



Facultad de Ciencias Económicas y Empresariales

Modelado de respuesta y optimización presupuestaria en campañas de Social Media Ads con un enfoque data-driven

Autor: Claudia García García

Director: Antonio Tena Blázquez

Resumen

El presente trabajo aborda el problema de la asignación óptima presupuestaria en campañas de *Social Media Ads*, incorporando la complejidad inherente del comportamiento de los entornos digitales. A partir del análisis realizado, se observa que el aumento de la inversión no se traduce en mejoras proporcionales en los resultados, evidenciando la existencia de rendimientos decrecientes y diferencias significativas entre canales.

Sobre esta base, se propone un modelo de optimización que no solo permite anticipar el comportamiento del sistema, sino también estructurar la toma de decisiones bajo objetivos y restricciones realistas. Los resultados ponen de manifiesto que no existe una asignación óptima universal, sino soluciones condicionadas por el equilibrio entre volumen y eficiencia, así como por la capacidad diferencial de cada canal para transformar la inversión en resultados. De este modo, el modelo no solo optimiza resultados, sino que redefine el proceso de asignación como un problema estructurado con múltiples criterios y restricciones.

En este sentido, el trabajo aporta un marco analítico que permite abordar esta complejidad de forma estructurada y fundamentada, facilitando una toma de decisiones más coherente en entornos caracterizados por la incertidumbre y la no linealidad.

Palabras clave: *Social Media Ads*, optimización presupuestaria, saturación publicitaria, modelos predictivos, eficiencia publicitaria

Abstract

This paper addresses the problem of optimal budget allocation in *Social Media Ads* campaigns, incorporating the inherent complexity of the behavior of digital environments. Based on the analysis carried out, it is observed that the increase in investment does not translate into proportional improvements in results, evidencing the existence of diminishing returns and significant differences between channels.

On this basis, an optimization model is proposed that not only allows anticipating the behavior of the system, but also structuring decision-making under realistic objectives and constraints. The results show that there is no universal optimal allocation, but rather solutions conditioned by the balance between volume and efficiency, as well as by the differential capacity of each channel to transform investment into results. In this way, the model not only optimizes results, but also redefines the assignment process as a structured problem with multiple criteria and constraints.

Keywords: *Social Media Ads*, budget optimization, advertising saturation, predictive models, advertising efficiency

Índice de contenidos

1. Introducción	8
1.1 Relevancia del tema	8
1.2 Objetivos del estudio.....	8
1.3 Metodología y estructura.....	9
2. Marco teórico y revisión bibliográfica.....	10
2.1 Evolución de la inversión en medios digitales y Social Media Ads	10
2.2 Fundamentos de marketing mix y elasticidades publicitarias.....	12
2.3 Dinámicas de respuesta publicitaria: persistencia, saturación y efectos cruzados.	14
2.4 Optimización de presupuesto y técnicas robusta	16
3. Descripción y preparación de los datos	18
3.1 Descripción del dataset	18
3.2 Limpieza y creación de variables	21
3.3 Análisis exploratorio	25
4. Metodología analítica	31
4.1 Modelos predictivos de conversión.....	31
4.2 Evaluación de modelos predictivos.....	36
4.3 Curvas de saturación y elasticidad	38
4.4 Simulación de escenarios	41
4.5 Análisis de duración.....	45
5. Aplicación práctica y recomendaciones.....	48
5.1 Formulación del problema de optimización presupuestaria	48
5.2 Resultados de la optimización presupuestaria.....	49
6. Limitaciones del estudio.....	51

7. Conclusiones.....	53
8. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado	55
9. Referencias bibliográficas.....	57
10. Anexos.....	61
10.1 Anexo 1. Glosario de términos en inglés	61
10.2 Anexo 2. Modelos clásicos de respuesta y asignación presupuestaria	61
10.3 Anexo 3. Interpretación gráficos boxplot.....	62
10.4 Anexo 4. Visualización de valores atípicos	62
10.5 Anexo 5. Código empleado para el trabajo	63

Índice de tablas

Tabla 1. Tipología de variables según su función analítica.....	20
Tabla 2. Resumen de las variables empleadas en el estudio.....	20
Tabla 3. Estructura inicial del dataset antes del preprocesamiento	22
Tabla 4. Estructura del dataset tras la estandarización de los nombres de las variables.....	22
Tabla 5. Estadística descriptiva de las variables clave del rendimiento publicitario	24
Tabla 6. Matriz de correlación de variables numéricas del modelo... ..	32
Tabla 7. Hiperparámetros del modelo Random Forest	33
Tabla 8. Hiperparámetros del modelo XGBoost.....	34
Tabla 9. Resumen de algoritmos con sus resultados comentados.....	36
Tabla 10. Comparación del rendimiento de los modelos predictivos.....	37
Tabla 11. Comparación de eficiencia y elasticidad por canal publicitario	39
Tabla 12. Resumen de la curva de saturación por tramos de inversión en Facebook.....	41
Tabla 13. Resultados de los escenarios de inversión por canal y métricas de eficiencia.....	43
Tabla 14. Escenario óptimo por canal en términos de eficiencia.....	44
Tabla 15. Escenario óptimo por canal según el equilibrio entre volumen de conversiones y eficiencia.....	44
Tabla 16. Conversiones estimadas en función del nivel de inversión y duración de la campaña.....	46
Tabla 17. Solución óptima del modelo de asignación presupuestaria bajo restricción de inversión.....	50
Tabla 18. Solución óptima del modelo de asignación presupuestaria bajo restricción de eficiencia.....	50
Tabla 19. Análisis de la capacidad explicativa de las variables categóricas.....	52

Índice de figuras

Figura 1. Dimensión del dataset tras su carga en Python	21
Figura 2. Comprobación de ausencia de valores nulos en el dataset	23
Figura 3. Aplicación de reglas de validación y acotación sobre variables clave del modelo...24	
Figura 4. Creación de los principales indicadores clave de rendimiento.....	24
Figura 5. Distribución del ROI por canal publicitario	25
Figura 6. Distribución del CPA por canal publicitario	26
Figura 7. Distribución conversiones por canal	27
Figura 8. Distribución de conversiones estimadas por audiencia objetivo.....	28
Figura 9. Distribución de variables numéricas y evidencia de asimetría.....	29
Figura 10. Síntesis del proceso de preparación y análisis de datos	30
Figura 11. Matriz de correlación entre variables de rendimiento publicitario.....	32
Figura 12. Flujo metodológico de los modelos predictivos.....	35
Figura 13. Relación entre inversión publicitaria y conversiones.....	38
Figura 14. Elasticidad del gasto publicitario en función del nivel de inversión en Facebook	40
Figura 15. Curva de saturación y punto óptimo de inversión en Facebook.....	41
Figura 16. Comparación de conversiones estimadas según escenarios de inversión en Pinterest.....	43
Figura 17. Flujo metodológico para la construcción y evaluación de escenarios de inversión publicitaria	45
Figura 18. Mapa de calor de conversiones estimadas según nivel de inversión y duración de la campaña	46
Figura 19. Análisis de modelos predictivos y principales hallazgos	47

1. Introducción

1.1 Relevancia del tema

La publicidad digital ha aumentado notablemente su importancia en el ecosistema publicitario global. Según las estimaciones más recientes, la inversión mundial en publicidad digital ha alcanzado aproximadamente 678.700 millones de dólares, lo que representa el 68.4% del gasto publicitario global (Dentsu, 2024). En este contexto, canales como *Social Media Ads* continúan mostrando un elevado dinamismo, dificultando la capacidad de toma de decisiones. Una misma inversión puede generar resultados muy distintos en función de múltiples factores, lo que evidencia la naturaleza no homogénea del rendimiento publicitario. Por ello, el modelado de respuesta publicitaria a partir de un enfoque *data-driven* resulta clave para transformar datos en decisiones de inversión más precisas, eficientes y estratégicas.

1.2 Objetivos del estudio

El objetivo general consiste en el diseño de un método analítico para la asignación óptima del presupuesto en campañas de *Social Media Ads* mediante modelos de respuesta y elasticidades, con uso de una herramienta de visualización de datos que facilite la conexión entre los resultados analíticos y la toma de decisiones operativas de inversión.

Los objetivos específicos consisten en:

- Construir, limpiar y validar métricas clave de desempeño (*KPIs*)¹ relevantes para el análisis.
- Desarrollar modelos predictivos que permitan estimar la relación entre inversión publicitaria y conversiones.
- Analizar las dinámicas de respuesta publicitaria, incluyendo efectos de saturación, elasticidad y posibles interacciones entre canales.
- Diseñar un modelo de optimización que permita identificar la asignación óptima del presupuesto entre canales.
- Documentar el proceso analítico completo, incluyendo *scripts* y fórmulas, así como exploraciones futuras.

¹ Los KPIs o Indicadores Clave de Rendimiento, son utilizados para medir el rendimiento de una estrategia o desempeño de una empresa en un área específica.

1.3 Metodología y estructura

La metodología de este trabajo se fundamenta en un enfoque multidisciplinario que combina una primera parte de investigación de la literatura académica y profesional, un análisis de tendencias recientes del mercado digital y redes sociales, y el desarrollo de una propuesta analítica específica basada en los resultados obtenidos. El objetivo principal es desarrollar un marco conceptual y empírico robusto que sirva de base para entender como varía el rendimiento de la inversión publicitaria y diseñar un sistema que optimice la asignación de presupuesto en campañas de *Social Media Ads*.

El *dataset* empleado de la plataforma de Kaggle, recoge campañas publicitarias durante el año 2022, por lo que ese año marca el punto de referencia del análisis empírico. Por lo tanto, es coherente hacer una revisión del estado del mercado en los años posteriores (2023-2025), ya que contextualizan el análisis dentro de la evolución más reciente y proyectada del entorno digital. Esta revisión abarca informes de referencia como *Global Ad Spend Outlook*² de WARC, en el que se proyecta el gasto global en publicidad, y el informe *AdEx Benchmark*³ de IAB Europe para mostrar como Europa no es ajena a las tendencias globales.

Posteriormente, se desarrolla un bloque teórico centrado en *Marketing Mix Modeling*, (MMM) utilizado para cuantificar el impacto de las variables de marketing sobre un resultado de negocio. Este constituye la base metodológica para medir la relación ente inversión y resultados en marketing, estimando parámetros como la elasticidad publicitaria y rendimientos decrecientes del gasto. Asimismo, se desarrolla un modelo de optimización que no solo busca maximizar los resultados esperados, sino también garantizar soluciones estables ante posibles cambios en los datos y en el entorno del mercado. Los principales términos técnicos en inglés empleados a lo largo del trabajo, han sido incluidos en el Anexo 1 con su respectiva definición.

Basándose en los conocimientos adquiridos a través de la revisión bibliográfica y del estudio de casos, el trabajo desarrolla una propuesta analítica que comienza con la construcción y depuración de un *dataset* y continúa con la generación de métricas y modelado de respuesta.

² Informe elaborado por WARC (*World Advertising Research Center*), que analiza la evolución y la previsión de la inversión publicitaria a nivel global, ampliamente utilizado como referencia para el análisis de tendencias del mercado.

³ Informe elaborado por IAB Europe que analiza la inversión en publicidad digital en Europa, desglosando los principales formatos, canales y tendencias del mercado.

2. Marco teórico y revisión bibliográfica

2.1 Evolución de la inversión en medios digitales y Social Media Ads

Durante las últimas dos décadas, la industria publicitaria ha experimentado una profunda transformación impulsada por la digitalización de los medios, el desarrollo tecnológico y los cambios en el comportamiento y poder del consumidor. La consolidación de Internet como red de comunicación esencial altera el paradigma tradicional de la comunicación publicitaria, al transformar de manera significativa la planificación y ejecución de las estrategias publicitarias, donde “los consumidores han ganado el control sobre dónde, cuándo y cómo interactúan con los medios de comunicación” (Meritxell Roca Sales, 2014, p.1). Como consecuencia de estos cambios, las empresas han comenzado a considerar e integrar la publicidad a través de medios y canales digitales como parte de su planificación comunicativa.

Tradicionalmente, la mayor parte de la inversión publicitaria se concentraba en medidas convencionales conocidas como medios pasivos (televisión, radio, prensa escrita y publicidad exterior). Estos canales ofrecen un alcance masivo, con una comunicación unidireccional, en la que el receptor, adopta un papel principalmente contemplativo, “solo se envía la información y dependerá del público cómo decida interpretarla” (Edwin Prieto, 2023), sin estímulos interactivos. Asimismo, los medios tradicionales presentan el contenido de manera estructurada y jerárquica, en la que la información se presenta de forma secuencial y priorizada, pero con limitaciones importantes en segmentación y medición de impacto al ser menos inmediatas las conversiones.

La irrupción masiva y transformadora de Internet ocurre a partir de los años 90, consolidándose a principios del siglo XXI, siendo un punto de inflexión en el proceso de reestructuración del modelo publicitario tradicional. Los medios digitales introducen un modelo de comunicación interactivo, participativo y multidireccional donde se trata de cautivar al consumidor sin imponer un mensaje, favoreciendo así la democratización del acceso a los contenidos y la diversidad de voces. El entorno digital, el cual se caracteriza por la actualización constante de contenidos personalizados y difusión inmediata a escala global, permite que la voz del consumidor sea fácilmente amplificadas. En consecuencia, se consolida la figura del prosumidor, concepto acuñado por Alvin Toffler para describir un usuario que combina los roles de productor y consumidor en el entorno digital (Kaufmann Argueta, 2014). En este marco, estos cambios no son meramente tecnológicos, sino que están estrechamente

relacionados con una nueva metodología de trabajo basada en estrategias de *engagement*⁴ orientadas a responder ante la ruptura del monopolio de la atención, la cual impide concentrar de manera exclusiva y prolongada la atención de las audiencias en único medio o mensaje publicitario (Meritxell Roca Sales, 2014). Su capacidad de segmentación avanzada facilita una mayor adecuación del mensaje publicitario al perfil del consumidor según variables demográficas, geográficas o psicográficas.

Estudios recientes muestran como la inversión en medios digitales basados en la interacción y participación del usuario han sufrido un crecimiento sostenido y acelerado. Esto es contrastado con el informe de previsión económica conocido como *Global Ad Spend Outlook 2024/2025* publicado por WARC (*World Advertising Research Center*), una de las fuentes más reconocidas en investigación sobre el gasto publicitario global, tanto en economías desarrolladas como en mercados emergentes. En este se proyecta el gasto global en publicidad con una cifra de casi el billón de dólares en 2024, con un crecimiento interanual estimado del 10.7% (WARC, 2024).

Desde una perspectiva estructural, los datos aportados por WARC ponen en manifiesto que el crecimiento del gasto publicitario no se distribuye de manera homogénea entre los distintos medios. La mayor parte del crecimiento del gasto publicitario global proviene de las grandes plataformas digitales como Google, Meta o Amazon que capturaron más de la mitad de todos los ingresos publicitarios (GroupM, 2024). Estas plataformas destacan por el empleo de grandes volúmenes de datos sobre el comportamiento e intereses de los consumidores, lo que mejora la eficiencia de las campañas con una medición del retorno de la inversión alta. A su vez, gestionan sistemas de compra programática a través de sus ecosistemas publicitarios donde los anunciantes pueden ajustar dinámicamente variables como el importe de las pujas, lo que incrementa la optimización en la asignación del presupuesto publicitario.

En el caso de las plataformas de carácter social, como Meta o TikTok, este crecimiento del gasto publicitario se articula fundamentalmente a través de una herramienta táctica conocida como *Social Media Ads*, que constituyen el eje central de su modelo de monetización. En su fase inicial, las redes sociales se consideraban “herramientas, aplicaciones y plataformas cuyo objetivo principal era facilitar las relaciones entre usuarios, la interacción, la colaboración y la distribución de contenidos” (José Sixto García, 2015), donde la presencia de las marcas tenía

⁴ El *engagement* es un término en inglés que se refiere al nivel de compromiso emocional o interacción activa que tiene el usuario con una marca.

un carácter secundario o experimental. Sin embargo, a medida que estas plataformas comenzaron a alcanzar una masa crítica de usuarios, sus datos se convirtieron en un recurso central para la generación de valor económico. Las plataformas sociales han evolucionado hacia sistemas híbridos en los que la interacción social y la lógica comercial se entrelazan, de modo que acciones aparentemente sociales usadas en las plataformas como compartir una foto, implican también la cesión de datos y la participación en un sistema mediático con intereses económicos propios. Como resultado, surge el concepto de *Social Media Ads* entendido como “anuncios pagados que se colocan en plataformas de redes sociales con el objetivo de llegar a audiencias específicas y generar interacción, clics o conversiones” (ESIC Business & Marketing School, s. f.).

Estos tienen un valor estratégico con un crecimiento destacado del 23,9% en 2024 (IAB Europe, 2025), siendo superior a otros segmentos de la publicidad digital como el *display*⁵ tradicional. La explicación de este fenómeno se encuentra en que son uno de los formatos más eficaces para alcanzar a los consumidores, al integrarse en el flujo habitual de contenidos que estos consumen en las redes sociales, combinando una segmentación avanzada sustentada en datos y algoritmos sumado a la capacidad de medición del rendimiento publicitario.

No obstante, este incremento de inversión en este tipo de publicidad en redes sociales genera la necesidad de encontrar métricas que evalúen de manera rigurosa y objetiva el rendimiento de las campañas. Por ello, el uso de datos empíricos obtenidos del *dataset Social Media Advertising*, resulta fundamental para el desarrollo de enfoques cuantitativos y modelos analíticos aplicados al ámbito publicitario capaces de identificar factores que influyen en el éxito de las campañas publicitarias. Se utilizan métricas como impresiones, clics, gasto y conversiones procedentes de campañas que son el resultado de las decisiones estratégicas asociadas al *marketing mix*.

2.2 Fundamentos de marketing mix y elasticidades publicitarias

El análisis del rendimiento publicitario en el entorno digital requiere que los resultados obtenidos sean interpretados dentro de un enfoque estructurado y sistemático para evitar las interpretaciones simplistas. Este tiene como punto de partida el *marketing mix* entendido como

⁵ El *display* es un formato de publicidad digital basado en anuncios visuales, que aparecen en páginas webs, apps o plataformas digitales.

el conjunto de herramientas y variables que tiene una organización para llegar a su público y alcanzar los objetivos planteados. “El mix se concreta en la planificación de las denominadas 4ps (en inglés, *product, price, place, promotion*) a partir del modelo formulado por J.E.McCarthy” (Fernández Marcial, 2015, p. 62), con el fin de facilitar la identificación de los puntos clave que requieran más atención para cumplir los objetivos de la empresa. Estas variables son interdependientes y permiten constituir un marco analítico que explique las decisiones estratégicas que toman las empresas, ya que cada dimensión condiciona de manera directa, la percepción de valor o comportamiento del consumidor. Neil H.Borden emplea la metáfora del mezclador de ingredientes para subrayar que las decisiones en marketing no son independientes, sino que se compone de múltiples procedimientos y políticas que se combinan para influir en el mercado (Borden, 1964).

Dentro de la publicidad digital, y particularmente en los *Social Media Ads*, el *marketing mix* adquiere una relevancia específica porque se materializa en decisiones muy concretas con una medición empírica en variables definidas y observables. En el entorno de la publicidad digital esta práctica no se limita a establecer únicamente el volumen total de la inversión, sino que es necesario incluir la distribución temporal y operativa entre campañas, conjuntos de anuncios o formatos de publicidad. Estas decisiones estratégicas condicionan directamente la capacidad de los *Social Media Ads* para llegar a la audiencia objetivo, incidiendo sobre variables como el alcance, clics, el número de impresiones y la frecuencia de exposición de los anuncios. Por lo tanto, existe una relación directa entre cualquier decisión estratégica del *marketing mix* y los resultados publicitarios observables.

No obstante, el marco del *marketing mix*, permite afirmar que una determinada decisión influye en los resultados, pero no indica en qué medida impacta cada decisión en los resultados, al no ser concebido como una herramienta cuantitativa, sino como un modelo orientado a la identificación y organización de las variables controlables del marketing. Para ello, son necesarias las herramientas analíticas que comprenden modelos estadísticos y econométricos que permiten medir la relación causa-efecto entre las variables. Entre estas herramientas se incluye el análisis de elasticidades publicitarias que desempeñan un puente metodológico entre el *marketing mix* y el rendimiento observable. Esto se consigue gracias a que las decisiones estratégicas y tácticas en el entorno digital se caracterizan por la posibilidad de modificar de manera continua las campañas de *Social Media Ads*.

Desde una perspectiva teórica, la elasticidad publicitaria representa cómo un cambio porcentual en la inversión en publicidad influye en el cambio porcentual en las ventas o en la demanda de un producto (Lifesight, s. f.). Tradicionalmente esta definición se ajustaba en términos de ventas o demanda porque surge del marketing clásico, donde esas eran las principales variables observables. Desde un punto de vista analítico, este concepto puede extenderse a cualquier métrica de rendimiento que responda a la inversión publicitaria, como a métricas de conversión.

Por lo tanto, se trata una medida estandarizada de la respuesta del mercado que es capaz de capturar el fenómeno de los rendimientos decrecientes de la inversión publicitaria. Esto es consistente con la idea de que en la práctica los incrementos en el gasto no generan aumentos proporcionales en los resultados, sino que el efecto marginal de la inversión tiende a disminuir a medida que el presupuesto crece. Con el análisis de elasticidades se identifica el punto de inflexión que permite una optimización del presupuesto para no agotar la audiencia más receptiva, se reduzca su eficiencia o incrementen los costes de captación.

El análisis de las métricas del *dataset* empleado durante el trabajo de fin de grado, como CPA (*Cost Per Acquisition*), permiten el cálculo de estas elasticidades, enseñando el efecto marginal de incrementar el presupuesto en un canal concreto. Este se convierte en una herramienta fundamental para la optimización presupuestaria en campañas de *Social Media Ads*, ya que facilita la comparación entre diferentes campañas, canales o periodos temporales distintos, independientemente de su escala o volumen de inversión.

En definitiva, los fundamentos del *marketing mix* y el concepto de elasticidad publicitaria constituyen los pilares teóricos fundamentales sobre el cual, se apoya el análisis cuantitativo de la inversión publicitaria en *Social Media Ads*, sirviendo de base para el modelado de la respuesta publicitaria y la posterior optimización del presupuesto.

2.3 Dinámicas de respuesta publicitaria: persistencia, saturación y efectos cruzados

Para modelar de forma correcta la respuesta publicitaria se requiere reconocer que los resultados derivados de la inversión en publicidad no son supuestos teóricos arbitrarios, sino que cambian de manera no lineal, y subraya la relevancia de considerar efectos temporales y variables dinámicas. Esta relación entre inversión y resultados se explica, en gran medida, por la presencia de tres fenómenos fundamentales que son:

- La persistencia temporal
- La saturación del impacto
- Los efectos cruzados entre canales

Son regularidades del comportamiento del consumidor y el funcionamiento de la publicidad que se repiten en distintos contextos y tienen especial relevancia en *Social Media Ads*.

En primer lugar, la persistencia temporal en *Social Media Ads* también conocida como *carryover effect* o efecto de arrastre se refiere al “lapso entre la exposición del consumidor al anuncio y su respuesta al mismo” (MBA Skool Team, s. f., traducción propia), es decir, la inversión publicitaria tiene un impacto que no termina con una única exposición, sino que persiste en el tiempo. La publicidad no solo informa, sino que genera y refuerza actitudes y asociaciones mentales que influyen en decisiones de compra futuras.

En el ámbito del *Marketing Mix Modeling* (MMM), esta memoria publicitaria se moldea a través del modelo *Adstock* que “representa el efecto acumulativo y retardado de una inversión publicitaria a lo largo del tiempo” (Ayala, s. f.). La principal ventaja es que refleja de forma realista la relación entre publicidad y resultados, especialmente relevante en este tipo de campañas donde el impacto puede manifestarse de manera retardada. No obstante, su identificación empírica requiere una estructura de datos con observaciones repetidas por campaña o unidad de análisis a lo largo del tiempo. Variables temporales como *date* permiten analizar patrones cronológicos y diferencias según momentos de activación, pero limita la estimación directa de efectos de arrastre en sentido estricto.

La saturación del impacto publicitario se puede definir como “la sobreabundancia de mensajes publicitarios en un medio determinado, pudiendo provocar así una indiferencia o rechazo hacia el anuncio perdiendo la efectividad de este” (Silva Vázquez, 2025, p. 19). Hace referencia al segundo fenómeno ampliamente documentado en la literatura de marketing y representa un reto para las empresas al tener que asegurarse la visibilidad sin sobrepasar el límite de la indiferencia del público, de ahí la importancia de la relevancia de los contenidos. En una primera fase, el incremento del gasto publicitario suele generar mejoras significativas en las métricas de rendimiento puesto que el mensaje alcanza principalmente a usuarios que aún no han sido expuestos previamente o que presentan elevada predisposición a responder. Sin embargo, a medida que la inversión se intensifica, los usuarios potenciales se reducen, lo que exige que el valor ofrecido vaya más allá de los incentivos comerciales, incorporando también la resolución de problemas y la atención a las necesidades del usuario.

Por ello, el anunciante debe abordarlo no desde una perspectiva cuantitativa, sino desde una perspectiva cualitativa que evite los rendimientos decrecientes. Desde una perspectiva empírica, este fenómeno se observa en métricas de rendimiento como el *CPA*, en relación con el gasto publicitario, cuyo análisis resultará eficiente para redistribuir el presupuesto hacia otros canales, formatos o segmentos de audiencia.

A ello, se le suman los efectos cruzados entendidos como posibles interdependencias entre acciones de marketing o publicidad en un canal o medio, que también influyen en los resultados obtenidos en otro canal o medio. Además, estas relaciones pueden ser complementarias o sustitutivas, sin tener por qué ser simétricas. En este caso, los canales de Instagram, X (antes Twitter), Facebook y Pinterest no deben analizarse por separado. Los esfuerzos de marketing en un canal pueden modificar los resultados de compra en otros canales. Esa influencia puede ser complementaria cuando el esfuerzo en uno aumenta los resultados en otro o puede ser sustitutiva si se produce el efecto inverso.

Estos efectos pueden explorarse a partir de combinaciones de las variables de la base de datos empleada, como *channel_used*, *target_audience*, *campaign_goal*, *customer_segment* observando cómo estas interacciones modifican indicadores como *conversion_rate* o *ROI*. Así, cuando no se consideren efectos cruzados, el modelo no tenderá a concentrar mucho más el esfuerzo en el canal con mayor elasticidad propia, pudiendo infra invertir en canales más informativos o equilibrados.

En conjunto, estas tres dimensiones ponen de manifiesto que la respuesta publicitaria no sea concebida como un proceso lineal o estático, sino como un problema analítico que exige el empleo de modelos que trasciendan una relación directa entre gasto y resultado.

2.4 Optimización de presupuesto y técnicas robustas

Estas dinámicas evidencian que la eficacia publicitaria no depende exclusivamente de la magnitud del gasto, sino también de su distribución, su intensidad y coherencia con el resto de las acciones del plan de marketing desarrolladas por la empresa. La optimización presupuestaria se considera una manifestación de la analítica prescriptiva, puesto que intenta recomendar la mejor acción futura y orientar la decisión hacia un objetivo concreto. En este sentido, constituye una extensión natural del análisis de marketing, al transformar la evidencia

empírica en decisiones de asignación óptima de recursos limitados entre distintas alternativas de inversión para maximizar resultados o minimizar costes (Lepenioti et al., 2020).

Anteriormente, los modelos clásicos de asignación presupuestaria, como Vidale-Wolfe (Anexo 2), explicaban la respuesta de las ventas al esfuerzo publicitario de forma agregada. La idea central del modelo es que las ventas de un producto en cada momento dependen de la fuerza expansiva de la publicidad y el efecto negativo del decaimiento vinculado al olvido (Caballero Fernández et al., s.f). La publicidad actúa como un mecanismo de captación y actividad de la demanda potencial, mientras que, de forma simultánea, el efecto alcanzado tiende a deteriorarse con el paso del tiempo debido a la presión competitiva o la pérdida de atención del consumidor. Por lo tanto, es un modelo que puede interpretarse como una representación agregada del comportamiento del mercado, con la existencia de la saturación de mercado mencionada en el capítulo anterior. De modo que, el impacto de la publicidad muestra rendimientos decrecientes a medida que se aumenta la proporción de consumidores alcanzados.

La relevancia del modelo reside en que ofrece una base teórica de la literatura de marketing dinámico y de la investigación operativa aplicada a la publicidad. Sin embargo, presenta una serie de limitaciones al tener un carácter agregado que no permite distinguir entre segmentos de clientes, audiencias o canales. En el caso de las campañas de *Social Media Ads*, la asignación del presupuesto debe entenderse como un problema dinámico, donde el modelo Vidale-Wolfe, resulta insuficiente para captar la complejidad de la publicidad digital contemporánea. Se exigen modelos más sofisticados, ya que su funcionamiento no se mide por ventas agregadas, sino por una cadena de resultados inmediatos procedentes de datos detallados de los usuarios e información en tiempo real sobre impresiones y respuestas.

En consecuencia, los modelos de optimización presupuestaria se estructuran teóricamente en torno a dos componentes principales. El primero es la función de respuesta, entendida como la relación cuantitativa entre un input de publicidad y un output de valor, como ventas, cuota o intención de compra, es decir, la relación funcional entre la presión publicitaria y el resultado esperado (Taylor et al., 2009). Esta función constituye la base analítica del problema de optimización, siendo estimada a partir de datos observados de campañas, de manera que la relación entre inversión y resultado no se deduzca de supuestos abstractos, sino de evidencia empírica. El segundo componente es la regla de asignación que determina cómo debe distribuirse el presupuesto disponible entre distintas alternativas en base a la función objetivo.

Optimizar estas decisiones puede generar incluso más beneficios para la compañía que un simple aumento del presupuesto sin mejoras en su estructura (Mantrala et al., 1992).

Junto a estos elementos, en un entorno de *Social Media Ads* aparece la incertidumbre porque el problema no se resuelve con valores totalmente fijos sino a partir de parámetros estimados con incertidumbre. En este tipo de campañas los costes de subasta⁶, la presión competitiva o el comportamiento de la audiencia pueden variar con rapidez. Esto conduce de forma natural a la necesidad de emplear técnicas robustas en el modelo de optimización que ayuden a mantener un buen desempeño incluso cuando las condiciones reales difieran de las previstas. Entre ellas destacan las técnicas de regularización o tratamiento de *outliers* para reducir el sobreajuste y la sensibilidad a observaciones extremas. Su aplicación en los modelos no sustituye al enfoque *data-driven*, sino que lo complementa, al reforzar la solidez de las decisiones presupuestarias. Asimismo, el análisis de sensibilidad y de escenarios ocupan un lugar importante como técnicas complementarias para evitar que la recomendación presupuestaria no dependa en exceso de un único supuesto rígido.

En definitiva, la optimización presupuestaria y las técnicas robustas aportan un marco conceptual para entender cómo puede orientarse la asignación eficiente del gasto publicitario con decisiones sustentadas en información cuantitativa y entornos cambiantes. Por ello, una vez establecido el marco teórico resulta necesaria para delimitar la base empírica del trabajo y contextualizar este análisis.

3. Descripción y preparación de los datos

3.1 Descripción del dataset

El presente trabajo de fin de grado utiliza un conjunto de datos que contiene métricas de rendimiento de *Social Media Ads* recogidos en un *dataset* ficticio llamado *Social Media Advertising Dataset*. Este procede de la plataforma en línea Kaggle, un repositorio ampliamente utilizado en el ámbito académico y profesional que permite analizar el comportamiento de los usuarios ante distintos anuncios publicitarios y evaluar la efectividad de las estrategias de marketing digital. El análisis parte de un conjunto inicial de 300.000 observaciones y 16 variables, susceptible de modificación tras la fase de depuración.

⁶ Los costes de subasta hacen referencia al importe asociado a la participación y adjudicación de un espacio publicitario.

Cada registro representa una campaña publicitaria identificada de forma única dirigido a un segmento concreto de la población. Las filas son cada unidad de análisis, es decir, *Social Media Ads*, mientras que las columnas corresponden a diferentes variables explicativas y de resultado que se pueden agrupar en varias categorías. El valor del *dataset* reside en la combinación de forma simultánea de variables de inversión, métricas de respuesta y características contextuales de las campañas para abordar el problema de optimización de forma completa.

En primer lugar, hay variables de identificación y metadatos, como *Campaign_ID*, *Company* y *Date*, para diferenciar entre distintas campañas, empresas y periodos temporales. Gracias a estas es posible tener un contexto y estructura para evaluar el desempeño entre los anuncios y analizar cómo evoluciona su rendimiento en el tiempo. En segundo lugar, se incluyen variables relacionadas con el diseño y la estrategia de la campaña como *Campaign_Goal*, *Channel_Used* y *Duration*. Estas variables permiten un aprendizaje transferible al representar decisiones directamente modificables en la planificación de acciones futuras. Asimismo, se incorporan variables relacionadas con la segmentación de la audiencia y el contexto de la campaña, tales como *Customer_Segment*, *Location*, *Language* y *Target_Audience*, que permiten caracterizar tanto el público objetivo como tener ciertas categorías que identifican la industria o vertical de mercado asociada a la campaña.

En cuanto a las métricas de interacción y rendimiento, se cuenta con *Clicks*, *Impressions* y *Engagement_Score* para reflejar el interés inicial del usuario. En el marco de publicidad digital orientada a resultados, se utiliza la métrica estándar conocida como *Clicks* para medir el número de veces que los usuarios interactúan activamente con el anuncio al hacer clic sobre él. Sin embargo, en el caso de que los usuarios no interactúen también se puede medir con la métrica de *Impressions*. La diferencia fundamental radica en que los clics reflejan una respuesta directa del usuario, mientras que las impresiones únicamente miden la exposición del anuncio, aunque no exista interacción (ADD Marketing, 2024).

Por último, las variables *Conversion_Rate*, *Acquisition_Cost* y *ROI* están vinculadas a los resultados económicos observables y eficiencia de las campañas publicitarias. Estas variables son las únicas que permiten vincular de forma directa la actividad publicitaria con resultados económicos observables y comparables. La tasa de conversión captura el comportamiento posterior del usuario a la exposición del anuncio desde el punto de vista del negocio. Las dos últimas permiten pasar de un análisis descriptivo del desempeño publicitario a un problema de

eficiencia económica, donde el objetivo es asignar recursos a aquellas combinaciones de variables que maximicen el retorno esperado.

Con el fin de facilitar la interpretación de las variables anteriormente descritas, se presenta la Tabla 1 con una clasificación en función de su papel dentro del análisis.

Tabla 1. *Tipología de variables según su función analítica*

Categoría	Variables incluidas
Identificación y metadatos	Campaign_ID, Company, Date
Diseño y estrategia de campaña	Campaign_Goal, Channel_Used, Duration
Segmentación y contexto	Customer_Segment, Location, Language, Target_Audience
Métricas de interacción	Clicks, Impressions, Engagement_Score
Resultados y eficiencia económica	Conversion_Rate, Acquisition_Cost, ROI

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

El carácter *data-driven* del presente trabajo se ve respaldado por la naturaleza cuantitativa y estructurada del *dataset*. Se cuenta con un gran volumen de datos como base empírica para tomar decisiones analíticas que no dependan de intuiciones o juicios subjetivos, sino de patrones observables en los datos. No obstante, para que esta información resulte verdaderamente útil en el contexto del modelado de respuesta publicitaria y la optimización presupuestaria, es necesario un proceso previo de preparación de datos y creación de variables derivadas.

A continuación, la Tabla 2 enseña un resumen de las variables incluyendo su denominación, tipo, unidades de medida y efecto esperado.

Tabla 2. *Resumen de las variables empleadas en el estudio*

Nombre	Tipo	Qué mide	Unidades	Cómo influye
campaign_id	Numérico (Int)	Identificador de la campaña	Número	Permite diferenciar cada campaña
target_audience	Catagórico	Público objetivo de la campaña	Categorías	Afecta al rendimiento según el segmento
campaign_goal	Catagórico	Objetivo de la campaña	Categorías	Define la estrategia de marketing
duration	Catagórico	Duración de la campaña	Días	Influye en el alcance y resultados
channel_used	Catagórico	Plataforma utilizada	Categorías	Impacta en el engagement y conversiones
conversion_rate	Numérico (Float)	Porcentaje de conversiones	Proporción	Indica la eficacia de la campaña

acquisition_cost	Catagórico porque contiene \$	Coste de adquisición de clientes	Dólares (\$)	Afecta a la rentabilidad
roi	Numérico (Float)	Retorno de la inversión	Ratio	Mide la rentabilidad de la campaña
location	Catagórico	Ubicación de la campaña	Ciudad	Puede influir en el comportamiento del público
language	Catagórico	Idioma de la campaña	Categorías	Afecta a la comprensión del anuncio
clicks	Numérico (Int)	Número de clics en el anuncio	Clics	Refleja el interés del público
impressions	Numérico (Int)	Número de visualizaciones del anuncio	Impresiones	Determina el alcance
engagement_score	Numérico (Int)	Nivel de interacción con el anuncio	Escala (1-10)	Mide el engagement generado
customer_segment	Catagórico	Tipo de cliente o sector	Categorías	Permite analizar resultados por segmento
date	Catagórico	Fecha de la campaña	Fecha	Permite analizar tendencias
company	Catagórico	Empresa anunciante	Nombre	Permite comparar campañas

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

3.2 Limpieza y creación de variables

La preparación comienza con la carga del archivo csv en la herramienta de Python y una inspección descriptiva. Se analizan aspectos básicos como la dimensión del *dataset* que se muestra en la Figura 1 con 300.000 observaciones y 16 variables, tipos de datos y estadísticos básicos.

Figura 1. Dimensión del dataset tras su carga en Python

```
===== SHAPE =====
(300000, 16)
```

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

En la Tabla 3 se permite detectar problemas habituales en *datasets* de marketing digital. Por ejemplo:

- *Acquisition_Cost* no está correctamente tipada para un análisis cuantitativo riguroso al estar inicialmente almacenada como texto (*object*) debido a la presencia del símbolo monetario \$.
- La variable *Date* también se encuentra como cadena de texto, a pesar de representar una variable temporal esencial para el análisis longitudinal de campañas. Debe convertirse al formato fecha estándar para poner orden cronológicamente las observaciones.

- La variable *Duration*, se encuentra en formato texto al tener la siguiente estructura “15 days”. Por ello, se transforma la variable creando una nueva denominada *duration_days* como variable cuantitativa adecuada para el análisis compuesta solo por el número.

Tabla 3. Estructura inicial del dataset antes del preprocesamiento

Campaign_ID	Target_Audience	Campaign_Goal	Duration	Channel_Used	Conversion_Rate	Acquisition_Cost	ROI	Location	Language	Clicks	Impressions	Engagement_Score	Customer_Segment	Date	Company	
0	529013	Men 35-44	Product Launch	15 Days	Instagram	0.15	\$500.00	5.790000	Las Vegas	Spanish	500	3000	7	Health	2022-02-25	Aura Align
1	275352	Women 45-60	Market Expansion	15 Days	Facebook	0.01	\$500.00	7.210000	Los Angeles	French	500	3000	5	Home	2022-05-12	Hearth Harmony
2	692322	Men 45-60	Product Launch	15 Days	Instagram	0.08	\$500.00	0.430000	Austin	Spanish	500	3000	9	Technology	2022-06-19	Cyber Circuit
3	675757	Men 25-34	Increase Sales	15 Days	Pinterest	0.03	\$500.00	0.909824	Miami	Spanish	293	1937	1	Health	2022-09-08	Well Wish
4	535900	Men 45-60	Market Expansion	15 Days	Pinterest	0.13	\$500.00	1.422828	Austin	French	293	1937	1	Home	2022-08-24	Hearth Harmony

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Con el objetivo de facilitar la reproducibilidad del análisis y minimizar errores en fases posteriores, se procede a la estandarización de los nombres de las columnas. Todas las variables se renombran utilizando letras minúsculas, eliminado espacios y caracteres especiales, por ejemplo, *Channel_Used* se transforma a *channel_used*. Esto genera la Tabla 4, donde se muestra un conjunto de variables homogéneo y compatible con buenas prácticas en proyectos estructurados cuyo objetivo es extraer conocimiento útil a partir de los datos.

Tabla 4. Estructura del dataset tras la estandarización de los nombres de las variables

campaign_id	target_audience	campaign_goal	duration	channel_used	conversion_rate	acquisition_cost	roi	location	language	clicks	impressions	engagement_score	customer_segment	date	company
529013	Men 35-44	Product Launch	15 Days	Instagram	0.15	\$500.00	5.790000	Las Vegas	Spanish	500	3000	7	Health	2022-02-25	Aura Align
275352	Women 45-60	Market Expansion	15 Days	Facebook	0.01	\$500.00	7.210000	Los Angeles	French	500	3000	5	Home	2022-05-12	Hearth Harmony
692322	Men 45-60	Product Launch	15 Days	Instagram	0.08	\$500.00	0.430000	Austin	Spanish	500	3000	9	Technology	2022-06-19	Cyber Circuit
675757	Men 25-34	Increase Sales	15 Days	Pinterest	0.03	\$500.00	0.909824	Miami	Spanish	293	1937	1	Health	2022-09-08	Well Wish
535900	Men 45-60	Market Expansion	15 Days	Pinterest	0.13	\$500.00	1.422828	Austin	French	293	1937	1	Home	2022-08-24	Hearth Harmony
...
565525	Men 25-34	Market Expansion	60 Days	Instagram	0.14	\$15000.00	4.950000	Los Angeles	English	39999	119998	7	Technology	2022-10-17	Code Crafters
539680	Women 25-34	Market Expansion	60 Days	Instagram	0.08	\$15000.00	3.410000	Austin	Spanish	39999	119998	8	Technology	2022-02-25	Innovate Infinity
140032	Women 35-44	Brand Awareness	60 Days	Pinterest	0.09	\$15000.00	0.904461	Miami	French	23492	77476	1	Food	2022-10-02	Gastronomy Guild
161067	Women 18-24	Market Expansion	60 Days	Instagram	0.13	\$15000.00	7.430000	Las Vegas	Spanish	39999	119999	10	Food	2022-01-24	Savor Street
420183	Men 25-34	Brand Awareness	60 Days	Facebook	0.08	\$15000.00	1.900000	Miami	English	40000	120000	10	Technology	2022-07-18	Innovate Infinity

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Por otro lado, los duplicados exactos son eliminados necesariamente para evitar doble conteo de inversión o resultados. En el conjunto de datos analizado no se identificaron registros duplicados exactos porque cada observación se encuentra asociada a un identificador único de campaña.

Según se indica la Figura 2, la exploración de valores ausentes muestra que ninguna de las variables que componen el *dataset* empleado presenta valores nulos, lo que refuerza la

consistencia del conjunto de datos a nivel de campañas publicitarias individuales. No obstante, se incluyó un bloque de verificación y limpieza de datos dentro del proceso de preprocesado, como medida preventiva para confirmar la integridad del conjunto de datos tras las transformaciones de variables realizadas.

Figura 2. *Comprobación de ausencia de valores nulos en el dataset*

```

Valores nulos restantes:
campaign_id      0
target_audience 0
campaign_goal    0
duration         0
channel_used     0
conversion_rate  0
acquisition_cost 0
roi             0
location        0
language        0
clicks          0
impressions     0
engagement_score 0
customer_segment 0
date           0
company        0
duration_days  0
acquisition_cost_usd 0
ctr           0
estimated_conversions 0
cpa          0
dtype: int64

```

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Antes de entrenar el modelo de optimización, un paso clave es aplicar las reglas de consistencia de la Figura 3, únicamente a aquellas variables para las que existen restricciones naturales, matemáticas o de negocio. En cambio, variables como *Company*, *Language* o *Location* no presentan dominios numéricos sobre los que puedan definirse rangos o relaciones aritméticas. Su tratamiento metodológico se basa fundamentalmente en procesos de estandarización y agrupaciones.

La primera regla de validación se centra en la coherencia lógica entre las variables cuantitativas como *Clicks* e *Impressions*, ambas consideradas como eventos observables dentro de una misma secuencia del *funnel* publicitario, donde la impresión precede necesariamente al clic. Para que se produzca un clic es necesario que previamente haya existido una impresión, al existir una relación jerárquica y lógica entre ambas variables. Esta relación se demuestra dado que el anuncio debe mostrarse al usuario antes de que este pueda interactuar con él, por lo tanto, el número de clics no puede superar al de impresiones.

En el conjunto de datos analizado, esta relación se cumple de forma sistemática, con valores de impresiones que oscilan entre aproximadamente 2.000 y 120.000, y valores de clics siempre inferiores a dicho umbral. No solo sirve con esta restricción lógica de negocio sino también es

necesario forzar no negatividad como en *acquisition_cost_usd* o *conversion_rate*, ya que no pueden tomar valores negativos sin perder su significado empírico.

Por otro lado, la variable *conversion_rate* es uno de los *KPIS* esenciales y se refiere al “porcentaje de usuarios que realizan una acción deseada en relación al total de visitantes”. Al ser una proporción solo puede tomar valores comprendidos entre 0 y 1 (equivalente a 0%-100%), donde 0 indica ausencia de conversiones y 1 la conversión de la totalidad de las oportunidades consideradas. En un contexto, donde la presencia de valores fuera de este intervalo carece de interpretación empírica válida debe aplicarse una operación de acotación que evita que entradas inválidas degraden estadísticas, visualizaciones o modelos.

Figura 3. *Aplicación de reglas de validación y acotación sobre variables clave del modelo*

```
df = df[df["clicks"] <= df["impressions"]]
df["clicks"] = df["clicks"].clip(lower=0)
df["impressions"] = df["impressions"].clip(lower=0)
df["acquisition_cost_usd"] = df["acquisition_cost_usd"].clip(lower=0)
df["conversion_rate"] = df["conversion_rate"].clip(0, 1)
df = df[df["conversion_rate"] > 0]
```

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Por último, a partir de las variables originales, se muestra en la Figura 4 como se construyeron indicadores clave de rendimiento como el *click-through rate* (CTR) y *cost-per-acquisition* (CPA), que permiten evaluar la eficiencia y eficacia de las campañas.

Figura 4. *Creación de los principales indicadores clave de rendimiento*

```
df["ctr"] = df["clicks"] / (df["impressions"] + eps)
df["estimated_conversions"] = df["clicks"] * df["conversion_rate"]
df["cpa"] = df["acquisition_cost_usd"] / (df["estimated_conversions"] + eps)
```

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Asimismo, los resultados de la Tabla 5 muestran que las métricas de rendimiento no siguen un comportamiento homogéneo, sino que representan variabilidad considerable entre campañas.

Tabla 5. *Estadística descriptiva de las variables clave del rendimiento publicitario*

Resumen variables clave:								
	count	mean	std	min	25%	50%	75%	max
acquisition_cost_usd	300000.0	7753.750796	4301.035148	500.000000	4121.915000	7755.900000	11385.075000	15000.000000
ctr	300000.0	0.314156	0.024657	0.151187	0.301698	0.325373	0.331266	0.333333
estimated_conversions	300000.0	1453.379235	1234.337706	3.330000	468.647500	1097.660000	2136.487500	5999.550000
cpa	300000.0	9.049368	9.469805	1.598721	3.779103	5.574987	9.749562	150.150150

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

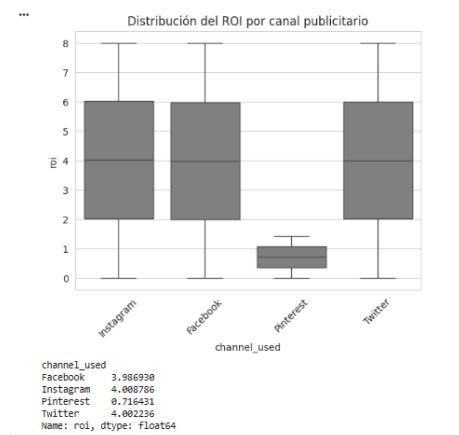
3.3 Análisis Exploratorio

El análisis exploratorio de los datos constituye una etapa preliminar a la modelización y optimización que permite comprender la estructura y calidad del *dataset* antes de iniciar con la aplicación de modelos analíticos más complejos. Tiene como finalidad identificar patrones comunes, distribuciones y relaciones iniciales sin modificar la información disponible, para ofrecer una primera aproximación a la relación entre las variables consideradas y la variable objetivo.

En primer lugar, el *dataset* analizado cuenta con ocho variables categóricas reales que permiten caracterizar las campañas publicitarias desde múltiples dimensiones. Con el objetivo de identificar las variables categóricas más relevantes para el análisis del rendimiento de las campañas, se realizó un análisis comparativo mediante visualizaciones que permiten evaluar su impacto sobre métricas clave como el retorno de la inversión (*ROI*), el coste por adquisición (*CPA*) y el número de conversiones.

Se analiza la variable *channel_used*, observándose diferencias en términos de eficiencia de las campañas. En particular, en la Figura 5, los canales de Instagram, Facebook y Twitter presentan valores de *ROI* elevados y distribuciones cercanas a 4, lo que indica un alto retorno de la inversión. En contraste, Pinterest muestra un rendimiento considerablemente menor, con valores de *ROI* por debajo de 1.5, lo que sugiere una menor capacidad para generar valor a partir del gasto publicitario (Anexo 3).

Figura 5. Distribución del ROI por canal publicitario

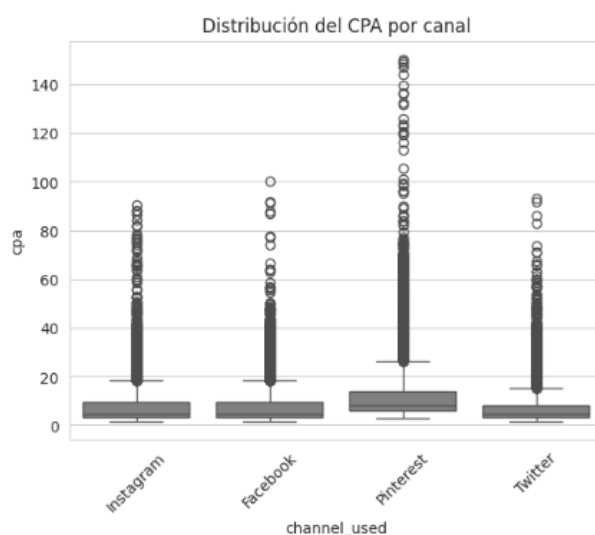


Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Estos resultados, se refuerzan al analizar el coste por adquisición (*CPA*) en la Figura 6, donde nuevamente se observan diferencias entre canales. Instagram, Facebook y Twitter presentan valores de *CPA* más bajos y con menor dispersión, lo que indica una mayor eficiencia en la conversión de usuarios. La presencia de valores atípicos en algunas variables derivadas, como el *CPA* o las conversiones estimadas, resulta esperable, ya que estas métricas se obtienen a partir de combinaciones de variables originales y pueden amplificar la dispersión existente en los datos. En este contexto, dichos valores no se interpretan necesariamente como errores, sino como observaciones extremas que reflejan la heterogeneidad propia del rendimiento de las campañas publicitarias.

Dado que los valores atípicos detectados aparecen en variables derivadas construidas a partir de métricas publicitarias, no se optó por su eliminación automática (Anexo 4).

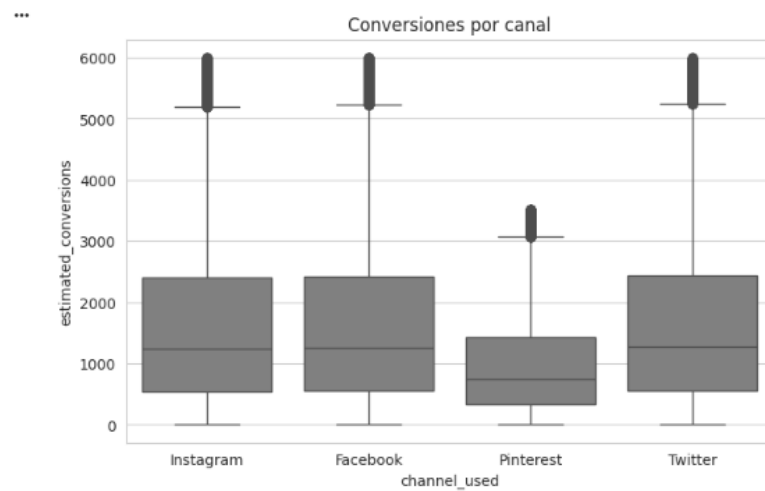
Figura 6. *Distribución del CPA por canal publicitario*



Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Asimismo, el análisis del gráfico de caja y bigotes de la Figura 7 sobre conversiones por canal muestra de nuevo como estos tres canales presentan medianas relativamente homogéneas en el rango de 1200 a 1400 conversiones.

Figura 7. *Distribución conversiones por canal*

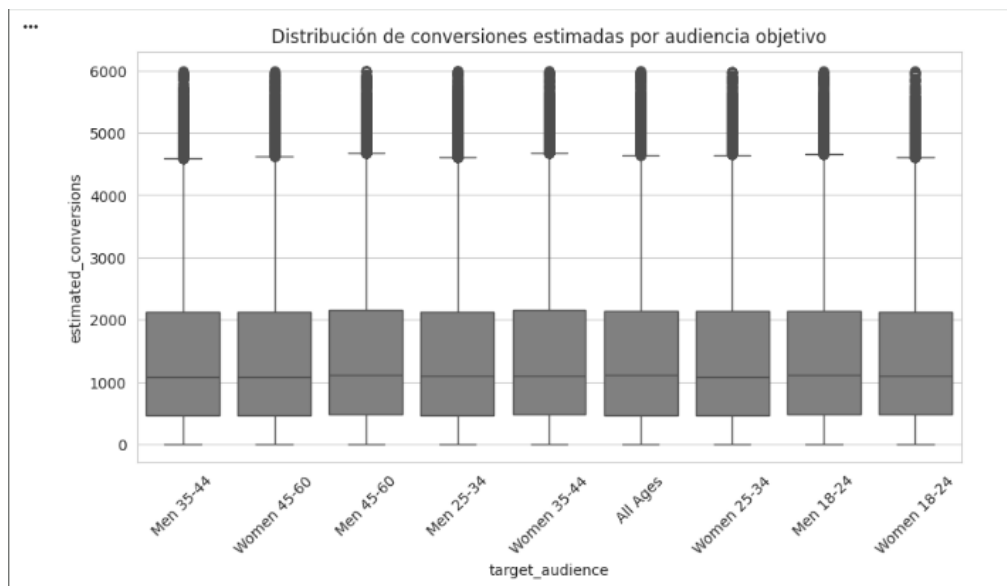


Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Desde el punto de vista de la optimización del presupuesto, estos resultados sugieren que la inversión debería de priorizar estos tres canales con mayor retorno y menor coste de adquisición. Además, muestran un patrón similar al observado en *benchmarks* recientes del mercado, en particular, sitúan el *CPA* de Meta Ads y Twitter/X en valores competitivos en el mercado.

Esta variable se identifica como lo más influyente debido a que presenta diferencias más claras en métricas clave como el *ROI*, el *CPA* y el número de conversiones. Sin embargo, variables como *target_audience* o *campaign_goal* muestran una variabilidad más limitada en los resultados como se observa en la Figura 8, con distribuciones similares entre categorías. Esto sugiere que, en este *dataset*, la segmentación de la audiencia u objetivo de campaña no genera diferencias significativas en el rendimiento de estas.

Figura 8. *Distribución de conversiones estimadas por audiencia objetivo*

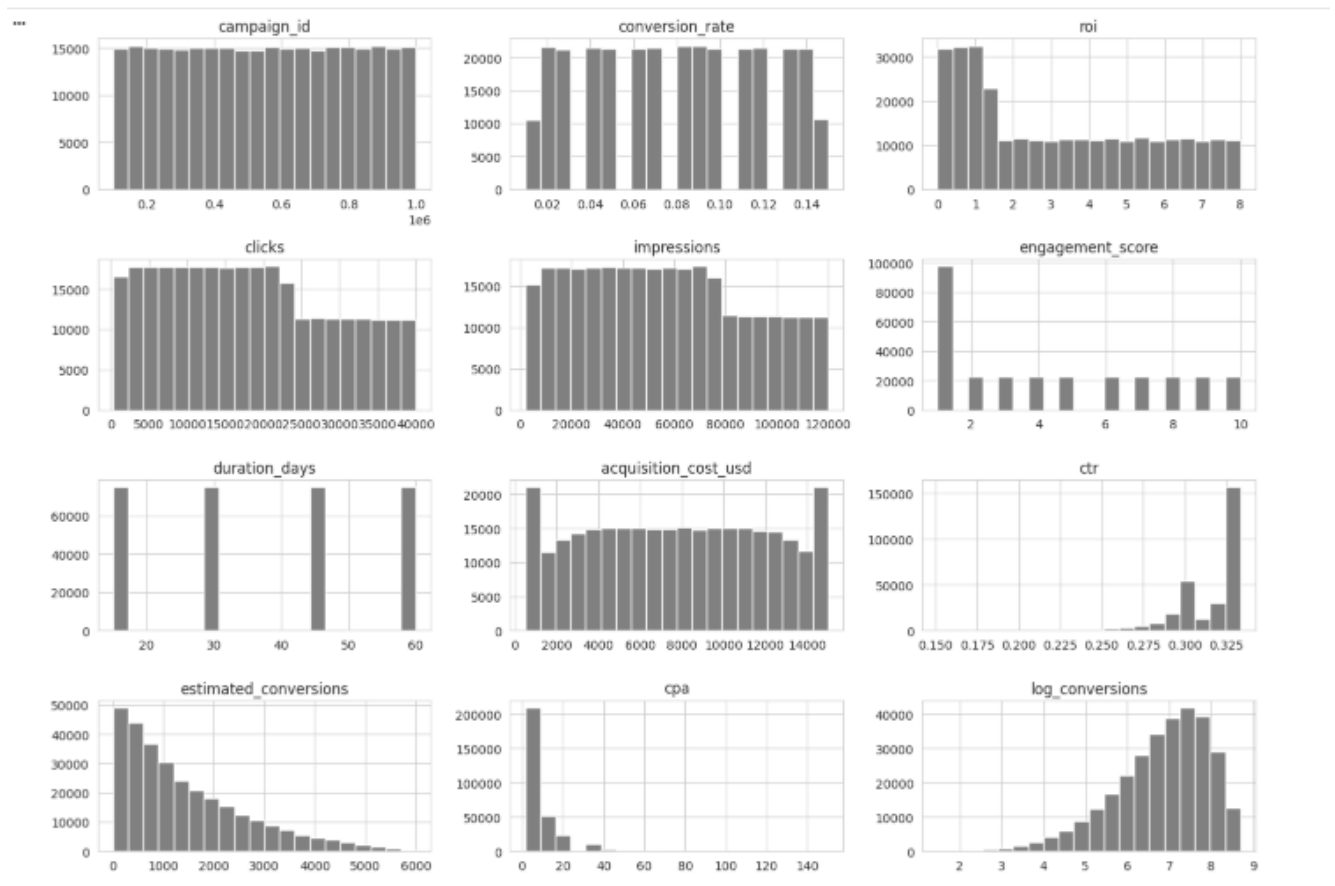


Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Por otro lado, las variables de *location*, *language* o *customer_segment* se consideran de carácter contextual, ya que su impacto en el rendimiento resulta menos significativo en comparación con las anteriores. Esta diferenciación resulta clave para centrar el análisis y el modelado en aquellas variables con mayor capacidad explicativa, facilitando así la posterior optimización del presupuesto publicitario.

En el análisis de las variables numéricas se observaron diferencias en términos de distribución, dispersión y asimetría. Las frecuencias de las variables muestran una distribución relativamente homogénea de las principales variables categóricas, lo que no favorece la comparación entre categorías. Variables como *clicks*, *impressions* y *acquisition_cost_usd* presentan una variabilidad considerable, mientras que *estimated_conversions* muestra una distribución claramente sesgada a la derecha, es decir, se observa asimetría positiva en la Figura 9. En este contexto, la transformación logarítmica aplicada a las conversiones permite obtener la variable *log_conversions* con una distribución más regular, lo que mejora su idoneidad para el análisis posterior.

Figura 9. Distribución de variables numéricas y evidencia de asimetría

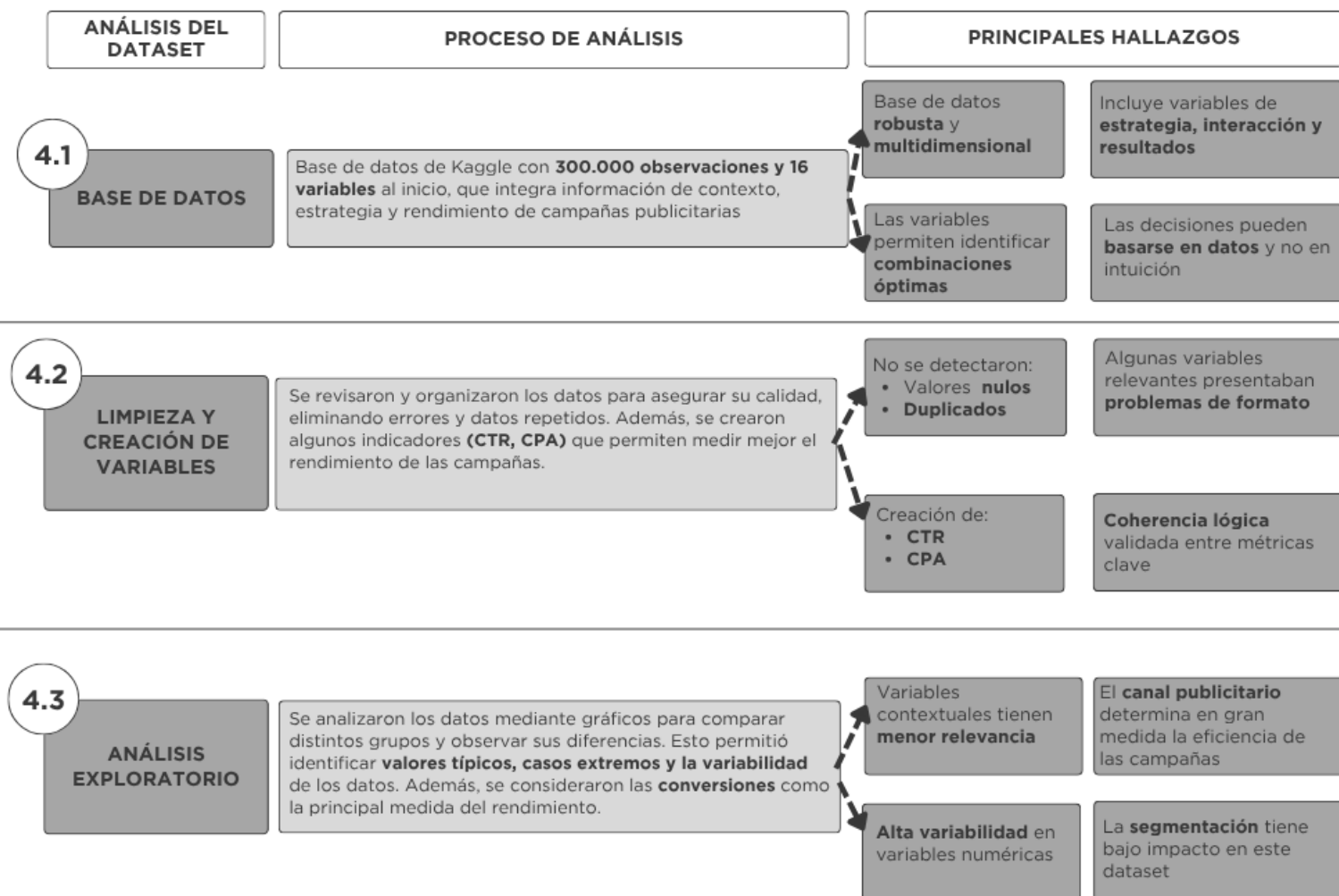


Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Dado que el *dataset* no incorpora una variable directa de conversiones observadas, se construyó una variable objetivo denominada *estimated_conversions*, calculada como el producto entre el número de clics y la tasa de conversión que si se encuentran creadas. Al ser las conversiones la medida directa del rendimiento de la inversión se define esta variable como objetivo, convirtiéndose en un problema de regresión lineal. A diferencia de métricas intermedias como los *clicks* o agregadas como el *ROI*, las conversiones reflejan una acción final de mayor valor para la empresa y permiten aproximar de forma más directa la efectividad de la inversión publicitaria al facilitar el análisis del impacto marginal de la inversión.

En definitiva, se pueden sacar las siguientes conclusiones expuestas en la Figura 10 sobre el análisis del *dataset*. Se confirma la calidad y consistencia del *dataset*, así como la relevancia de métricas para evaluar el rendimiento. Destacando el canal publicitario como principal factor explicativo, lo que refuerza la necesidad de una asignación de inversión basada en datos.

Figura 10. Síntesis del proceso de preparación y análisis de datos



Fuente: elaboración propia

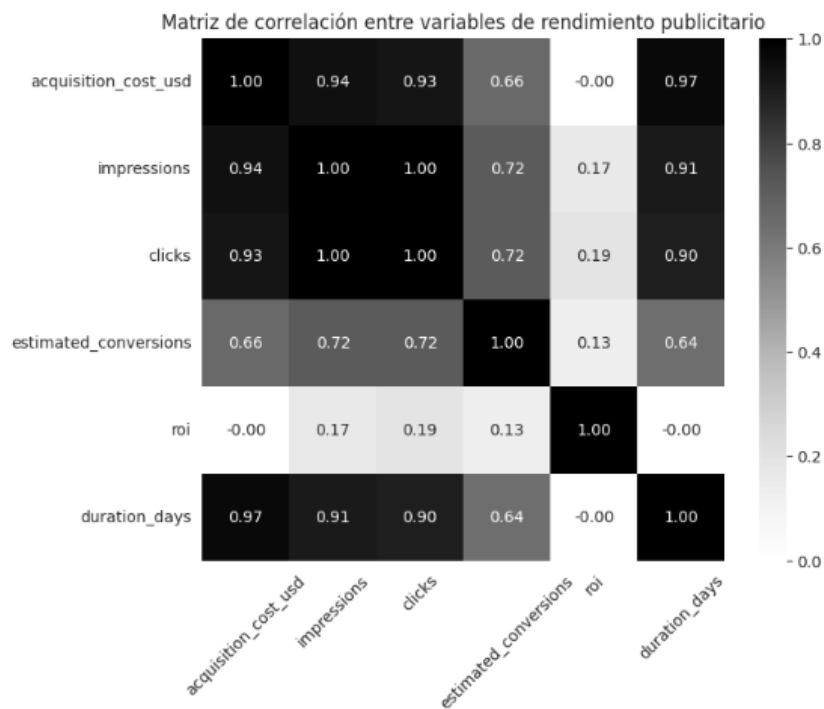
4. Metodología analítica

4.1 Modelos predictivos de conversión

Con el fin de modelizar la variable objetivo en función de las características de las campañas publicitarias, se han desarrollado distintos modelos de aprendizaje supervisado. El motivo de emplear este tipo de técnicas de aprendizaje supervisado se debe a la disponibilidad de una variable objetivo claramente definida. Se trata de “un modelo de aprendizaje automático que usa datos de entrenamiento etiquetados (datos estructurados) para asignar un atributo específico a una etiqueta” (Google Cloud, s. f., párr. 9). Por el contrario, técnicas de aprendizaje no supervisado, como el *clustering*, no resultan adecuadas para este problema, ya que su objetivo principal es identificar patrones o agrupaciones en los datos, permitiendo la segmentación de campañas, pero no modelar la relación que se busca.

En primer lugar, la regresión lineal constituye un modelo estadístico que permite “identificar la relación entre una variable dependiente continua y una o más variables independientes” (IBM, s. f., párr. 3). Los hiperparámetros internos del modelo se han mantenido en su configuración predeterminada. En este caso, la variable dependiente es *log_conversions* y como variables predictoras se seleccionan *acquisition_cost_usd*, *duration_days*, *ctr* y *channel_used*. En el caso de esta última, se escoge puesto que es la variable que según el análisis exploratorio presenta una relación potencial con la variable objetivo al cambiar el valor medio de conversiones según la categoría.

Por su parte, en la Figura 11 *duration_days* y *acquisition_cost_usd* muestran una correlación alta de 0.97, puesto que ambas variables contienen información similar dentro del conjunto de datos. En términos prácticos, esto sugiere que las campañas con mayor duración tienden a presentar mayores costes de adquisición. No obstante, se decidieron mantener ambas en el modelo porque su eliminación empeora el rendimiento predictivo, por lo tanto, pese a la colinealidad, aportan información relevante para la predicción.

Figura 11. Matriz de correlación entre variables de rendimiento publicitario

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Por otro lado, en la Tabla 6, la variable *CTR* presenta una correlación positiva moderada con la variable objetivo, que implica que tengan una relación relevante. Desde el punto de vista conceptual, esta relación resulta coherente, al reflejar el nivel de interacción generado por la campaña con un equilibrio entre capacidad predictiva y estabilidad del modelo. La inclusión de *CTR* mejora el rendimiento predictivo del modelo, obteniéndose mejores métricas de ajuste que al utilizar conjuntamente *clicks* e *impressions*.

Tabla 6. Matriz de correlación de variables numéricas del modelo

	ctr	cpa	estimated_conversions
duration_days	0.507920	-0.020155	0.641895
conversion_rate	0.002979	-0.698383	0.597959
acquisition_cost_usd	0.568000	-0.020141	0.663179
roi	0.292593	-0.141341	0.134281
clicks	0.695254	-0.099844	0.716391
impressions	0.684816	-0.090812	0.715495
engagement_score	0.312635	-0.149513	0.143016
ctr	1.000000	-0.166583	0.497589
cpa	-0.166583	1.000000	-0.461841
estimated_conversions	0.497589	-0.461841	1.000000

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

La regresión lineal se seleccionó como modelo base por su simplicidad, transparencia e interpretabilidad, cualidades especialmente valiosas para una primera aproximación al

problema predictivo. Con la incorporación de *PolynomialFeatures* se permiten interacciones de segundo grado, y no patrones no lineales aprendidos de manera más libre. Por ello, se prueban estas mismas variables seleccionadas con el modelo *Random Forest* basado en un conjunto de árboles de decisión. Es una técnica de ensamblado que combina múltiples modelos individuales para mejorar la precisión predictiva. Gracias a esto se consigue una mayor flexibilidad para modelar relaciones no lineales e interacción complejas sin necesidad de especificarlas previamente. Los principales hiperparámetros del algoritmo son aquellos presentados en la Tabla 7.

Tabla 7. *Hiperparámetros empleados en el modelo Random Forest*

Hiperparámetro	Valores Considerados	Descripción
n_estimators	[100, 200, 300]	Número de árboles del bosque. Un mayor número de árboles suele mejorar la estabilidad del modelo, aunque incrementa el coste computacional.
max_depth	[10, 20, None]	Profundidad máxima de los árboles. Valores bajos reducen la complejidad del modelo, mientras que valores altos pueden provocar sobreajuste.
min_samples_split	[2, 5, 10]	Número mínimo de muestras necesarias para dividir un nodo. Un valor muy bajo permite más divisiones y más profundidad, mientras que uno alto evita el sobreajuste.
min_samples_leaf	[1, 2, 4]	Número mínimo de muestras en cada hoja. Valores mayores reducen la variabilidad del modelo y ayudan a evitar sobreajuste.
max_features	["sqrt", "log2"]	Número de variables consideradas en cada división. Introduce aleatoriedad y reduce la correlación entre árboles.
bootstrap	[True]	Indica si se utiliza muestreo con reemplazo para construir cada árbol, aumentando la diversidad del modelo.

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Por último, el modelo XGBoost (*Extreme Gradient Boosting*) es un algoritmo predictivo de aprendizaje supervisado basado en el uso de múltiples árboles de decisión construidos de forma secuencial empleando la técnica de optimización del descenso de gradiente. Su estrategia de *boosting* lo diferencia del anterior modelo, ya que emplea un enfoque aditivo para corregir los errores cometidos por el conjunto de árboles construido hasta el momento. De forma que el modelo va corrigiendo desviaciones entre las conversiones observadas y las conversiones predichas. Asimismo, se requieren incorporar hiperparámetros adicionales comentados en la Tabla 8. Con el fin de evitar una elección arbitraria de estos, se definió un espacio de búsqueda con distintas configuraciones plausibles para los modelos de Random Forest y XGBoost. Esta estrategia permitió identificar, mediante validación cruzada, la combinación que ofrecía el mejor rendimiento predictivo.

Tabla 8. *Hiperparámetros empleados en el modelo XGBoost*

Hiperparámetro	Valores Considerados	Descripción
n_estimators	[150,250]	Indica el número de árboles que forman el modelo. Un mayor número de árboles permite mejorar la precisión, ya que el modelo corrige progresivamente los errores, aunque también muestra el tiempo de entrenamiento.
learning_rate	[0.05, 0.08, 0.1]	Tasa de aprendizaje. Valores bajos hacen que el modelo aprenda de forma más gradual, lo que suele mejorar la generalización, aunque requiere más árboles para alcanzar un buen rendimiento.
max_depth	[3, 4, 5]	Limita la profundidad de los árboles, una profundidad muy baja puede llevar al infraajuste mientras que una muy alta puede crear modelos demasiado complejos.
subsample	[0.8,1.0]	Indica la proporción de observaciones utilizadas para entrenar cada árbol.
colsample_bytree	[0.8,1.0]	Controla la proporción de variables utilizadas en cada árbol. Este parámetro reduce la dependencia entre árboles y contribuye a contrarlar el sobreajuste

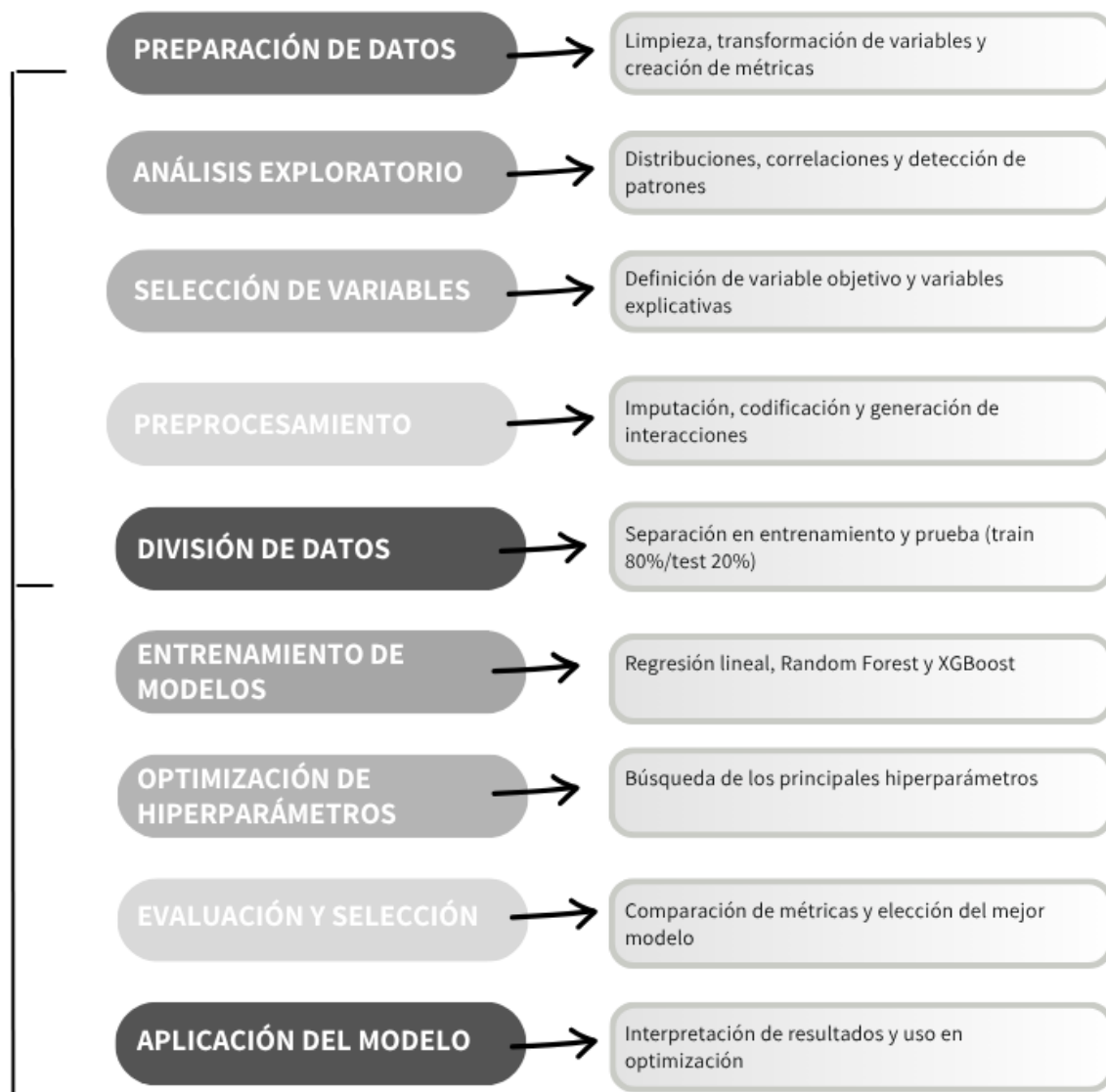
Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Aunque modelos como KNN o redes neuronales pueden aplicarse a problemas de regresión, no se han considerado en este análisis debido a sus limitaciones en este contexto. El algoritmo KNN, presenta un elevado coste computacional en *datasets* de gran tamaño como este con 300.000 observaciones. Asimismo, las redes neuronales ofrecen menor interpretabilidad porque su funcionamiento interno es más complejo y difícil de explicar y no suele ofrecer mejoras sistemáticas frente a modelos basados en árboles.

Por último, con el fin de sintetizar el procedimiento seguido en la construcción de los modelos predictivos, en la Figura 12 se presenta el flujo metodológico aplicado.

Figura 12. Flujo metodológico de los modelos predictivos

**Flujo
Metodológico**



Fuente: elaboración propia

4.2 Evaluación de modelos predictivos

En el ámbito de aprendizaje supervisado, los distintos modelos predictivos empleados para estimar el modelo de respuesta han de ser evaluados mediante diversas métricas cuantitativas, ya que permiten determinar si un modelo generaliza adecuadamente a datos no observados. En este problema de regresión lineal, donde se predicen valores continuos, las métricas más habituales son MAE, MSE, RMSE y R^2 . Los resultados resumidos en la Tabla 9 muestran que los tres modelos presentan un desempeño muy similar en términos de error. Sin embargo, se observan ligeras diferencias que permiten establecer comparaciones relevantes.

Tabla 9. *Resumen de algoritmos con sus resultados comentados*

Algoritmo	Resultados	Comentarios
XGBoost	MAE = 0.546 MSE = 0.456 R2 = 0.638	Presenta el mejor rendimiento global. Su capacidad para modelizar relaciones no lineales y la incorporación de regularización permiten obtener predicciones más precisas.
Random Forest	MAE = 0.546 MSE = 0.457 R2 = 0.637	Modelo con alto rendimiento y gran capacidad de generalización. La similitud con XGBoost se debe a que ambos son métodos de ensamblado basados en árboles.
Regresión Lineal	MAE = 0.547 MSE = 0.457 R2 = 0.637	A pesar de su simplicidad, ofrece un rendimiento competitivo. Indica que la relación entre variables puede ser en gran parte aproximada mediante un modelo lineal.

Fuente: elaboración propia

En primer lugar, el modelo XGBoost comete menos errores de predicción en comparación con los demás. Como en estas tres métricas un valor más bajo implica un mejor ajuste, esto significa que las predicciones generadas por el modelo están, en promedio, más cerca de los valores reales de conversiones. Un resultado cuya justificación se basa en la capacidad para modelizar relaciones complejas e incorporar mecanismos de regularización (Tianqi Chen & Carlos Guestrin, 2016).

En contraste, Random Forest presenta un desempeño cercano al de XGBoost con diferencias mínimas en las métricas. Una posibilidad de esta similitud se debe a que ambos modelos se basan en métodos de ensamblado de árboles de decisión, con la combinación de modelos simples que permiten reducir la varianza del modelo y mejorar su capacidad de generalización. Por último, la regresión lineal presenta una diferencia marginal con respecto a los otros modelos sofisticados, al ser capaz de aproximar la relación entre las variables con un nivel de precisión muy similar.

El análisis conjunto de las métricas de la Tabla 10 permite extraer las siguientes conclusiones:

- El error absoluto medio, MAE, es una métrica que mide el promedio de los errores absolutos entre los valores predichos y los valores reales observados, y es prácticamente idéntico entre los modelos, lo que indica estabilidad en las predicciones.
- El error cuadrático medio o MSE en inglés, se obtiene del promedio de los errores al cuadrado cuya escasa diferencia observada sugiere que ninguno de los modelos presenta desviaciones notablemente superiores (García et al., 2024). De la raíz de esta métrica se obtiene el valor del RMSE, por lo que ambos contienen la misma información sobre el error del modelo, aunque expresada en distinta escala (Sanz et al., 2023).
- El coeficiente de determinación, denotado como R^2 , se emplea como medida del grado de fiabilidad del modelo ajustado a un conjunto de datos (Martínez Rodríguez, 2005). Aproximadamente el 63% de la variación observada en las conversiones queda explicada por el modelo. El margen de mejora hace referencia a la proporción de variabilidad que el modelo no consigue explicar, la cual podría reducirse mejorando la calidad de los datos o incorporando nuevas variables.

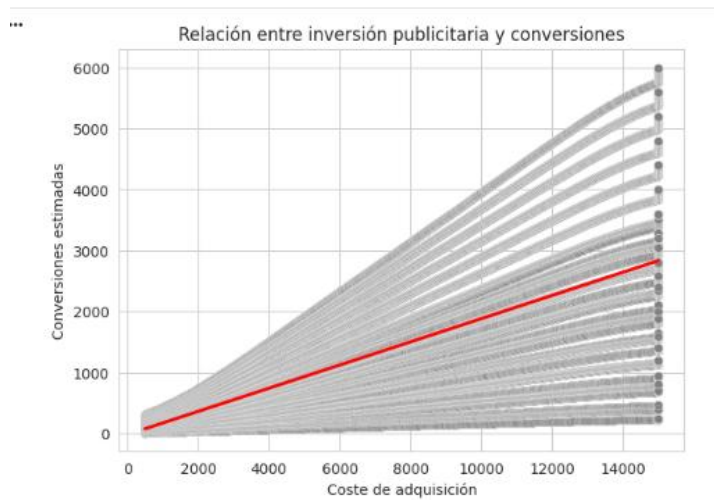
Tabla 10. Comparación del rendimiento de los modelos predictivos

===== COMPARATIVA DE MODELOS =====				
	MAE	MSE	RMSE	R2
XGBoost	0.546297	0.455820	0.675145	0.637675
RandomForest	0.546415	0.456539	0.675677	0.637103
LinearRegression	0.546590	0.456865	0.675918	0.636844

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

En este caso, aunque el análisis de la Figura 13 sugiere una relación aproximadamente lineal entre la inversión publicitaria y las conversiones, los resultados de evaluación muestran que XGBoost presenta el mejor desempeño global. Se opta por seleccionar este modelo de referencia por la importancia de maximizar la precisión de las predicciones.

Figura 13. *Relación entre inversión publicitaria y conversiones*



Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

El siguiente análisis se realiza de forma individual para cada canal, por lo tanto, los resultados obtenidos en términos de saturación, elasticidad y escenarios no constituyen todavía una decisión de asignación presupuestaria global. Servirá como base para determinar la asignación conjunta del presupuesto entre todos los canales. Para ello, se construye un escenario controlado en el que variables como *duration_days*, *engagement_score* y *ctr* se mantienen constantes en valores de referencia para aislar el efecto de la inversión sobre las conversiones. Estas últimas no se pueden fijar antes del lanzamiento de una campaña, sino que su resultado depende de su comportamiento tras ponerse en ejecución.

4.3 Curvas de saturación y elasticidad

Antes de la formulación del modelo de optimización es necesario emplear el modelo predictivo de referencia para estudiar empíricamente la relación entre el gasto y las conversiones esperadas. En primer lugar, en el marco teórico se definió el efecto de saturación en las campañas publicitarias como un fenómeno por el cual, a medida que aumenta la presión publicitaria, el impacto adicional de cada nueva exposición disminuye. Este comportamiento es consistente con la existencia de rendimientos decrecientes, cuya presencia se analiza mediante la simulación de distintos escenarios bajo el supuesto *ceteris paribus*⁷ y tres indicadores específicos para su detección.

⁷ *Ceteris paribus* es una expresión en latín que significa literalmente “manteniendo todo lo demás constante”.

El primer indicador es el coste en que se incurre para adquirir un nuevo cliente o también conocido como *CPA*. En la Tabla 11, se presentan valores parecidos en los canales de Facebook, Instagram y Twitter, sin embargo, su combinación con la elasticidad con valor inferior a 1, sugiere que estos canales se encuentran en una fase cercana a la saturación. “La elasticidad publicitaria de la demanda se calcula dividiendo el cambio porcentual en la cantidad demandada entre el cambio porcentual en el gasto publicitario” (Investopedia, s.f., párr. 1, traducción propia). Eso significa, interpretado de forma aproximada, que un aumento del 1% en la inversión se asocia con un aumento cercano al 0.5% en conversiones en los tres canales. Por el contrario, Pinterest registra una elasticidad y *CPA* con valores más altos, lo que indica que, aunque sea menos eficiente en términos de coste medio, conserva un retorno marginal superior. Mantener un nivel alto en este último indicador demuestra una situación de menor saturación relativa.

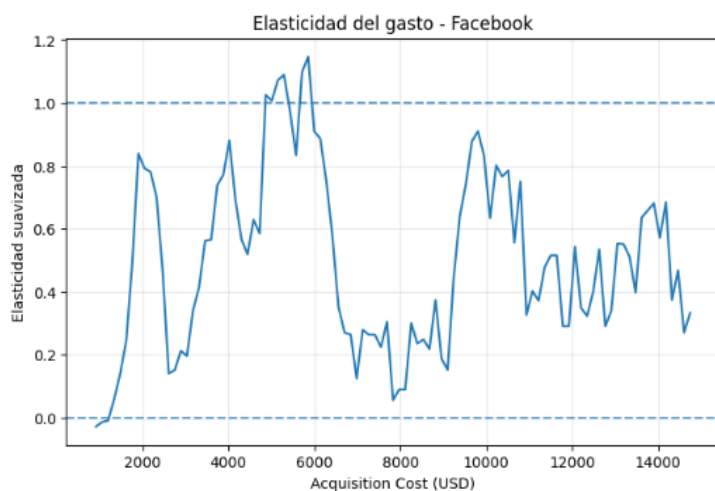
Tabla 11. Comparación de eficiencia y elasticidad por canal publicitario

channel_used	estimated_cpa	elasticity	marginal_return
Facebook	6.957	0.504	0.078
Instagram	6.970	0.503	0.078
Pinterest	7.269	0.507	0.076
Twitter	6.924	0.496	0.078

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Si se analiza la elasticidad por separado en tramos se puede identificar con mayor precisión las fases del proceso de saturación, evidenciando que la pérdida de sensibilidad ocurre de forma progresiva. El patrón se repite de forma casi idéntica en Instagram, Twitter y Facebook con una fase inicial de respuesta moderada, con mayor capacidad de respuesta en niveles intermedios de inversión entre 4,3 mil y 11,2 mil USD, como se ve en la Figura 14. Posteriormente, existe una reducción en la capacidad de respuesta a medida que aumenta la inversión.

Figura 14. Elasticidad del gasto publicitario en función del nivel de inversión en Facebook

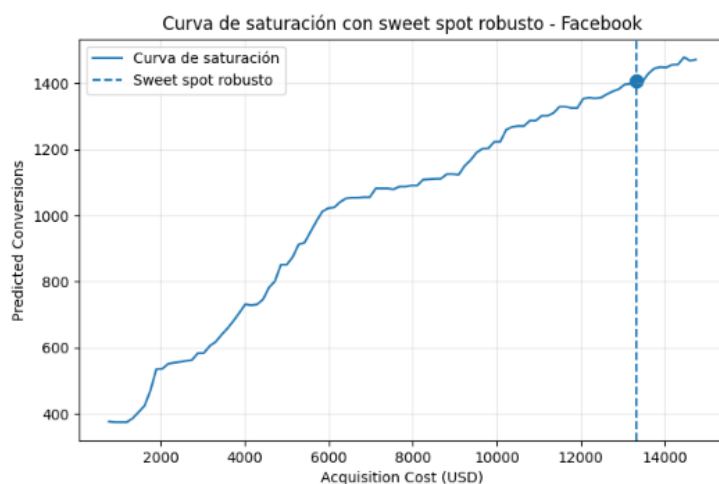


Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

El análisis conjunto de estos indicadores muestra que los canales no están en la misma fase de su curva de saturación, y que ningún indicador por sí solo basta para diagnosticarla. La saturación es útil porque transforma las predicciones del modelo en información accionable para la toma de decisiones. Permite identificar el punto a partir del cual los incrementos adicionales de gasto generan beneficios marginales cada vez menores.

Tomando la Figura 15 y Tabla 12 como ejemplo, se observa una relación no lineal entre la inversión publicitaria y las conversiones predichas. En los niveles iniciales de inversión, la curva presenta una pendiente pronunciada, lo que indica una alta eficiencia del gasto. Sin embargo, a medida que aumenta la inversión, la pendiente se reduce y se evidencia la presencia de rendimientos decrecientes. Asimismo, se definió el punto marcado como *sweet spot* robusto entendido como el primer nivel de inversión a partir del cual se alcanza una proporción elevada del rendimiento máximo. En este caso, se elige un umbral del 95% por su carácter intermedio entre eficiencia y exhaustividad (Rothenberg, 2005).

El máximo de conversiones se sitúa en torno a las 1.470 conversiones para un nivel de inversión cercano a 15.000 USD. Sin embargo, el *sweet spot* robusto se alcanza en los 13.500 USD, con unas 1.400 conversiones. Esto implica que un incremento adicional de 1.500 USD en inversión solo genera en torno a 70 conversiones adicionales. Por lo tanto, el resultado sugiere que el canal ya se encuentra en una fase avanzada de maduración. Desde una perspectiva de decisión, esto implica que seguir invirtiendo puede justificarse si el objetivo es maximizar volumen, pero no necesariamente si la prioridad es optimizar la eficiencia del gasto.

Figura 15. Curva de saturación y punto óptimo de inversión en Facebook

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Tabla 12. Resumen de la curva de saturación por tramos de inversión en Facebook

Tramo de Inversión (USD)	Conversiones	Incremento de inversión	Incremento de conversiones	Eficiencia marginal
1.500 → 4.000	380 → 730	+2.500	+350	Alta
4.000 → 6.000	730 → 1.030	+2.000	+300	Alta
6.000 → 10.000	1.030 → 1.220	+4.000	+190	Media
13.300 → 15.000	1.406 → 1.470	+1.700	+64	Baja

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

4.4 Simulación de escenarios

Tras analizar la respuesta del mercado a la inversión publicitaria desde el concepto de saturación, así como de la sensibilidad de dicha relación mediante la elasticidad, el siguiente paso consiste en aplicar estos resultados a situaciones reales con el fin de orientar la toma de decisiones más adecuadas. A partir del *sweet spot* robusto entendido como un punto de referencia que captura la mayor parte del potencial del canal, se definieron cuatro escenarios de inversión para cada canal. Si el *sweet spot* robusto indica el nivel de gasto a partir del cual ya se han capturado aproximadamente el 95% del máximo de conversiones estimadas, resulta útil analizar que ocurre si la inversión se sitúa con otro valor.

Los escenarios se construyen sobre el *acquisition_cost* al punto de base, no sobre el 95% directamente.

- 70%: escenario prudente o conservador
- 85% : escenario cercano al óptimo, pero más eficiente en coste
- 100% : punto de referencia, es igual al *sweet spot* robusto

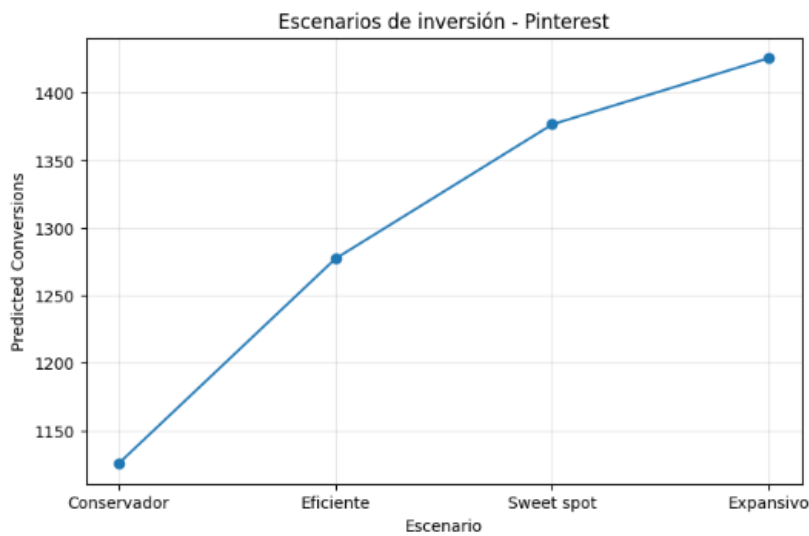
- 115%: escenario de expansión para comprobar si compensa seguir invirtiendo

Estos escenarios son creados siguiendo la técnica de modelización financiera conocida como *what-if-analysis* (Corporate Finance Institute, 2024). Esta permite evaluar cómo varían los resultados de un modelo cuando se modifican una o varias variables de entrada. En el ámbito de inversión publicitaria, esta herramienta permite responder a la pregunta ¿qué pasaría si se invierte menos, lo recomendado o algo más? Si continuamos con el ejemplo anterior del caso de Facebook, el *sweet spot* robusto se sitúa en 13.500 USD, sobre el que se construyen los escenarios de la siguiente manera:

- Escenario conservador corresponde a invertir 70% de 13.500 USD
- Escenario eficiente representa el 85% de 13.500 USD
- Escenario *sweet spot* mantiene el 100% que sería 13.500 USD
- Escenario expansivo supone el 115% de 13.500 USD

Una vez definidos los presupuestos, el modelo XGBoost estima para cada uno de ellos el volumen de conversiones esperado y el *CPA* asociado. Los resultados evidencian un patrón común en todos los canales analizados, puesto que el aumento de la inversión genera un incremento en el volumen de conversiones, pero a costa de una pérdida progresiva de eficiencia. En el caso de Pinterest, en el epígrafe de la saturación se observó como todavía mejora su capacidad de respuesta cuando pasa de niveles bajos a medios de inversión, entrando en su zona de mejor aprovechamiento algo más tarde que el resto. Mientras que en la Tabla 13, una vez superado ese umbral, se observa una reducción más acusada en los niveles altos de inversión, especialmente en el paso del escenario *sweet spot* al expansivo, donde el rendimiento marginal desciende de forma notable. Por lo tanto, como se observa en la Figura 16, el escenario expansivo, aunque maximiza el volumen de conversiones, lo hace a costa de una menor eficiencia marginal, lo que pone de manifiesto la existencia de un *trade-off* entre volumen y coste.

Figura 16. Comparación de conversiones estimadas según escenarios de inversión en Pinterest



Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Tabla 13. Resultados de los escenarios de inversión por canal y métricas de eficiencia

===== TABLA EJECUTIVA DE ESCENARIOS =====

	channel_used	scenari	acquisition_cost_usd	predicted_conversions	estimated_cpa	delta_cost_vs_prev	delta_conv_vs_prev	marginal_conv_per_usd_vs_prev
0	Facebook	Conservador	9331.783254	1167.958618	7.989823	NaN	NaN	NaN
1	Facebook	Eficiente	11331.451094	1311.025635	8.643196	1999.667840	143.067017	0.071545
2	Facebook	Sweet spot	13331.118934	1406.424805	9.478728	1999.667840	95.399170	0.047708
3	Facebook	Expansivo	15330.786774	1480.261963	10.356807	1999.667840	73.837158	0.036925
4	Instagram	Conservador	9331.783254	1165.545898	8.006363	NaN	NaN	NaN
5	Instagram	Eficiente	11331.451094	1308.317627	8.661086	1999.667840	142.771729	0.071398
6	Instagram	Sweet spot	13331.118934	1404.310425	9.493000	1999.667840	95.992798	0.048004
7	Instagram	Expansivo	15330.786774	1476.234985	10.385058	1999.667840	71.924561	0.035968
8	Pinterest	Conservador	9529.352873	1125.349609	8.467904	NaN	NaN	NaN
9	Pinterest	Eficiente	11571.357060	1277.100220	9.060650	2042.004187	151.750610	0.074315
10	Pinterest	Sweet spot	13613.361247	1376.590576	9.889187	2042.004187	99.490356	0.048722
11	Pinterest	Expansivo	15655.365435	1425.737915	10.980536	2042.004187	49.147339	0.024068
12	Twitter	Conservador	9331.783254	1173.808105	7.950007	NaN	NaN	NaN
13	Twitter	Eficiente	11331.451094	1315.099121	8.616424	1999.667840	141.291016	0.070657
14	Twitter	Sweet spot	13331.118934	1407.533936	9.471259	1999.667840	92.434814	0.046225
15	Twitter	Expansivo	15330.786774	1478.096436	10.371981	1999.667840	70.562500	0.035287

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Desde una perspectiva de decisión, un *trade-off* es una situación en la que la mejora de un objetivo implica el deterioro parcial de otro (Zhang et al., 2021). En este caso, si se considera exclusivamente la eficiencia, medida a través del coste por adquisición (*CPA*), se observa en la Tabla 14 que el escenario conservador resulta óptimo en todos los canales analizados. Esta evidencia resulta coherente con la estructura de las curvas de saturación estimadas, donde los niveles iniciales de inversión se asocian a una mayor productividad marginal del gasto.

Tabla 14. *Escenario óptimo por canal en términos de eficiencia*

```
===== MEJOR ESCENARIO POR EFICIENCIA (MENOR CPA) =====
```

	channel_used	scenario	acquisition_cost_usd	predicted_conversions	estimated_cpa
0	Facebook	Conservador	9331.783254	1167.958618	7.989823
1	Instagram	Conservador	9331.783254	1165.545898	8.006363
2	Pinterest	Conservador	9529.352873	1125.349609	8.467904
3	Twitter	Conservador	9331.783254	1173.808105	7.950007

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Sin embargo, si el objetivo es encontrar un equilibrio entre volumen y eficiencia, la referencia debe buscarse en la Tabla 15. Los niveles intermedios de inversión son, en general, los más adecuados desde una perspectiva de rendimiento global, poniendo en manifiesto que no coincide necesariamente con el de menor *CPA*. Por lo tanto, si el objetivo del anunciante es decidir dónde y cuánto invertir, la evidencia sugiere priorizar Facebook, Instagram y Twitter en niveles cercanos al escenario eficiente, evitando la sobreinversión en escenarios expansivos.

Tabla 15. *Escenario óptimo por canal según el equilibrio entre volumen de conversiones y eficiencia*

```
===== MEJOR ESCENARIO POR EQUILIBRIO VOLUMEN / EFICIENCIA =====
```

	channel_used	scenario	acquisition_cost_usd	predicted_conversions	estimated_cpa	balance_score
0	Facebook	Eficiente	11331.451094	1311.025635	8.643196	0.182067
1	Instagram	Eficiente	11331.451094	1308.317627	8.661086	0.184288
2	Pinterest	Sweet spot	13613.361247	1376.590576	9.889187	0.270732
3	Twitter	Eficiente	11331.451094	1315.099121	8.616424	0.189178

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Con el fin de clarificar el procedimiento seguido en la construcción de escenarios de inversión y su evaluación la Figura 17 presenta de forma esquemática las principales etapas del proceso.

Figura 17. Flujo metodológico para la construcción y evaluación de escenarios de inversión publicitaria



Fuente: elaboración propia

4.5 Análisis de duración

La duración de la campaña constituye otra de las variables del modelo que determina el periodo en el que el presupuesto se distribuye y genera impacto sobre la audiencia. En primer lugar, en los niveles bajos de presupuesto, las conversiones estimadas permanecen prácticamente invariables dentro cada tramo de duración. Por ejemplo, para las campañas de 7 y 14 días el modelo predice aproximadamente 1.246,07 conversiones con 13.000 USD de presupuesto. La ausencia de diferencias relevantes en el tramo de duraciones más cortas puede interpretarse como que la campaña no está generando un procesamiento suficiente, como para traducirse en cambios apreciables en la respuesta. Esta interpretación se alinea con el estudio de Sang Yeal Lee y Yong-Suk Cho, quienes encuentran que la relación entre repetición y efectividad no es lineal y que, en marcas no familiares, el recuerdo comienza a emerger solo

después de varias repeticiones (Lee, 2010). En nuestro caso, a partir de los 30 días si se observa una mejora sostenida en las conversiones.

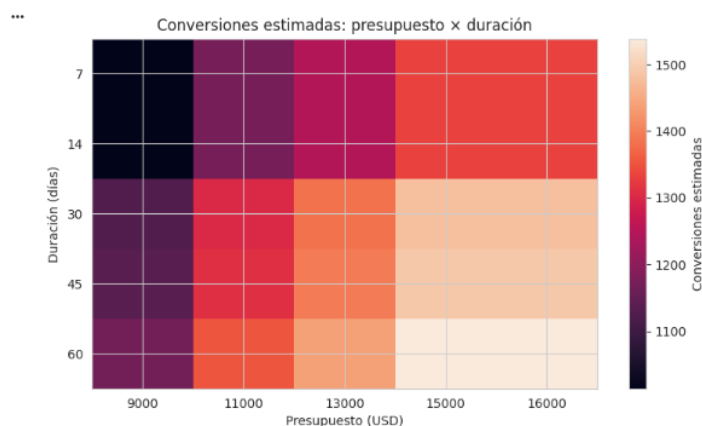
Asimismo, los resultados de la Tabla 16 y Figura 18 apuntan a una interacción entre las variables *acquisition_cost* y *duration_days*. La segunda comienza a adquirir mayor capacidad explicativa cuando el presupuesto alcanza un nivel suficiente para sostener una presencia publicitaria continuada. Por ejemplo, la diferencia entre una campaña de 7 días y una de 60 días es de aproximadamente 155,74 conversiones con un presupuesto de 9.000 USD, mientras que supera las 206 conversiones con un presupuesto de 15.000 USD. Desde una perspectiva de optimización, cuando el presupuesto es alto, alargar la campaña es la estrategia más rentable, porque la mayor duración sí se traduce en un aumento relevante de conversiones. Sin embargo, si el presupuesto es bajo, la mejor opción es una campaña corta y concentrada para no perder intensidad.

Tabla 16. Conversiones estimadas en función del nivel de inversión y duración de la campaña

acquisition_cost_usd	9000	11000	13000	15000	16000
duration_days					
7	1013.891846	1171.858521	1246.070435	1332.003540	1332.003540
14	1013.891846	1171.858521	1246.070435	1332.003540	1332.003540
30	1125.661621	1302.305420	1384.771240	1480.261963	1480.261963
45	1132.985352	1310.777222	1393.779175	1490.495605	1490.495605
60	1169.626953	1353.163818	1438.847656	1538.689331	1538.689331

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Figura 18. Mapa de calor de conversiones estimadas según nivel de inversión y duración de la campaña



Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

En definitiva, en los diferentes puntos abordados en la metodología del trabajo, se han utilizado diferentes modelos para predecir los resultados de las campañas publicitarias, obteniendo en general, un rendimiento similar, aunque el modelo XGBoost destaca por ofrecer una mayor precisión. Los resultados muestran que variables como el coste, la duración de la campaña y el nivel de interacción influyen directamente en las conversiones. Sin embargo, aumentar la inversión no siempre indica mejores resultados proporcionales, ya que a partir de cierto punto los beneficios crecen cada vez menos. Por ello, es importante identificar un nivel de inversión adecuado que permita obtener buenos resultados sin gastar de más. Se pueden sacar las siguientes conclusiones expuestas en la Figura 19.

Figura 19. *Análisis de modelos predictivos y principales hallazgos*

	ANÁLISIS DEL DATASET	PROCESO DE ANÁLISIS	PRINCIPALES HALLAZGOS	
5.1	MODELOS PREDICATIVOS	Se utilizaron distintos modelos para predecir las conversiones a partir de las características de las campañas	Modelos con resultados parecidos	XGBoost es el más preciso
5.2	EVALUACIÓN DE MODELOS	Se compararon los modelos mediante métricas para medir su capacidad de predicción	XGBoost comete menos errores	El modelo explica gran parte de los resultados
5.3	SATURACIÓN Y ELASTICIDAD	Se analizó cómo cambian las conversiones al aumentar la inversión en cada canal	Más inversión = Más conversiones pero cada vez aporta menos resultados	Algunos canales están cerca de su límite de saturación
5.4	ESCENARIOS DE INVERSIÓN	Se simularon distintos niveles de inversión para ver cómo cambian los resultados	Invertir más aumenta el volumen pero reduce la eficiencia	Los niveles intermedios son lo más equilibrados
5.5	DURACIÓN DE CAMPAÑA	Se analizó el efecto de la duración en las conversiones según el presupuesto	Con poco presupuesto mejor las campañas cortas	Con más presupuesto mejor campañas largas La duración influye más cuando la inversión es alta

Fuente: elaboración propia

5. Aplicación práctica y recomendaciones

5.1 Marco de optimización presupuestaria

Con el análisis del modelo predictivo y de los principales patrones de comportamiento del sistema se procede a la construcción de un marco de optimización presupuestaria orientado a transformar dichos resultados en una recomendación operativa de asignación de inversión en *Social Media Ads*.

El objetivo de este marco es la toma de decisión sobre la combinación de inversión más adecuada entre los distintos canales disponibles. Para hacerlo, no solo se tiene en cuenta el rendimiento esperado en términos de conversiones en los canales de *Social Media*, sino también las condiciones y límites reales del entorno real de gestión publicitaria. A diferencia de las etapas anteriores de análisis, centradas en entender cómo se comporta el modelo predictivo ante distintas configuraciones de inversión, esta fase introduce explícitamente un problema de decisión, en el que se busca identificar la solución óptima entre un conjunto de alternativas. De esta manera se avanza hacia un análisis prescriptivo alineado con las necesidades reales de planificación.

Un modelo de optimización suele estar compuesto de tres elementos básicos: función objetivo, variables de decisión y restricciones. (Martinich, s.f)

En primer lugar, la función objetivo no busca únicamente obtener un mayor número de conversiones, sino encontrar una combinación que de buenos resultados sin que el coste por acción, *CPA*, tenga resultados poco realistas. Para ello, el modelo premia las opciones que generan conversiones de forma eficiente y penaliza por saturación las alternativas que parecen buenas por volumen, pero en realidad son poco eficientes.

En segundo lugar, la variable de decisión representa una cantidad controlable del sistema, en este caso, consiste en elegir un único escenario para cada canal publicitario para tener una asignación correcta del presupuesto. La diferencia con el elemento anterior radica en que las variables de decisión representan la elección entre las distintas alternativas disponibles, mientras que la función objetivo define el criterio de evaluación que permite identificar la solución óptima. La optimización no incorpora todas las variables utilizadas en el modelo predictivo, variables como *CTR* o *engagement* no se incluyen como variables de decisión, dado que no son directamente controlables, sino que su efecto ya está recogido en las predicciones generadas por el modelo.

Por último, las restricciones operativas del modelo permiten delimitar el espacio de soluciones puesto que no todas ellas son admisibles. El modelo incorpora una restricción de presupuesto total de modo que la suma de las inversiones seleccionadas no puede exceder el presupuesto disponible. Asimismo, dado que para cada canal se han construido varios escenarios, el modelo incorpora una restricción de selección única por canal. Para evitar forzar una distribución artificial del presupuesto entre todos los canales, se incluye la posibilidad de no invertir.

Conviene señalar que la optimización realizada se apoya en las predicciones del modelo seleccionado. Por tanto, los resultados obtenidos deben interpretarse como recomendaciones basadas en patrones aprendidos de datos históricos, y no como relaciones causales exactas. Este enfoque presenta como principal ventaja su aplicabilidad práctica y flexibilidad para analizar distintos escenarios. No obstante, las recomendaciones son orientativas y no constituyen una demostración causal exacta del efecto de la inversión.

5.2 Resultados de la optimización presupuestaria

El modelo de optimización desarrollado permite abordar el problema de asignación de presupuesto desde una perspectiva conjunta, considerando simultáneamente todos los canales bajo una serie de restricciones. En primer lugar, en la Tabla 17, se obtienen los resultados bajo un presupuesto máximo de 50.000 USD y utilizando como función objetivo la maximización del volumen total de conversiones dentro del presupuesto disponible. Desde una perspectiva empresarial, este resultado recomienda adoptar una estrategia de asignación presupuestaria diferenciada por canal. En concreto, la mayor asignación a Instagram y Pinterest de 27%, indica que estos canales presentan una mayor capacidad para transformar incrementos de inversión en conversiones adicionales. En el caso de Pinterest, el análisis de saturación presentaba un menor volumen medio de conversiones y un CPA más elevado, pero con una elasticidad superior, por ello, su inclusión en la solución óptima se puede justificar al margen que tiene de seguir generando conversiones adicionales ante aumentos de inversión.

Por otro lado, Facebook y Twitter, aunque siguen siendo relevantes, reciben una inversión más moderada, en torno al 23%, al encontrarse en una fase con menor impacto marginal en el resultado global.

Tabla 17. Solución óptima del modelo de asignación presupuestaria bajo restricción de inversión

	channel_used	scenari	acquisition_cost_usd	predicted_conversions	estimated_cpa	share_of_budget	share_of_conversions
0	Facebook	Eficiente	11331.451094	1311.025635	8.643196	0.228423	0.242467
1	Instagram	Sweet spot	13331.118934	1404.310425	9.493000	0.268733	0.259720
2	Pinterest	Sweet spot	13613.361247	1376.590576	9.889187	0.274422	0.254593
3	Twitter	Eficiente	11331.451094	1315.099121	8.616424	0.228423	0.243220

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

¿Por qué se elige el presupuesto de 50.0000 USD? Dado que los escenarios de inversión por canal se sitúan entre 11.000 y 13.000 USD, se ha seleccionado este límite máximo. Es un dato coherente con el rango de inversión observado, permitiendo combinar distintos escenarios entre canales sin exceder niveles realistas de gasto.

En segundo lugar, cuando el objetivo del modelo no se limita a maximizar conversiones, sino que también se incorpora la eficiencia de la inversión, la solución óptima cambia de forma significativa. En este caso, en la Tabla 18 se muestra que la mejor solución es mantener niveles de inversión más bajos, sin gastar todo el presupuesto inicial. Se prefiere renunciar a parte del volumen potencial para preservar una mejor eficiencia en la asignación. Como se ha visto en puntos anteriores, existen rendimientos decrecientes que hacen que deje de ser necesariamente óptimo agotar todo el presupuesto disponible. Este comportamiento evidencia la existencia de un *trade-off* entre crecimiento y rentabilidad.

Tabla 18. Solución óptima del modelo de asignación presupuestaria bajo restricción de eficiencia

	channel_used	scenari	acquisition_cost_usd	predicted_conversions	estimated_cpa	share_of_budget	share_of_conversions
0	Facebook	Conservador	9331.783254	1167.958618	7.989823	0.248684	0.252114
1	Instagram	Conservador	9331.783254	1165.545898	8.006363	0.248684	0.251593
2	Pinterest	Conservador	9529.352873	1125.349609	8.467904	0.253949	0.242916
3	Twitter	Conservador	9331.783254	1173.808105	7.950007	0.248684	0.253377

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Más allá de los resultados obtenidos, el modelo desarrollado permite extraer una serie de recomendaciones relevantes para la gestión de la inversión publicitaria:

- Asignar el presupuesto de forma diferenciada entre canales, priorizando aquellos con mayor capacidad de generar conversiones adicionales. No todos los canales responden igual a la inversión.

- No agotar necesariamente el presupuesto disponible, especialmente cuando los incrementos de inversión generan una pérdida de eficiencia.
- Adaptar la estrategia al comportamiento de cada canal, teniendo en cuenta su grado de saturación y su capacidad de respuesta.

En este sentido, el modelo aporta un marco estructurado que facilita la toma de decisiones, permitiendo priorizar entre alternativas y reducir la arbitrariedad en la asignación de recursos.

6. Limitaciones del estudio

Una de las principales limitaciones del estudio reside en la homogeneidad observada en las variables categóricas. Esto no solo limita la capacidad explicativa del modelo, sino también la utilidad de la segmentación que apenas permite identificar patrones de rendimiento diferenciados. Esto se observa en la siguiente Tabla 19, donde existen reducidos valores en las columnas *between_total_ratio* y *eta_squared* obtenidos para la mayoría de las variables categóricas. Estas columnas se obtienen de comparar la variabilidad entre grupos con la variabilidad total de una métrica.

- *Between_total_ratio*: indica cuánto cambian los resultados al comparar unos grupos con otros.
 - Si da un valor pequeño, significa que separar en grupos apenas cambia nada.
 - Si da un valor grande, significa que separar en grupos sí cambia bastante los resultados.
- *Eta_squared*: indica cuánto explica esa variable de grupo las diferencias en los resultados
 - Si da cerca de 0, esa variable casi no explica nada.
 - Si da más alto, esa variable sí que ayuda a explicar las diferencias.

La mayor parte de la variabilidad entre grupos se explica por canal, mientras que el resto de las variables categóricas con valores cercanos a 0.00, muestran un comportamiento muy homogéneo y apenas discriminan el rendimiento.

Tabla 19. *Análisis de la capacidad explicativa de las variables categóricas*

group_col	value_col	n_groups	global_mean	total_std	between_std	between_total_ratio	eta_squared
4	channel_used	cpa	9.049	9.470	2.692	0.2843	0.0606
24	company	cpa	9.049	9.470	0.115	0.0122	0.0001
0	customer_segment	cpa	9.049	9.470	0.031	0.0033	0.0000
8	target_audience	cpa	9.049	9.470	0.058	0.0061	0.0000
12	campaign_goal	cpa	9.049	9.470	0.032	0.0034	0.0000
16	location	cpa	9.049	9.470	0.033	0.0035	0.0000
20	language	cpa	9.049	9.470	0.020	0.0021	0.0000
6	channel_used	ctr	0.314	0.025	0.015	0.5886	0.2600
26	company	ctr	0.314	0.025	0.000	0.0122	0.0001
2	customer_segment	ctr	0.314	0.025	0.000	0.0053	0.0000
10	target_audience	ctr	0.314	0.025	0.000	0.0042	0.0000
14	campaign_goal	ctr	0.314	0.025	0.000	0.0013	0.0000
18	location	ctr	0.314	0.025	0.000	0.0029	0.0000
22	language	ctr	0.314	0.025	0.000	0.0023	0.0000

Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

En segundo lugar, la estimación de efectos dinámicos, como los de arrastre comentados en el apartado del marco teórico, requieren de datos longitudinales con observaciones repetidas para ser identificados, es decir, no sirve con tener una fecha. En este caso, la variable *date* no muestra una evolución repetida de la misma campaña en distintos momentos, por lo tanto, el estudio no puede estimar de forma rigurosa si la inversión publicitaria realizada en un período continúa generando efectos en periodos posteriores, ni estimar la intensidad de dicha persistencia.

Asimismo, algunas variables utilizadas en el modelo no son independientes, puesto que se han calculado usando otras variables del propio *dataset*, como es el caso de las conversiones estimadas o el coste por adquisición. Esto implica que pueden generarse dependencias estructurales en los datos, lo que facilita que el modelo obtenga buenas predicciones. Sin embargo, parte de ese rendimiento puede estar influido por la propia construcción de las variables. Por otro lado, dentro de la fuente pública de Kaggle se menciona que este *dataset* es una base de datos ficticia. Esta naturaleza sintética implica que el conjunto de datos no reproduce completamente la complejidad, competencia y variabilidad en el entorno publicitario real. En consecuencia, se limita la generalización de los resultados a contextos reales.

Por último, este análisis presenta la limitación de asumir que el resto de los factores permanecen constantes. En la práctica, el rendimiento de una campaña no depende solo de la inversión, sino también de otros elementos que pueden cambiar al mismo tiempo.

7. Conclusiones

El desarrollo de este trabajo pone de manifiesto la necesidad de replantear la forma en la que se aborda la inversión publicitaria en entornos digitales. Durante mucho tiempo, las decisiones se han basado en intuición o en reglas simples heredadas de experiencias pasadas. Pero en el presente análisis se revela que este enfoque ya se queda corto. Los resultados obtenidos evidencian que el comportamiento de las campañas digitales responde a dinámicas complejas, no lineales y heterogéneas entre canales.

Uno de los principales hallazgos es que la relación entre inversión y resultados no es proporcional. En la práctica, esto significa que llega un punto en el que seguir incrementando la inversión no solo deja de ser eficiente, sino que puede implicar un uso poco óptimo de los recursos. Más inversión ya no se traduce en un crecimiento significativo, sino en un aumento del coste por cada resultado obtenido. Esto rompe con una de las creencias más arraigadas en la publicidad y obliga a replantear por completo el enfoque estratégico. Ya no se trata simplemente decidir cuánto invertir, sino identificar con precisión el momento en el que el beneficio adicional deja de justificar el coste.

En este contexto, el análisis revela que la gestión de la inversión publicitaria está atravesada por una tensión constante entre dos objetivos que, aunque relacionados, no siempre avanzan en la misma dirección. Aumentar el volumen (más conversiones, más alcance, más impacto) suele exigir incrementar la inversión, pero este crecimiento viene acompañado, como se ha observado de una pérdida progresiva de eficiencia. El modelo desarrollado evidencia que, bajo un enfoque orientado al crecimiento, la solución óptima tiende a concentrar recursos en aquellos canales con mayor elasticidad, como Instagram y Pinterest, capaces de seguir generando conversiones adicionales ante incrementos de inversión.

Por el contrario, optimizar la eficiencia (reducir el coste por conversión o maximizar el retorno) implica, en muchos casos, contener o incluso limitar el nivel de gasto. Esta dualidad introduce un matriz fundamental, que consiste en que la inversión publicitaria deja de ser una cuestión meramente técnica u operativa, para convertirse en una decisión profundamente estratégica.

Otro aspecto relevante es la limitación de ciertas variables categóricas para explicar las diferencias en el rendimiento de las campañas. A pesar de que, en la práctica, la segmentación suele apoyarse en múltiples dimensiones, los resultados muestran que muchas de estas variables apenas aportan información útil para discriminar qué es lo que funciona mejor. Dentro de este marco, el canal destaca como una de las pocas variables con una capacidad para

segmentar el rendimiento. Los canales publicitarios no solo difieren en su rendimiento, sino también en la forma en la que responden a la inversión. Algunos canales muestran signos de saturación, al alcanzar rápidamente un punto en el que aumentar la inversión apenas genera mejoras adicionales en los resultados. Esto suele ocurrir en entornos muy competitivos o con audiencias limitadas, donde la presión publicitaria es elevada y las oportunidades de captación incremental son reducidas. Por ello, la diferenciación entre canales no es solo una recomendación, sino una condición necesaria para optimizar el rendimiento global.

Desde una perspectiva metodológica, se integran modelos predictivos con técnicas de optimización como vía para avanzar desde el análisis hacia la toma de decisiones informadas. Mientras que los modelos predictivos permiten estimar el comportamiento esperado de las campañas ante distintos escenarios, la optimización introduce un marco formal que permite seleccionar, entre las alternativas posibles, aquellas que mejor se ajustan a un conjunto de objetivos y restricciones reales. El análisis deja de ser meramente descriptivo o exploratorio para convertirse en un instrumento operativo, capaz de identificar soluciones que son simultáneamente coherentes desde el punto de vista teórico y viables en la práctica.

Su principal aportación no reside únicamente en las soluciones específicas que genera, sino en la lógica de decisión que introduce, permitiendo abordar la asignación de recursos de forma coherente y basada en evidencia.

En esta línea, resulta fundamental reconocer que el modelo desarrollado no elimina la incertidumbre inherente al entorno publicitario, sino que la gestiona. Las recomendaciones obtenidas deben interpretarse como soluciones óptimas dentro de un marco de supuestos y condiciones específicas, lo que implica que su validez está condicionada por la estabilidad de los patrones observados. La toma de decisiones no puede ser estática. Aplicar el modelo de optimización de forma rígida equivale a asumir que el contexto no evoluciona, cuando en realidad lo hace constantemente.

En conclusión, se confirma que la optimización de la inversión publicitaria no puede abordarse desde enfoques parciales o simplificados, sino que exige una comprensión integrada del sistema en su conjunto. No se trata de proponer una solución única o universal, sino se crea un modelo como marco analítico flexible, capaz de adaptarse a distintos contextos, restricciones y prioridades. Esta aproximación contribuye a una gestión más rigurosa, transparente y alineada con los objetivos del negocio en la inversión en *Social Media Ads*.

8. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Claudia García García, estudiante de Business Analytics y Relaciones Internacionales de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Modelado de respuesta y optimización presupuestaria en campañas de Social Media Ads con un enfoque data-driven", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
6. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
7. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
8. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
9. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.

10. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
11. **Generador de datos sintéticos de prueba:** Para la creación de conjuntos de datos ficticios.
12. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
13. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
14. **Generador de encuestas:** Para diseñar cuestionarios preliminares.
15. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 22/04/2026

Firma: Claudia García García

9. Bibliografía

- Ayala, M. (s. f.). *Comprendiendo el rol del adstock en los modelos de marketing mix modeling (MMM)*. Bunker Help Center. <https://help.bunkerdb.com/es/articles/11906124-comprendiendo-el-rol-del-adstock-en-los-modelos-de-marketing-mix-modeling-mmm>
- Borden, N. H. (1964). *The concept of the marketing mix*. *Journal of Advertising Research*, 4(2), 2–7. https://www.guillaumenicaise.com/wp-content/uploads/2013/10/Borden-1984_The-concept-of-marketing-mix.pdf
- Caballero Fernández, R., Castrodeza Chamorro, M. C., & Gómez Núñez, T. (1988). *Políticas óptimas de producción y publicidad*. *Investigaciones Económicas*, 12(1), 87–108. https://www.academia.edu/34626689/POLITICAS_OPTIMAS_DE_PRODUCCION_Y_PUBLICIDAD
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*. <https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf>
- Corporate Finance Institute. (2024). *What-if analysis*. <https://corporatefinanceinstitute.com/resources/financial-modeling/what-if-analysis/>
- ESIC Business & Marketing School. (s. f.). *Social ads: Qué son, principales plataformas y ejemplos*. <https://www.esic.edu/rethink/marketing-y-comunicacion/social-ads-que-son-c>
- Fernández Marcial, V. (2015). *Marketing Mix de servicios de información: valor e importancia de la P de producto*. *Bibliotecas. Anales de Investigación*, 11, 64–78. <https://revistasbnjm.sld.cu/index.php/BAI/article/view/196/205>
- García, J., Rodríguez, M., & Pérez, L. (2024). *Predicción de lectura en instrumento patrón para una empresa de metrología a través de un dashboard empleando*

algoritmos de regresión. Revista de Investigación en Tecnologías de la Información (RITI), 12(27), 26–39. <https://riti.es/index.php/riti/article/view/304>

- Google Cloud. (s. f.). *What is machine learning?* <https://cloud.google.com/learn/what-is-machine-learning>
- GroupM. (2024). *This year next year: End of year 2024* [Documento en PDF]. Scribd. <https://es.scribd.com/document/879774348/GroupM-TYNY-EOY-Dec-2024>
- IAB Europe. (2025, mayo 21). *IAB Europe AdEx Benchmark 2024 report*. https://iab europe.eu/knowledge_hub/iab-europe-adex-benchmark-2024-report/
- IBM. (s. f.). *¿Qué es el aprendizaje supervisado?* <https://www.ibm.com/es-es/think/topics/supervised-learning>
- Investopedia. (s.f.). *Advertising elasticity of demand*. <https://www.investopedia.com/terms/a/advertising-elasticity-of-demand.asp>
- Kaufmann Argueta, J. (2014). *De la publicidad tradicional a la publicidad digital: Desafíos para agencias y profesionales* (Documento de investigación). Universidad de Navarra. https://d1wqtxts1xzle7.cloudfront.net/35319621/Publicidad_digital-desafios_para_agencias_y_profesionales-libre.pdf?1414565688=&response-content-disposition=inline%3B+filename%3DDe_la_publicidad_tradicional_a_la_public.pdf&Expires=1768328597&Signature=hQ0YeSzkJ19CieDOPdbS5BHeN3fAKrQcGiB12ka97lzQ9vigVt4W00g8LP4-ARZ-o84dh6qrrK11MoIBVmM9d6yHPOQt3PpyaMUL2R~yPCPKPWOjbjj8-r2qt9Sln4PXuoGDdAjHPg4dVHZ61aWHMJURxFa4ESa0VbbXTLHY~yJevWW-2zOcxLprSS~0ZEGH-aH-tv8Qpa35HY9surD1MAYGts7s5CGc8pE~-JmQe5J-Ic7LbO-3y1ZdL48gKwoEN10uC58RJ061x7v7OSboLJ42qTQBKtVyV7jBmWb76pwx61WF3pLZQF9~kvBFDNSJRnAu7~Vt5nypJKKEgBCRiQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- Lee, S. Y. (2010). *Exploring wearin and wearout in web advertising*. <https://files01.core.ac.uk/download/pdf/237003718.pdf>

- Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). *Prescriptive analytics: Literature review and research challenges*. *International Journal of Information Management*, 50, 57–70. <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>
- Lifesight. (s. f.). *Advertising elasticity*. <https://lifesight.io/glossary/advertising-elasticity/>
- Mantrala, M. K., Sinha, P., & Zoltners, A. A. (1992). *Impact of resource allocation rules on marketing investment-level decisions and profitability*. *Journal of Marketing Research*, 29(2), 162–175.
<https://seekscholar.com/sites/default/files/resource%20allocation.pdf>
- Martínez Rodríguez, E. (2005). *Errores frecuentes en la interpretación del coeficiente de determinación lineal*. *Anuario Jurídico y Económico Escurialense*, (38), 315–331. <https://dialnet.unirioja.es/servlet/articulo?codigo=1143023&>
- Martinich, J. S. (s.f.). *Linear programming* (Supplement B). University of Kentucky. <https://www.uky.edu/~dsianita/300/online/LP.pdf>
- MBA Skool Team. (s. f.). *Carryover effects: Definition & meaning*. MBA Skool. <https://www.mbaskool.com/business-concepts/marketing-and-strategy-terms/13881-carryover-effects.html>
- Prieto, E. (2023, agosto 29). *Diferencias entre los medios tradicionales y los nuevos medios*. Southern New Hampshire University. <https://es.snhu.edu/blog/diferencias-entre-los-medios-tradicionales-y-los-nuevos-medios>
- Roca Sales, M. (2014). *Percepciones de los profesionales en Estados Unidos: retos y oportunidades*
<https://telos.fundaciontelefonica.com/archivo/numero099/percepciones-de-los-profesionales-en-estados-unidos-retos-y-oportunidades/?output=pdf>

- Rothenberg, S. J. (2005). *Testing the dose–response specification in epidemiology*. *Environmental Health Perspectives*, 113(10), 1456–1461. <https://pmc.ncbi.nlm.nih.gov/articles/PMC1280400/>
- Sanz, I., et al. (2023). *Predicción de lectura en instrumento patrón para una empresa*. Dialnet. <https://dialnet.unirioja.es/servlet/articulo?codigo=10342385>
- Silva Vázquez, D. (2025). *Saturación publicitaria: actitud del público ante la repetición excesiva de anuncios en plataformas de vídeo y streaming*. (Trabajo de fin de grado). Universidad de Valladolid. <https://uvadoc.uva.es/bitstream/handle/10324/77581/TFG-N.%202683.pdf>
- Sixto García, J. (2015). *Desarrollo de las redes sociales como herramienta de marketing: Estado de la cuestión hasta 2015*. *Revista Científica General José María Córdova*, 13(15), 179–196. http://www.scielo.org.co/scielo.php?pid=S1692-25222015000100010&script=sci_arttext
- Taylor, J., Kennedy, R., & Sharp, B. (2009). *Is once really enough? Making generalizations about advertising’s convex sales response function*. *Journal of Advertising Research*, 49(2), 198–200. https://www.researchgate.net/publication/250168821_Is_Once_Really_Enough_Making_Generalizations_about_Advertising's_Convex_Sales_Response_Function
- WARC. (2024). *Global ad spend outlook 2024/25*. <https://page.warc.com/global-ad-spend-outlook-2024-25.html>
- Zhang, X., et al. (2021). *Optimal trade-offs in decision-making for sustainability and resilience in manufacturing supply chains*. *Journal of Cleaner Production*, 323, 129177. <https://www.sciencedirect.com/science/article/pii/S095965262101814X>

10. Anexos

Anexo 1. *Glosario de términos en inglés*

Adstock	Magnitud del impacto residual que un anuncio genera en el consumidor, reflejando la acumulación y la progresiva depreciación del efecto publicitario a lo largo del tiempo
Benchmarks	Indicadores de referencia para evaluar y comparar el desempeño de una variable
Carryover effect	Efecto persistente de la publicidad a lo largo del tiempo, mediante el cual la inversión realizada en un determinado periodo sigue influyendo en los resultados futuros
Clustering	Algoritmo de machine learning no supervisado que permite agrupar elementos similares entre si dentro de un conjunto de datos
Data-driven	Modelo de gestión basado en el uso de los datos para lograr un determinado objetivo, desplazando la dependencia exclusiva de la intuición o la experiencia
Dataset	Conjuntos de datos organizados de manera estructurada, que sirven como base para el análisis
Display	Formato de publicidad digital para difundir mensajes publicitarios visuales en entornos digitales con el fin de aumentar la visibilidad de la marca.
Engagement	Nivel de compromiso emocional e interacción activa del usuario con contenido, anuncio o campaña digital
Funnel	Abarca el conjunto de etapas que recorre el usuario desde las primeras fases de conocimiento hasta la conversión final
KPI	Entendido como indicador clave de rendimiento, es decir, una métrica cuantitativa empelada para medir y evaluar el rendimiento de una campaña o estrategia
Marketing Mix Modeling	Técnica de modelado estadístico que estima el efecto de distintas variables de marketing sobre resultados del negocio
Scripts	Fragmento de código que se ejecuta automáticamente para realizar una tarea específica
Social Media Ads	Contenido publicitario de pago distribuido a través de plataformas digitales orientados a aumentar la visibilidad, la interacción o las conversiones
Sweet Spot	Punto óptimo en el que se obtiene el mejor equilibrio entre variables
Trade-off	Situación en la que mejorar una variable puede suponer renunciar parcialmente a otra

Fuente: elaboración propia

Anexo 2. *Modelos clásicos de respuesta y asignación presupuestaria*

En el apartado de optimización de presupuesto y técnicas robustas se ha desarrollado el modelo *Vidale-Wolfe model*, como referencia principal para el análisis de la relación entre inversión publicitaria y ventas. Con el objetivo de complementar este marco teórico, a continuación, se presentan otros modelos clásicos de la literatura que han abordado el problema de la optimización desde diferentes enfoques.

En primer lugar, el modelo de *Nerlove-Arrow* intenta explicar que la publicidad no actúa solo en el momento en el que se invierte, sino que deja un efecto acumulativo en el tiempo. Sirve para explicar el efecto dinámico de la publicidad: cómo se construye, cómo se mantiene y cómo se deteriora. Se trata de un modelo útil para entender la persistencia del efecto publicitario, pero sigue siendo bastante agregado y simplificado. No distingue bien entre canales, audiencias o comportamientos más complejos.

En segundo lugar, el modelo o condición de *Dorman-Steiner* es un criterio de optimización que indica el nivel óptimo de gasto de publicidad de una empresa para maximizar sus beneficios. No intenta explicar solo si la publicidad aumenta la demanda, sino qué proporción de los ingresos conviene destinar a la publicidad. Es un modelo útil como referencia teórica, pero simplifica bastante la realidad. Asume relaciones estables y no recoge bien cuestiones como saturación, dinámicas entre canales o incertidumbre.

La diferencia principal de estos modelos radica en la dimensión del fenómeno publicitario que cada uno prioriza. El modelo de *Vidale-Wolfe model* explica cómo la inversión se traduce en ventas considerando efectos de saturación, el modelo de *Nerlove-Arrow model* introduce una dimensión temporal al recoger el carácter acumulativo y decreciente del impacto publicitario. Por su parte, el modelo de *Dorfman-Steiner condition* se centra en determinar el nivel óptimo de inversión desde una perspectiva económica.

Anexo 3. Interpretación gráficos boxplot

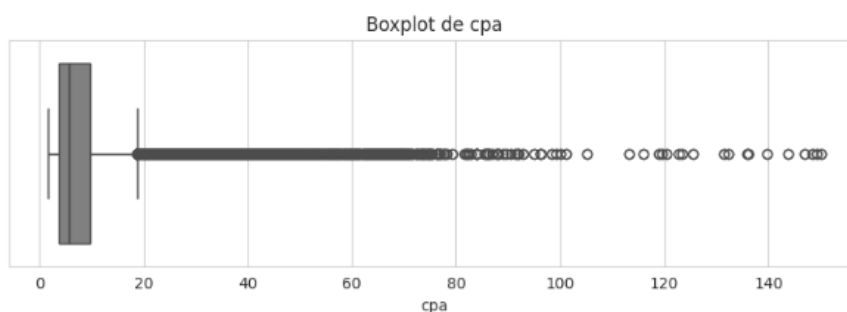
El diagrama de caja de bigotes o boxplot es una herramienta gráfica utilizada para observar la distribución de los datos de una variable. Con este gráfico se pueden observar tres cosas principalmente:

- La tendencia central o mediana, que indica el valor que queda en medio cuando los datos se ordenan de menor a mayor. En el gráfico, aparece como una línea dentro de la caja.
- La dispersión de los datos se observa en el tamaño de las cajas. La caja representa el rango donde se encuentra la mitad central de los valores, es decir, los valores que están entre el primer y tercer cuartil.
- Los valores atípicos, aparecen como puntos individuales fuera de los bigotes del diagrama e indican datos que se alejan mucho del resto.

Durante el trabajo de fin de grado, este tipo de gráfico se emplea para facilitar la comparación entre variables o grupos de estudio, proporcionando una visión sintética de las distribuciones.

Anexo 4. Visualización de valores atípicos

Se observan valores atípicos representados como puntos a la derecha del gráfico. Esto indica que se alejan de forma notable del comportamiento general del conjunto de datos. En el caso del CPA, los valores atípicos representan campañas con costes de adquisición significativamente superiores a los habituales.



Fuente: elaboración propia basada en el contenido de Social Media Advertising Dataset

Anexo 5. Código empleado para el trabajo

El código empleado para este trabajo, así como la base de datos se pueden encontrar en el siguiente repositorio de Github o en la carpeta de Google Drive. El *script* contiene esquemas y todo el código comentado.

- Acceso a carpeta de Google Drive con base de datos y script completo: https://drive.google.com/drive/folders/1yRTuAF_PUWu5R9RITKXcNHxl-tRXkJmO?usp=drive_link
- Acceso al script en Github: <https://github.com/2003cgg-cell/Trabajo-Fin-De-Grado---Business-Analytics-Claudia-Garc-a-Garc-a>
- Link al dataset en kaggle: <https://www.kaggle.com/datasets/jsonk11/social-media-advertising-dataset>