. El objetivo principal es identificar las campañas que son más relevantes en términos de generación de ingresos, considerando tanto el costo invertido como la eficiencia de cada una. Analizar estas métricas nos permitirá comprender qué campañas nos brindan un mayor nivel de ingresos y cuáles pueden ser consideradas como las más importantes desde el punto de vista financiero. Esto nos ayudará a tomar decisiones estratégicas sobre cómo asignar nuestros recursos y enfocar nuestros esfuerzos para maximizar nuestros resultados comerciales.

A través del gráfico 6, podremos visualizar las campañas y su desempeño en términos de costo y en términos de Adcost, que previamente definimos como nuestra medida de eficiencia en nuestra

sección 1.3. Al identificar las categorías que generan un mayor retorno de la inversión, podremos enfocar nuestros esfuerzos en potenciar esas campañas y optimizar nuestros recursos para impulsar aún más nuestros ingresos.

Gráfico 6: Porcentaje de Costo sobre el Total y Adcost por Campaña



*Fuente: Elaboración Propia

Haciendo un análisis del gráfico 6 y considerando el impacto de cada campaña en nuestras conversiones, hemos tomado la decisión de clasificar las campañas en dos grupos con el objetivo de priorizar nuestras inversiones y enfocar nuestros recursos en las categorías que generan un mayor retorno.

En función del porcentaje de contribución a nuestras conversiones y el análisis del adcost (costo/ventas), hemos determinado que aquellas campañas que representan el 4% o más del total de conversiones y tengan un adcost más bajo se considerarán como más importantes. Estas categorías no solo tienen una gran participación en nuestras conversiones, sino que también muestran una eficiencia más alta en términos de costos en comparación con otras categorías. Representan aproximadamente el 85% de nuestros costos totales (véase en el gráfico como las categorías con columnas con mayor tamaño). Reconocemos la importancia crítica de estas campañas en la generación de ingresos y, por lo tanto, les otorgamos una atención especial.

Por otro lado, hemos identificado que existen categorías que generan una cantidad menor de conversiones y presentan un adcost más alto. Estas categorías no parecen ser tan relevantes para nuestros ingresos. Aunque asignaremos recursos a estas categorías, no las priorizaremos en términos de inversión y esfuerzo. Se considerarán de menor importancia dentro de nuestra estrategia en curso. Cabe aclarar que, asignar al menos un porcentaje de costo a las categorías menos importantes nos permite diversificar nuestras inversiones, obtener aprendizajes y datos valiosos, explorar oportunidades de crecimiento y brindar una experiencia completa a nuestros clientes. Aunque estas categorías pueden tener una contribución limitada en el presente, su potencial de crecimiento y los beneficios a largo plazo justifican su inclusión en nuestra estrategia de e-commerce. Al formar parte

de la estrategia de la empresa, es necesario asignarles, aunque sea poco representativo del total, un presupuesto de marketing.

Con esta clasificación estratégica, buscamos maximizar nuestro retorno de inversión al enfocar nuestros recursos y esfuerzos en las categorías que tienen un mayor impacto en las conversiones y los ingresos, al tiempo que consideramos la eficiencia en términos de costos. Esto nos permitirá optimizar nuestra estrategia publicitaria y asignar los recursos de manera más efectiva para obtener resultados positivos en nuestro desempeño financiero.

Para facilitar la clasificación y seguimiento de las campañas en nuestro problema de asignación que se verá más adelante, hemos asignado índices numéricos del 1 al 23 a cada una de ellas. Cabe destacar que las primeras 12 categorías corresponden a las más importantes, ya que representan una gran parte de nuestras conversiones y costos. Por otro lado, las últimas 11 categorías (índices del 13 al 23) se consideran menos prioritarias en nuestra estrategia de inversión. Esta decisión nos permitirá enfocar nuestros esfuerzos en las campañas más rentables y generar un mayor impacto en términos de generación de ingresos.

El gráfico 7 presenta la distribución del GMV (Gross Merchandise Value) total del sitio por categorías, con el objetivo de proporcionar una visión integral del aporte de cada categoría a las ventas generadas en el sitio. Este análisis va más allá del presupuesto de marketing y nos permite comprender la importancia de cada categoría en el contexto global del sitio a modo de justificación de nuestra decisión sobre dividir las categorías en grupos según importancia.

Gráfico 7: Distribución GMV (Ventas) total sitio por categorías

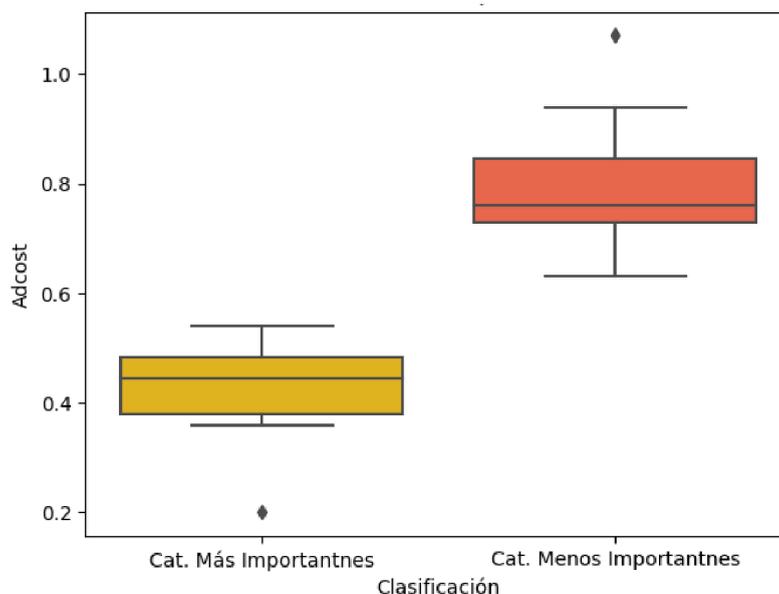


**Fuente: Extracto del reporte anual de la ecommerce*

El gráfico 7 nos brinda una perspectiva sobre la relevancia de cada categoría en términos de generación de ventas para toda la empresa. Podemos observar que algunas categorías presentan una contribución importante, lo que respalda nuestra decisión de priorizar estas campañas en nuestras estrategias de inversión. Es importante destacar que este análisis considera el panorama global del sitio, y no se limita únicamente al impacto de nuestras actividades de marketing. Esto nos brinda una visión más completa y nos ayuda a tomar decisiones fundamentadas en la asignación de recursos para poder ir en línea con los objetivos de la empresa.

En el gráfico 8 se presenta un boxplot de la variable "Adcost" para los dos grupos propuestos previamente: "Cat. Más Importantes" y "Cat. Menos Importantes".

Gráfico 8: BoxPlot de Adcost según Importancia Categoría



*Fuente: Elaboración Propia

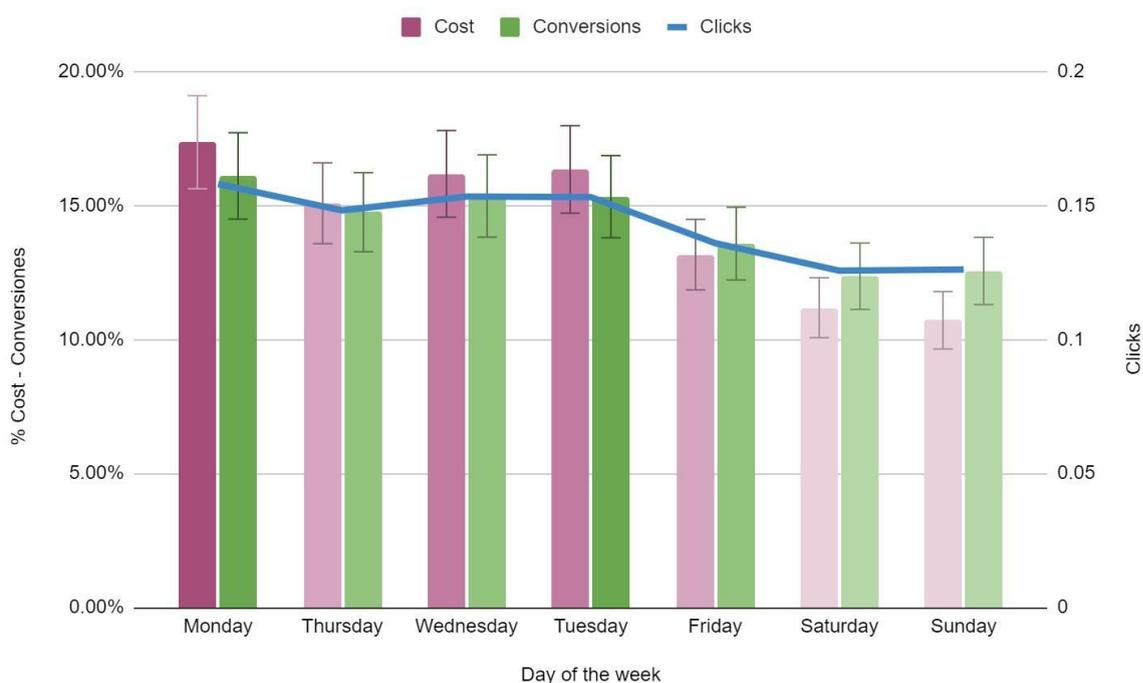
A partir del gráfico 8 podemos hacer observaciones sobre la distribución y variabilidad de la métrica de eficiencia en relación a cada categoría. Se puede observar que en general, las categorías "Cat. Más Importantes" presentan valores más bajos de Adcost en comparación con las "Cat. Menos Importantes". Esto indica que las campañas pertenecientes a las categorías más importantes tienen una relación más favorable entre el costo y las ventas. Esto demuestra y afirma la importancia de tener en cuenta estos valores al tomar decisiones relacionadas con la asignación de recursos y la inversión en marketing, ya que permiten identificar las categorías que tienen un mejor desempeño en términos de costo-ventas y aquellas que requieren una atención especial para mejorar su eficiencia publicitaria.

3.3.3.b Análisis de Distribución y Comportamiento de los Días de la Semana

En esta sección, se presenta un análisis detallado de la distribución y comportamiento de métricas a lo largo de los días de la semana. El objetivo es comprender la variabilidad y tendencias asociadas a cada día de la semana, lo que nos permitirá tomar decisiones informadas sobre la asignación de inversión en base a estos patrones.

En primer lugar, se han examinado tres métricas de mercado: los clicks, el costo y las conversiones. Estas métricas nos brindan una visión integral del rendimiento de nuestras campañas de marketing. Al analizar cómo se distribuyen y comportan a lo largo de la semana, podemos identificar posibles patrones y entender cómo afectan a nuestros resultados

Gráfico 9: Distribución Costo y Conversiones y comportamiento de los Clicks a lo largo de la semana

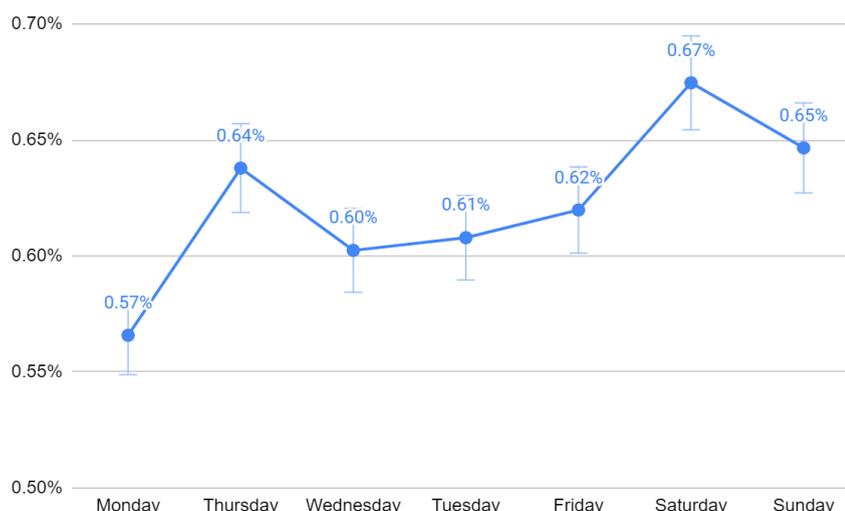


*Fuente: Elaboración Propia

El gráfico 9 muestra la distribución de las métricas costo, clicks y conversiones a lo largo de los días de la semana para el período de nuestro análisis. Los porcentajes indicados representan la proporción de cada métrica con respecto al total. Se puede observar que el día de la semana que presenta el mayor porcentaje de costo es el lunes, con un 17.36% del total. Le siguen el miércoles y el martes con porcentajes similares alrededor del 16%. Por otro lado, el domingo tiene el menor porcentaje de costo, con un 10.71%. A su vez, el patrón de distribución de los clicks parece ser similar al del costo. El lunes también lidera en términos de porcentaje, con un 15.80% del total. Le siguen el miércoles y el martes, con porcentajes cercanos al 15%. El domingo tiene el menor porcentaje de clicks, con un 12.61%.

Por último en términos de conversiones, nuevamente notamos que el lunes se destaca como el día con mayor porcentaje de conversiones, con un 16.10% del total, seguido por el miércoles y el martes, con porcentajes cercanos al 15%. Al igual que en las métricas anteriores, el domingo muestra el menor porcentaje de conversiones, con un 12.55%.

Gráfico 10: Adcost a lo largo de los días de la Semana



*Fuente: Elaboración Propia

El gráfico 10 presenta la tasa de eficiencia representada por el cociente "Adcost" (Costo/Convvalue). En el análisis de los valores del cociente "Adcost" para cada día de la semana, se observa que los días con mayor eficiencia son los Lunes(0.57%), Miércoles (0.60%)y Jueves (0.61%). En este contexto, una mayor eficiencia se traduce en un menor costo de ventas generado, lo que sugiere que los días con un cociente "Adcost" más bajo resultan más eficaces para la campaña publicitaria.

Estos hallazgos ayudan a comprender cómo el costo y las conversiones varían a lo largo de la semana, y cómo esa variación impacta en la eficacia de la campaña. La visualización de los datos a través de este gráfico permite identificar patrones específicos y tomar decisiones informadas para optimizar el rendimiento publicitario y maximizar la métrica de eficiencia de la inversión en publicidad.

Podemos observar que los días lunes, miércoles y martes suelen tener mayores niveles de costo, clicks y conversiones en comparación con el resto de los días. Esto indica que estos días son potencialmente más relevantes en términos de interacción de los usuarios y generación de resultados. Por otro lado, los fines de semana (sábado y domingo) presentan porcentajes más bajos en todas las métricas, lo que sugiere una menor actividad y participación de los usuarios en esos días.

La observación de que los lunes suelen generar el mayor volumen de clics en los anuncios de publicidad en línea es interesante y puede ser justificada por varios factores. Algunas investigaciones respaldan estas tendencias y proporcionan un contexto adicional sobre el comportamiento de los usuarios. Un estudio relevante en este contexto es el paper titulado "Click-Through Rate Prediction in Online Advertising: A Literature Review". En este paper, los autores analizan el impacto del día de la semana en las tasas de clicks en la publicidad en línea. Los resultados revelan que los lunes tienden a tener tasas de clicks más altas en comparación con otros días de la semana. Además, los autores proponen varias explicaciones para este fenómeno, que respaldan las razones mencionadas anteriormente.²¹

²¹ Yang y Zhai (2022)

Según el estudio, el regreso al trabajo después del fin de semana puede influir en el aumento de la actividad en línea, lo que se traduce en una mayor interacción con los anuncios publicitarios. Además, la planificación de las actividades de la semana, incluyendo compras y gastos, puede aumentar el interés de los usuarios en los anuncios y, en consecuencia, generar más clicks los lunes. El estudio también destaca la posibilidad de que las empresas ofrezcan incentivos especiales durante el fin de semana, lo que podría despertar el interés de los usuarios y llevar a un mayor número de clicks los lunes. Por último, explica que los lunes suelen ser días más activos en términos de trabajo y actividades, lo que significa que más personas están en línea y, por lo tanto, más expuestas a los anuncios publicitarios.

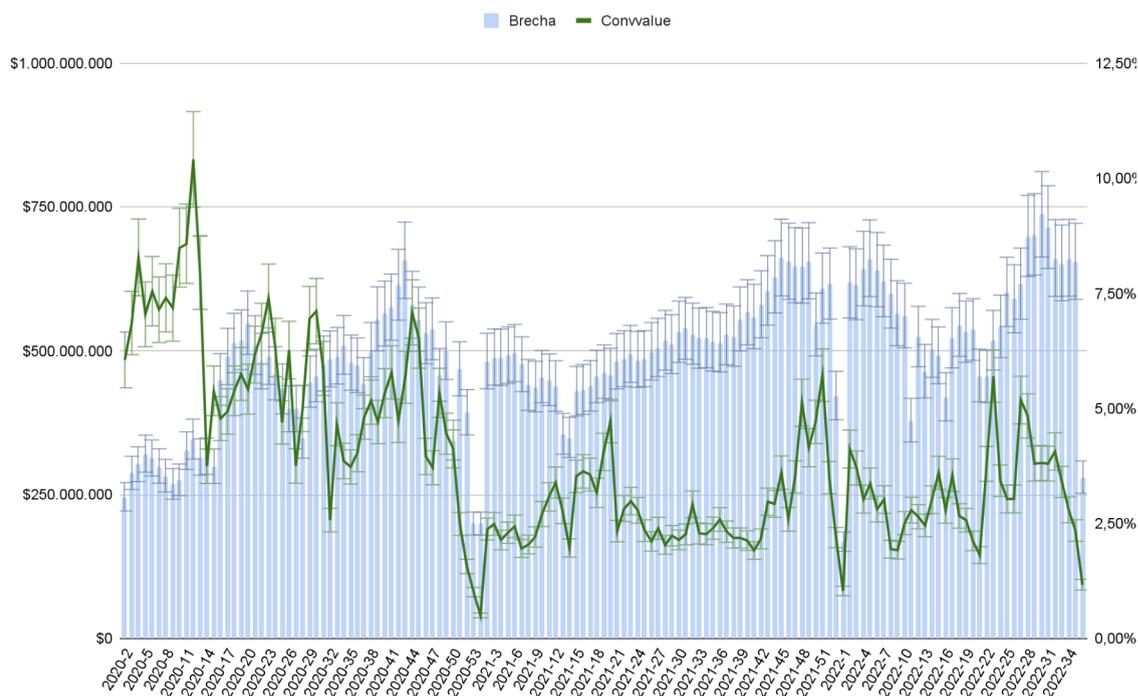
Estos hallazgos respaldan la idea de que los lunes son días importantes para la publicidad en línea, y brindan una base teórica para comprender el comportamiento de los usuarios y la distribución de los clicks y las ventas a lo largo de la semana. Esto nos proporciona un indicio de los resultados que esperamos ver en nuestra regresión y un fuerte incentivo a generar restricciones adicionales que contemplen el día de la semana al momento de asignar inversión.

3.3.3.c Análisis de la Brecha del Tipo de Cambio y las Ventas

En esta sección, se analiza la relación entre la Brecha del Tipo de Cambio (dólar oficial versus CCL) y las Ventas de nuestra empresa. Es importante destacar que la Brecha del Tipo de Cambio refleja el cociente entre el valor del dólar oficial y el valor del dólar en el mercado (CCL). Como fue mencionado previamente, esta brecha puede ser influenciada por diversos factores económicos, políticos y financieros, y puede tener un impacto en las operaciones comerciales, especialmente en el contexto de un comercio electrónico que depende de importaciones/ exportaciones y en donde los consumidores adelantan sus compras por expectativas de devaluación.

El gráfico 11 muestra el comportamiento semanal de la Brecha y el valor de las ventas. Se observa que la tendencia de la brecha es positiva, lo que indica un aumento gradual de la Brecha a medida que transcurre el tiempo:

Gráfico 11: Relación entre la Brecha del Tipo de Cambio y las Ventas



*Fuente: Elaboración Propia

Al analizar la relación entre la Brecha del Tipo de Cambio y las Ventas en el gráfico 11, se observa un patrón interesante. Se evidencia que cuando la Brecha es mayor, es decir, cuando la diferencia entre los tipos de cambio es más amplia, las ventas tienden a incrementarse. Este hallazgo sugiere que los cambios en la Brecha del Tipo de Cambio pueden afectar positivamente las ventas de la empresa de comercio electrónico. Además, se destaca que existe un rezago temporal en el comportamiento de las ventas con respecto a la brecha del Tipo de Cambio. Cuando se observa un pico en la Brecha en una determinada semana, se puede anticipar que la semana siguiente o en un plazo cercano, las ventas también alcanzarán un pico. Esta relación rezagada puede ser explicada por el tiempo que toma para que las variaciones en el Tipo de Cambio se reflejen en las decisiones de compra de los consumidores, así como en las estrategias de la empresa para ajustar sus precios y promociones. Un ejemplo de esto en el gráfico es en la semana 50 del 2021 donde vemos que el siguiente comportamiento del conv value luego de un salto en la brecha, es subir y tener un pico.

La Brecha del Tipo de Cambio puede influir en el comportamiento de los consumidores y en las decisiones de compra, y entender esta relación puede ser fundamental para la toma de decisiones estratégicas en el negocio. La identificación de patrones y la comprensión de cómo los cambios en la Brecha del Tipo de Cambio afectan las ventas pueden proporcionar a la empresa una ventaja competitiva al adaptar su estrategia de precios, promociones y gestión de inversión en canales de marketing, como es nuestro caso.

3.4 Preprocesamiento de las Variables

En esta sección, se abordarán los pasos del preprocesamiento de las variables utilizadas en el estudio. Para lograrlo, se realizaron tres importantes procesos: ajuste por inflación de las variables nominales, transformación de variables no estacionarias y abordaje de la multicolinealidad. A continuación, se detallarán los métodos y las técnicas específicas utilizadas en cada proceso, junto con los resultados obtenidos y su relevancia para el desarrollo del estudio

3.4.1 Ajuste por Inflación de las Variables Nominales

En el caso particular de las variables expresadas en pesos argentinos, llevamos a cabo un proceso de corrección por inflación de estas series con el fin de obtener una mejor visión de los datos subyacentes. En el caso de Argentina, la inflación ha sido históricamente un desafío constante y significativo, lo que implica que los valores observados en pesos argentinos pueden estar distorsionados por el efecto de la inflación (en nuestro caso, la inversión y las ventas que están expresadas en pesos argentinos).²²

El ajuste por inflación de las series de tiempo es un procedimiento estadístico que tiene como objetivo principal eliminar o ajustar el efecto de la inflación en los datos, de manera que sea posible analizar y comparar variables a lo largo del tiempo. Este proceso implica tomar en consideración el Índice de Precios al Consumidor (IPC) y otros indicadores de inflación para realizar una corrección de los valores nominales, transformándolos en términos reales o constantes. Para realizar nuestra corrección por inflación, utilizaremos los datos mensuales del INDEC.²³

3.4.1.a Metodología utilizada

Para llevar a cabo el proceso de ajuste por inflación de las series de tiempo en pesos argentinos, nos basamos en el siguiente enfoque. En primer lugar, obtuvimos el Índice de Precios al Consumidor (IPC) del INDEC. Luego, para cada mes en el conjunto de datos, tomamos el valor nominal de la variable y lo dividimos por el valor del IPC correspondiente a ese mes para luego multiplicarlo por 100. Por ejemplo, si en Julio de 2021 las ventas fueron de \$1.099.554 y el IPC para ese mes fue de 175,6, el cálculo sería: $1.099.554 / 175,6 * 100 = \626.306 .²⁴ Este resultado representa el valor de las ventas del periodo expresadas en pesos del periodo base del índice de precios. Realizamos este cálculo para cada mes en el conjunto de datos, lo que nos permite obtener los valores ajustados por inflación para cada punto de la serie de tiempo. La fórmula que aplicamos para los costos y las ventas es la siguiente:

$$\text{Inflation adjusted value} = \frac{\text{Actual Value}}{\text{Index Value}} * 100$$

²² Vitelli (1990)

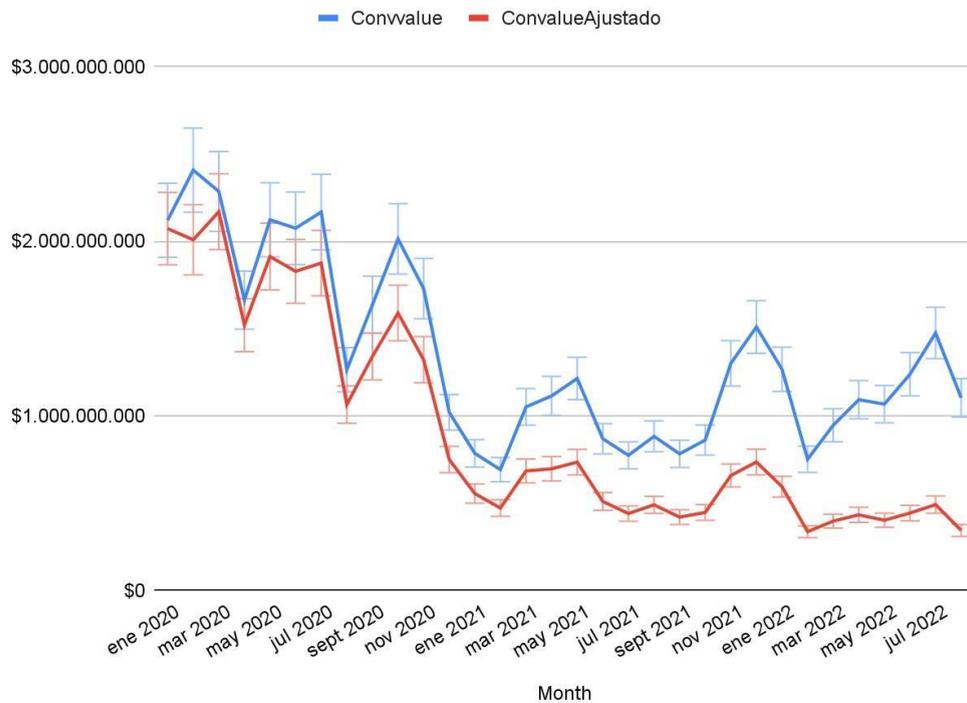
²³ Información Extraída de <https://www.indec.gov.ar/>

²⁴ Información Extraída de

<https://towardsdatascience.com/the-what-and-why-of-inflation-adjustment-5eedb496e080>

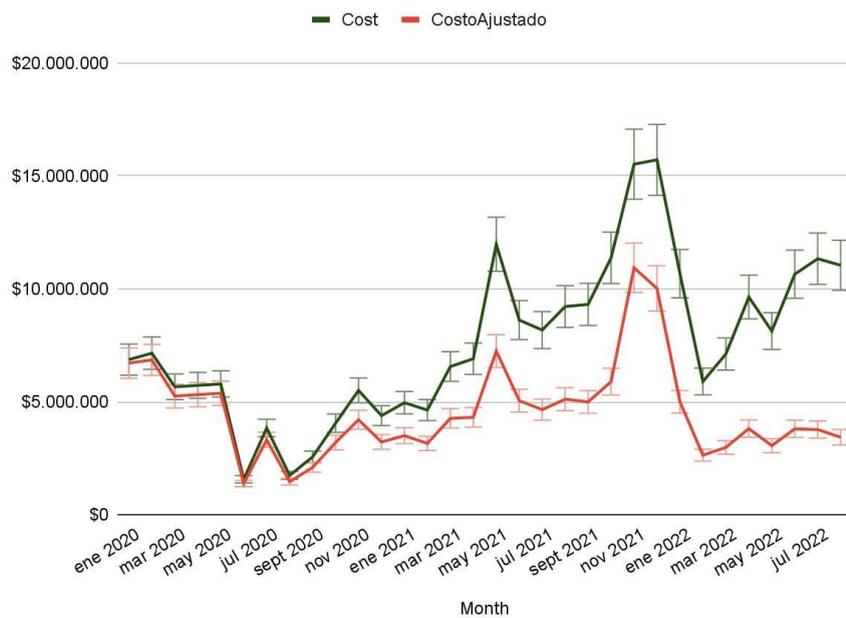
Con el propósito de ilustrar el impacto de la inflación en nuestras series de tiempo, se procede a representar gráficamente tanto las variables ajustadas como las variables no ajustadas en los gráficos 12 y 13:

Gráfico 12: Ajuste por Inflación en la variable Convalue



*Fuente: Elaboración Propia

Gráfico 13: Ajuste por Inflación en la variable Cost



*Fuente: Elaboración Propia

3.4.2 Estacionariedad de las Series de Tiempo

Antes de construir modelos de predicción, es necesario evaluar la estacionariedad de las series de tiempo. El test Dickey-Fuller Aumentado (ADF) se utiliza para determinar si una serie de tiempo es estacionaria o no, lo que proporciona información para la selección de los modelos adecuados.

La estacionariedad débil implica que las propiedades estadísticas de la serie de tiempo, como la media, la varianza y las autocorrelaciones se mantienen constantes en el tiempo. Esta característica es importante para poder realizar pronósticos, ya que permite asumir que las relaciones y patrones observados en el pasado se mantendrán en el futuro. Por lo tanto, la estacionariedad de las series de tiempo es una condición necesaria para evitar regresiones espúreas.²⁵

El test de Dickey-Fuller proporciona una herramienta estadística para evaluar la estacionariedad de una serie de tiempo. Los resultados se presentaron en la Tabla 4, donde se muestran los valores del estadístico ADF y el valor p para cada variable.

Tabla 4 : Resultados Test ADF

Variable	ADF Statistic	p-value
ConvalueAjustado	-6.12764	8.55251e-08
Clicks	-7.64506	1.85203e-11
CostoAjustado	-7.74433	1.04169e-11
Brecha	-1.93177	0.317268
AvgCPC	-6.47933	1.30692e-08
Week_Num	-3.61757	0.00543395
Month_Num	-3.53154	0.00721312
Impr	-7.15415	3.08583e-10
Adcost	-9.63197	1.60326e-16

**Fuente: Elaboración Propia*

Para determinar la estacionariedad de las variables, se comparan los valores del estadístico ADF con el valor crítico correspondiente al nivel de significancia del 5%. En este caso, el valor crítico utilizado fue de -2.861672609732128. Si el valor del estadístico ADF era menor que el valor crítico, se rechaza la hipótesis nula de raíces unitarias en favor de la alternativa de que la serie temporal era estacionaria. Además, se consideró el valor p asociado a cada variable. Un valor p bajo indica una fuerte evidencia en contra de la hipótesis nula, lo que sugiere la estacionariedad de la serie temporal. En este estudio, se utilizó un nivel de significancia de 0.05, lo que implica que un valor p menor que 0.05 se consideró como una fuerte evidencia en contra de la hipótesis nula.

²⁵ Ocerin (2012)

En base a los resultados presentados en la Tabla 4, se encontró que las variables 'ConvalueAjustado', 'Clicks', 'CostoAjustado', 'AvgCPC', 'Week_Num', 'Month_Num', 'Impr' y 'Adcost' presentaron valores del estadístico ADF por debajo del valor crítico y valores p extremadamente bajos. Esto indica que estas variables son estacionarias y no tienen raíces unitarias. Por otro lado, la variable 'Brecha' mostró un valor del estadístico ADF por encima del valor crítico y un valor p relativamente alto, lo que sugiere la presencia de raíces unitarias y la no estacionariedad de la serie temporal.

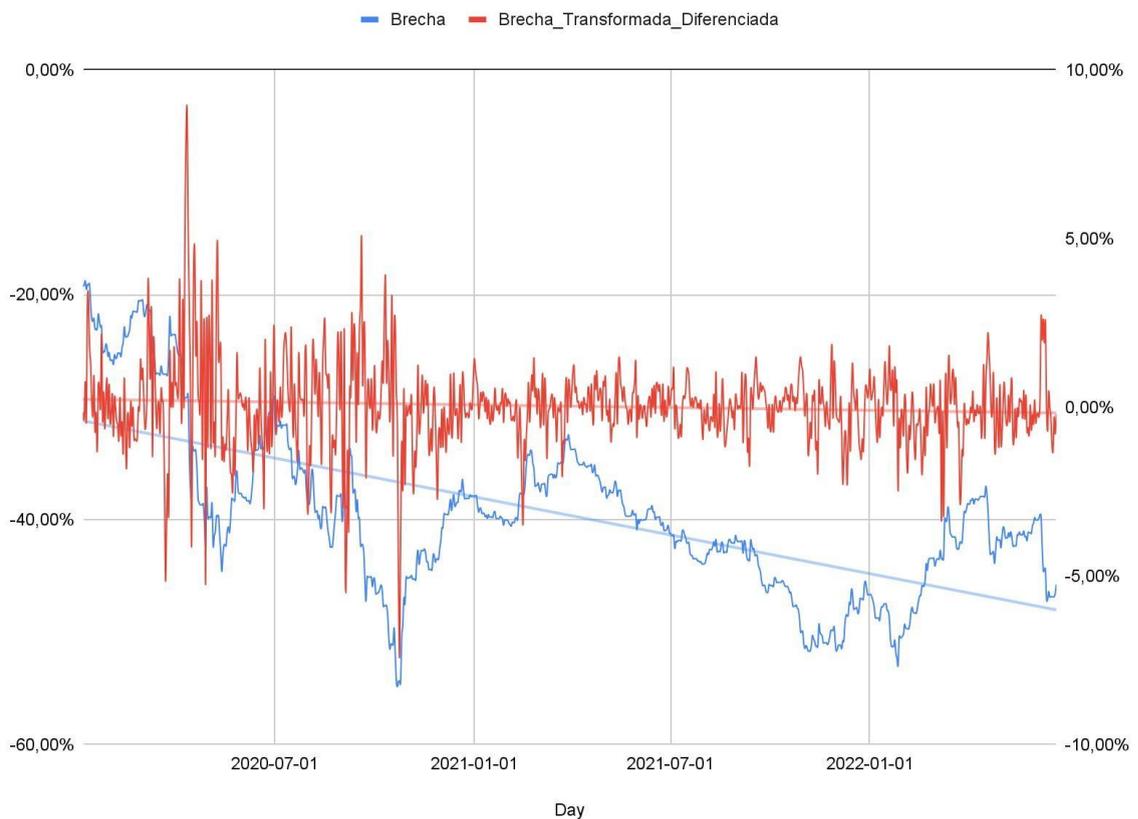
Cuando una serie de tiempo no es estacionaria, existen varias técnicas para transformarla y lograr la estacionariedad. Algunas opciones comunes incluyen la diferenciación de la serie para estabilizar en media la serie, o la aplicación de transformaciones logarítmicas o exponenciales para estabilizar la varianza. En este caso elegimos la diferenciación.

3.4.2.a Diferenciación Para Variable no Estacionarias

Se realizó un proceso de transformación en la variable "Brecha" con el objetivo de lograr la estacionariedad necesaria para aplicar técnicas de análisis de series de tiempo. La variable "Brecha" originalmente presentaba valores nulos debido a que los fines de semana no hay cotización oficial y no cumplía con los requisitos de estacionariedad, lo que dificulta su modelado y predicción. Para abordar este desafío, se aplicó una transformación específica a la variable "Brecha". En primer lugar, se reemplazaron todos los valores nulos por el valor del último día hábil previo a la fecha. De esta forma, el valor que representa la variable es la diferencia nominal del tipo de cambio oficial contra el tipo de cambio valor CCL sin valores nulos.

Posteriormente, se procedió a realizar la diferenciación en la variable. La diferenciación consistió en calcular la diferencia entre los valores sucesivos de la serie. Al aplicar esta técnica, se obtuvo una nueva serie de tiempo, "Brecha_transformada_diferenciada", que mostró estacionariedad en comparación con la serie original como podemos ver en el gráfico 14:

Gráfico 14: Variable Brecha Original vs Variable Brecha Transformada y Diferenciada



*Fuente: Elaboración Propia

3.4.3 Multicolinealidad entre Variables

En el desarrollo de nuestro análisis, es importante evaluar la presencia de multicolinealidad entre las variables independientes incluidas en el modelo. La multicolinealidad se refiere a la existencia de altas correlaciones entre las variables o entre combinaciones lineales de variables, lo cual puede tener un impacto en la interpretación de los coeficientes de regresión y en la capacidad predictiva de nuestro modelo.

La alta correlación entre variables puede dificultar la interpretación de los efectos individuales y la confiabilidad de los coeficientes de regresión. Además, la presencia de multicolinealidad puede afectar la estabilidad del modelo y provocar cambios drásticos en los resultados ante pequeñas variaciones en los datos. Esto se debe a la introducción de redundancias en la información y la amplificación del impacto de las variaciones en los datos.

Para abordar la multicolinealidad, se utilizará el método del Factor de Inflación de la Varianza (VIF, por sus siglas en inglés). El VIF proporciona una medida cuantitativa de la multicolinealidad entre las variables independientes en un modelo de regresión. Valores de VIF superiores a ciertos umbrales establecidos indican la presencia de multicolinealidad significativa (en este caso establecemos como

umbral el valor de VIF mayor a 5). El método VIF permite identificar las variables que contribuyen a la multicolinealidad y, en consecuencia, tomar decisiones informadas para abordar este problema.²⁶

3.4.3.a Resultados VIF

Se aplicó el cálculo del VIF a las variables predictoras en nuestro conjunto de datos, que incluyeron las siguientes variables: 'ConvalueAjustado', 'Clicks', 'CostoAjustado', 'Brecha_Diff', 'AvgCPC', 'Impr', 'Adcost' y 'Intercept'. Cada una de estas variables fue evaluada individualmente para determinar su grado de multicolinealidad con las demás variables (o con combinaciones de estas). Los resultados revelaron que algunas variables presentan multicolinealidad, mientras que otras mostraron una correlación moderada o no presentaron multicolinealidad.

Tabla 5 : Resultados VIF

Variable	VIF
Impr	19.158212
Clicks	16.153067
Intercept	4.964694
CostoAjustado	4.740608
ConvalueAjustado	2.245334
AvgCPC	1.986677
Adcost	1.481638
Brecha Diff	1.002175

*Fuente: Elaboración Propia

Las variables "Clicks" e "Impr" exhiben multicolinealidad, indicado por sus altos valores de VIF (16.15 y 19.15, respectivamente). Esto sugiere que estas variables están altamente correlacionadas con otras variables en el modelo, lo que puede dificultar su interpretación individual y afectar la capacidad predictiva del modelo.

Las variables "ConvalueAjustado", "CostoAjustado" y "AvgCPC" muestran una multicolinealidad leve, con valores de VIF de 2.24, 4.74 y 2.2, respectivamente. Estas variables exhiben cierta correlación con otras variables en el modelo, aunque no se considera problemática en gran medida. Es importante tener en cuenta esta correlación al interpretar los efectos individuales de estas variables y al evaluar la capacidad predictiva del modelo.

Las variables "Brecha_Diff" y "Adcost" muestran una baja correlación con otras variables en el modelo, como se refleja en sus valores de VIF de 1.0 y 1.48, respectivamente. Esto indica que no hay una multicolinealidad apreciable en estas variables, lo que facilita su interpretación individual y minimiza la distorsión en las estimaciones de los coeficientes. Estas variables pueden considerarse independientes y no requieren un tratamiento especial debido a la multicolinealidad.

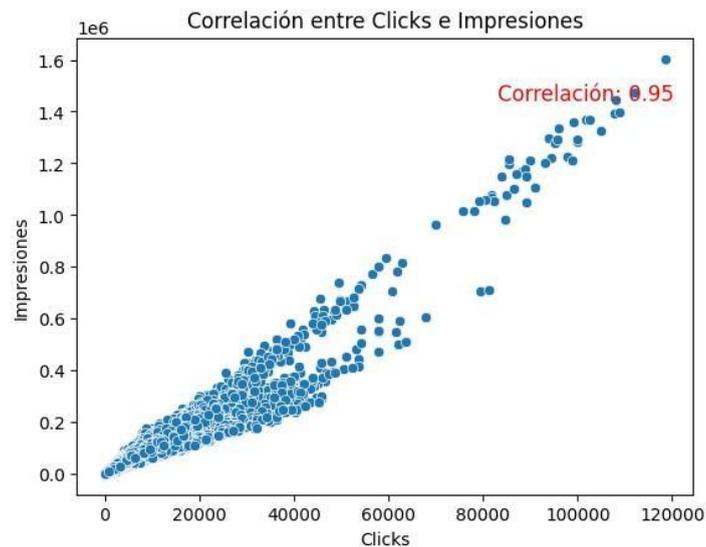
3.4.3.a Abordaje Multicolinealidad

A raíz de los resultados de VIF, se decidió analizar la correlación entre las variables de Clicks e Impresiones con el objetivo de poder determinar si es necesario tener ambas dentro de nuestra

²⁶ Merino y Díaz (2005)

predicción. Los resultados arrojaron una alta correlación entre sí como puede verse en el gráfico 15, lo que indica que ambas métricas están relacionadas y capturan información similar. La presencia de alta multicolinealidad de estas dos variables y su alta correlación, puede generar problemas en el modelo, ya que distorsiona las estimaciones de los coeficientes y dificulta la interpretación individual de cada variable. Por lo tanto, se tomó la decisión de eliminar la variable "Impr" del modelo de predicción.

Gráfico 15: Correlación Clicks e Impresiones



*Fuente: Elaboración Propia

Después de la eliminación de una de las variables problemáticas en términos de multicolinealidad, se realizó nuevamente el cálculo del VIF para las variables restantes. Los resultados mostraron una mejora en los valores de VIF. Ahora, los nuevos valores de VIF para las variables son los que se pueden observar en la tabla 6:

Tabla 6 : Resultados VIF sin Impresiones

	Variable	VIF
6	Intercept	4.847694
1	Clicks	3.793392
2	CostoAjustado	3.393008
0	ConvalueAjustado	2.084744
4	AvgCPC	1.983119
5	Adcost	1.481497
3	Brecha Diff	1.001960

*Fuente: Elaboración Propia

Estos nuevos valores de VIF indican que la multicolinealidad en el modelo se ha reducido. Un valor de VIF menor a 5 generalmente se considera aceptable, lo que sugiere que las variables restantes no están altamente correlacionadas entre sí. Estos resultados son alentadores y respaldan la idea de que el modelo de regresión ajustado es más confiable para abordar la predicción de ventas.

4. Predicción de Ventas y Optimización de Presupuesto

En esta sección se busca utilizar el modelo estimado de conversiones para encontrar la asignación óptima de inversión entre las campañas para cada día de la semana. El output de esta sección es la asignación de inversión óptima de lunes a domingo entre las distintas categorías. Primero se estiman las ventas con el modelo SUR y luego se plantea un modelo de asignación para el cual definir cuánto invertir según el límite de presupuesto.

4.1. SUR

En esta sección, se presentará el enfoque que se utilizó para abordar nuestro problema puntual de predicción de ventas, empleando el método de SUR. Además, se describirán los factores que se tuvieron en cuenta durante el proceso y los resultados obtenidos como resultado de esta metodología.

4.1.2 Variables del Modelo de Predicción

A continuación, se detalla brevemente la interpretación de la relación de cada feature con el target (ConvValue Ajustado).

- **`Campaign`**: Hace referencia a la categoría de la campaña publicitaria en la que se generan las conversiones. Esta, puede afectar el valor de las conversiones, por lo que es importante considerarla como variable en la predicción. En la sección 8.3 del Anexo se especifica el significado de cada campaña. Crearemos una ecuación de regresión por categoría para estimar las ventas endógenas de cada campaña
- **`Dayoftheweek`**: El día de la semana en que se generan las conversiones puede ser un factor importante, como vimos en la sección 2.3.3 b, ya que es posible que ciertos días tengan mayores ventas en comparación con los días de semana.
- **`Clicks`**: El número de clics en un anuncio parece estar relacionado con la probabilidad de conversión, por lo que esta variable nos es útil para la predicción. Tiene sentido que los clicks estén relacionados con las ventas en el sitio ya que al mostrarse el anuncio, la intención de compra se hace más evidente si el usuario efectivamente hace click.
- **`CostoAjustado`**: El costo de una campaña publicitaria influye en las ventas generadas ya que impacta directamente en la visualización de los anuncios. Al invertir más dinero en una campaña cierto día, nuestros anuncios se visualizarán más aún por lo que la exposición al usuario será mayor y la probabilidad de compra crece. La variable, como vimos en la sección 2.4.1, está ajustada por inflación.
- **`BrechaDiff`**: Como mencionamos antes, la brecha del tipo de cambio parece ser importante para predecir el comportamiento del consumidor y su especulación sobre los precios por lo cual es sumada a la predicción de las conversiones. Como vimos en la sección 2.4.2, es una serie diferenciada por problemas con la estacionariedad.

- **`AvgCPC`**: El costo promedio por clic en un anuncio puede estar relacionado con la probabilidad de conversión es por eso que se tiene en cuenta para la predicción. No solo depende de cuánto dinero estamos invirtiendo como empresa en cierta categoría sino que también nos da información sobre el mercado. Esta variable puede impactar en la conversión ya que a clicks más caros menos anuncios son mostrados a un nivel de inversión fija, por lo que podemos estar mostrando menos anuncios a un CPC más caro y estar disminuyendo nuestra probabilidad de venta. Cabe aclarar que esta métrica está puntualmente armada en base al mercado en el que está la empresa y es mayormente controlada por Google Ads por lo que se hace difícil predecir pero nos da una noción de mercado para nuestra predicción de ventas.

- **`Week Num`** y **`Month Num`**: El número de la semana y del mes en que se generan las conversiones también parecen ser factores importantes a considerar, por ejemplo, puede haber patrones de conversión en función del momento del mes o de la semana del año.

-**Adcost**: Es la relación Costo/Ventas que nos indica la eficiencia de las campañas. Esta métrica, que representa el costo de publicidad en relación con las ventas generadas, nos proporciona información valiosa sobre la eficiencia de nuestras campañas publicitarias. Es una métrica promedio que nos brinda la plataforma, es decir, no es el resultado exacto de hacer costo/ventas diarios sino que es tomado en base a data pasada de la plataforma (en general promedio de semanas anteriores). Al ser la medida de éxito del proyecto y la métrica que la empresa da más relevancia al momento de tomar decisiones, se hace necesario justificar la inclusión de esta variable en nuestro modelo. Incluir el Adcost en nuestro modelo de predicción de ventas nos brinda una visión más actualizada de la eficiencia de nuestras campañas.

Esto último, nos permite tomar decisiones informadas sobre la asignación de presupuesto por campaña y día de la semana, maximizando nuestros esfuerzos y recursos en aquellas campañas que históricamente han mostrado mayor eficiencia en términos de costos y ventas. Además, nos proporciona una base sólida para realizar análisis comparativos y evaluaciones continuas de nuestras estrategias publicitarias.

4.1.2.a Creación de Variables Dummy utilizando One-Hot Encoding

En la predicción de datos, es común encontrarnos con variables categóricas que no pueden ser directamente utilizadas en modelos de regresión. Una técnica ampliamente utilizada para convertir estas variables categóricas en representaciones numéricas es el one-hot encoding. El objetivo de esta sección es definir cómo creamos y utilizamos las variables dummy generadas mediante one-hot encoding en el modelo para mejorar nuestras predicciones de ventas en el contexto de nuestra ecommerce.

El one-hot encoding es una técnica que convierte variables categóricas en una serie de variables binarias (0 o 1), donde cada variable representa una categoría única de la variable original. Esta técnica nos permite utilizar información valiosa de variables categóricas en nuestros modelos de regresión.²⁷ Estas variables dummy capturan la influencia de los días de la semana para las ventas de

²⁷ Bird et al. (2009)

cada categoría y nos permiten una mejor comprensión de los factores que influyen en el comportamiento de los usuarios.

Las variables dummy generadas en nuestro estudio, fueron para el día de la semana con el objetivo de analizar cómo influye cada día en las conversiones. Observamos que estas variables dummy nos podrían brindar información sobre las preferencias y comportamientos de los usuarios en diferentes días de la semana. Por ejemplo, podemos determinar si los usuarios tienden a realizar más conversiones los fines de semana o si existen variaciones relevantes en los patrones de compra según el día de la semana.

4.1.3 Entrenamiento del Modelo

Se presenta el proceso detallado de construcción del modelo SUR para analizar y predecir las ventas de cada categoría en particular en base a múltiples variables explicativas. Al emplear el modelo SUR, podremos identificar y cuantificar la influencia que cada variable tiene sobre las ventas de cada categoría, teniendo en cuenta cómo se relacionan entre sí sus errores.

Para proporcionar una visión general de la base de datos utilizada en este estudio, se presentará un ejemplo con las primeras cinco observaciones. Esto nos permitirá familiarizarnos con la estructura de los datos y comprender cómo se presentan las variables explicativas y la variable dependiente (ventas).

Tabla 7 : Vista de Base de Datos para SUR

	Campaign	Clicks	CostoAjustado	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Brecha Diff	AvgCPC	Week_Num	Month_Num	Adcost
18662	ARTE	597.0	198.553	1	0	0	0	0	0	0	0.140	0.34	2.0	1.0	0.11
18666	ARTE	548.0	176.022	0	1	0	0	0	0	0	0.330	0.33	2.0	1.0	0.31
18669	ARTE	601.0	216.647	0	0	1	0	0	0	0	-0.730	0.37	2.0	1.0	0.05
18670	ARTE	641.0	238.280	0	0	0	1	0	0	0	0.400	0.38	2.0	1.0	0.14
18672	ARTE	444.0	166.540	0	0	0	0	1	0	0	-3.060	0.38	2.0	1.0	0.19

**Fuente: Elaboración Propia*

Luego de definir la base en nuestro modelo, dividimos los datos en variables dependientes (y) e independientes (X) que se utilizarán para entrenar el modelo SUR. Las fórmulas de regresión de nuestro modelo SUR tiene el siguiente aspecto:

$$y_{ti} = \sum_{j=1}^{k_i} X_{tij} \beta_{ij} + \varepsilon_{ti} \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, M; \quad j = 1, 2, \dots, K_i$$

Donde y hace referencia a la variable dependiente, X a las variables independientes, β a los coeficientes asociados con X y por último ε que hace referencia al valor del componente de error aleatorio. y_{ti} hace referencia a la t^{th} observación en la i^{th} variable dependiente que será explicada por la i^{th} ecuación de la regresión. X_{tij} es la t^{th} observación de la j^{th} variable explicatoria en la i^{th}

ecuación. β_{ij} es el coeficiente asociado con X_{tij} en cada observación y ε_{ti} es el t^{th} valor del error aleatorio asociado con la ecuación i^{th} del modelo.²⁸

Para nuestro caso, T representa la cantidad de categorías (23) y K representaría la cantidad de variables explicativas en el modelo (Clicks + CostoAjustado + 7 días de la semana dummies + Brecha + AvgCPC + Week_Num + Month_Num + Adcost).

Por lo tanto la fórmula de regresión para el modelo SUR para las ventas de cada categoría quedaría de la siguiente manera:

$$\begin{aligned} ConvvalueAjustado_{ti} = & \beta_0 + Campaign * \beta_{1i} + Clicks_{ti} * \beta_{2i} + CostoAjustado_{ti} * \beta_{3i} \\ & + Lunes_{ti} * \beta_{4i} + Martes_{ti} * \beta_{5i} + Miercoles_{ti} * \beta_{6i} + Jueves_{ti} * \beta_{7i} + Viernes_{ti} * \beta_{8i} \\ & + Sabado_{ti} * \beta_{9i} + Domingo_{ti} * \beta_{10i} + BrechaDiff * \beta_{11i} + AvgCPC_{ti} * \beta_{12i} + WeekNum_{ti} * \beta_{13i} \\ & + MonthNum_{ti} * \beta_{14i} + Adcost_{ti} * \beta_{15i} + \varepsilon_{ti} \end{aligned}$$

La idea es utilizar las correlaciones entre los errores de las variables dependientes para mejorar las predicciones. Es por eso que, tomamos las variables seleccionadas y las organizamos en estos dos dataframes mencionados: **Y**, que contiene las variables dependientes y **X**, que contiene las variables independientes. De esta forma, se pueden utilizar estos dataframes para entrenar el modelo SUR. Es decir, ajustar múltiples regresiones para las variables dependientes de cada categoría, teniendo en cuenta la correlación entre los errores de las variables dependientes. Para tener un mejor panorama de cómo funcionan las distintas ecuaciones en nuestro modelo SUR proponemos la siguiente matriz:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & X_M \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{pmatrix}$$

*Fuente: Billio (2017)

En nuestro caso, cada Y serían las ventas de cada categoría y las X las variables explicativas mencionadas previamente. El objetivo es encontrar una combinación de las variables predictoras que explique de la mejor manera las variaciones en la variable objetivo Conv value Ajustado, teniendo en cuenta la relación entre las variables dependientes. Al ajustar múltiples regresiones, el modelo SUR busca mejorar la capacidad predictiva de las estimaciones de los coeficientes en comparación con la estimación de n modelos de regresión lineal simple por separado.

4.1.3.a Codificación de referencia completa para la interpretación directa de variables

Como mencionamos en la sección 4.1.2.a, al utilizar variables dummy del día de la semana y de las campañas en el modelo de predicción, estamos re-codificando información que nos permite capturar y cuantificar la influencia de estos factores en las conversiones de la empresa. Esto nos ayuda a, posiblemente, comprender mejor los patrones de comportamiento de los usuarios y a personalizar

²⁸ Billio (2017)

nuestras estrategias de marketing en función de estos insights. A su vez, para el modelo de optimización, nos parece importante distinguir el impacto individual de cada día de la semana y de cada campaña para que nuestra asignación de presupuesto sea óptima.

Para evitar que las variables dummy se comparan con una categoría de referencia fija en el intercepto, se puede utilizar un enfoque conocido como "codificación de referencia completa" o "codificación de efectos fijos"²⁹. Este enfoque implica crear variables dummy para cada categoría y eliminar la variable intercept, que es el enfoque se abordará en nuestro análisis. En lugar de tener una única variable intercepto que representa una categoría de referencia, se crean variables dummy para cada categoría. Esto permite interpretar directamente el coeficiente asociado a cada categoría, sin tener que compararlo con una categoría fija. Luego, la interpretación de los coeficientes de cada variable dummy será directa y se referirá al efecto de cada categoría.

4.1.3.b Ajuste del modelo

Tras aplicar el modelo SUR a nuestro conjunto de datos, se obtuvieron diversos indicadores que permiten evaluar la calidad y la capacidad explicativa del modelo. El coeficiente de determinación global, denominado Overall R-squared, reveló un valor de 0.7213. Esto indica que alrededor del 72.13% de la variabilidad de las variables endógenas puede ser explicada por las variables exógenas incluidas en el modelo. Este resultado sugiere una relación significativa entre las variables y respalda la idoneidad del enfoque SUR en nuestro análisis. En la tabla 8 podemos visualizar los resultados:

Tabla 8 : Resultados SUR (primera tabla)

System OLS Estimation Summary	
Overall R-squared:	0.7213
McElroy's R-squared:	0.7396
Judge's (OLS) R-squared:	0.7213
Berndt's R-squared:	0.8220
Dhrymes's R-squared:	0.6213

**Fuente: Elaboración Propia*

Al considerar medidas adicionales de ajuste, se observa que McElroy's R-squared y Dhrymes's R-squared presentan valores de 0.7396 y 0.6213, respectivamente. Estas medidas, que tienen en cuenta la covarianza entre las ecuaciones del modelo y la estimación de mínimos cuadrados ordinarios para cada ecuación individualmente, respaldan la validez y la robustez de nuestro enfoque metodológico. Asimismo, Judge's (OLS) R-squared, que proporciona una medida adicional del coeficiente de determinación basada en la estimación de mínimos cuadrados ordinarios para cada ecuación individual, arrojó un resultado de 0.7213. Este valor indica una buena capacidad del modelo para explicar la variabilidad de las variables endógenas de manera individual.

Además, Berndt's R-squared, que ajusta el R-cuadrado de McElroy por el número de parámetros estimados en el modelo, muestra un resultado de 0.8220. Este valor destaca la capacidad del modelo para explicar la variabilidad de las variables endógenas, considerando la complejidad del sistema y el número de variables incluidas en el análisis. La capacidad del modelo para explicar una buena parte

²⁹ Johnson y Neyman (1936)

de la variabilidad de las variables endógenas mediante las variables exógenas seleccionadas demuestra su utilidad en la comprensión del impacto de las campañas publicitarias en nuestra e-commerce.³⁰

En la Tabla 9 de la salida, se presentan los coeficientes estimados, los errores estándar, los valores t, los valores p y los intervalos de confianza del 95% para cada variable independiente incluida en el modelo. Al ser 23 ecuaciones diferentes, una para cada estimación de ventas por categoría, se propone a modo de ejemplo dos salidas de dos categorías diferentes que hacen referencia a las primeras 2 ecuaciones, la ecuación 1, que corresponde a la categoría de computación, y la ecuación 2, que corresponde a la categoría de hogar. Se ha utilizado un nivel de significancia de 0.05 (o 5%) para evaluar la significancia estadística de los coeficientes en las ecuaciones. Este nivel de significancia es comúnmente utilizado en análisis estadísticos para determinar si los coeficientes tienen un impacto significativo en las variables dependientes.

Tabla 9 : Coeficientes de la regresión SUR

Equation: equation1, Dependent Variable: ConvalueAjustado_COMP						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	205.27	4.2469	8.335	0.0000	196.95	213.60
CostoAjustado	14.86	3.7801	3.922	0.0001	0.41	22.35
Brecha Diff	-71.47	3.5424	1.580	0.0039	370.8	467.95
AvgCPC	-2.09	7.1014	2.954	0.0031	62.76	158.74
Week_Num	-4.79	7.2248	6.638	0.0000	2.74	28.40
Month_Num	11.93	3.1223	6.190	0.0000	2.94	11.35
Adcost	-22.87	1.6715	1.717	0.0018	22.78	45.67
Monday	23.12	3.7881	9.667	0.0000	1.25	2.35
Tuesday	11.03	3.8820	6.659	0.0000	9.58	20.11
Wednesday	15.02	3.9247	6.198	0.0000	4.44	11.45
Thursday	19.84	4.1044	4.000	0.0000	4.29	11.53
Friday	9.30	3.9643	3.479	0.0000	2.05	9.34
Saturday	6.55	3.7940	2.539	0.0000	7.80	9.29
Sunday	6.09	3.7215	1.750	0.0000	7.36	8.82

Equation: equation2, Dependent Variable: ConvalueAjustado_HOGAR						
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	147.39	4.3286	4.050	0.0000	138.91	155.88
CostoAjustado	6.10	3.6086	1.692	0.0006	0.9664	13.179
Brecha Diff	-53.76	4.0504	1.051	0.0054	139.44	2063.09
AvgCPC	-4.39	5.8913	2.454	0.0000	25.54	133.23
Week_Num	-5.58	6.9766	6.060	0.0000	6.95	24.21
Month_Num	12.20	3.0156	7.322	0.0000	1.61	9.79
Adcost	-11.91	1.3475	1.423	0.0047	24.55	77.22
Monday	17.23	3.9641	3.191	0.0000	1.15	5.31
Tuesday	5.17	4.0833	2.682	0.0000	1.09	4.25
Wednesday	2.18	4.1615	2.421	0.0000	6.10	11.26
Thursday	10.14	4.3027	6.676	0.0000	8.06	9.23
Friday	10.09	4.1696	6.354	0.0000	1.01	9.18
Saturday	-1.32	3.8214	2.004	0.0070	9.56	21.10
Sunday	9.52	3.8823	2.528	0.0000	8.76	10.02

³⁰ Fomby et al. (2012)

A nivel general, ambas ecuaciones muestran que todas las variables tienen un impacto significativo (al 0.05) en las ventas ajustadas de cada categoría. Ambas ecuaciones incluyen variables como "Clicks", "CostoAjustado" y "Month_Num" (número de mes), las cuales tienen coeficientes positivos y significativos al 0.05. Esto sugiere que un mayor número de clics, un mayor costo ajustado, semanas específicas del año y meses específicos están asociados con mayores ventas en ambas categorías.

Además, ambas ecuaciones incluyen variables dummy para cada día de la semana. En ambas ecuaciones, todas las variables que hacen referencia a cada día tienen coeficientes positivos y significativos (al 0.05), lo que indica que estos días de la semana tienen un impacto positivo en las ventas en ambas categorías. En el caso de la variable "Week_Num" (número de semana) notamos un coeficiente negativo lo que sugiere que mayor sea la semana del año, se tiene un impacto menor en la ventas.

Sin embargo, también hay algunas diferencias notables entre las dos ecuaciones. En la ecuación de la categoría de computación, la variable "Brecha Diff" (la brecha del tipo de cambio diferenciada) tiene un coeficiente negativo significativo al 0.05, lo que indica que una mayor brecha entre el tipo de cambio oficial y el CCL está asociada con menores ventas en esta categoría. Lo cual puede justificarse debido a que el consumidor especula basado en cuestiones de importaciones, que un impacto en la brecha cambiaria impacta más a las categorías de electrónica o tecnología. Por otro lado, en la ecuación de la categoría de hogar, la variable "Brecha Diff" también tiene un coeficiente negativo significativo en términos estadísticos al 0.05, pero de menor impacto en términos nominales al valor de conversión.

En cuanto a las variables "AvgCPC" (costo promedio por clic) y "Adcost" (costo de publicidad), ambas tienen coeficientes negativos significativos en ambas ecuaciones (con un alpha del 0.05). Esto sugiere que un menor costo promedio por clic y una mayor eficiencia (menor adcost) están asociados con mayores ventas tanto en la categoría de computación como en la categoría de hogar.

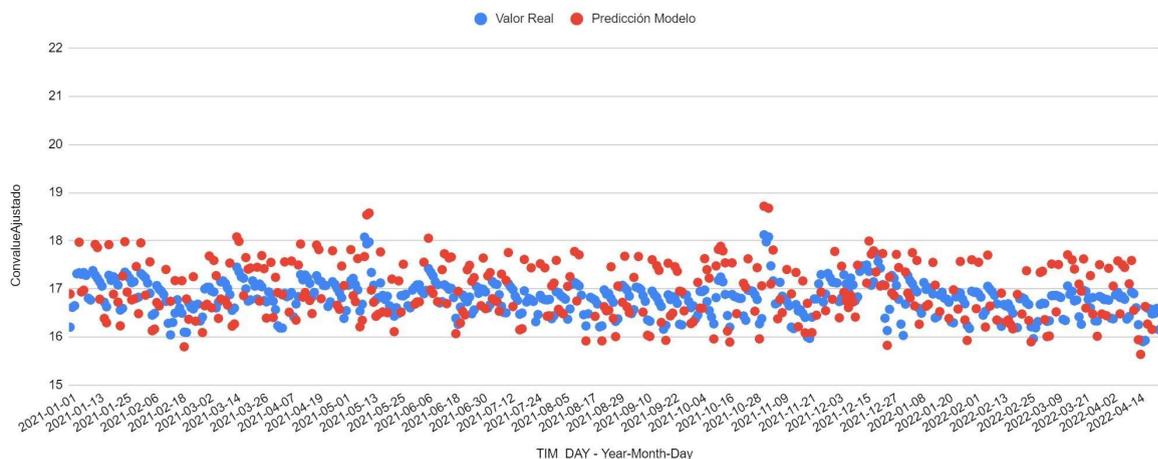
En cuanto a la significancia estadística, podemos considerar la variable "Adcost" (costo de publicidad) en la ecuación de la categoría de hogar. Esta variable muestra un valor de p-valor de 0.0047, lo cual indica una significancia estadística a un nivel de confianza del 95%. Dado que el p-valor es menor que el nivel de significancia establecido (0.05), podemos concluir que el coeficiente asociado con la variable "Adcost" es estadísticamente significativo en relación con las ventas ajustadas en la categoría de hogar. De manera similar, consideremos la variable "Brecha Diff" (diferencia de brecha) en la misma ecuación de la categoría de hogar. Esta variable muestra un valor de p-valor de 0.0054, también por debajo del nivel de significancia establecido (0.05). Esto indica que el coeficiente asociado con la variable "Brecha Diff" es estadísticamente significativo y tiene un impacto importante en las ventas ajustadas de la categoría de hogar.

4.1.4 Predicción

Se realiza una evaluación de la predicción del modelo SUR al comparar las predicciones generadas por el modelo (representadas en rojo) con los valores reales observados (representados en azul).

Este análisis nos permite comprender qué tan bien se ajusta el modelo SUR a los datos reales. En el Gráfico 16, podemos visualizar la proximidad entre las predicciones y los valores reales para los datos de validación. Al observar el gráfico, podemos notar el grado de ajuste del modelo SUR a los datos reales.

Gráfico 16: Ajuste SUR



*Fuente: Elaboración propia

La línea roja, que representa las predicciones, parece seguir de cerca la línea azul, que representa los valores reales. Esto sugiere que el modelo SUR es capaz de capturar ciertos patrones y tendencias presentes en los datos observados. Como mencionamos previamente, SUR tiene en cuenta la posible correlación de los errores entre las ventas de los distintos productos al estimar las variables simultáneamente. De esta manera, intentamos mejorar nuestro modelo de predicción al considerar las interacciones entre las variables y evitar la subestimación o sobreestimación de la demanda. Utilizamos SUR para predecir las ventas de la e-commerce con distintas categorías de productos porque este modelo tiene en cuenta la posible correlación entre las ventas de los distintos productos, lo cual puede mejorar las predicciones y evitar subestimaciones o sobreestimaciones de la demanda. Más adelante en la sección 5 de resultados, generamos una comparación con modelo más sencillo AR(1) como benchmark para la capacidad predictiva de este modelo.

Utilizando datos históricos y variables relevantes, aplicamos el modelo SUR para analizar la relación entre las ventas diarias de cada campaña y los presupuestos asignados. Consideramos el impacto de las diferentes campañas con el objetivo de comprender cómo los presupuestos afectan las ventas y maximizar la eficiencia en términos del ratio costo/ventas. Una vez que definamos el problema de optimización en ZIMPL, incorporaremos las estimaciones obtenidas del modelo SUR en la función objetivo correspondiente

4.1.4.a RECM

El resultado de nuestro modelo de predicción de ventas en pesos argentinos arrojó un valor de RECM (Root Mean Squared Error) de 1257. Este valor nos proporciona una medida de la precisión

promedio de las predicciones del modelo en relación con los valores reales de las ventas en la moneda local.³¹ Al analizar el significado de un RECM de 1257 en el contexto de nuestro estudio, es importante considerar la escala de los datos y la naturaleza del problema. Dado que nuestro modelo está en pesos argentinos, un RECM de 1257 indica que, en promedio, las predicciones del modelo difieren en 1257 pesos argentinos del valor real de las ventas.

La interpretación del valor del RECM puede variar dependiendo de diversos factores, como la magnitud de las ventas y las expectativas de precisión establecidas para el modelo. En general, un RECM más bajo indica una mayor precisión y una menor diferencia entre las predicciones y los valores reales.

4.2 Descripción Modelo de Asignación

El problema de asignación de presupuesto publicitario es un desafío importante en el ámbito empresarial, especialmente en el área de marketing y ventas. En este contexto, el objetivo es determinar la cantidad de recursos financieros que deben ser asignados a cada categoría de productos que ofrece la empresa, a fin de maximizar la efectividad de su campaña publicitaria.

Este problema puede ser abordado a través de la programación lineal, que permite la optimización de una función objetivo sujeta a un conjunto de restricciones. En el presente trabajo se utilizará este enfoque para formular el problema de asignación de presupuesto publicitario en una empresa con múltiples categorías de productos.

Se quiere desarrollar un plan de inversión semanal en las 23 categorías diferentes de nuestra e-commerce para cada día de la semana. 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 número de ventas posible con un gasto mínimo. Para lograr este objetivo, se plantea un modelo de optimización lineal donde se 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.

4.2.1 Función Objetivo

La función objetivo se puede escribir de la siguiente manera:

$$obj: \min \sum_{i=1}^{23} \sum_{d=1}^7 \frac{X_{id}}{C_{id}}$$

Donde:

- X_{id} representa la cantidad de pesos a invertir en la categoría i el día d

³¹ Andrade-Zeballos, M. A. (2022). Comparativa de Técnicas de Aprendizaje Automático para Análisis y Predicción de Sequía (Master's thesis)

- C_{id} representa la cantidad de conversiones en el día d para la categoría i

Esta última variable C_{id} de nuestro modelo, sale de la predicción de nuestra regresión SUR. Esta función objetivo busca minimizar el ratio total de inversión (X_{id}) respecto 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 día para la semana que queremos evaluar

Buscamos minimizar la relación entre el gasto realizado y el rendimiento obtenido en términos de conversiones. Este enfoque permite asignar recursos de manera estratégica y priorizar aquellas categorías que generan un mayor impacto con menor inversión los días de la semana que son más eficientes.

4.2.2 Restricciones

El objetivo principal de este problema de optimización es encontrar la asignación óptima que minimice el ratio total de X_{id}/C_{id} , es decir, minimizar la relación entre la inversión asignada y las conversiones generadas para cada categoría i y día d. Para lograr este objetivo, se han establecido una serie de restricciones:

1. **Restricción de inversión mínima semanal:** Se debe invertir un monto mínimo de \$70,000 en cada categoría i, asegurando que no se asignen cantidades nulas de presupuesto a categorías que tengan peor performance o que aporten menos a nuestra eficiencia.

$$\sum_{d=1}^7 X_{id} \geq 70000 \quad \forall i \in I(c1)$$

Esto también ayuda a diversificar el riesgo, ya que cada categoría tiene la oportunidad de recibir una asignación mínima de presupuesto. Además, al establecer un umbral mínimo, se evita asignar cantidades muy pequeñas de presupuesto que podrían no generar un impacto visible en la campaña publicitaria. A su vez, al establecer un umbral mínimo, se puede optimizar la utilización del presupuesto y evitar dispersar los recursos de manera fragmentada. Esto permite asignar montos relevantes a cada categoría y maximizar el impacto de la inversión en la generación de conversiones.

- Este monto puede ser modificado fácilmente a medida que la empresa lo requiera o la inflación lo indique.
- Se establece el monto de \$70,000 para que como mínimo, de la inversión máxima, represente un 0.4% de inversión cada categoría (que es lo observado como mínimo tiene una categoría de bajo rendimiento en la sección 3.4), es decir, para que en promedio se gasta \$10.000 por día.

2. **Restricción de inversión máxima diaria:** La sumatoria de la inversión asignada en cada categoría i para cada día d no puede superar los \$2,500,000, lo cual evita un desbalance excesivo en la distribución del presupuesto diario. Es decir, la suma total de lo invertido por día no puede superar este monto

$$\sum_{i=1}^{23} X_{id} \leq 2500000 \quad \forall d \in D \text{ (c2)}$$

Este monto viene ligado al presupuesto decretado por el sector de finanzas a nivel mensual. El cálculo está hecho en base a decisiones estratégicas propias de la empresa que proponen un monto final por mes, que, teniendo en cuenta nuestra data y nuestros montos, da como máximo este valor promedio por día.

- Establecer un límite de inversión máxima diaria nos permite tener un control preciso sobre el gasto en publicidad. Al fijar un tope en la cantidad de dinero que se puede invertir por día, se evita la posibilidad de exceder el presupuesto asignado y se mantiene una gestión financiera responsable. Esto es especialmente importante para nuestro escenario ya que se busca maximizar la eficiencia y evitar gastos innecesarios.
 - Esto también nos impulsa a una asignación de recursos inteligente. Al tener un límite establecido, se obliga a realizar una distribución estratégica del presupuesto, centrándonos en las categorías más importantes o en las campañas con mejor rendimiento. Esto permite maximizar el impacto de la inversión en cada día y obtener resultados más favorables. Esto significa que se asignará más presupuesto a aquellas campañas que estén generando mejores resultados y se limitará la inversión en aquellas que no estén cumpliendo con los objetivos esperados.
 - Este monto puede ser modificado fácilmente a medida que la empresa lo requiera o la inflación lo indique. En el caso de que, por cuestiones fuera del área de marketing, se quiera recortar presupuesto o asignar presupuesto, se puede modificar en nuestro problema de asignación.
3. **Restricción de inversión máxima en categorías menos importantes:** El gasto total en las categorías menos importantes (categorías con índices i del 13 al 23) no puede superar el 20% de la inversión total para cada día. Esto asegura que no se destine un porcentaje desproporcionado del presupuesto a estas categorías.

$$\sum_{i=13}^{23} X_{id} \leq 0.2 * X_{id} \sum_{i=1}^{23} \quad \forall d \in D \text{ (c3)}$$

La restricción de inversión máxima en categorías menos importantes contribuye a un enfoque estratégico en la asignación de recursos. Permite centrar la inversión en las áreas prioritarias, optimizar los recursos disponibles, brindar flexibilidad para adaptarse a los cambios del entorno y

promover una priorización estratégica. Se invierte una proporción menor en categorías menos relevantes, lo que impulsa una evaluación más rigurosa de su desempeño y rentabilidad. Esto promueve una mayor disciplina en la toma de decisiones y una gestión más eficiente del presupuesto para estas categorías que, como vimos en la sección 2.4, no parecen ser importantes para la empresa en general. De esta manera, nos alineamos a los objetivos de la empresa y no invertimos montos excesivos en categorías que a nivel gerencial no parecen ser relevantes.

- Aunque está restricción parece limitar la inversión en cierto tipo de categorías, está fuertemente ligada a la restricción de inversión mínima diaria ya que son conceptos relacionados pero que abordan aspectos diferentes en la gestión de inversiones. Aunque del presupuesto finalmente asignado no puedan representar más del 20% estas categorías, ya sabemos que como mínimo tienen un monto a invertir asignado en la restricción 2°.
- Planteamos el valor del 20% ya que es aproximadamente el valor sobre el costo total que se le asignan hoy en día a estas categorías como vimos en la sección 3.4.

4. **Restricción de inversión máxima en categorías importantes:** El gasto total en las categorías importantes (categorías con índices i del 1 al 12) debe ser como mucho el 80% de la inversión total para cada día. Esto garantiza que las categorías estratégicas reciban una asignación importante mínima de presupuesto.

$$\sum_{i=1}^{12} X_{id} \leq 0.8 * X_{id} \sum_{i=1}^{23} \quad \forall d \in D (c4)$$

Como mencionamos en la sección 3.4 las categorías más importantes son aquellas que generan un mayor ingreso a la empresa. Al imponer la restricción de inversión mínima en estas categorías, se garantiza que reciban una asignación mínima de presupuesto, lo cual refuerza la estrategia de priorización y enfoque en las áreas más rentables y estratégicas del negocio.

- Con esta restricción se asegura estar alineados con los objetivos y metas comerciales de la empresa. Esto incluye el impulso de ventas, el crecimiento de la participación de mercado y la consolidación de la presencia en las categorías más relevantes. La restricción garantiza que se destine un nivel mínimo del total de nuestros recursos para lograr estos objetivos estratégicos.
- Planteamos el valor del 80% ya que es aproximadamente el valor sobre el costo total que se le asignan hoy en día a estas categorías y para que no genere problema con la restricción del 20% que asignamos a las categorías menos importantes. Forzamos a que el porcentaje de costo que representan estas categorías sea el 80% como máximo.

5. **Restricción de no invertir cero:** Se prohíbe asignar cantidades nulas de presupuesto en todas las categorías i y días d .

$$X_{id} > 0 \quad \forall i \in I \text{ y } d \in D \text{ (c5)}$$

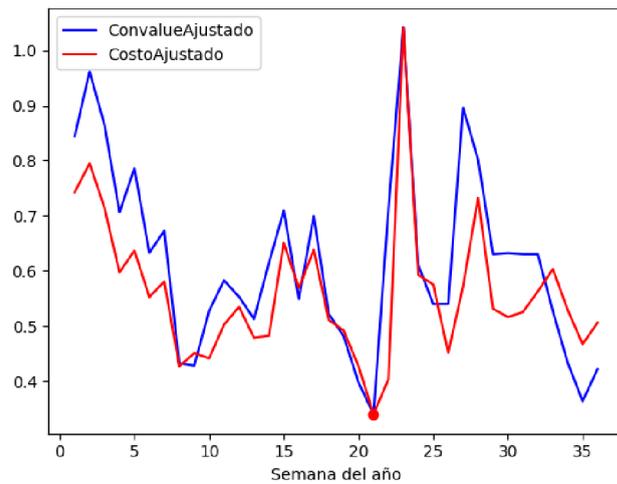
- Con esta restricción nos aseguramos una utilización racional de los recursos, maximizando las oportunidades de inversión, promoviendo la equidad en la asignación y facilitando la evaluación adecuada del rendimiento. Es una medida importante para garantizar que todos los recursos asignados se utilicen de manera óptima y contribuyan al éxito global del negocio.
- Al no permitir inversiones nulas, se facilita la evaluación y el análisis del rendimiento de cada categoría y día. Si se permitieran asignaciones de cero, sería difícil distinguir entre las categorías y días que realmente no requieren presupuesto y aquellos que podrían haberse beneficiado de una inversión. Al mantener todas las asignaciones mayores a cero, se obtiene una visión clara y precisa del rendimiento de cada área y se facilita la toma de decisiones basada en datos.
- Con esta restricción evitamos situaciones en las que una categoría o día reciba una asignación nula mientras que otras reciben recursos excesivos. Esto contribuye a mantener un equilibrio y una distribución justa de los recursos, evitando posibles sesgos o desequilibrios en la estrategia de inversión.

6. **Restricción de conversión:** Para generar una conversión mínima de \$1, se debe invertir al menos \$0.4 en cada categoría *i* y día *d*. Esto asegura una relación adecuada entre la inversión y las conversiones.

$$C_{id} * 0.4 \leq X_{id} \quad \forall i \in I \text{ y } d \in D \text{ (c6)}$$

- Esta restricción de conversión nos garantiza una relación adecuada entre la inversión y las conversiones, promueve la eficiencia y el retorno de la inversión, optimiza el presupuesto y ayuda a controlar el gasto. Es una medida importante para asegurar que los recursos asignados se utilicen de manera estratégica y generen resultados que contribuyan al éxito del negocio.
- Relación inversión-conversión: La restricción garantiza una relación adecuada entre la inversión realizada (*X*) y las conversiones generadas (*C*). Establece que para generar una conversión mínima de \$1, se debe invertir al menos \$0.4 en cada categoría y día correspondiente. Esto ayuda a asegurar que la inversión esté alineada con los objetivos de conversión y que se asignen recursos suficientes para obtener resultados notables. Al establecer un umbral mínimo de inversión para lograr una conversión mínima, se busca maximizar el retorno de la inversión al asignar recursos a áreas estratégicas y de alto rendimiento.
- En definitiva esta restricción nos limita a que para convertir al menos \$1 se debe invertir al menos \$0.4 con el objetivo de que la optimización no asigne un presupuesto muy bajo pensando que sin invertir no se convierte. Esto lo podemos ver de nuestra data donde:

Gráfico 17: Relación Ventas con Inversión



*Fuente: Elaboración propia

El valor de \$0.04 nos da un indicio de que se requiere una inversión mínima relativamente baja para obtener una conversión de \$1. Esto podría interpretarse como una señal positiva de eficiencia en la asignación de recursos, ya que se logra un retorno apreciable con una inversión modesta. Consideramos esta cifra como el umbral mínimo de inversión para que una campaña sea rentable. Si se invierte menos de esa cantidad, es probable que no se alcance el objetivo mínimo de generación de ingresos y estemos siendo lo suficientemente óptimos. Al identificar que una inversión de \$0.04 es necesaria para obtener \$1 en conversiones, se puede evaluar la viabilidad y escalabilidad del negocio. Si se desea generar mayores ingresos, se requerirá una inversión proporcionalmente mayor.

Además de estas restricciones, se establecen límites (bounds) en las variables de decisión X_{id} , donde se establece que la inversión asignada en cada categoría i y día d debe ser mayor o igual a cero y no debe exceder los \$17,500,000.

$$0 \leq X_{id} \leq 17500000 \quad \forall i \in I \text{ y } d \in D$$

En conjunto, estas restricciones y límites buscan optimizar la asignación de presupuesto en función de las características específicas de cada categoría y día de la semana de la empresa. Esto quizá nos permita maximizar el retorno de inversión y alcanzar los objetivos comerciales de la e-commerce.

4.2.3 Modelo de Optimización Final

Conjuntos:

$$I = \{1, \dots, 23\} \quad (1)$$

$$D = \{1, \dots, 7\} \quad (2)$$

(1) Cantidad de campañas para distribuir inversión (23)

(2) Cantidad de días para distribuir la inversión y optimizar la eficiencia (Lunes a Domingo)

Parámetros:

C_{id} : Conversiones en Pesos en el día d para la categoría i - Ventas predichas por el Modelo SUR

Variables:

X_{id} : Cantidad de pesos a invertir en la categoría i el día d

Función Objetivo

$$\min \sum_{i=1}^{23} \sum_{d=1}^7 \frac{X_{id}}{C_{id}}$$

s.a.

$$\sum_{d=1}^7 X_{id} \geq 70000 \quad \forall i \in I (c1)$$

$$\sum_{i=1}^{23} X_{id} \leq 2500000 \quad \forall d \in D (c2)$$

$$\sum_{i=13}^{23} X_{id} \leq 0.2 * X_{id} \sum_{i=1}^{23} \quad \forall d \in D (c3)$$

$$\sum_{i=1}^{12} X_{id} \leq 0.8 * X_{id} \sum_{i=1}^{23} \quad \forall d \in D (c4)$$

$$X_{id} > 0 \quad \forall i \in I \text{ y } d \in D (c5)$$

$$C_{id} * 0.4 \leq X_{id} \quad \forall i \in I \text{ y } d \in D (c6)$$

$$0 \leq X_{id} \leq 17500000 \quad \forall i \in I \text{ y } d \in D$$

4.2.3.a Consideraciones del Modelo de Optimización de Presupuesto

En el desarrollo del modelo de optimización de presupuesto para nuestra empresa de e-commerce, se deben tener en cuenta algunas consideraciones importantes. En primer lugar, es necesario destacar que el enfoque se realiza a nivel semanal, dado que el presupuesto de la empresa está estructurado de esta manera. Esto implica que al inicio de cada semana, particularmente los días lunes, se debe definir la cantidad de inversión que se destinará para el resto de los días de la semana. Esto es debido a cómo se maneja la estructura de hoy en día. La idea de este modelo es poder

acelerar la asignación de presupuesto a principio de la semana teniendo en cuenta la data pasada y las posibles conversiones que se puede tener en la semana.

Es relevante mencionar que, si bien existen métodos más sofisticados que pueden incorporar predicciones de clicks u otros indicadores, en el presente modelo se ha optado por una aproximación más sencilla. La razón detrás de esta elección radica en la disponibilidad de datos, ya que no siempre es posible obtener información precisa y confiable de proveedores externos, como Google, para realizar predicciones precisas. Asimismo, un análisis más exhaustivo conlleva un mayor tiempo y recursos, lo cual puede no ser factible en el contexto de la empresa.

Durante el proceso de desarrollo de este modelo de optimización de presupuesto, se encontraron situaciones en las que fue necesario realizar ajustes en las restricciones y probar diferentes valores para lograr obtener una solución factible. Específicamente, se identificó que en ciertas semanas donde los cambios de inversión eran más abruptos, el modelo no conseguía una solución óptima. Ante esta situación, se llevaron a cabo iteraciones y se realizaron pruebas con distintos límites y rangos de valores en las restricciones pertinentes siempre teniendo en cuenta las posibilidades de la empresa y con cuidado de plantear restricciones que no fueran cumplibles en la realidad.

Es importante resaltar que la adaptabilidad del modelo es clave para hacer frente a eventos especiales o situaciones particulares que puedan surgir en el entorno empresarial. Durante eventos especiales, como lanzamientos de productos o campañas promocionales, es posible que las restricciones e, incluso, la función objetivo deban ser modificadas para reflejar las condiciones específicas y los objetivos comerciales establecidos para dichos eventos. Esto permite que el modelo se ajuste y pueda ofrecer soluciones óptimas en contextos cambiantes y dinámicos.

Por último, consideramos que es válido aclarar que el modelo de optimización de presupuesto planteado está diseñado específicamente para la empresa que nos estamos enfocando y su estructura particular. Las características y necesidades de esta empresa se han tenido en cuenta al definir las restricciones y la función objetivo del modelo. Esto garantiza que el modelo sea adecuado y se ajuste a los requerimientos específicos de la empresa en cuanto a la asignación eficiente de su presupuesto publicitario.

En resumen, el modelo de optimización de presupuesto para la e-commerce se ha desarrollado considerando la perspectiva semanal, tomando decisiones de inversión al inicio de cada semana. Se ha optado por una aproximación más sencilla, dada la disponibilidad limitada de datos externos y la necesidad de un análisis ágil. Este enfoque está específicamente adaptado a las características y estructura de la empresa, lo que permite una asignación eficiente de su presupuesto publicitario en función de sus objetivos comerciales.

4.2.3.b Aplicación en ZIMPL

A través del uso del lenguaje de programación ZIMPL, fue posible traducir la formulación matemática del problema en un formato comprensible y utilizable para su implementación en un entorno de programación. El código generado en ZIMPL representa el modelo de programación lineal completo,

que incluye la función objetivo, las variables de decisión y las restricciones asociadas. En la siguiente imagen puede verse la formulación del modelo en ZIMPL³²:

```

set I := 1..23; // Conjunto de campañas
set D := 1..7; // Conjunto de días

var X[I][D] >= 0; // Variables de decisión: inversión en cada campaña i, cada día d

param C[I][D] >= 0; // Datos de conversiones predictas para cada campaña i, cada día d

minimize Z: sum(i in I, d in D) X[i][d] / C[i][d]; // Función objetivo: minimizar la relación costo/inversión

subject to:

// Restricción 1
sum(d in 1..17) X[i][d] >= 70000, for all i in I;

// Restricción 2
sum(i in I, d in D) X[i][d] <= 2500000;

// Restricción 3
sum(i in 13..23, d in D) X[i][d] <= 0.2 * sum(i in I, d in D) X[i][d];

// Restricción 4
sum(i in 1..12, d in D) X[i][d] <= 0.8 * sum(i in I, d in D) X[i][d];

// Restricción 5
X[i][d] >= 0, for all i in I, d in D;

// Restricción 6
C[i][d] * 0.4 <= X[i][d], for all i in I, d in D;

// Bounds
0 <= X[i][d] <= 17500000, for all i in I, d in D;

```

Al implementar y resolver este modelo, nos fue posible obtener los valores óptimos de inversión en cada campaña y día, maximizando así el retorno de la inversión y optimizando la asignación de recursos en el ámbito del marketing.

4.2.3.d Solución

Al utilizar el Solver ZIMPL para resolver el modelo de programación lineal, es importante tener en cuenta que la solución óptima encontrada está sujeta a las restricciones y limitaciones establecidas en el problema. Sin embargo, en algunos casos, puede ser necesario ajustar los límites y restricciones debido a limitaciones presupuestarias u otras consideraciones prácticas.

En nuestro caso, al implementar el modelo en ZIMPL, nos encontramos con una situación en la que el presupuesto asignado inicialmente no permitía encontrar una solución factible. Para abordar esta limitación, tuvimos que realizar modificaciones en los límites establecidos en el modelo. Específicamente, se ajustaron los límites de inversión de las variables X_{id} para cada campaña i y día d , de acuerdo con el presupuesto disponible. Por ejemplo, en lugar de permitir inversiones ilimitadas, se estableció un límite superior de 17500000 para garantizar que las soluciones encontradas fueran financieramente viables dentro de las restricciones presupuestarias.

Una vez encontradas las soluciones óptimas para cada campaña cada día de la semana, se obtiene un reporte donde figura cuáles son los valores de las Xs para cada caso y la solución final de Adcost (costo/inversión)

³² Koch (2001)

4.3 Incorporación de predicciones del modelo SUR en el método ZIMPL

En este contexto, ZIMPL se ha destacado como una herramienta efectiva para resolver problemas de optimización lineal. Sin embargo, consideramos necesario incorporar datos predictivos que reflejen las tendencias y patrones actuales del mercado. En esta sección, se describe cómo integramos las predicciones generadas por el modelo SUR (Seemingly Unrelated Regressions) en el marco de nuestro modelo de asignación con ZIMPL.

4.3.1 Datos del Modelo Sur

4.3.1.a Obtención de las predicciones del modelo SUR

Una vez validado y ajustado el modelo SUR como lo vimos en la sección 4.1, lo utilizamos para generar predicciones de las conversiones en las campañas de marketing. Estas predicciones se basan en los datos históricos de las variables seleccionadas y las relaciones establecidas por el modelo SUR. Específicamente, generamos las predicciones para cada día (d) y categorías (i) en el horizonte de tiempo deseado, en este caso iremos semana a semana para asignar los presupuestos. Estas predicciones representan estimaciones de las conversiones futuras, teniendo en cuenta las tendencias y patrones históricos capturados por nuestro modelo.

Antes de generar las predicciones, es necesario preparar los datos de entrada requeridos por el modelo SUR. Estos datos incluyen las variables seleccionadas, que ya mencionamos antes. Es por eso que obtenemos los datos necesarios para poder evaluarlos en el modelo y así generar las predicciones. En principio los datos serán obtenidos de google ads donde se calculará, mediante las funciones de nuestro código, los promedios de las 3 semanas anteriores en caso de ser necesario para cada variable. A su vez, el dato de la brecha cambiaria se obtendrá de la base que tendremos del sitio oficial de cotización para cada semana, donde luego mediante las funciones de nuestro código se calcula el promedio de la última semana (sin tener en cuenta sábados y domingos).

4.3.2 Ventanas Temporales de las Variables

Es importante tener en cuenta que nuestro objetivo final es crear un modelo capaz de predecir el valor de nuestras conversiones y asignar el presupuesto de manera eficiente, utilizando los datos proporcionados. Para lograr esto, es necesario suministrar al modelo los datos de semanas previas. Esto se debe a que queremos que el modelo esté actualizado con la información más reciente cuando lo utilicemos para predecir en una semana futura.

Algunas de las características (features) que consideraremos en nuestro modelo para predecir las conversiones de una semana específica, tanto para cada día como para cada categoría, son: Clicks, Costo, CPC, Adcost y Brecha cambiaria. Sin embargo, es importante tener en cuenta que estos datos no están disponibles en el momento actual, ya que dependen del rendimiento diario y se conocen al finalizar cada día. Es decir, al momento de predecir las conversiones de la semana no se tienen cuantos clicks se obtendrán esa semana por ejemplo.

Los datos provenientes exclusivamente de Google Ads (Clicks, Costo, CPC y Adcost) se basan en información propia de la plataforma, lo cual dificulta su predicción. Además, la Brecha cambiaria de cada día se determinará en ese mismo día y posiblemente cuando los mercados abran a las 10am, lo que nos limita en términos de tiempo para incluirlo en nuestro análisis.

Es importante tener en cuenta estas limitaciones en la disponibilidad y la naturaleza de los datos al desarrollar nuestro modelo de predicción de conversiones. A pesar de ello, buscaremos soluciones alternativas y estrategias para maximizar la veracidad de nuestras predicciones utilizando los datos disponibles y considerando las características relevantes.

En las siguientes secciones, se realizarán comparaciones de modelos de regresión para definir el tiempo previo a tener en cuenta para la predicción de nuestro target. Resulta necesario realizar tests por separado, por un lado evaluar las métricas Clicks, Costo, CPC y Adcost y por otro evaluar la Brecha cambiaria, ya que permite considerar mejor su impacto para con las conversiones.

4.3.2.a Definición Ventana Temporal para las Variables propias de Google Ads

Como ya mencionamos previamente, es necesario considerar la influencia de diversas métricas, como Clicks, Costo, CPC y Adcost, que pueden afectar el rendimiento de las ventas. Sin embargo, surge una pregunta fundamental: ¿es más relevante utilizar la información de la última semana o un promedio de varias semanas anteriores para realizar estas predicciones? En esta sección, se examina y se compara la importancia de la data más reciente de la última semana frente al promedio de semanas anteriores en la predicción de ventas.

La elección adecuada de las métricas y la ventana temporal utilizada para realizar predicciones de ventas puede tener un impacto en los modelos. Datos más recientes pueden proporcionar información valiosa sobre las tendencias y cambios recientes en el comportamiento del mercado, lo que podría mejorar la capacidad del modelo para capturar patrones actuales y realizar pronósticos más precisos. Por otro lado, el promedio de varias semanas anteriores podría suavizar las fluctuaciones temporales y brindar una visión del rendimiento que podría ser valiosa.

Para testear esto, se realizará un experimento comparativo que consiste en dividir los datos en dos conjuntos: uno que contenga únicamente la información de la última semana y otro que incluya un promedio de las métricas de varias semanas anteriores (en este caso las últimas 3 semanas). Luego, se entrenará modelos de predicción de ventas utilizando ambos conjuntos y se evaluará su rendimiento.

Se dividen los datos en dos conjuntos: el conjunto A que contiene las métricas de la última semana y el conjunto B que contiene el promedio de las últimas 3 semanas. Luego, se entrenan modelos de regresión lineal utilizando los conjuntos A y B por separado, y se realizan predicciones para cada uno de ellos. Nuestra métrica de evaluación será el R cuadrado ajustado y el AIC para comparar el rendimiento de los modelos en la predicción de ventas y poder definir los parámetros a utilizar. El R2 ajustado es una medida de la calidad del ajuste del modelo que tiene en cuenta la cantidad de variables incluidas en el modelo y penaliza la inclusión de variables irrelevantes. Esta métrica nos permitirá determinar qué enfoque utilizar. La información de la última semana o un promedio de

semanas anteriores. A su vez, AIC es una herramienta valiosa para seleccionar variables, como otros hiperparámetros, es un estimador consistente del error de generalización del modelo, que toma en cuenta una penalización por la inclusión de variables adicionales.

Cabe aclarar que tanto para este testeo como para el testeo de la brecha, consideramos los otros predictores también. Esto significa que no solo se evalúa la relevancia de una variable en particular, sino que se toman en cuenta todas las variables del conjunto de datos. Al considerar otros predictores, se pueden capturar las interacciones y efectos conjuntos entre las variables, lo que puede mejorar la capacidad predictiva del modelo.

El Conjunto A (última semana) presenta un R2 ajustado de 0.51. El R2 ajustado de 0.51 indica que el modelo de regresión, utilizando solo los datos de la última semana, explica aproximadamente el 51% de la variabilidad de las ventas. Un valor más cercano a 1 indica un mejor ajuste del modelo a los datos. Por otro lado, el Conjunto B (promedio de las 3 semanas anteriores) presenta un R2 ajustado de 0.69. En este caso, utilizando el promedio de 3 semanas anteriores, el modelo de regresión explica aproximadamente el 69% de la variabilidad de las ventas.

Comparando los valores de R2 ajustados, podemos observar que el conjunto B (promedio de semanas anteriores) presenta un valor más alto que el conjunto A (última semana). Esto indica que el modelo que utiliza el promedio de las semanas anteriores tiene un potencial mejor ajuste y explica más la variabilidad de las ventas en comparación con el modelo que solo utiliza los datos de la última semana.

Además del R2 ajustado, obtuvimos un AIC =1555 para el conjunto A y un AIC =1023 para el conjunto B. Podemos determinar que el conjunto B (promedio de semanas anteriores) tiene un valor más bajo en AIC con el conjunto A (última semana). Esto sugiere que el modelo que utiliza el promedio de semanas anteriores tiene un mejor ajuste y una menor complejidad en comparación con el modelo que utiliza solo los datos de la última semana. Además, al analizar el error cuadrático medio de las predicciones, el conjunto B obtuvo un RECM de 2670, mientras que el conjunto A registró un RECM de 3560. Estos resultados respaldan la elección del conjunto B como el modelo correcto para predecir las ventas y optimizar la estrategia de marketing en el contexto del presente estudio.

4.3.2.b Definición Periodo de Tiempo Anterior para la Brecha del tipo de cambio

Nos proponemos entonces analizar la influencia de la brecha cambiaria en las ventas, específicamente comparando la brecha cambiaria de la última semana con el promedio de la brecha de las últimas tres semanas. Utilizamos un enfoque basado en regresión lineal y evaluamos la calidad de las predicciones mediante el cálculo del R cuadrado ajustado y del AIC como vimos en la sección 4.3.2.a. Nuestro objetivo es determinar cuál de estas dos variables es más relevante para una mejor predicción de las ventas.

Como hicimos con las variables anteriores, organizamos los datos en un DataFrame, separando las variables independientes (brecha cambiaria) y la variable objetivo (ventas) creando dos conjuntos de datos: uno para la última semana y otro para el promedio de las últimas tres semanas de brecha cambiaria. Luego, implementamos dos modelos donde uno se ajustó al conjunto de datos de la

última semana, mientras que el otro se ajustó al conjunto de datos del promedio de las últimas tres semanas. Entrenamos los modelos con los conjuntos de datos correspondientes y calculamos el R cuadrado ajustado comparando las predicciones obtenidas con los valores reales de las ventas.

El R cuadrado ajustado de la brecha de la última semana fue de 0.413 y el R cuadrado ajustado del promedio de las últimas tres semanas fue de 0.23. Estos resultados indican que la predicción de ventas utilizando el promedio de la brecha de las últimas tres semanas tiene un R cuadrado ajustado menor en comparación con la predicción utilizando la brecha de la última semana. Un R cuadrado ajustado mayor sugiere que las predicciones están más cerca de los valores reales de las ventas, lo que implica una mejor capacidad predictiva. Por otro lado, el AIC del modelo A (promedio de la última semana) fue de 563 y el modelo B (promedio de semanas anteriores) de 912, lo que nos indica que el modelo A tiene una mejor calidad y ajuste en comparación con el modelo B.

Al evaluar el error cuadrático medio de las predicciones (RECM), se encontró que el modelo B tiene un RECM de 5470, mientras que el modelo A tiene un RECM de 4790. Estos resultados muestran que el modelo A presenta evidencia de un menor error en las predicciones, lo que indica una mayor capacidad predictiva en comparación con el modelo B.

Es por esto que, a diferencia de las anteriores métricas, nos daría un potencial mejor resultado en nuestra predicción de ventas utilizando el promedio de la brecha cambiara de la semana inmediatamente anterior a la que queremos predecir. Una vez definida la venta temporal de los features numéricos, se describen la totalidad de las variables a tomar para realizar nuestra predicción de ventas.

Cabe destacar que este ejercicio realizado de selección de variables se realiza utilizando los otros predictores también al igual que el anterior.

El modelo no solo tomará en cuenta las métricas de la semana anterior como la brecha, o del promedio de las últimas 3 semanas, como el costo por clic (CPC), el costo, los clics y el adcost, sino también el contexto de lo que se desea predecir. Esto incluye la semana del año, el mes, el año, el día de la semana y la campaña específica. Al considerar estos factores adicionales, el modelo podrá generar predicciones más ajustadas a las circunstancias específicas en las que se realizará la asignación de presupuesto. Al proporcionar datos actualizados y contextualmente relevantes al modelo, nos aseguramos de que esté bien equipado al momento de hacer las predicciones para semanas futuras

4.3.3 Incorporación de las predicciones del modelo SUR en ZIMPL

Para incorporar estas predicciones de conversiones en nuestro modelo de optimización, hemos utilizado el lenguaje de modelado ZIMPL. En el siguiente fragmento de código, podemos ver cómo hemos incluido los datos de conversiones predichas para cada campaña *Cid* en una semana específica).

```
# Incorporación de los datos de conversiones predichas
param Cid :=
2 1 2276275.69,
2 2 721602.1,
2 3 5596465.1,
2 4 10561013.66,
2 5 2859457.59,
2 6 3781385.8,
2 7 384842.23;
```

Tomamos para una semana en particular, las primeras siete predicciones que corresponden a la categoría de computación (i = 2) y a cada día de la semana (d = 1 a 7), se realiza el reemplazo de las C_{id} en nuestra función objetivo por las predicciones correspondientes:

$$Z = \sum_{i=1}^{23} \sum_{d=1}^7 \frac{X_{id}}{\text{Predicciones } C_{id}}$$

Donde *Predicciones Cid* representa la columna de constantes que contiene las predicciones de conversiones Cid para cada campaña y día de la semana. Cada línea del parámetro Cid representa una campaña específica y su respectiva predicción de conversiones para esa semana. Este ejemplo contiene sólo 7 predicciones para una categoría, pero SUR genera esto para todas las categorías todos los días y luego es aplicado a nuestro modelo como fue mencionado.

El reporte final del método ZIMPL representa los resultados obtenidos al aplicar el algoritmo de optimización a nuestro modelo de asignación para una semana específica. Este reporte proporciona información valiosa sobre las asignaciones óptimas de inversión en las campañas de marketing, teniendo en cuenta las predicciones de conversiones y las restricciones establecidas.

A continuación, se presenta un ejemplo del reporte final utilizando una semana determinada:

Semana: Del 9 al 15 de enero de 2023.

Asignaciones óptimas de inversión:

Las inversiones óptimas en cada campaña y día de la semana se muestran a continuación:

Campaña 1:	Campaña 2:	Campaña 23:
Lunes: X11 = 53240	Lunes: X21 = 119000	Lunes: X231 = 48700
Martes: X12 = 101213	Martes: X22 = 78200	Martes: X232 = 62800
Miércoles: X13 = 98670	Miércoles: X23 = 81900	Miércoles: X233 = 60340
Jueves: X14 = 99800	Jueves: X24 = 79500	Jueves: X234 = 61400
Viernes: X15 = 60900	Viernes: X25 = 84300	Viernes: X235 = 73100
Sábado: X16 = 48710	Sábado: X26 = 52100	Sábado: X236 = 51560
Domingo: X17 = 49340	Domingo: X27 = 52800 Domingo: X237 = 50920

Valor objetivo:

La función objetivo Z, formulada de manera negativa para minimizar X/C, presenta el siguiente valor:
Z = 0.71

Si comparamos este valor objetivo con el promedio real de la semana, que fue de 0.93, podemos observar que el adcost obtenido a través de la asignación óptima de inversión propuesta por el

modelo es menor que el adcost promedio real. En otras palabras, el modelo logró generar un retorno mayor en comparación con la inversión realizada en la semana en cuestión. Esta justificación se basa en la premisa de que un adcost más bajo indica una mayor eficiencia en la utilización de los recursos invertidos. En este caso, el adcost de 0.71 obtenido a través del modelo sugiere que la asignación propuesta permitió obtener un mayor retorno en relación al costo de inversión. Sin embargo, este es solo un ejemplo puntual. Para tener un mejor panorama de cómo funciona nuestro modelo, proponemos tomar data por fuera de nuestro modelo de predicción y ver qué tan bueno es en comparación tanto con la realidad como con otros modelos más sencillos para la predicción de ventas. Esto lo veremos en la Sección 5.

Es importante destacar que los ajustes realizados en las asignaciones de inversión cumplen con las restricciones mencionadas, como el presupuesto máximo, el mínimo de inversión por campaña y la priorización de las categorías del 1 al 12. Este reporte final del método ZIMPL proporciona una visión completa y detallada de las asignaciones óptimas teniendo en cuenta todas las restricciones establecidas, permitiendo una evaluación precisa del rendimiento y la eficacia del modelo de asignación para la semana específica analizada.

En el Anexo 8.4 se observa un ejemplo del reporte final realizado para tres semanas donde se genera una tabla para ver inversión de la semanal y luego un reporte con todas las soluciones por día por semana.

4.4 Restricción Adicional a nuestro modelo de Asignación

En la sección 3.3.3 b, hemos discutido detalladamente la importancia que tienen los días de la semana para las ventas de la empresa. Es por esto que, hemos identificado la necesidad de introducir una nueva restricción en nuestro modelo de asignación de presupuesto. Esta restricción se refiere al gasto en los días de la semana, y tiene como objetivo garantizar una distribución equilibrada de los recursos a lo largo de la semana, teniendo en cuenta las particularidades de cada día.

Específicamente, hemos establecido dos aspectos clave en relación al gasto en los días de la semana. En primer lugar, hemos determinado que el día $d=1$, que corresponde al lunes en nuestro caso, debe tener asignado como mínimo un 15% del total de la inversión. Esta medida busca asegurar que se destine una cantidad importante de recursos al inicio de la semana, lo que puede ser especialmente relevante para proyectos que requieren un impulso inicial o que dependen de actividades programadas para los días laborables.

En segundo lugar, hemos establecido que los días sábado y domingo, en conjunto, no pueden superar el 20% de la inversión total. Esta limitación se basa en la comprensión de que los fines de semana suelen tener una dinámica diferente a los días laborables, con potenciales restricciones operativas y menor disponibilidad de recursos humanos. Al asegurarnos de que no se destine un exceso de inversión en los días no laborables, estamos evitando una potencial descompensación en la distribución de recursos a lo largo de la semana.

4.4.1 Incorporación Nuevas restricciones

Las ecuaciones para introducir en ZIMPL serán entonces las siguientes:

Restricción 7: Gasto mínimo en el día d (lunes)

$$\sum_{i=1}^{23} X_{i1} \geq 0.15 * \sum_{i=1}^{23} \sum_{d=1}^7 X_{id} \quad \forall i \in I \text{ y } d \in D$$

En la restricción 7, se introduce la nueva restricción que establece que en el día d (lunes en este caso, correspondiente al índice 1) se debe gastar como mínimo un 15% del total de la inversión.

Restricción 8: Gasto máximo en fines de semana (sábado y domingo)

$$\sum_{i=1}^{23} \sum_{d=6}^7 X_{id} \leq 0.2 * \sum_{i=1}^{23} \sum_{d=1}^7 X_{id} \quad \forall i \in I \text{ y } d \in D$$

En la restricción 8, se establece la limitación de que la suma del gasto en los días sábado y domingo (índices 6 y 7) no debe exceder el 20% del total de la inversión.

La notación en ZIMPL sería la siguiente:

```
// Restricción 7: Gasto mínimo en el día d (lunes)
sum(i in I) X[i][1] >= 0.15 * sum(i in I, d in D) X[i][d];

// Restricción 8: Gasto máximo en fines de semana (sábado y domingo)
sum(i in I, d in {6, 7}) X[i][d] <= 0.2 * sum(i in I, d in D) X[i][d];
```

4.4.2 Validación Nuevas Restricciones en nuestro modelo

En el marco de nuestra investigación, hemos aplicado nuestro modelo de asignación de presupuesto con restricciones adicionales a la data extraída originalmente para validar nuestros resultados. Nuestro objetivo era evaluar el impacto de estas restricciones en la eficiencia y la viabilidad del modelo en comparación con el modelo original que no las tenía. Sin embargo, los resultados obtenidos no fueron alentadores. En la tabla 10 podemos ver los resultados parciales de correr estas asignaciones en ZIMPL:

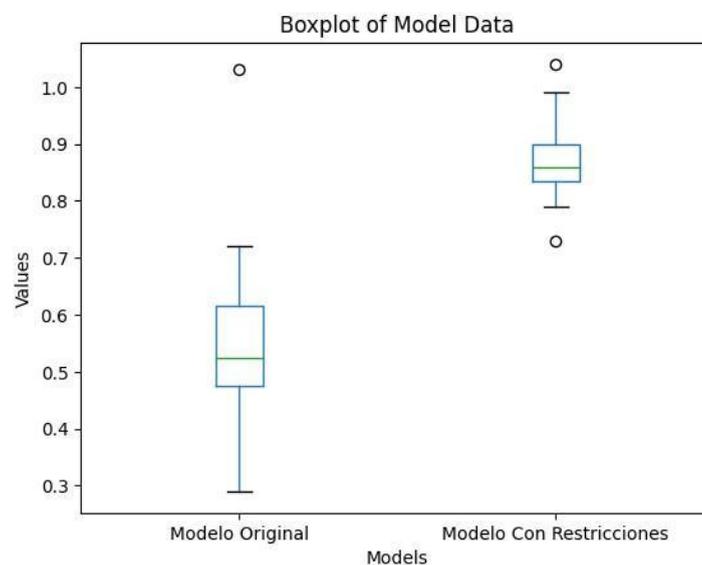
Tabla 10 : Comparación Salidas ZIMPL

Modelo	Adcost Promedio	Semanas Sin solución	P valor Dif-Medias
Original	0.560	8.45%	6.265390896e-04
Rest. Adicionales	0.924	38.36%	6.265390896e-04

Al comparar el modelo con restricciones adicionales con el modelo original, observamos que el nivel de *adcost* promedio fue menor en el modelo original, es decir, el modelo sin las restricciones adicionales. Esto indica que, en general, el modelo original logró asignar el presupuesto de manera más eficiente, sin la necesidad de imponer restricciones específicas sobre el gasto en los días de la semana. Además, otro aspecto preocupante fue que muchas semanas no obtuvieron solución utilizando el modelo con las restricciones adicionales. Esto plantea dificultades a la hora de aplicar este modelo en la realidad, ya que no sería viable ni práctico contar con un modelo que no puede ofrecer una solución satisfactoria en una cantidad importante de semanas.

Para testear la existencia de diferencias entre estas series, se llevó a cabo un test de medias apareadas, también conocido como prueba t de Student para muestras relacionadas. Este tipo de prueba estadística se utiliza para determinar si hay una diferencia significativa, en este caso al 0.05, entre las medias de dos grupos de datos que están relacionados entre sí. El valor p indica la probabilidad de que los resultados observados pueden deberse sólo al azar, y se calcula como la chance de obtener una estadística de prueba t igual o más extrema que la observada en la muestra, asumiendo que la hipótesis nula es verdadera. En este caso, el valor p es $6.265390896e-04$, lo que indica que hay una probabilidad extremadamente baja de que los resultados sean solo al azar y sugiere que la hipótesis nula puede rechazarse. A su vez, se muestra un box plot para ver la dispersión de los resultados de cada modelo:

Gráfico 18 : Boxplot Modelo con restricciones Originales vs Modelo con Restricciones Adicionales



**Fuente: Edición propia*

En base a estos hallazgos, debemos concluir que la incorporación de estas restricciones adicionales en el modelo de asignación de presupuesto no resultó beneficioso en términos de eficiencia y viabilidad. El modelo original, sin las restricciones, demostró ser más efectivo y aplicable para nuestros casos en la asignación de recursos financieros. Estos resultados nos plantean la necesidad de analizar nuestros resultados finales con el modelo de optimización de presupuesto original sin

tener en cuenta la distribución del costo en los días de las semanas como restricción en nuestro modelo.

5. Validación de Resultados

En esta sección, se presentan los métodos para evaluar los resultados del modelo de asignación de presupuesto propuesto: el Test de Giacomini y White y la comparación de los resultados con los datos reales. Además, con el objetivo de garantizar la robustez y confiabilidad de nuestros resultados, se implementará una técnica de validación cruzada específica utilizando los datos de entrenamiento.

El Test de Giacomini y White es un enfoque estadístico para evaluar la capacidad predictiva del modelo y determinar si el modelo de asignación de presupuesto propuesto es superior a otros modelos más sencillos. Específicamente, es un test de capacidad predictiva condicional que evalúa el rendimiento predictivo de dos modelos competidores.³³

Por otro lado, al comparar nuestros resultados con los datos reales, podremos evaluar cómo nuestro modelo se desempeña en situaciones del mundo real y compararlo con el enfoque actual de asignación de presupuesto utilizado. Evaluaremos nuestra metodología en comparación con datos reales resultados de las decisiones tomadas por la empresa. Asimismo, evaluaremos únicamente nuestro modelo de optimización de presupuesto en dos escenarios distintos, utilizando en nuestro estimador por un lado los datos reales de las conversiones para c , y por otro los datos predichos para cid de SUR, para detectar el impacto que tiene nuestra predicción.

Al final de esta sección, se espera proporcionar una evaluación completa del modelo propuesto, evaluando su performance predictiva en comparación con el enfoque actual utilizado para la asignación de presupuesto.

5.1 Validación Cruzada con Semanas Excluidas

Como fue mencionado en la sección 3.2.1, se utilizó una técnica de validación cruzada donde se excluyeron las semanas más recientes de los datos para formar un conjunto de validación independiente. La finalidad de este enfoque fue evaluar adecuadamente la capacidad predictiva de los modelos generados.

Al dividir el grupo de entrenamiento en conjuntos "train" y "validation" en diferentes pruebas, se pudo minimizar el riesgo de sobreajuste, lo que garantiza que los modelos no estén ajustados exclusivamente a los datos de entrenamiento, sino que sean capaces de generalizar en diferentes contextos. El proceso de excluir semanas recientes de la data y utilizarlas como conjunto de validación permitió realizar predicciones independientes y evaluar la verdadera performance de cada modelo. Esta técnica nos brindó una visión más precisa del comportamiento de los modelos en

³³ McCracken (2020)

situaciones futuras, lo que fortaleció la confiabilidad de nuestros resultados y respaldó las conclusiones obtenidas a lo largo de este estudio.³⁴

Cabe destacar que, para los análisis de las secciones 5.2 y 5.3, se utilizaron estos datos excluidos en la validación cruzada. La utilización de estos datos excluidos en los análisis posteriores añade una capa adicional de confianza en nuestras conclusiones y nos permite tomar decisiones informadas y confiables en el contexto de la optimización de campañas publicitarias.

5.2 Evaluación del modelo de asignación de presupuesto mediante el Test de Giacomini y White

El presente trabajo se centra en el análisis de la predicción de ventas de una empresa utilizando un modelo de Regresión de Variables Simultáneas (SUR). En este contexto, uno de los objetivos principales de este trabajo es evaluar la capacidad predictiva del modelo SUR, lo cual se logra mediante la realización del Test de Giacomini y White. Este test se utiliza para comparar las predicciones del modelo SUR con un modelo más sencillo, como podría ser un modelo de Autoregresión de Orden 1 (AR(1)) para cada una de las series de ventas.

5.2.1 Metodología Utilizada

Para poder comparar los modelos se realizó un test de Tests of Conditional Predictive Ability (TCPA), es una metodología utilizada para evaluar y comparar la capacidad predictiva condicional de diferentes modelos. Estos modelos se utilizan para predecir el valor futuro de una variable dependiente (en este caso nuestra ventas) utilizando información histórica y otras variables independientes. El TCPA se basa en la idea de dividir los datos en submuestras y evaluar el rendimiento predictivo de los modelos en cada submuestra. Esto permite evaluar cómo los modelos se desempeñan en diferentes períodos de tiempo y bajo diferentes condiciones económicas.³⁵

Los TCPA son útiles para evaluar modelos en situaciones en las que las relaciones entre las variables pueden cambiar a lo largo del tiempo o en diferentes condiciones. Estos tests permiten seleccionar el modelo que mejor se ajuste a los datos y que tenga una mayor capacidad de predicción.

Una vez calculadas las predicciones para ambos modelos, se comparan las predicciones de los dos modelos con los valores reales de la variable dependiente en la submuestra de validación. Como mencionamos utilizamos el test de Giacomini y White que compara las diferencias en las predicciones de los dos modelos utilizando pruebas estadísticas. El test se basa en la idea de que si un modelo tiene una mayor capacidad predictiva que el otro, las diferencias entre las predicciones de los dos modelos serán significativas.

Se busca entonces determinar si el modelo SUR posee una performance predictiva significativamente mayor (al 0.05) que el modelo AR(1), es decir, si el uso de un modelo más complejo con variables interdependientes mejora significativamente la capacidad predictiva de las predicciones. En caso de que el modelo SUR no parezca tener una performance predictiva significativamente mejor que el

³⁴ Ochoa (2012)

³⁵ McCracken (2020b)

modelo AR(1), se podría concluir que el uso de un modelo más complejo no justifica la complejidad adicional del modelo.³⁶ En la tabla 11 se presentarán los resultados obtenidos al realizar el test de Giacomini y White en el contexto del modelo SUR utilizado para la predicción de ventas de la empresa en estudio.

Tabla 11 : Model Comparison Results

	SUR	AR(1)
Campaign	0.8	0.6
CostoAjustado	0.7	0.4
Dayoftheweek	0.7	0.5
Clicks	0.6	0.4
BrechaDiff	0.3	0.1
AvgCPC	0.9	0.7
Week_Num	0.4	0.2
Month_Num	0.2	0.1
Acost	0.5	0.4

**Fuente: Elaboración Propia*

En la tabla 11, se muestran los coeficientes de determinación (R cuadrado) obtenidos por el modelo SUR y el modelo AR(1) para cada variable en estudio. Los valores más altos indican una mayor capacidad del modelo para explicar la variabilidad en la variable dependiente.

En este caso, se puede observar que el modelo SUR obtiene valores más altos que el modelo AR(1) para todas las variables, lo que sugiere que el modelo SUR es una mejor opción para predecir la variable de interés en comparación con el modelo AR(1). En términos del Test de Giacomini y White, una diferencia significativa en los coeficientes de determinación entre los dos modelos indicaría una discrepancia significativa en la capacidad predictiva de los modelos.

Una vez realizado este test, se obtiene un p valor de 0.00024687533456. Esto sugiere que hay suficiente evidencia estadística para rechazar la hipótesis nula de que los modelos son iguales y concluir que hay una diferencia significativa en la performance predictiva de los modelos. Dado que el valor p es menor que el nivel de significancia elegido (0.05), podemos rechazar la hipótesis nula y tener suficiente evidencia estadística para concluir que el modelo SUR es estadísticamente superior al modelo AR(1) en términos de capacidad predictiva para la variable dependiente. En otras palabras, el modelo SUR es más apropiado para la tarea de predicción de la variable de interés que el modelo AR(1) para cada variable en estudio.

Un aspecto relevante es que, al considerar datos apartados para la evaluación de los modelos, se observaron resultados concretos. El modelo AR(1) arrojó un valor de RECM mayor, específicamente 7650, en comparación con el valor de RECM obtenido por el modelo SUR con estos mismos datos, el cual fue de 2009. Esta diferencia en los valores de RECM respalda aún más la conclusión de que el

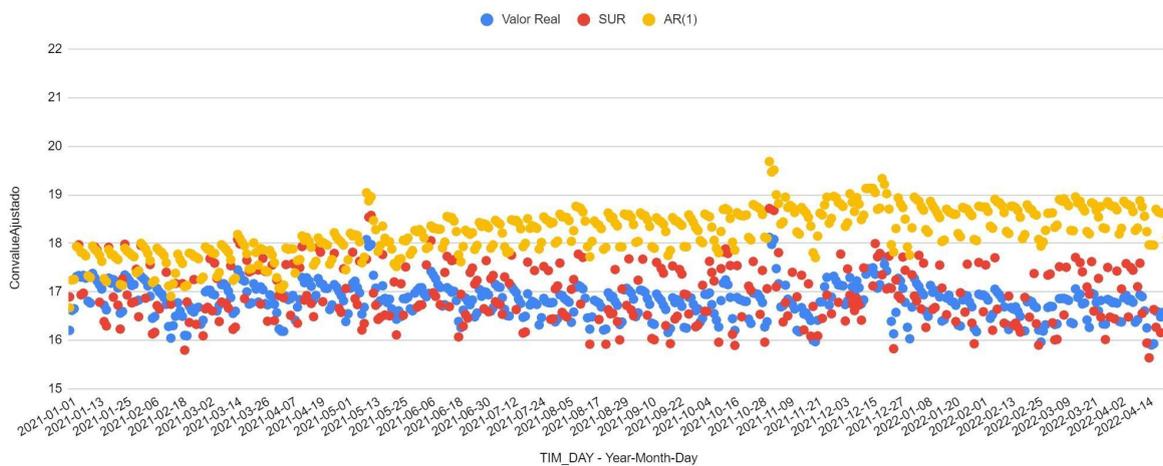
³⁶ McCracken (2020b)

modelo SUR es más apropiado y eficaz para la tarea de predicción de la variable de interés en estudio que el modelo AR(1).

5.2.2 Comparación Gráfica de los Modelos

El Gráfico 19, que se muestra a continuación, incluye puntos amarillos adicionales a los que vimos en la sección 4.1.4, que representan una comparación entre el modelo SUR y el modelo más sencillo AR(1) en términos de su capacidad predictiva. Específicamente, se explorará cómo el modelo AR(1) tiende a sobrestimar los valores en comparación con el modelo SUR.

Gráfico 19: Relación Ventas con Inversión



*Fuente: Elaboración Propia

Al observar el Gráfico 19, podemos notar una clara diferencia entre los resultados obtenidos por el modelo SUR y el modelo AR(1) de las ventas semanales. Los puntos amarillos, que representan las estimaciones generadas por el modelo AR(1), tienden a estar por encima de la línea azul, que representa los valores reales. Esto indica que el modelo AR(1) tiene una tendencia a sobrestimar los valores de las ventas en comparación con el modelo SUR.

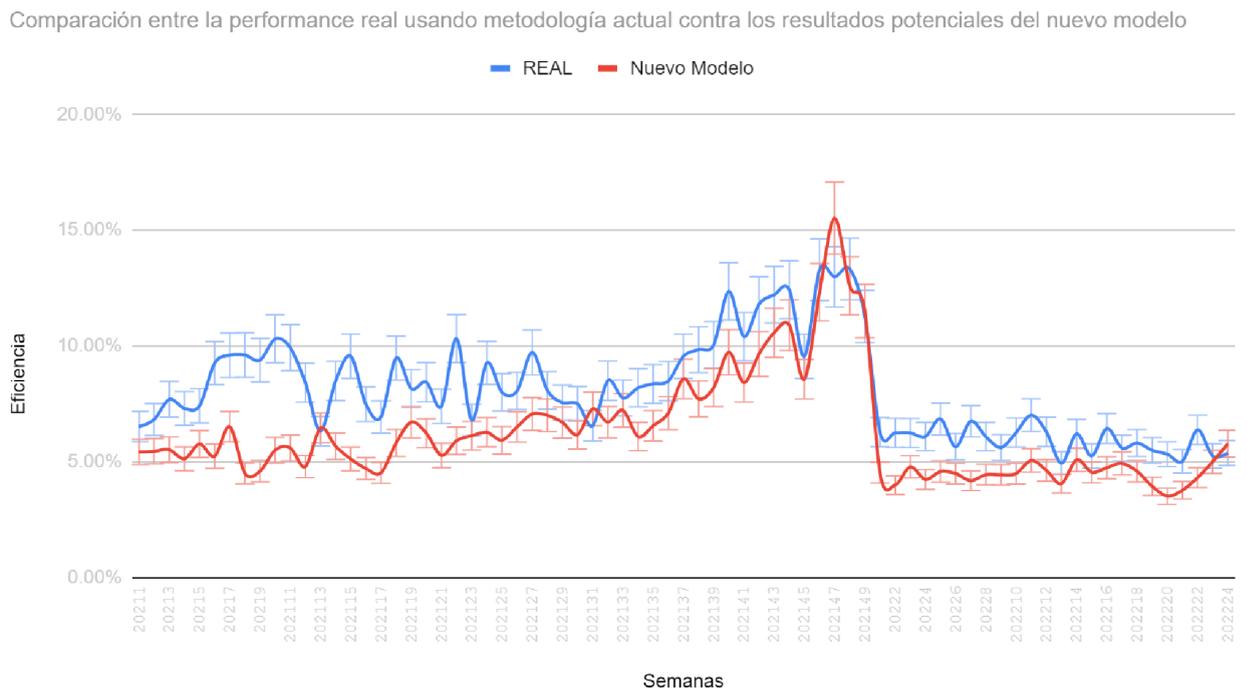
Esta diferencia en las estimaciones es un factor importante a la hora de seleccionar el modelo más adecuado. En este caso, el modelo SUR muestra un mejor desempeño en términos de ajuste a los datos reales y capacidad predictiva. Al considerar la complejidad y las interrelaciones de las variables endógenas en el modelo SUR, podemos capturar de manera más precisa los patrones y las tendencias subyacentes en los datos.

5.3 Benchmark

En esta sección del proyecto se comparará el modelo de asignación de presupuesto propuesto con lo sucedido en la realidad. El modelo se basa en una predicción SUR que asigna presupuesto según el día y la categoría durante la semana. Se tomaron ciertas semanas que fueron excluidas del modelo de aprendizaje para poder ver como la predicción y el modelo de asignación hubiesen funcionado en el caso de haberlos aplicado. Las semanas excluidas son las semanas que comprenden el período de tiempo desde Septiembre del 2022 al 6 de marzo del 2023.

El objetivo de esta comparación es demostrar cómo funcionó el modelo según nuestras estimaciones y cómo efectivamente fueron los resultados con la metodología actual de la empresa para la asignación de presupuesto en la realidad. Para ello, se analizarán los resultados obtenidos por ambos casos, uno con ventas y asignaciones reales y otro con ventas predichas por SUR y asignaciones en base a nuestro modelo. A su vez, se discutirán las posibles razones por las que el modelo propuesto habría obtenido mejores resultados. Este análisis permitirá evaluar la performance predictiva del modelo y su potencial aplicación en situaciones similares en el futuro.

Gráfico 20: Eficiencia Real vs Eficiencia Modelo Propuesto



*Fuente: Elaboración propia

Tabla 12 : Summary Statistics

	Adcost	
	Modelo Nuevo	Modelo Ventas Reales
min	0.041000	0.034000
max	0.077400	0.073200
median	0.056100	0.052500
mean	0.057699	0.051695

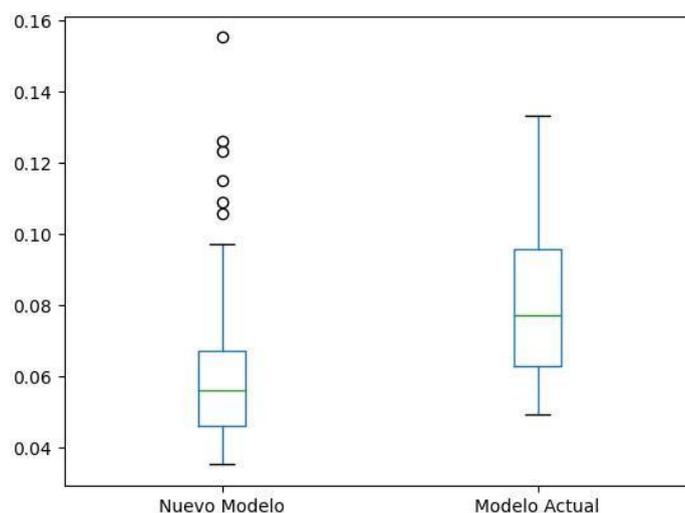
*Fuente: Elaboración propia

En relación al gráfico 20 presentado, se puede observar que con la nueva metodología se obtiene un mejor resultado en la métrica de eficiencia definida en sección 2.5 en comparación al comportamiento real en la mayoría de los casos (con un margen de error del 10% según lo calculado). Notamos que la eficiencia visualmente parece ser mejor ya que es menor en la mayoría de los casos, por lo que se puede inferir que la relación entre la inversión y la ganancia parecería ser mejor.

Para analizar la diferencia entre estas series, se llevó a cabo un test de medias apareadas, también conocido como prueba t de Student para muestras relacionadas. Este tipo de prueba estadística se utiliza para determinar si hay una diferencia significativa entre las medias de dos grupos de datos que están relacionados entre sí. En el contexto de este proyecto, se utilizó el test de medias apareadas para comparar las asignaciones de presupuesto realizadas en la realidad, con las asignaciones predichas por nuestro modelo³⁷. La idea es evaluar si existe una diferencia significativa entre estas dos series de asignaciones.

El estadístico de prueba utilizado en el test de medias apareadas es el estadístico t. Este valor se calcula considerando la diferencia entre las observaciones emparejadas y la variabilidad de esas diferencias. Cuanto mayor sea el valor absoluto del estadístico t, mayor será la evidencia de una diferencia significativa entre las medias de los dos grupos. Además del estadístico t, la salida de la prueba también incluye otros dos valores importantes: el valor p y los grados de libertad. El valor p representa la probabilidad de obtener una diferencia tan grande o más grande que la observada, bajo la suposición nula de que no hay diferencia entre las medias de los dos grupos. Si el valor p es menor que un umbral de significancia predefinido (en este caso 0.05), se considera que la diferencia entre las medias es estadísticamente significativa. Los grados de libertad están relacionados con el tamaño de la muestra y se utilizan para determinar la distribución del estadístico t.³⁸

Gráfico 21: Boxplot Nuevo Modelo vs Boxplot Modelo Actual



*Fuente: Elaboración propia

En este caso, el valor estadístico t indica la magnitud de la diferencia entre las dos muestras y se calcula dividiendo la diferencia entre las dos medias por el error estándar de la diferencia. En este caso, el valor del estadístico de prueba es -11.895218359621913.

El valor p indica la probabilidad de que los resultados observados pueden deberse sólo al azar, y se calcula como la probabilidad de obtener una estadística de prueba t igual o más extrema que la observada en la muestra, asumiendo que la hipótesis nula es verdadera. En este caso, el valor p es

³⁷ Grigelionis (2012)

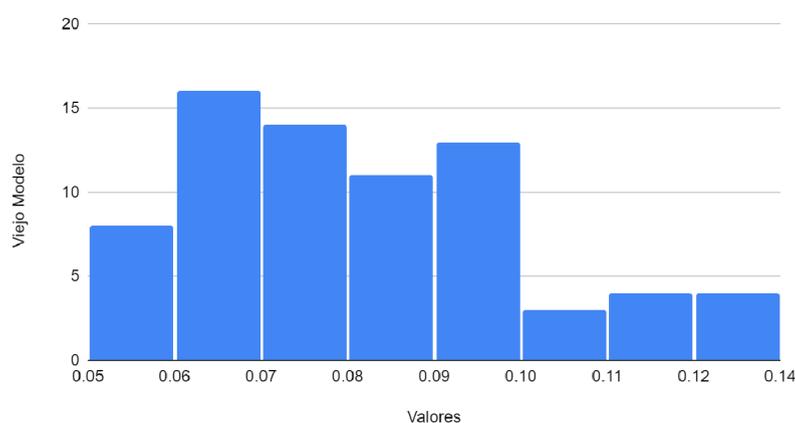
³⁸ Cuadras (2016)

1.1589733691390896e-18, lo que indica que hay una probabilidad extremadamente baja de que los resultados sean solo al azar y sugiere que la hipótesis nula puede rechazarse.

Los grados de libertad se utilizan para calcular el valor p y se calculan como el número de observaciones menos el número de parámetros que se deben estimar. En este caso, los grados de libertad son 72.

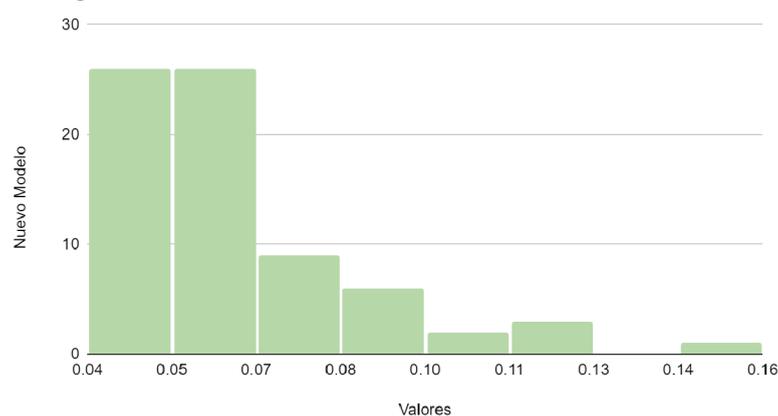
En resumen, el resultado de la prueba t indica una diferencia significativa (al 0.05) entre las medias de las dos muestras, con un valor p extremadamente bajo, lo que sugiere que la hipótesis nula se puede rechazar con confianza y que la diferencia entre las dos medias no se debe solo al azar.

Gráfico 22: Histograma valor Adcost Modelo SUR+ZIMPL



**Fuente: Elaboración Propia*

Gráfico 23: Histograma valores Adcost Modelo Actual



**Fuente: Elaboración Propia*

Como chequeo adicional, se realiza un test no paramétrico donde no se asume que los datos sigan una distribución normal. Algunos ejemplos de pruebas no paramétricas son la prueba de Wilcoxon, la prueba de Mann-Whitney y la prueba de Kruskal-Wallis. En este caso, corremos la prueba de Mann-Whitney

La prueba de Mann-Whitney es una prueba no paramétrica utilizada para comparar las distribuciones de dos grupos independientes. Al estar comparando la asignación de presupuesto

entre dos tratamientos diferentes, esta prueba es apropiada ya que no es necesario asumir que los datos siguen una distribución normal. A su vez, la prueba de Mann-Whitney es adecuada para datos en una escala ordinal o de intervalo. Como nuestros datos de asignación de presupuesto se pueden clasificar en categorías y tienen un orden específico basado en las fechas, esta prueba puede ser más apropiada que la prueba de Wilcoxon o la prueba de Kruskal-Wallis. Dado que nuestro objetivo es analizar la asignación de presupuesto de dos modelos, la prueba de Mann-Whitney parece ser la más apropiada para este proyecto, ya que me permite evaluar las diferencias entre dos grupos independientes y sin hacer suposiciones sobre la distribución de los datos.³⁹

En este caso, el valor de la estadística U es 1239.0, lo que indica la suma de los rangos asignados al grupo más pequeño. El valor p es 2.44e-08, que es menor que 0.05. Por lo tanto, se puede concluir que hay evidencia suficiente para rechazar la hipótesis nula H_0 al 5% de significancia de que las medianas de los dos grupos son significativamente diferentes. Esto indica que el modelo propuesto parece ser una herramienta valiosa para la empresa en cuestión, ya que permitiría mejorar en gran parte su desempeño y obtener mejores resultados en sus operaciones. Es necesario seguir evaluando los resultados y considerar otros factores que puedan afectar a la eficiencia de la empresa, pero estos hallazgos son prometedores y sugieren que la nueva metodología es una opción a considerar para la optimización del presupuesto y la planificación de operaciones.

5.3.1 Impacto de la Predicción de Ventas

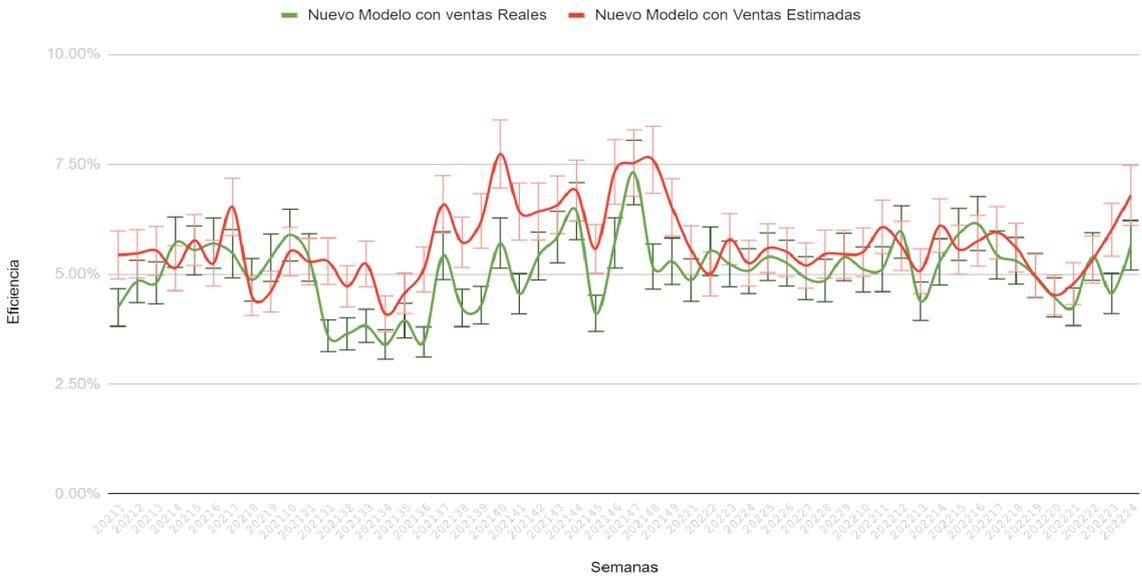
Es importante tener en cuenta que, aunque los resultados obtenidos con el modelo SUR parecen superiores a lo que se hizo en la realidad, es posible que la predicción de las ventas esté afectando la eficiencia propuesta. Por esta razón, se ha decidido hacer un nuevo chequeo utilizando estas semanas excluidas, es decir, utilizando una ventana de datos que no haya sido utilizada para entrenar el modelo SUR ni para el modelo utilizado en ZIMPL.

En esta nueva comparación, utilizaremos en ambos casos nuestro modelo de asignación, pero, en un caso utilizaremos las ventas predichas por SUR y en el otro las ventas reales. Esto permitirá evaluar cuánto ha afectado el pronóstico de ventas en los resultados obtenidos con el modelo SUR y determinar si es necesario hacer ajustes en la metodología propuesta. En definitiva, este nuevo análisis permitirá obtener una visión más completa y precisa de la eficacia del modelo de asignación propuesto y su capacidad para mejorar la eficiencia de la empresa.

³⁹ Levin (2010).

Gráfico 24: Eficiencia con Ventas Reales vs Eficiencia con Ventas SUR

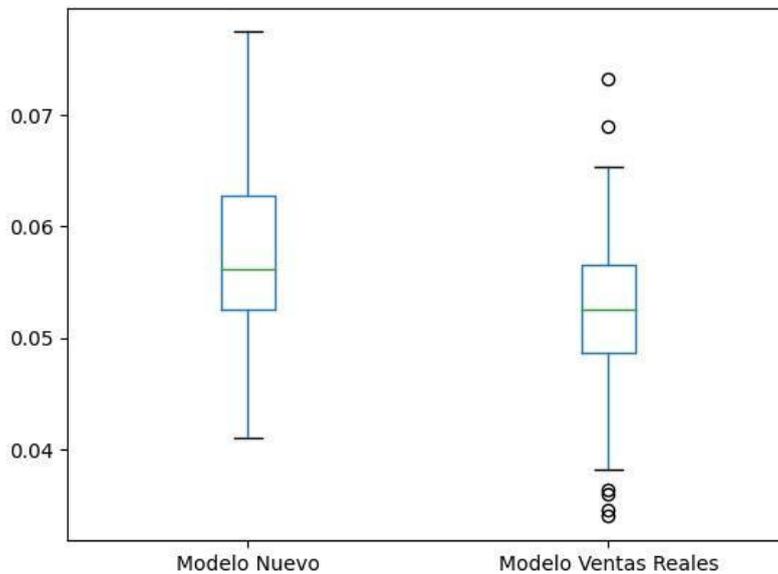
Comparación entre nuestro modelo tomando las ventas reales contra nuestro modelo tomando las ventas pronosticadas por SUR



*Fuente: Elaboración Propia

Al analizar el gráfico 24, se puede observar que la eficiencia del modelo de asignación propuesto tiende a ser menor cuando se utilizan las ventas pronosticadas por SUR en lugar de las ventas reales. Lo cual tiene sentido ya que nuestro modelo de estimación de ventas SUR puede mejorarse ampliamente. Sin embargo, para ver si la diferencia entre los resultados son significativos al 0.05 , se realiza un test de Mann-Whitney como con en la sección 5.1.2 para ver si efectivamente la performance está afectada por nuestra estimación con SUR.

Gráfico 25: Boxplot Nuevo Modelo con Ventas Sur vs Boxplot Modelo con Ventas Reales



*Fuente: Elaboración Propia

El p valor para este caso es de $p = 0.9999851859673662$. En este caso, si se establece un nivel de significancia de 0.05, como el valor p es mayor que este nivel, no se rechaza la hipótesis nula de que no hay diferencia entre las dos muestras. Es decir, al testear las hipótesis, no se puede concluir que son estadísticamente diferentes. Esto sugiere que, aunque el modelo no siempre performa mejor que cuando se utilizan las ventas reales, su desempeño no es significativamente peor en la mayoría

de los casos. Además, como se vio en el gráfico anterior, el modelo propuesto ya es un avance relevante en comparación con lo que se realiza actualmente en la empresa. En términos de porcentaje, la nueva propuesta presenta un adcost un 23% menor en promedio que el método utilizado hoy en día, que en puntos porcentuales representa un 1.86pp menos.

6. Conclusiones

En el desarrollo del proyecto, se llevaron a cabo una serie de ejercicios para abordar de manera integral el tema de estudio. A continuación, se describen los principales pasos emprendidos.

En primer lugar, se procedió a eliminar los outliers presentes en la data recopilada. Esta tarea resultó crucial para mejorar los fundamentos de nuestro modelo predictivo, permitiendo así obtener resultados más confiables en las predicciones futuras. Posteriormente, se enfrentó el desafío de seleccionar un modelo adecuado capaz de predecir múltiples variables dependientes, tomando en cuenta la posible correlación existente entre ellas debido a las posibles correlaciones en los términos de error (SUR). Esta decisión fue determinante para la construcción de un enfoque completo en el análisis de los datos.

Otro aspecto relevante abordado en este proyecto fue el armado de un modelo de asignación de presupuesto que considerara todas las limitaciones y restricciones inherentes al problema, así como las particularidades propias de la empresa en cuestión. La elaboración de este modelo ayudó a optimizar los recursos y maximizar los resultados en el contexto del estudio.

Finalmente, se llevó a cabo un análisis de la validez de nuestro modelo, comparándolo con un enfoque más sencillo (AR(1)), y se realizaron pruebas para respaldar nuestras predicciones de ventas. Esta evaluación nos brindó una base para confiar en nuestro modelo de predicción.

En resumen, los hallazgos obtenidos respaldan la utilidad del modelo de asignación propuesto para optimizar el presupuesto y la planificación de operaciones basados en predicciones de conversiones. Es importante destacar que la evaluación periódica de la calidad de las predicciones generadas por el modelo es fundamental. Esto implica comparar las predicciones con los datos reales observados en períodos anteriores y utilizar métricas de evaluación. Si las predicciones no son satisfactorias, se pueden realizar ajustes adicionales al modelo, como la inclusión de variables adicionales o la consideración de patrones estacionales, para mejorar la calidad de las predicciones. Además, es posible volver a correr el modelo con datos más recientes para obtener predicciones más precisas.

En cuanto a los modelos utilizados, se observó que el modelo SUR (Seemingly Unrelated Regression) parece ser más apropiado para la tarea de predicción de la variable de interés en comparación con el modelo AR(1) para cada variable estudiada. Sin embargo, cabe mencionar que existen otros modelos que podrían explorarse en futuras investigaciones para comparar su desempeño y determinar cuál es el más adecuado en contextos específicos.

Es importante resaltar que, si bien el modelo propuesto representa un avance en comparación con los enfoques existentes, una mejora adicional en la capacidad predictiva puede tener un impacto sustancial. La comparación con un modelo de optimización basado en ventas reales demuestra que aún se puede lograr una mejor predicción con nuestro modelo propuesto. Por lo tanto, se sugiere que futuras líneas de investigación se enfoquen en aspectos como la incorporación de datos más detallados, la exploración de diferentes técnicas de modelado y la consideración de otros factores relevantes que puedan influir en las conversiones y el presupuesto asignado.

En conclusión, los resultados obtenidos respaldan la utilidad y confiabilidad del modelo de asignación de inversión propuesto basado en predicciones de conversiones. Sin embargo, se

reconoce la necesidad de evaluar periódicamente la capacidad predictiva del modelo y explorar otras posibles mejoras. Las futuras investigaciones podrían centrarse en el uso de diferentes estimadores, la inclusión de datos más detallados y la consideración de otros factores relevantes para mejorar las predicciones y avanzar en el campo de la asignación de presupuesto basada en modelos predictivos.

7. Referencias

1. Alamo, J. B. (1952b). Máximos y mínimos de los coeficientes de asimetría y curtosis en poblaciones finitas.
2. Billio, M. (2017). Bayesian nonparametric sparse seemingly unrelated regression model (SUR).
3. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: Analyzing Text with the Natural Language Toolkit. "O'Reilly Media, Inc."
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-excellence-dave-chaffey-pr-smith>
5. Cuadras, C. M. (2016). Problemas de probabilidades y estadística. Vol. 2. Inferencia estadística. Edicions Universitat Barcelona.
6. Dave Bussière (2011) The Day of the Week Effect in Consumer Behavior: Analyzing Utilitarian and Hedonistic Consumer Modes, *Journal of Promotion Management*, 17:4, 418-425, DOI: 10.1080/10496491.2011.620477
7. Del Valle, E. (2016). Google AdWords: Trucos y estrategias para el éxito. Alfaomega Grupo Editor.
8. Ferrán, M. M. (1999). Introducción al SPSS: manipulación de datos y estadística descriptiva. Edicions Universitat Barcelona.
9. Fomby, T. B., Hill, R. C., & Johnson, S. R. (2012). Advanced econometric methods. Springer Science & Business Media.
10. Gorai, A. K. (2022). Optimization techniques and their applications to mine systems.
11. Grigelionis, B. (2012). Student's t-Distribution and related stochastic processes. Springer Science & Business Media.
12. Hayashi, F. (2011). Econometrics. Princeton University Press.
13. Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems.
14. Koch, T. (2001). ZIMPL User Guide. -. <https://opus4.kobv.de/opus4-zib/files/646/ZR-01-20.pdf>
15. Kolman, B., & Beck, R. E. (2014). Elementary Linear programming with applications. Elsevier.
16. Levin, R. I. (2010). Estadística para Administración y Economía.
17. Loureiro, A. I., Miguéis, V. L., & Da Silva, L. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81-93. <https://doi.org/10.1016/j.dss.2018.08.010>
18. Manuals, M. P., & MicroStrategy. (2013). Sales and Distribution Analysis Module Reference for MicroStrategy 9.3.1. MicroStrategy.
19. Marqués, M. P. (2014). Business Intelligence: Técnicas, herramientas y aplicaciones.

20. Martín, D. R. (2022). Estadística inferencial aplicada.: Segunda edición revisada y aumentada. Universidad del Norte.
21. McCracken, M. W. (2020). Tests of conditional predictive ability: Existence, Size, and Power.
22. McCracken, M. W. (2020b). Tests of conditional predictive ability: Existence, Size, and Power.
23. Redondo, A. C. (2019). Google AdWords y sus aplicaciones publicitarias (IFCM008PO). Especialidades formativas. EDITORIAL CEP.
24. Roldán, R. C., Martín, E. M., & Rodríguez, A. R. (2019). Big data con Python: recolección, almacenamiento y proceso.
25. Singh, K., Booma, P. M., & Eaganathan, U. (2020). E-Commerce system for sale prediction using machine learning technique. *Journal of physics*, 1712(1), 012042. <https://doi.org/10.1088/1742-6596/1712/1/012042>
26. Srivastava, V. K., & Giles, D. E. (2020). *Seemingly unrelated regression equations models: Estimation and Inference*. CRC Press.
27. Viana, L. (1988). Los efectos de la brecha cambiaria sobre una economía pequeña y abierta: en las terceras Jornadas Anuales de Economía (Montevideo, 7 al 9 de noviembre de 1988).
28. Vitelli, G. (1990). Las lógicas de la economía argentina: Inflación y crecimiento. <http://ci.nii.ac.jp/ncid/BB19817017>
29. 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>
30. Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov models for time series: An Introduction Using R*, Second Edition. CRC Press.

8.Anexo

8.1 Código

El código de este proyecto tiene la siguiente estructura:

- Exploration Analysis + SUR: colab que contiene todos los scripts que se utilizan para crear las perspectivas y gráficos del capítulo de análisis exploratorio y la estimación del modelo SUR para las ventas
- Problema_Asig_Python.py: donde se presenta el problema de asignación para ser cargado en python
- ZIMPL_Final_xc.txt: archivo que se usa para correr ZIMPL con la data de la estimación de las ventas del modelo predictivo SUR
- Test de Medias.ipynb: donde se realizan todos los test de medias y análisis sobre los resultados.

8.2 Data

Los datos son descargados mediante un informe de google ads de la ecommerce en cuestión. Estos informes son generados por el sistema de seguimiento de conversiones de Google, que utiliza cookies para rastrear la actividad del usuario y la interacción con los anuncios. Se hace un informe de rendimiento de la campaña que nos proporciona toda la información sobre el rendimiento que pedimos para hacer el análisis en cuestión (como por ejemplo la cantidad de impresiones, clics, conversiones, costo por clic (CPC), entre otras) .

Debido a la confidencialidad de la empresa que proporcionó los datos sin procesar, no se puede presentar la información bruta utilizada en este estudio. La empresa ha expresado su preocupación acerca de la exposición de sus números y estrategias a personas ajenas a la misma. Por lo tanto, se presentarán únicamente los resultados del análisis y las inferencias derivadas de ellos. Aunque no se puede proporcionar acceso a los datos, se asegura que los resultados obtenidos son confiables y se han obtenido siguiendo las prácticas éticas y profesionales adecuadas

Tabla Anexo 1: Resumen Base de Datos para Primera etapa de análisis

```

RangeIndex: 21844 entries, 0 to 21843
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Campaign              21844 non-null  object
1   Day                   21844 non-null  object
2   Dayoftheweek         21844 non-null  object
3   N°Day                 21844 non-null  int64
4   Month                 21844 non-null  object
5   Year                  21844 non-null  int64
6   Clicks                21844 non-null  float64
7   Impr                  21844 non-null  float64
8   Cost                  21844 non-null  float64
9   AvgCPC                21844 non-null  float64
10  Adcost                21844 non-null  float64
11  Convvalue              21844 non-null  float64
12  Brecha                 21844 non-null  float64
13  Week_Num              21844 non-null  int64
14  Month_Num              21844 non-null  int64
15  Campaign_Num           21844 non-null  int64
dtypes: float64(7), int64(5), object(4)

```

La tabla 1, generada mediante el uso de la función `df.info()` en Python, proporciona una descripción detallada de los tipos de datos de cada variable en el conjunto de datos. Esta información es fundamental para comprender la naturaleza de las variables y asegurar la integridad de los datos utilizados en nuestro análisis. La ausencia de valores nulos o NA en todas las variables nos proporciona un grado de certeza en cuanto a la integridad de nuestros datos, lo cual es esencial para llevar a cabo análisis confiables y obtener resultados válidos. La presencia de datos nulos o NA puede tener un impacto significativo en el rendimiento y los resultados de nuestros análisis, ya que la falta de información puede afectar la validez de las conclusiones obtenidas. Al garantizar la ausencia de datos faltantes, nos aseguramos de que los análisis posteriores se basen en una base de datos completa y coherente.

8.3 Categorías

Para dar mayor contexto sobre las campañas a las que hacemos referencia durante los análisis, a continuación, se describen brevemente las categorías de productos a las que corresponden dichas campañas:

- Accesorios para Vehículos: Productos relacionados con automóviles, como llantas, herramientas, repuestos, entre otros.
- Alimentos y Bebidas: Productos alimenticios, como comestibles, bebidas, especias, entre otros.
- Animales y Mascotas: Productos relacionados con animales y mascotas, como alimentos, juguetes, accesorios, entre otros.
- Belleza y Cuidado Personal: Productos de belleza y cuidado personal, como maquillaje, cremas, perfumes, entre otros.
- Computación: Productos relacionados con tecnología y computadoras, como laptops, tablets, impresoras, entre otros.

- Consolas y Videojuegos: Productos relacionados con videojuegos y consolas, como juegos, consolas, accesorios, entre otros.
- Deportes y Fitness: Productos relacionados con deportes y ejercicio, como equipamiento deportivo, ropa deportiva, entre otros.
- Electrodomésticos y Aires Acondicionados: Productos electrodomésticos, como neveras, aires acondicionados, lavadoras, entre otros.
- Instrumentos Musicales: Productos relacionados con la música, como instrumentos, accesorios, equipos de sonido, entre otros.
- Juegos y Juguetes: Juguetes y juegos para todas las edades, como muñecas, carros, juegos de mesa, entre otros.
- Libros, Revistas y Comics: Productos relacionados con la lectura, como libros, revistas, cómics, entre otros.
- Salud y Equipamiento Médico: Productos relacionados con la salud, como productos de primeros auxilios, equipamiento médico, entre otros.
- Ropa y Accesorios: Ropa y accesorios de moda, como ropa para hombres, mujeres y niños, accesorios, entre otros.
- Electrónica, Audio y Video: Productos electrónicos, como televisores, equipos de sonido, altavoces, entre otros.
- Cámaras y Accesorios: Productos relacionados con cámaras y fotografía, como cámaras, lentes, trípodes, entre otros.
- Hogar, Muebles y Jardín: Productos relacionados con el hogar, como muebles, decoración, jardinería, entre otros.
- Herramientas y Construcción: Productos relacionados con herramientas y construcción, como herramientas manuales, eléctricas, de medición, entre otros.

8.4 Reporte Soluciones

Presentamos un ejemplo de cómo se ven las soluciones para una semana dada, este reporte se arma en base a las soluciones arrojadas por ZIMPL y a la performance definida por nuestro modelo de predicción SUR. En la tabla 2 se observa un resumen armado de todas las soluciones por día donde solo se ve cuánto se asigna cada día en términos de costo y cual es el monto esperado de ventas para cada día y para toda la semana. A su vez, se define un adcost por día esperado y un adcost total esperado para toda la semana. Agregamos para poder tener la calidad del resultado estimado de la predicción una columna con el R2. Por último, para tener noción de cuanto se le asigna a cada día en términos porcentuales se arma un columna de share de inversión total por día para esta semana dada. En la tabla 3 especifica las soluciones obtenidas para la semana 6 del 2023 por campaña, por día de la semana. Se incluyen las columnas de costo asignado y ventas esperadas de SUR .

Tabla Anexo 2: Resumen Soluciones

Solucion	ConvValue Fcast	Costo Asig	Adcost Fcast	Day	% Inv.	R2
Mon Solucion 1	278976672,2	1.721.689	0,62%	Lunes	18,39%	0.7657107861010687
Tues 1 Solucion 2	269459190,2	1.607.553	0,60%	Martes	17,17%	0.7657107861010687
Wed 1 Solucion 3	258650810,6	1.500.983	0,58%	Miercoles	16,03%	0.7657107861010687
Thurs 1 Solucion 4	229123544,7	1.231.615	0,54%	Jueves	13,16%	0.7657107861010687
Frid 1 Solucion 5	226260408,3	1.306.983	0,58%	Viernes	13,96%	0.7657107861010687
Satur 1 Solucion 6	217821914,9	936.003	0,43%	Sabado	10,00%	0.7657107861010687
Sund 1 Solucion 7	232013124,4	1.056.750	0,46%	Domingo	11,29%	0.7657107861010687
Total Week T-1	1712305665	9.361.577	0,55%		100,00%	0.7657107861010687

*Fuente: Elaboración Propia

Tabla Anexo 3 Soluciones por Categoría por Día

*Fuente: Elaboración Propia

Campaign	Day	Costo Asig	ConvValue Fcast	Adcost Fcast
CEL	Monday	148086,52	32139763,85	0,46%
MODA	Monday	64223,35	7655841,53	0,84%
COMP	Monday	74829,45	8670626,08	0,86%
HERRA	Monday	58865,34	7865676,23	0,75%
ELECT	Monday	44503,21	7481162,56	0,59%
ALIM	Monday	39053,87	5701460,66	0,68%
ACCVE	Monday	38932,84	6106844,81	0,64%
HOGAR	Monday	38220,65	5425834,14	0,70%
BELLE	Monday	26452,36	3941350,18	0,67%
DEPO	Monday	24801,84	4302532,11	0,58%
EAV	Monday	20969,42	4231037,87	0,50%
SALU	Monday	20357,53	3357721,72	0,61%
MASC	Monday	19712,95	3536587,09	0,56%
CAM	Monday	19336,68	2404052,25	0,80%
INST	Monday	13924,47	3044384,03	0,46%
LIBRO	Monday	6042,43	1186187,34	0,51%
CONS	Monday	5858,72	1178761,44	0,50%
INDUST	Monday	4796,34	849239,73	0,56%
JOYAS	Monday	4164,53	774681,33	0,54%
ARTE	Monday	2892,67	507724,08	0,57%
JUE	Monday	1885,23	780270,75	0,24%
BEBES	Monday	635,65	207098,86	0,31%
COLECC	Monday	515,72	184301,74	0,28%
CEL	Tuesday	148086,52	32139763,86	0,46%
MODA	Tuesday	57300,63	7212355,15	0,79%
COMP	Tuesday	44651,22	6392510,76	0,70%
HERRA	Tuesday	34931,47	5816103,46	0,60%
ELECT	Tuesday	44503,21	7481162,56	0,59%

ALIM	Tuesday	30285,96	4900737,81	0,62%
ACCVE	Tuesday	36045,55	5833106,22	0,62%
HOGAR	Tuesday	26236,75	4363643,3	0,60%
BELLE	Tuesday	24704,2	3775112,18	0,65%

DEPO	Tuesday	24801,84	4302532,11	0,58%
EAV	Tuesday	20969,42	4231037,87	0,50%
SALU	Tuesday	20357,53	3357721,73	0,61%
MASC	Tuesday	19712,95	3536587,09	0,56%
CAM	Tuesday	10419,3	1650809,57	0,63%
INST	Tuesday	13924,47	3044384,03	0,46%
LIBRO	Tuesday	6042,43	1186187,34	0,51%
CONS	Tuesday	5858,72	1178761,45	0,50%
INDUST	Tuesday	3918,16	769008,34	0,51%
JOYAS	Tuesday	3843,95	744426,05	0,52%
ARTE	Tuesday	2892,67	507724,08	0,57%
JUE	Tuesday	1885,23	780270,77	0,24%
BEBES	Tuesday	635,65	207098,86	0,31%
COLECC	Tuesday	515,72	184301,75	0,28%
CEL	Wednesday	148086,52	32139763,86	0,46%
MODA	Wednesday	57300,63	7212355,15	0,79%
COMP	Wednesday	44651,22	6392510,76	0,70%
HERRA	Wednesday	31696,72	5498188	0,58%
ELECT	Wednesday	43002,23	7325686,37	0,59%
ALIM	Wednesday	24889,48	4360480,8	0,57%
ACCVE	Wednesday	29623,35	5190163,21	0,57%
HOGAR	Wednesday	21724,37	3911874,58	0,56%
BELLE	Wednesday	19927,91	3297002,43	0,60%
DEPO	Wednesday	20976,58	3918443,89	0,54%
EAV	Wednesday	20969,42	4231037,87	0,50%
SALU	Wednesday	20357,53	3357721,73	0,61%
MASC	Wednesday	17840,51	3345667,62	0,53%
CAM	Wednesday	10412,06	1650112,42	0,63%
INST	Wednesday	13924,47	3044384,03	0,46%
LIBRO	Wednesday	5950,18	1176603,26	0,51%
CONS	Wednesday	5361,82	1128103,79	0,48%
INDUST	Wednesday	3352,31	712343,71	0,47%
JOYAS	Wednesday	3282,06	688159,93	0,48%
ARTE	Wednesday	2892,67	507724,08	0,57%

JUE	Wednesday	1885,23	780270,77	0,24%
BEBES	Wednesday	635,65	207098,86	0,31%
COLECC	Wednesday	515,72	184301,75	0,28%
CEL	Thursday	148086,52	32139763,86	0,46%
MODA	Thursday	57300,63	7212355,15	0,79%
COMP	Thursday	44651,22	6392510,76	0,70%
HERRA	Thursday	31696,72	5498188	0,58%
ELECT	Thursday	32968,41	6226074,07	0,53%

ALIM	Thursday	21029,01	3944272,01	0,53%
ACCVE	Thursday	22965,25	4460435,53	0,51%
HOGAR	Thursday	20580,35	3791223,83	0,54%
BELLE	Thursday	15079,97	2765777,17	0,55%
DEPO	Thursday	16611,84	3439980,04	0,48%
EAV	Thursday	17504,92	3847710,77	0,45%
SALU	Thursday	16869,28	2967595,6	0,57%
MASC	Thursday	14144,49	2940504,9	0,48%
CAM	Thursday	10412,06	1650112,42	0,63%
INST	Thursday	13924,47	3044384,03	0,46%
LIBRO	Thursday	4784,46	1048800,36	0,46%
CONS	Thursday	4370,9	1019451,75	0,43%
INDUST	Thursday	2738,13	644999,54	0,42%
JOYAS	Thursday	2673,65	621448,81	0,43%
ARTE	Thursday	2892,67	507724,08	0,57%
JUE	Thursday	1885,23	780270,77	0,24%
BEBES	Thursday	635,65	207098,86	0,31%
COLECC	Thursday	515,72	184301,75	0,28%
CEL	Friday	148086,52	32139763,86	0,46%
MODA	Friday	57300,63	7212355,15	0,79%
COMP	Friday	44651,22	6392510,76	0,70%
HERRA	Friday	31696,72	5498188	0,58%
ELECT	Friday	32041,31	6118310,57	0,52%
ALIM	Friday	21029,01	3944272,01	0,53%
ACCVE	Friday	22346,11	4388469,75	0,51%
HOGAR	Friday	20580,35	3791223,83	0,54%
BELLE	Friday	14635,39	2714101,07	0,54%
DEPO	Friday	16200,98	3392222,7	0,48%
EAV	Friday	17101,14	3800776,77	0,45%
SALU	Friday	16333,64	2905334,51	0,56%

MASC	Friday	13796,34	2900037,83	0,48%
CAM	Friday	10412,06	1650112,42	0,63%
INST	Friday	13924,47	3044384,03	0,46%
LIBRO	Friday	4673,77	1035933,9	0,45%
CONS	Friday	4276,06	1008428,42	0,42%
INDUST	Friday	2679,28	638159,26	0,42%
JOYAS	Friday	2615,44	614683,15	0,43%
ARTE	Friday	2892,67	507724,08	0,57%
JUE	Friday	1885,23	780270,77	0,24%
BEBES	Friday	635,65	207098,86	0,31%
COLECC	Friday	515,72	184301,75	0,28%
CEL	Saturday	148086,52	32139763,86	0,46%

MODA	Saturday	57300,63	7212355,15	0,79%
COMP	Saturday	44651,22	6392510,76	0,70%
HERRA	Saturday	31696,72	5498188	0,58%
ELECT	Saturday	29270,7	5788810,15	0,51%
ALIM	Saturday	21029,01	3944272,01	0,53%
ACCVE	Saturday	20963,84	4224821,19	0,50%
HOGAR	Saturday	20580,35	3791223,83	0,54%
BELLE	Saturday	14243,58	2668075,14	0,53%
DEPO	Saturday	14964,26	3245139,54	0,46%
EAV	Saturday	15881,45	3655717,24	0,43%
SALU	Saturday	14745,22	2716433,07	0,54%
MASC	Saturday	12748,15	2775375,87	0,46%
CAM	Saturday	10412,06	1650112,42	0,63%
INST	Saturday	13924,47	3044384,03	0,46%
LIBRO	Saturday	4339,45	996172,95	0,44%
CONS	Saturday	3988,76	974259,04	0,41%
INDUST	Saturday	2582,64	626758,51	0,41%
JOYAS	Saturday	2439,14	593715,09	0,41%
ARTE	Saturday	2693,6	483906,81	0,56%
JUE	Saturday	1885,23	780270,77	0,24%
BEBES	Saturday	635,65	207098,86	0,31%
COLECC	Saturday	515,72	184301,75	0,28%
CEL	Sunday	148086,52	32139763,85	0,46%
MODA	Sunday	57300,63	7212355,15	0,79%
COMP	Sunday	44651,22	6392510,76	0,70%
HERRA	Sunday	31696,72	5498188	0,58%

<i>ELECT</i>	<i>Sunday</i>	33912,04	6334558,03	0,54%
<i>ALIM</i>	<i>Sunday</i>	21029,01	3944272,02	0,53%
<i>ACCVE</i>	<i>Sunday</i>	23594,68	4532797,68	0,52%
<i>HOGAR</i>	<i>Sunday</i>	20580,35	3791223,83	0,54%
<i>BELLE</i>	<i>Sunday</i>	15533,11	2817872,21	0,55%
<i>DEPO</i>	<i>Sunday</i>	17028,59	3487891,43	0,49%
<i>EAV</i>	<i>Sunday</i>	17913,79	3894715,76	0,46%
<i>SALU</i>	<i>Sunday</i>	17416,5	3030506,93	0,57%
<i>MASC</i>	<i>Sunday</i>	14497,57	2981097,39	0,49%
<i>CAM</i>	<i>Sunday</i>	10412,06	1650112,42	0,63%
<i>INST</i>	<i>Sunday</i>	13924,47	3044384,03	0,46%
<i>LIBRO</i>	<i>Sunday</i>	4896,56	1061687,3	0,46%
<i>CONS</i>	<i>Sunday</i>	4466,79	1030476,47	0,43%
<i>INDUST</i>	<i>Sunday</i>	2797,6	651837,38	0,43%
<i>JOYAS</i>	<i>Sunday</i>	2732,51	628215,64	0,43%
<i>ARTE</i>	<i>Sunday</i>	2892,67	507724,08	0,57%
<i>JUE</i>	<i>Sunday</i>	1885,23	780270,77	0,24%
<i>BEBES</i>	<i>Sunday</i>	635,65	207098,86	0,31%
<i>COLECC</i>	<i>Sunday</i>	515,72	184301,75	0,28%

8.5 Salidas Modelo Predictivo

System OLS Estimation Summary

```

=====
Overall R-squared:          0.7213
McElroy's R-squared:      0.7396
Judge's (OLS) R-squared:   0.7213
Berndt's R-squared:       0.8220
Dhrymes's R-squared:      0.6213
=====

```

Equation: equation1, Dependent Variable: ConvalueAjustado_COMP

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	205.27	4.2469	8.335	0.0000	196.95 213.60
CostoAjustado	14.86	3.7801	3.922	0.0001	0.41 22.35
Brecha Diff	-71.47	3.5424	1.580	0.0039	370.8 467.95
AvgCPC	-2.09	7.1014	2.954	0.0031	62.76 158.74
Week_Num	-4.79	7.2248	6.638	0.0000	2.74 28.40
Month_Num	11.93	3.1223	6.190	0.0000	2.94 11.35
Adcost	-22.87	1.6715	1.717	0.0018	22.78 45.67
Monday	23.12	3.7881	9.667	0.0000	1.25 2.35
Tuesday	11.03	3.8820	6.659	0.0000	9.58 20.11
Wednesday	15.02	3.9247	6.198	0.0000	4.44 11.45
Thursday	19.84	4.1044	4.000	0.0000	4.29 11.53
Friday	9.30	3.9643	3.479	0.0000	2.05 9.34
Saturday	6.55	3.7940	2.539	0.0000	7.80 9.29
Sunday	6.09	3.7215	1.750	0.0000	7.36 8.82

Equation: equation2, Dependent Variable: ConvalueAjustado_HOGAR

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	147.39	4.3286	4.050	0.0000	138.91 155.88
CostoAjustado	6.10	3.6086	1.692	0.0006	0.9664 13.179
Brecha Diff	-53.76	4.0504	1.051	0.0054	139.44 2063.09
AvgCPC	-4.39	5.8913	2.454	0.0000	25.54 133.23
Week_Num	-5.58	6.9766	6.060	0.0000	6.95 24.21
Month_Num	12.20	3.0156	7.322	0.0000	1.61 9.79
Adcost	-11.91	1.3475	1.423	0.0047	24.55 77.22
Monday	17.23	3.9641	3.191	0.0000	1.15 5.31
Tuesday	5.17	4.0833	2.682	0.0000	1.09 4.25
Wednesday	2.18	4.1615	2.421	0.0000	6.10 11.26
Thursday	10.14	4.3027	6.676	0.0000	8.06 9.23
Friday	10.09	4.1696	6.354	0.0000	1.01 9.18
Saturday	-1.32	3.8214	2.004	0.0070	9.56 21.10
Sunday	9.52	3.8823	2.528	0.0000	8.76 10.02

Equation: equation3, Dependent Variable: ConvalueAjustado_CEL

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	150,57	4,2106	7,852	0,0000	138,59	446,55
CostoAjustado	6,27	4,4452	8,464	0,0008	1,15	15,71
Brecha Diff	-128,86	6,3422	2,505	0,0000	166,00	2459,30
AvgCPC	-6,27	8,7031	4,447	0,0000	25,7	158,82
Week_Num	-3,76	6,2170	6,554	0,0000	8,28	28,86
Month_Num	15,06	4,4189	9,188	0,0000	1,92	11,67
Adcost	-13,80	2,7856	2,828	0,0015	29,26	92,05
Monday	25,10	4,9741	4,083	0,0000	1,37	6,33
Tuesday	6,27	5,1237	3,365	0,0000	1,30	5,07
Wednesday	12,55	5,2218	2,123	0,0000	7,27	13,42
Thursday	12,55	5,3990	8,377	0,0000	9,61	11,00
Friday	6,27	8,9107	5,890	0,0000	1,20	10,94
Saturday	-31,37	7,3923	2,792	0,0151	11,40	25,15
Sunday	6,27	4,5280	4,299	0,0000	10,44	11,94

Equation: equation4, Dependent Variable: ConvalueAjustado_ACCVE

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	101.35	3.3556	6.258	0.0000	110.45	355.88
CostoAjustado	50.10	3.5426	6.745	0.0006	0,98	12,55
Brecha Diff	-13.60	5.0544	1.996	0.0594	132,47	1959,94
AvgCPC	-6.85	6.8913	3.544	0.0000	24,26	126,57
Week_Num	-4.58	4.9546	5.223	0.0000	6.60	23.52
Month_Num	2.21	3.5216	7.322	0.0000	1.61	9.79
Adcost	-8.45	2.3475	2.254	0.0012	24.55	77.22
Monday	6.88	3.9641	3.254	0.0000	1.15	5.31
Tuesday	2.69	4.0833	2.682	0.0000	1.09	4.25
Wednesday	2.54	4.1615	1.692	0.0000	6.10	11.26
Thursday	5.25	4.3027	6.676	0.0000	8.06	9.23
Friday	7.87	7.1014	4.694	0.0000	1.01	9.18
Saturday	1.05	5.8913	2.225	0.0120	9.56	21.10
Sunday	1.56	3.6086	3.426	0.0000	8.76	10.02

Equation: equation5, Dependent Variable: ConvalueAjustado_ELECT

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	94,97	2,6558	4,953	0,0000	87,42 281,66
CostoAjustado	93,96	2,8038	5,338	0,0005	0,73 9,91
Brecha Diff	-88,20	4,0003	1,580	0,0470	104,84 1551,20
AvgCPC	-3,96	5,4541	2,805	0,0000	19,20 100,17
Week_Num	-2,37	3,9213	4,134	0,0000	5,23 18,20
Month_Num	9,50	2,7872	5,795	0,0000	1,21 7,36
Adcost	-8,71	1,8579	1,784	0,0009	18,46 58,06
Monday	15,83	3,1374	2,575	0,0000	0,86 3,99
Tuesday	3,96	3,2317	2,123	0,0000	0,82 3,20
Wednesday	7,91	3,2936	1,339	0,0000	4,59 8,47
Thursday	6,85	3,4054	5,284	0,0000	6,06 6,94
Friday	3,96	5,6204	3,715	0,0000	0,76 6,90
Saturday	-1,79	4,6627	1,761	0,0095	7,19 15,86
Sunday	3,96	2,8560	2,712	0,0000	6,59 7,53

Equation: equation6, Dependent Variable: ConvalueAjustado_MASC

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	61,81	2,3306	4,346	0,0000	76,71 247,17
CostoAjustado	38,89	2,4605	4,685	0,0004	0,64 8,70
Brecha Diff	-15,97	3,5105	1,386	0,0413	92,01 1361,26
AvgCPC	-3,47	4,7863	2,461	0,0000	16,85 87,91
Week_Num	-2,08	3,4412	3,628	0,0000	4,59 15,97
Month_Num	8,33	2,4459	5,085	0,0000	1,06 6,46
Adcost	-7,64	1,6304	1,566	0,0008	16,20 50,95
Monday	10,42	2,7532	2,260	0,0000	0,76 3,50
Tuesday	3,95	2,8360	1,863	0,0000	0,72 2,80
Wednesday	6,95	2,8904	1,175	0,0000	4,02 7,43
Thursday	6,04	2,9884	4,637	0,0000	5,32 6,09
Friday	3,19	4,9322	3,260	0,0000	0,67 6,06
Saturday	1,39	4,0918	1,545	0,0083	6,31 13,92
Sunday	2,24	2,5063	2,380	0,0000	5,78 6,61

Equation: equation7, Dependent Variable: ConvalueAjustado_EAV

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	120,02	4,5250	8,439	0,0000	148,94	479,90
CostoAjustado	75,52	4,7772	9,096	0,0008	1,24	16,88
Brecha Diff	-31,02	6,8158	2,692	0,0801	178,63	2642,95
AvgCPC	-6,74	9,2928	4,779	0,0000	32,72	170,68
Week_Num	-4,05	6,6812	7,043	0,0000	8,90	31,01
Month_Num	16,18	4,7488	9,874	0,0000	2,06	12,54
Adcost	-14,83	3,1656	3,039	0,0016	31,45	98,92
Monday	20,23	5,3455	4,388	0,0000	1,47	6,80
Tuesday	7,66	5,5063	3,617	0,0000	1,40	5,44
Wednesday	13,48	5,6117	2,282	0,0000	7,81	14,42
Thursday	11,72	5,8021	9,003	0,0000	10,33	11,82
Friday	6,20	9,5762	6,330	0,0000	1,29	11,76
Saturday	2,70	7,9444	3,000	0,0162	12,25	27,03
Sunday	4,34	4,8662	4,620	0,0000	11,22	12,84

Equation: equation8, Dependent Variable: ConvalueAjustado_DEPO

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	94,00	3,5439	6,609	0,0000	116,65	375,85
CostoAjustado	59,14	3,7414	7,124	0,0006	0,97	13,22
Brecha Diff	-24,29	5,3381	2,108	0,0627	139,90	2069,94
AvgCPC	-5,28	7,2781	3,743	0,0000	25,62	133,67
Week_Num	-3,17	5,2327	5,516	0,0000	6,97	24,29
Month_Num	12,67	3,7193	7,733	0,0000	1,62	9,82
Adcost	-11,62	2,4793	2,381	0,0013	24,63	77,48
Monday	15,84	4,1866	3,437	0,0000	1,15	5,33
Tuesday	6,00	4,3125	2,833	0,0000	1,09	4,26
Wednesday	10,56	4,3951	1,787	0,0000	6,12	11,30
Thursday	9,18	4,5442	7,051	0,0000	8,09	9,26
Friday	4,86	7,5000	4,957	0,0000	1,01	9,21
Saturday	2,11	6,2220	2,350	0,0127	9,59	21,17
Sunday	3,40	3,8111	3,618	0,0000	8,79	10,05

Equation: equation9, Dependent Variable: ConvalueAjustado_CAM

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	109,99	4,1470	7,734	0,0000	136,50 439,81
CostoAjustado	72,21	4,3781	8,336	0,0007	1,13 15,47
Brecha Diff	-28,42	6,2465	2,467	0,0734	163,71 2422,18
AvgCPC	-6,18	8,5166	4,380	0,0000	29,99 156,42
Week_Num	-3,71	6,1231	6,455	0,0000	8,16 28,02
Month_Num	14,83	4,3522	9,049	0,0000	1,89 11,49
Adcost	-13,59	2,9011	2,786	0,0015	28,82 90,66
Monday	18,54	4,8990	4,021	0,0000	1,35 6,23
Tuesday	7,02	5,0463	3,315	0,0000	1,28 4,99
Wednesday	12,36	5,1430	2,091	0,0000	7,16 13,22
Thursday	10,74	5,3175	8,251	0,0000	9,46 10,84
Friday	5,68	8,7762	5,801	0,0000	1,19 10,78
Saturday	2,47	7,2807	2,750	0,0148	11,22 24,77
Sunday	3,98	4,4597	4,234	0,0000	10,28 11,76

Equation: equation10, Dependent Variable: ConvalueAjustado_EAV

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	137,67	5,1906	9,680	0,0000	170,85 550,49
CostoAjustado	86,62	5,4799	10,433	0,0009	1,42 19,37
Brecha Diff	-65,58	7,8184	3,088	0,0000	204,91 3031,72
AvgCPC	-7,73	10,6598	5,482	0,0000	37,53 195,78
Week_Num	-4,64	7,6640	8,079	0,0000	10,21 35,58
Month_Num	18,56	5,4474	11,326	0,0000	2,37 14,39
Adcost	-17,02	3,6312	3,487	0,0019	36,08 113,48
Monday	23,20	6,1319	5,033	0,0000	1,69 7,80
Tuesday	8,79	6,3162	4,149	0,0000	1,60 6,25
Wednesday	15,47	6,4372	2,617	0,0000	8,96 16,55
Thursday	13,44	6,6556	10,327	0,0000	11,84 13,56
Friday	7,12	10,9848	7,261	0,0000	1,48 13,49
Saturday	3,09	9,1129	3,442	0,0186	14,05 31,01
Sunday	4,98	5,5819	5,299	0,0000	12,87 14,72

Equation: equation11, Dependent Variable: ConvalueAjustado_BELLE

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	82,37	3,1056	5,792	0,0000	102,22	329,36
CostoAjustado	51,83	3,2786	6,242	0,0006	0,85	11,59
Brecha Diff	-21,29	4,6778	1,847	0,0550	122,60	1813,90
AvgCPC	-4,63	6,3778	3,280	0,0000	22,46	117,14
Week_Num	-2,78	4,5854	4,834	0,0000	6,11	21,29
Month_Num	11,11	3,2592	6,776	0,0000	1,42	8,61
Adcost	-10,18	2,1726	2,086	0,0011	21,58	67,89
Monday	13,88	3,6687	3,012	0,0000	1,01	4,67
Tuesday	5,26	3,7790	2,482	0,0000	0,96	3,74
Wednesday	9,25	3,8514	1,566	0,0000	5,36	9,90
Thursday	8,04	3,9821	6,179	0,0000	7,09	8,12
Friday	4,26	6,5723	4,344	0,0000	2,89	8,07
Saturday	1,85	5,4523	2,059	0,0111	8,41	18,55
Sunday	2,98	3,3397	3,171	0,0000	7,70	8,81

Equation: equation12, Dependent Variable: ConvalueAjustado_MODA

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	89,43	3,3717	6,288	0,0000	110,98	357,58
CostoAjustado	56,27	3,5596	6,777	0,0006	0,92	12,58
Brecha Diff	-23,11	5,0786	2,006	0,0597	133,10	1969,31
AvgCPC	-5,02	6,9243	3,561	0,0000	24,38	127,17
Week_Num	-3,01	4,9783	5,248	0,0000	6,63	23,11
Month_Num	12,06	3,5385	7,357	0,0000	1,54	9,35
Adcost	-11,05	2,3587	2,265	0,0012	23,43	73,71
Monday	15,07	3,9831	3,270	0,0000	1,10	5,07
Tuesday	5,71	4,1028	2,695	0,0000	1,04	4,06
Wednesday	10,05	4,1814	1,700	0,0000	5,82	10,75
Thursday	8,73	4,3233	6,708	0,0000	7,69	8,81
Friday	4,62	7,1354	4,716	0,0000	0,96	8,76
Saturday	2,01	5,9195	2,236	0,0121	9,13	20,14
Sunday	3,24	3,6259	3,442	0,0000	8,36	9,56

Equation: equation13, Dependent Variable: ConvalueAjustado_SALU

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	58,24	2,1960	4,095	0,0000	72,28	232,90
CostoAjustado	36,65	2,3184	4,414	0,0004	0,60	8,19
Brecha Diff	-5,05	3,3078	1,306	0,0389	86,69	1282,65
AvgCPC	-3,27	4,5099	2,319	0,0000	15,88	82,83
Week_Num	-1,96	3,2425	3,418	0,0000	4,32	15,05
Month_Num	7,85	2,3047	4,792	0,0000	1,00	6,09
Adcost	-7,20	1,5363	1,475	0,0008	15,26	48,01
Monday	9,82	2,5942	2,130	0,0000	0,71	3,30
Tuesday	3,72	2,6723	1,755	0,0000	0,68	2,64
Wednesday	6,54	2,7234	1,107	0,0000	3,79	7,00
Thursday	5,69	2,8158	4,369	0,0000	5,01	5,74
Friday	3,01	4,6474	3,072	0,0000	0,63	5,71
Saturday	1,31	3,8555	1,456	0,0079	5,94	13,12
Sunday	2,11	2,3616	2,242	0,0000	5,45	6,23

Equation: equation14, Dependent Variable: ConvalueAjustado_INST

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	31,54	1,1893	2,218	0,0000	39,15	126,14
CostoAjustado	19,85	1,2556	2,391	0,0002	0,33	4,44
Brecha Diff	-8,15	1,7915	0,707	0,0211	46,95	694,67
AvgCPC	-1,77	2,4425	1,256	0,0000	8,60	44,86
Week_Num	-1,06	1,7561	1,851	0,0000	2,34	8,15
Month_Num	4,25	1,2482	2,595	0,0000	0,54	3,30
Adcost	-3,90	0,8320	0,799	0,0004	8,27	26,00
Monday	5,32	1,4050	1,153	0,0000	0,39	1,79
Tuesday	2,01	1,4473	0,951	0,0000	0,37	1,43
Wednesday	3,54	1,4750	0,600	0,0000	2,05	3,79
Thursday	3,08	1,5250	2,366	0,0000	2,71	3,11
Friday	1,63	2,5170	1,664	0,0000	0,34	3,09
Saturday	0,71	2,0881	0,789	0,0043	3,22	7,10
Sunday	1,14	1,2790	1,214	0,0000	2,95	3,37

Equation: equation15, Dependent Variable: ConvalueAjustado_JOYAS

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	59,11	2,2287	4,156	0,0000	73,36	236,37
CostoAjustado	37,19	2,3529	4,480	0,0004	0,61	8,32
Brecha Diff	-15,28	3,3570	1,326	0,0395	87,98	1301,75
AvgCPC	-3,32	4,5771	2,354	0,0000	16,11	84,06
Week_Num	-1,99	3,2907	3,469	0,0000	4,39	15,28
Month_Num	7,97	2,3390	4,863	0,0000	1,02	6,18
Adcost	-7,31	1,5592	1,497	0,0008	15,49	48,72
Monday	9,96	2,6329	2,161	0,0000	0,73	3,35
Tuesday	3,77	2,7120	1,781	0,0000	0,69	2,68
Wednesday	6,64	2,7640	1,124	0,0000	3,85	7,10
Thursday	5,77	2,8578	4,434	0,0000	5,09	5,82
Friday	3,06	4,7166	3,118	0,0000	0,64	5,79
Saturday	1,33	3,9129	1,478	0,0080	6,03	13,31
Sunday	2,14	2,3968	2,275	0,0000	5,53	6,32

Equation: equation16, Dependent Variable: ConvalueAjustado_JUE

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	89,51	3,3747	6,294	0,0000	111,08	357,90
CostoAjustado	56,32	3,5627	6,783	0,0006	0,92	12,59
Brecha Diff	-23,13	5,0831	2,007	0,0000	133,22	1971,08
AvgCPC	-5,03	6,9305	3,564	0,0000	24,40	127,29
Week_Num	-3,02	4,9828	5,253	0,0000	6,64	23,13
Month_Num	12,07	3,5416	7,364	0,0000	1,54	9,35
Adcost	-11,06	2,3608	2,267	0,0012	23,46	73,78
Monday	15,09	3,9866	3,273	0,0000	1,10	5,07
Tuesday	5,71	4,1065	2,697	0,0000	1,04	4,06
Wednesday	10,06	4,1852	1,702	0,0000	5,83	10,76
Thursday	8,74	4,3272	6,714	0,0000	7,70	8,82
Friday	4,63	7,1418	4,721	0,0000	0,96	8,77
Saturday	2,01	5,9248	2,238	0,0121	9,13	20,16
Sunday	3,24	3,6291	3,445	0,0000	8,37	9,57

Equation: equation17, Dependent Variable: ConvalueAjustado_ALIM

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	72,30	2,7258	5,083	0,0000	89,72	289,09
CostoAjustado	45,49	2,8777	5,479	0,0005	0,75	10,17
Brecha Diff	-18,68	4,1058	1,621	0,0000	107,61	592,08
AvgCPC	-4,06	5,5979	2,879	0,0000	19,71	102,81
Week_Num	-2,44	4,0247	4,243	0,0000	5,36	18,68
Month_Num	9,75	2,8606	5,948	0,0000	1,24	7,55
Adcost	-8,94	1,9069	1,831	0,0010	18,95	59,59
Monday	12,18	3,2201	2,643	0,0000	0,89	4,10
Tuesday	4,61	3,3169	2,179	0,0000	0,84	3,28
Wednesday	8,12	3,3804	1,374	0,0000	4,71	8,69
Thursday	7,06	3,4951	5,423	0,0000	6,22	7,12
Friday	3,74	5,7686	3,813	0,0000	0,78	7,08
Saturday	-1,62	4,7856	1,807	0,0097	7,38	16,28
Sunday	2,62	2,9313	2,783	0,0000	6,76	7,73

Equation: equation18, Dependent Variable: ConvalueAjustado_COLLEC

Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI	
Clicks	22,83	5,8607	1,605	0,0000	28,33	91,28
CostoAjustado	14,36	2,9086	1,730	0,0002	0,24	3,21
Brecha Diff	-5,90	2,2964	0,512	6,0580	33,98	502,70
AvgCPC	-1,28	3,7675	0,909	0,0000	6,22	32,46
Week_Num	-0,77	2,2708	1,340	0,0000	1,69	5,90
Month_Num	3,08	5,9032	1,878	0,0000	0,39	2,39
Adcost	-2,82	5,6021	0,578	0,0003	5,98	18,82
Monday	3,85	6,0167	0,835	0,0000	0,28	1,29
Tuesday	1,46	4,0473	0,688	0,0000	0,27	1,04
Wednesday	2,56	2,0674	0,434	0,0000	1,49	2,74
Thursday	2,23	1,1036	1,712	0,0000	1,96	2,25
Friday	1,18	2,8214	1,204	0,0000	0,25	2,24
Saturday	0,51	1,5110	0,571	0,0031	2,33	5,14
Sunday	0,83	3,9256	0,879	0,0000	2,13	2,44

Equation: equation19, Dependent Variable: ConvalueAjustado_CONS

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	49,53	1,8673	3,482	0,0000	61,46	198,04
CostoAjustado	31,16	1,9714	3,754	0,0003	0,51	6,97
Brecha Diff	-12,80	2,8127	1,111	0,0331	73,72	1090,68
AvgCPC	-2,78	3,8349	1,972	0,0000	13,50	70,43
Week_Num	-1,67	2,7572	2,907	0,0000	3,67	12,80
Month_Num	6,68	1,9597	4,075	0,0000	0,85	5,18
Adcost	-6,12	1,3064	1,254	0,0007	12,98	40,82
Monday	8,35	2,2060	1,811	0,0000	0,61	2,81
Tuesday	3,16	2,2723	1,492	0,0000	0,58	2,25
Wednesday	5,56	2,3158	0,942	0,0000	3,22	5,95
Thursday	4,84	2,3944	3,715	0,0000	4,26	4,88
Friday	2,56	3,9518	2,612	0,0000	0,53	4,85
Saturday	1,11	3,2784	1,238	0,0067	5,05	11,15
Sunday	1,79	2,0081	1,907	0,0000	4,63	5,30

Equation: equation20, Dependent Variable: ConvalueAjustado_INSTU

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	22,83	2,2029	4,108	0,0000	72,51	233,63
CostoAjustado	14,36	2,3257	4,428	0,0004	0,60	8,22
Brecha Diff	-5,90	3,3181	1,310	0,0390	86,96	1286,67
AvgCPC	-1,28	4,5240	2,327	0,0000	15,93	83,09
Week_Num	-0,77	3,2526	3,429	0,0000	4,33	15,10
Month_Num	3,08	2,3119	4,807	0,0000	1,00	6,11
Adcost	-2,82	1,5411	1,480	0,0008	15,31	48,16
Monday	3,85	2,6024	2,136	0,0000	0,72	3,31
Tuesday	1,46	2,6806	1,761	0,0000	0,68	2,65
Wednesday	2,56	2,7320	1,111	0,0000	3,80	7,02
Thursday	2,23	2,8247	4,383	0,0000	5,03	5,76
Friday	1,18	4,6620	3,082	0,0000	0,63	5,73
Saturday	2,51	3,8676	1,461	0,0079	5,96	13,16
Sunday	1,83	2,3690	2,249	0,0000	5,46	6,25

Equation: equation21, Dependent Variable: ConvalueAjustado_LIBRO

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	51,84	2,5208	4,701	0,0000	82,97	267,35
CostoAjustado	42,07	2,6613	5,067	0,0005	0,69	9,41
Brecha Diff	-17,28	3,7970	1,499	0,0446	99,51	472,37
AvgCPC	-3,76	5,1770	2,662	0,0000	18,23	95,08
Week_Num	-2,25	3,7221	3,924	0,0000	4,96	17,28
Month_Num	9,01	2,6456	5,501	0,0000	1,15	6,99
Adcost	-8,26	1,7635	1,693	0,0009	17,52	55,11
Monday	11,27	2,9780	2,445	0,0000	0,82	9,79
Tuesday	4,27	3,0675	2,015	0,0000	0,78	3,03
Wednesday	7,51	3,1263	1,271	0,0000	4,35	8,04
Thursday	6,53	3,2323	5,015	0,0000	2,75	6,59
Friday	3,46	5,3348	3,526	0,0000	0,72	6,55
Saturday	-2,75	4,4258	1,671	0,0090	6,82	15,06
Sunday	-1,76	2,7109	2,574	0,0000	6,25	7,15

Equation: equation22, Dependent Variable: ConvalueAjustado_ARTE

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	24,25	1,1794	2,200	0,0000	38,82	125,09
CostoAjustado	19,68	1,2452	2,371	0,0002	0,32	4,40
Brecha Diff	-8,08	1,7765	0,702	0,0209	46,56	288,89
AvgCPC	-1,76	2,4222	1,246	0,0000	8,53	44,49
Week_Num	-1,05	1,7415	1,836	0,0000	2,32	8,08
Month_Num	4,22	1,2378	2,574	0,0000	0,54	3,27
Adcost	-3,87	0,8251	0,792	0,0004	8,20	25,78
Monday	5,27	1,3933	1,144	0,0000	0,38	1,77
Tuesday	2,00	1,4352	0,943	0,0000	0,36	1,42
Wednesday	3,51	1,4627	0,595	0,0000	2,04	3,76
Thursday	3,05	1,5123	2,347	0,0000	2,69	3,08
Friday	1,62	2,4960	1,650	0,0000	0,34	3,07
Saturday	-0,70	2,0707	0,782	0,0042	3,19	7,05
Sunday	-1,13	1,2684	1,204	0,0000	2,93	3,35

Equation: equation23, Dependent Variable: ConvalueAjustado_BEBES

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Clicks	31,14	1,5142	2,824	0,0000	49,84	160,59
CostoAjustado	25,27	1,5985	3,044	0,0003	0,41	5,65
Brecha Diff	-10,38	2,2807	0,901	0,0268	59,77	884,39
AvgCPC	-2,26	3,1096	1,599	0,0000	10,95	57,11
Week_Num	-1,35	2,2357	2,357	0,0000	2,98	10,38
Month_Num	5,41	1,5891	3,304	0,0000	0,69	4,20
Adcost	-4,96	1,0593	1,017	0,0005	10,52	33,10
Monday	6,77	1,7887	1,468	0,0000	0,49	2,28
Tuesday	2,56	1,8425	1,210	0,0000	0,47	1,82
Wednesday	4,51	1,8778	0,763	0,0000	2,61	4,83
Thursday	3,92	1,9415	3,012	0,0000	3,46	3,96
Friday	2,08	3,2044	2,118	0,0000	0,43	3,94
Saturday	-1,45	2,6584	1,004	0,0054	4,10	9,05
Sunday	-2,26	1,6283	1,546	0,0000	3,76	4,30