



Facultad de Ciencias Económicas y Empresariales

MARKETING DIGITAL

Optimización de campañas mediante analítica avanzada

Autor: Paula García González

Director: Antonio Tena Blázquez

Agradecimientos

A mi abuelo José, que aunque no me va a poder acompañar el día de mi graduación, me ha escuchado y ayudado a conseguir todo lo que quería estos últimos años.

Resumen:

En una era profundamente marcada por la digitalización y el avance de las nuevas tecnologías, hemos sido testigos de cómo el marketing tradicional ha ido evolucionando hacia el marketing digital. Este enfoque permite monitorizar con gran precisión métricas y KPIs, facilitando la evaluación del rendimiento de las campañas en tiempo real y una optimización más eficiente de los recursos invertidos. El presente trabajo se estructura en dos partes. En primer lugar, se desarrollará un marco teórico en el que se analizará la evolución del marketing digital, así como los conceptos clave, las métricas y las herramientas más relevantes. En segundo lugar, se presentará un caso práctico en el que, a partir de una base de datos, se construirá un modelo de regresión con el objetivo de establecer la probabilidad de conversión de un usuario e identificar las variables que puedan resultar más significativas en dicha probabilidad.

Palabras clave: *marketing digital, conversión, métricas, KPIs, regresión, campañas publicitarias, marketing funnel*

Abstract:

In an era deeply shaped by digitalization and the advancement of new technologies, we have witnessed how traditional marketing has progressively evolved into digital marketing. This approach enables precise monitoring of metrics and KPIs, facilitating the real-time evaluation of campaign performance and a more efficient optimization of invested resources. This paper will be structured in two parts. First, a theoretical framework is developed in which the evolution of digital marketing is analyzed, along with some key concepts, metrics, and the most relevant tools. Second, a practical case is presented in which, based on a dataset, a regression model will be built with the aim of determining the probability of user conversion and identifying the variables that may be most significant in that conversion.

Keywords: *digital marketing, conversion, metrics, KPIs, regression, advertising campaigns, marketing funnel*

Índice de contenido

1.	Introducción y contextualización.....	6
1.1.	Contexto relevancia y surgimiento del marketing digital	6
1.2.	Metodología	7
1.3.	Objetivo y motivación del trabajo.....	8
2.	Marco teórico	8
2.1.	Evolución del marketing digital	8
2.2.	Marketing Funnel	11
2.3.	Full Funnel:	13
2.5.	Métricas y KPIs en marketing digital	16
3.	Caso Práctico:.....	20
3.1.	Lectura y limpieza de los datos	20
3.2.	Análisis exploratorio por segmentos	26
3.2.1.	Aspectos sociodemográficos.....	27
3.2.2.	Comportamiento del consumidor.....	29
3.2.3.	Características de la campaña	37
3.3.	Modelo explicativo:	40
3.4.	Modelo predictivo	45
3.5.	Conclusiones y recomendaciones según los resultados obtenidos	54
4.	Reflexión crítica:	55
	Declaración uso responsable de herramientas de IA	59
	Bibliografía:	61
	Anexo.....	64
	Anexo 01. Archivo código caso práctico	64
	Anexo 02: Implementación de One-Hot-Encoding para la conversión de variables categóricas	64
	Anexo 03: Eliminación de variables	64
	Anexo 04: Unificación formato numérico	64
	Anexo 05: estandarización de variables	64
	Anexo 06: Valores nulos	65

Índice de ilustraciones

Ilustración 1: Empresas que tienen una página web, 2023 (% de empresas)	6
Ilustración 2: Empresas que adoptaron aplicaciones de negocio electrónico (e-business) (% de empresas)	6
Ilustración 3: Lectura del dataset	21
Ilustración 4: Análisis descriptivo del conjunto de datos y variable objetivo.....	22
Ilustración 5: Ajuste del dataset para equilibrar las clases.....	22
Ilustración 6: Tipo de variables.....	26
Ilustración 7: Tasa de Conversión Promedio por Género	27
Ilustración 8: Tasa de conversión promedio por edad.....	28
Ilustración 9: tasa de conversión promedio por rango salarial	28
Ilustración 10: tasa de conversión promedio por CTR	30
Ilustración 11: Tasa de conversión promedio por Conversion Rate	30
Ilustración 12: Tasa de conversión promedio por website visits	31
Ilustración 13: tasa de conversión por medio por páginas por visita.....	31
Ilustración 14: tasa de conversión promedio por time on site	32
Ilustración 15: tasa de conversión promedio por social shares.....	32
Ilustración 16: tasa de conversión promedio por email opens	33
Ilustración 17: tasa de conversión promedio por email clicks	33
Ilustración 18: Tasa de Conversión Promedio por Previous Purchases	34
Ilustración 19: Tasa de conversión promedio por loyalty points	34
Ilustración 20: tasa de conversión promedio por campaign type.....	37
Ilustración 21: tasa de conversión promedio por campaign channel	38
Ilustración 22: tasa de conversión promedio por ad spend	38
Ilustración 23: Resultados Matriz de correlación	42
Ilustración 24: Curva Característica Operativa del Receptor (ROC).....	50
Ilustración 25: Importancia de las variables en el modelo predictivo Random Forest.....	51
Ilustración 26: Resultados matriz de confusión Random Forest	53
Ilustración 27: Implementación de One-Hot-Encoding para la conversión de variables categóricas.....	64
Ilustración 28: Eliminación de variables.....	64
Ilustración 29: Unificación formato numérico.....	64
Ilustración 30: estandarización de variables	64
Ilustración 31: Búsqueda de valores nulos en las variables	65

Índice de tablas de contenido

Tabla 1: KPIs por etapa del Marketing Funnel.....	20
Tabla 2: Resultados modelo explicativo de clasificación	40
Tabla 3: Resultados modelo explicativo de clasificación II.....	43
Tabla 4: Resultados modelo predictivo Random Forest	48

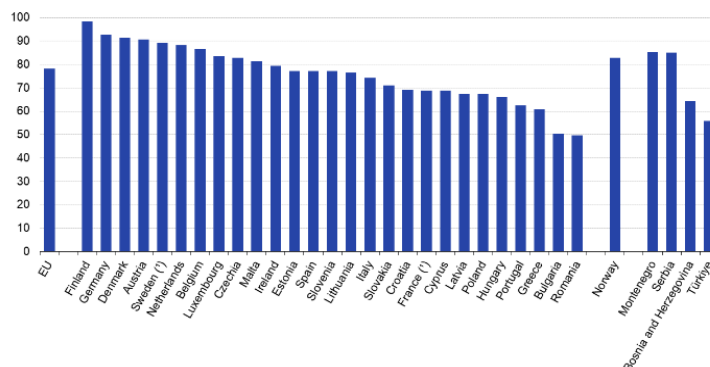
1. Introducción y contextualización

1.1. Contexto relevancia y surgimiento del marketing digital

Nos encontramos en una etapa profundamente marcada por la expansión del comercio electrónico y la digitalización. Esto ha sido impulsado por el desarrollo de nuevas tecnologías y el auge de nuevas formas y medios de comunicación como las redes sociales, páginas web o motores de búsqueda. Todos estos avances han generado que las empresas hayan cambiado de manera radical la forma en la que se relacionan con sus clientes.

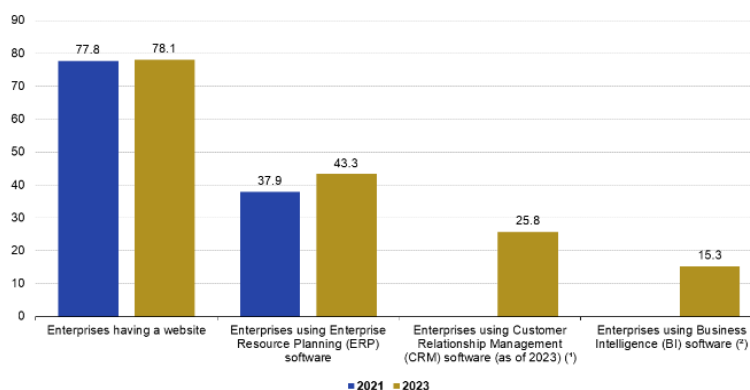
Según las estadísticas del Eurostat, en 2023, prácticamente el 80% de las empresas de la Unión Europea contaron con su propio sitio web. Según datos de la misma fuente, ese mismo año más del 25 % de las empresas europeas declararon utilizar algún tipo de software de gestión de relaciones con clientes (CRM). En el caso de España, el porcentaje se situó por encima de la media europea, con casi un tercio de las empresas afirmando emplear este tipo de herramientas.

Ilustración 1: Empresas que tienen una página web, 2023 (% de empresas)



Fuente: EUROSTAT 2024

Ilustración 2: Empresas que adoptaron aplicaciones de negocio electrónico (e-business) (% de empresas)



Fuente: EUROSTAT 2024

Según Philip Kotler, Hermawan Katajaya e Iwan Setiawan (2017), el marketing digital emerge como una evolución natural del marketing tradicional, utilizando canales y herramientas digitales para conectar con audiencias de manera más directa, personalizada y medible. Aunque el objetivo de ambas disciplinas, tanto marketing tradicional como marketing digital sea el mismo, es decir, guiar a los consumidores hacia la compra o adquisición de un bien o servicio específico, el marketing digital se diferencia por el uso de medios y tecnologías digitales que permiten medir y optimizar con mayor precisión las estrategias empleadas.

Es muy importante tener en cuenta que, el marketing digital, no solo facilita la promoción y publicidad de productos y servicios en entornos digitales, sino que también ofrece la posibilidad de segmentar a la audiencia objetivo y analizar en tiempo real el comportamiento de los usuarios, de esta forma, siendo capaces de optimizar al máximo las estrategias comerciales y mejorando profundamente la eficacia de las campañas.

Tener la capacidad de medir, analizar y predecir el rendimiento de las acciones de marketing constituye una ventaja competitiva clave, ya que permite a las empresas y organizaciones basar sus decisiones en datos reales, para poder adaptar sus estrategias de forma precisa, anticiparse a las necesidades de los consumidores y responder de manera eficaz en un entorno cada vez más competitivo y dinámico.

1.2. Metodología

Este trabajo tiene como objetivo comprender los conceptos, estrategias y herramientas clave que se emplean en el entorno del marketing digital hoy en día, así como analizar su impacto en la relación entre las empresas y los consumidores en la actualidad.

Para ello, el análisis se ha estructurado en distintos apartados. En primer lugar, se llevará a cabo una introducción y contextualización del tema, con el fin de situarnos en el marco del marketing digital y comprender su relevancia en el entorno empresarial actual. A continuación, se desarrollará un marco teórico, en el que se abordarán distintos puntos clave como: la evolución del marketing tradicional al marketing digital, la diferenciación entre los tipos de medios y formatos publicitarios, los enfoques del *marketing funnel*, o el *full finnel* y la definición de algunas de las métricas y KPIs más utilizados en la actualidad, sentando así las bases conceptuales necesarias y consiguiendo comprender los cambios significativos que han experimentado las estrategias de marketing los últimos años.

La tercera parte del trabajo se centrará en un análisis práctico sobre una base de datos. Mediante un dos modelos, uno explicativo y otro predictivo ambos de clasificación, utilizando el lenguaje

de programación de Python, se intentará establecer la probabilidad de conversión de un usuario, así como identificar las variables más significativas, es decir, aquellas que tienen mayor influencia sobre dicha conversión, además de la posible correlación entre ellas. En este apartado, el primer paso fue la búsqueda del conjunto de datos (*dataset*), seguido de un proceso de limpieza y preparación del mismo asegurando así tanto la calidad como la coherencia y consistencia del análisis. Posteriormente se llevará a cabo un análisis exploratorio de las variables, para, finalmente, desarrollar el modelo predictivo y el estudio y análisis de este.

Por último, se interpretarán y contrastarán los resultados obtenidos en el análisis práctico con el marco teórico, y se expondrán las principales conclusiones y los hallazgos más relevantes derivados del estudio.

1.3. Objetivo y motivación del trabajo

El objetivo principal de este trabajo es comprender la importancia real del marketing digital y analizar cómo se ha convertido en un eje clave en las estrategias comerciales de las empresas. Se pretende identificar qué variables de una campaña son capaces de influir de manera significativa en la probabilidad de conversión de un consumidor, entendiendo cómo distintos factores de interacción y comportamiento afectan la efectividad de las campañas digitales.

La motivación de este trabajo surge del entorno empresarial altamente dinámico y competitivo en el que nos encontramos. Las preferencias y necesidades de los consumidores cambian de forma constantemente, y es imprescindible que las empresas sean capaces de adaptarse rápidamente a este escenario para mantener su competitividad. Comprender qué factores impulsan la conversión permite a las organizaciones diseñar estrategias más efectivas, optimizar recursos y mejorar la experiencia del cliente.

Considero que el marketing es uno de los pilares fundamentales de cualquier organización, ya que no solo impulsa las ventas, sino que también promueve el reconocimiento, la fidelización y la confianza de los consumidores.

2. Marco teórico

2.1. Evolución del marketing digital

Nos encontramos en una era profundamente digital, caracterizada tanto por el desarrollo de las nuevas tecnologías, como por la expansión y consolidación de medios de comunicación digitales y comercio electrónico. Todo esto ha generado un entorno empresarial profundamente dinámico y competitivo, en el que las empresas se han visto obligadas a desarrollar una

capacidad de adaptación prácticamente inmediata a las cambiantes demandas y exigencias de los consumidores. Por ello, durante los últimos años, las organizaciones han cambiado de forma radical la manera en la que se comunican e interactúan con su público objetivo.

El marketing digital se ha consolidado como un pilar fundamental dentro de las estrategias empresariales de las compañías. Ya no basta con únicamente conocer y complacer las necesidades del mercado, sino que es imprescindible comprender cómo los consumidores expresan dichas necesidades, qué canales utilizan y qué tendencias siguen a lo largo de sus procesos de consumo. En las últimas décadas, los hábitos de compra han cambiado sustancialmente. A principios del siglo XXI, la mayoría de la población consumía información y publicidad a través de medios tradicionales como la radio, televisión y prensa. Sin embargo, acompañado del nacimiento de plataformas como Facebook, YouTube, Twitter, Instagram, TikTok u otros entornos digitales durante los últimos años, el panorama del marketing y la comunicación se ha transformado drásticamente. Hoy en día resulta impensable concebir el mundo empresarial fuera del entorno digital.

Según el Kemp (2024), prácticamente dos tercios de la población mundial ya tienen acceso a Internet. Por otro lado, más de 5.000 millones de personas usan redes sociales, lo que representa más del 60% de la población global. Este fenómeno obliga a las empresas a integrar nuevos canales digitales en sus estrategias comunicativas para conectar de manera eficaz con su audiencia.

El marketing digital no surgió como tal, sino que fue el resultado de la evolución y adaptación del marketing tradicional al nuevo entorno digital y electrónico. Aunque resulta complicado establecer una fecha exacta para el nacimiento del marketing digital, uno de los primeros hitos de esta disciplina, fue la campaña de email marketing realizada en 1987 por *Digital Equipment Corporation (DEC)*. La campaña consistió en el envío de 400 correos electrónicos a través de *ARPANET*¹¹ (*Advanced Research Project Network*), y fue capaz de generar en torno a 13 millones de dólares en ventas, demostrando el potencial comercial del canal. Desde entonces, poco a poco las campañas de email marketing se han ido consolidando como una de las herramientas más eficaces de comunicación digital. Según el Informe de *Email Marketing SMS en España 2025*, cerca del 70% de las empresas españolas que utilizan este canal afirman estar satisfechas con los resultados obtenidos, y el 85% de ellas prevé aumentar o mantener su

¹ Primera red de computadoras descentralizada, creada en 1969 por el Departamento de Defensa de EE. UU.

inversión hacia este tipo de campañas. Joanna Milliken, vicepresidenta senior de *Salesforce Marketing Cloud*, señala que, el email marketing es una de las formas de campaña de marketing digital que no solo ha sido capaz de mantenerse, sino que ha sido también capaz de adaptarse.

Otro hito relevante en la historia del marketing digital fue la aparición del primer banner publicitario publicado por la revista *HotWired* en 1994. El anuncio contenía un único mensaje: “*Have you ever clicked your mouse right here? You will*”. El banner consiguió un CTR² del 44%, lo que significó que casi la mitad de la gente que lo vio hizo clic en él. No se sabe si este éxito fue a raíz de la novedad absoluta del formato o curiosidad pura, pero, lo que sí está claro es que consiguió evidenciar el fuerte potencial de los formatos publicitarios en internet.

Paralelamente, el posicionamiento en motores de búsqueda, a través de estrategias SEO (*Search Engine Optimization*) y SEM (*Search Engine Marketing*), comenzó a consolidarse como otra de las herramientas esenciales para aumentar la visibilidad online y atraer tráfico cualificado. Desde entonces, la publicidad digital ha evolucionado hacia múltiples formatos y estrategias que buscan maximizar el retorno de la inversión mediante la segmentación y el análisis de datos. A diferencia de los medios tradicionales, la publicidad digital permitía medir con precisión el comportamiento del usuario y optimizar las campañas en tiempo real.

Con el auge de las redes sociales, las empresas comenzaron no solo a abrirse perfiles corporativos, sino también a incorporar estrategias de *influencers* mediante la colaboración con diferentes perfiles con el objetivo de promocionar su producto y llegar a nuevas audiencias. Como ya hemos mencionado anteriormente, más del 60% de la población mundial utiliza redes sociales, con un tiempo medio diario de uso de 2,3 horas. Según *Social Media Marketing Statistics 2025*, dentro de este 60%:

- El 82% de los usuarios afirma que las redes sociales influyen en sus decisiones de compra.
- El 74% busca información sobre marcas en redes sociales antes de adquirir un producto.
- Cerca del 60% descubre nuevas marcas a través de estas plataformas.
- El 92% de las empresas utiliza redes sociales con fines de marketing.
- El 78% de los profesionales del marketing afirma que estas han incrementado el reconocimiento de marca.

² Click-Through Rate (CTR): porcentaje de los usuarios que han visto el anuncio han hecho clic en él

- El 69% de los profesionales considera que el uso de redes sociales por parte de las marcas contribuye directamente a la generación de leads.

Todos estos datos reflejan que las redes sociales funcionan como un canal de comunicación, y juegan un papel clave a la hora de influir en la decisión de compra de los usuarios.

Por último, en los últimos años hemos sido testigos de la incorporación progresiva de nuevas tecnologías como la realidad aumentada, la realidad virtual y la inteligencia artificial en nuestra vida cotidiana. Estas innovaciones han abierto un amplio abanico de posibilidades para ofrecer experiencias más inmersivas, interactivas y personalizadas al cliente, transformando la manera en que las marcas se relacionan con su público.

Un ejemplo destacado de la aplicación de estas herramientas es la reciente implementación de Zara en su página web y en su *app*, que permite a los usuarios visualizar cómo les quedarían determinadas prendas sin necesidad de probárselas físicamente ni acudir a la tienda. Este tipo de soluciones no solo mejoran la experiencia de compra, sino que también reducen barreras en el proceso de decisión y refuerzan la integración entre el entorno digital y el físico.

En conclusión, podemos afirmar que, una de las principales ventajas del marketing digital frente al marketing tradicional es su elevada capacidad de medición y monitorización. A diferencia de los medios convencionales, el entorno digital permite analizar en tiempo real el rendimiento de las campañas, optimizar estrategias sobre la marcha y asignar los recursos de manera más eficiente. En este contexto, el uso de KPIs (*Key Performance Indicators*) resulta imprescindible para evaluar el desempeño de las acciones digitales y garantizar que los objetivos planteados se estén cumpliendo de forma efectiva.

2.2. Marketing Funnel

El embudo de conversión, conocido también como *Marketing Funnel*, es el proceso por el cual un cliente transita desde el momento en que conoce una marca hasta que realiza una compra. Este concepto, formulado por Elias St. Elmo Lewis a finales del siglo XIX, tiene como objetivo principal aportar orden y estructura a la incertidumbre inherente al proceso de compra que experimentan los consumidores.

La teoría del *Marketing Funnel* o *Customer Conversion Funnel* ayuda a las empresas a alinear sus esfuerzos y contenidos con el comportamiento del cliente. La estrategia consiste en ofrecer

el mensaje adecuado, en el momento oportuno, a la persona correcta. Este embudo se divide en cinco etapas: conciencia (*awareness*), consideración (*consideration*), conversión (*conversion*), fidelización (*loyalty*) y recomendación (*advocacy*).

- i. Conciencia (*Awareness*):** los clientes toman conciencia de una necesidad o problema que desean resolver. El objetivo de la marca es lograr que el cliente se familiarice con la empresa. Para ello, se utilizan herramientas como publicidad *display*, menciones de influencers, contenido viral en redes sociales y estrategias de posicionamiento SEO. Los formatos más habituales incluyen videos cortos, anuncios digitales y publicaciones informativas. En esta fase, la empresa se dirige a audiencias amplias, buscando captar la atención del mayor número posible de personas.
- ii. Consideración (*Consideration*):** los usuarios ya son conscientes de su necesidad y empiezan a investigar distintas opciones, comparando productos o servicios que puedan ofrecer una solución. La marca busca generar interés y construir confianza. Para esto, se emplean herramientas como campañas de *retargeting*, secuencias de emails y guías comparativas. Los contenidos más comunes son casos de estudio, *webinars* y series de correos informativos. El objetivo es educar al consumidor, dirigir tráfico cualificado y fortalecer la percepción de marca entre quienes ya muestran cierto interés.
- iii. Conversión (*Conversion*):** el usuario toma finalmente la decisión de compra convirtiéndose en cliente. El objetivo de la empresa es motivar al usuario mediante estímulos adecuados para que realice la adquisición. Las herramientas más usadas incluyen ofertas por tiempo limitado, *landing pages* optimizadas y pruebas gratuitas. Los contenidos más efectivos son videos demostrativos, cupones de descuento, ofertas temporales y páginas de aterrizaje optimizadas. Se busca maximizar las ventas, mejorar la tasa de conversión y optimizar el retorno, enfocándose exclusivamente en usuarios que están activamente considerando la compra.
- iv. Fidelización (*Loyalty*):** la relación con el cliente no termina con la compra. En esta etapa, la marca trabaja para mantener y fortalecer la relación, fomentando que el cliente siga comprando y se convierta en usuario habitual. Para ello, se emplean programas de fidelización, newsletters y experiencias personalizadas a través de *apps* u otros canales digitales. El objetivo es aumentar la lealtad del cliente y prolongar su vínculo con la marca.
- v. Recomendación (*Advocacy*):** finalmente, se busca transformar a los clientes satisfechos en embajadores de la marca, que recomienden el producto o servicio a otros consumidores. Para ello, se utilizan programas de *referrals*, campañas con contenido generado por

usuarios (UGC, por sus siglas en inglés) y estrategias de creación de comunidades. Los contenidos más comunes incluyen testimonios, reseñas y herramientas que facilitan la difusión en redes sociales, aprovechando el principio de *Social Proof*, donde la recomendación y validación de otros usuarios generan confianza y credibilidad. El objetivo es ampliar el alcance de la marca a través de la credibilidad y confianza que generan las recomendaciones orgánicas.

Es importante destacar que, en el entorno digital, el proceso de compra no sigue un recorrido estrictamente lineal. Los consumidores pueden avanzar y retroceder entre las distintas etapas antes de tomar una decisión definitiva. Por esta razón, la teoría del *funnel* ha evolucionado hacia el modelo de *Flywheel*, el cual le presta especial importancia a la retención y satisfacción del cliente. Esta nueva teoría entiende cada interacción con el cliente como una oportunidad para impulsar el crecimiento de la empresa en la que los clientes satisfechos pasan a convertirse en promotores activos de la marca. El enfoque del *Flywheel* se apoya en tres componentes principales que son: atraer, interactuar y deleitar. Estos tres pilares trabajan de forma conjunta para mantener tanto el movimiento como el crecimiento del negocio.

2.3. Full Funnel:

A partir del modelo del *marketing funnel* y de su evolución hacia enfoques más dinámicos como el *flywheel*, surge el concepto de *Full Funnel Marketing*, que responde a la necesidad de abordar de forma integral el comportamiento y las necesidades del consumidor en cada etapa de su recorrido.

Mientras que el *funnel* tradicional permite entender de una forma conceptual el camino que recorre un cliente desde el descubrimiento de la marca hasta la decisión final de compra del producto o servicio, el enfoque *full funnel* se centra en cómo las empresas deben actuar en cada una de esas fases. Es decir, no solo analiza el recorrido del usuario, sino que también define estrategias específicas para influir y acompañar al consumidor de forma continua, teniendo en cuenta la conexión entre todas las etapas.

La estrategia de Full Funnel combina distintas tácticas de marketing que permiten guiar a los usuarios a lo largo del embudo de ventas manteniendo su interés y mejorando de forma progresiva la tasa de conversión.

En la parte superior del embudo, también conocido por sus siglas en inglés como *TOFU*, el objetivo principal es generar visibilidad y reconocimiento de marca. En esta fase se utilizan estrategias de redes sociales y SEO, con el fin de atraer a una audiencia amplia y despertar el interés inicial del consumidor.

En la fase intermedia, *MOFU*, el objetivo es conseguir construir relaciones más sólidas mediante acciones como *email marketing*, *retargeting*, *webinars* o guías, que ayuden a los potenciales clientes a evaluar diferentes opciones y considerar la propuesta de valor de la empresa.

Por último, en la parte inferior del *funnel*, *BOFU*, el objetivo de la empresa es conseguir la conversión del cliente. Por ello, en esta fase, se aplican tácticas como ofertas personalizadas, testimonios de clientes o demostraciones del producto con el fin de resolver posibles dudas que hayan podido surgir y facilitar la decisión de compra.

Este enfoque aporta numerosos beneficios a las empresas. En primer lugar, aumenta el reconocimiento de la marca, alcanzando audiencias mucho más amplias, ya que se llevan a cabo estrategias de visibilidad mucho mejor estructuradas. En segundo lugar, mejora la nutrición de leads, al mantener una comunicación constante y relevante, fortaleciendo así la confianza del usuario en la empresa.

Además, contribuye también a incrementar la tasa de conversión, ya que se acompaña al cliente con los mensajes adecuados en cada etapa del proceso de compra, aumentando así de forma significativa la probabilidad de conversión de este. También es capaz de proporcionar mejores *insights* sobre el comportamiento de los consumidores, lo que permite no solo personalizar mejor las campañas sino también optimizar las diferentes estrategias.

Por último, las estrategias de *full funnel* favorecen también a una mayor fidelización de clientes, ya que el acompañamiento no termina con la decisión final de compra, sino que continúa después, fomentando la repetición de conversión y la lealtad de los consumidores con la marca. Finalmente, ayuda a mejorar la alineación entre los equipos de marketing y ventas, al trabajar con objetivos comunes y procesos unificados.

En definitiva, una estrategia de *full funnel marketing* permite a las empresas acompañar al cliente de forma completa durante todo el proceso de compra. Esto no solo mejora la conversión

y el retorno de la inversión (ROI), sino que también fortalece las relaciones con los clientes y contribuye a un crecimiento de las empresas más sostenible a largo plazo.

2.4. Paid Media, Owned Media y Earned Media

En el entorno digital actual, la estrategia comunicativa de una empresa se articula en torno a tres tipos de formatos publicitarios: *Paid Media*, *Owned Media* y *Earned Media*. Este modelo se conoce con el nombre de POEM o POSE.

- a) **Medios pagados (*Paid Media*):** engloba todos aquellos anuncios publicitarios que impliquen algún tipo de pago, ya sea monetario o de otro tipo, por parte de la marca en contraprestación de una campaña publicitaria. Se caracteriza por ofrecer resultados relativamente rápidos, posibilidad de escalamiento según el presupuesto disponible y un alto nivel de control de la marca sobre el mensaje. Dentro de *Paid Media* se distinguen dos grandes categorías: *paid search* y publicidad *display*. *Paid search* son todos aquellos anuncios que aparecen cuando los usuarios están buscando activamente un producto, servicio o marca determinada. Por otro lado, la publicidad *display* consiste en aquellos anuncios dirigidos a aquellos usuarios que no están buscando como tal un producto, servicio o marca concreta, pero sí podrían estar interesados, aunque no los estén buscando de forma activa. La publicidad pagada en YouTube móvil tiene un 84 % más de probabilidades de captar la atención de los usuarios en comparación con la publicidad tradicional en televisión (Google, 2016), lo que refuerza la relevancia creciente de los formatos digitales.
- b) **Medios propios (*Owned Media*):** son todos aquellos contenidos publicitarios creados y gestionados directamente desde la marca. Incluye todas las publicaciones en redes sociales desde el perfil corporativo, la página web, blogs, correos electrónicos o cualquier otro canal propio de la empresa. El principal objetivo es fortalecer el reconocimiento de la marca a la vez que se construye una relación directa con la audiencia. Los medios propios se caracterizan por disponer de un alto control de la marca, y no depender de terceros para su difusión. Sin embargo, requiere una inversión constante en el tiempo, recursos humanos y producción de contenido de calidad.
- c) **Medios ganados (*Earned Media*):** hace referencia a la publicidad que recibe una marca de manera orgánica, es decir, aquella por la que no paga, pero sobre la que tampoco tiene control directo sobre el mensaje. Incluye recomendaciones entre consumidores, reseñas, valoraciones, menciones en redes sociales realizadas por usuarios sin contraprestación, así

como cobertura mediática espontánea. Aunque a menudo se considera “publicidad gratuita”, en realidad es el resultado de una experiencia positiva del cliente o de una estrategia de marca sólida. El *Earned Media* destaca por su alta credibilidad y su capacidad de amplificarse rápidamente, aunque implica un bajo nivel de control por parte de la empresa.

Es importante destacar que para conseguir que una campaña de marketing tenga éxito, es necesario combinar los tres tipos de medios, pagados, propios y ganados, para maximizar el alcance, reforzar la credibilidad y generar resultados sostenibles a largo plazo.

2.5. Métricas y KPIs en marketing digital

En el ámbito del marketing digital, es fundamental medir el progreso y la evolución de nuestras campañas. Es exactamente en este punto cuando entran en juego las métricas e indicadores clave de desempeño, mejor conocidos como KPIs por sus siglas en inglés (*Key Performance Indicators*). Es importante destacar, que una métrica no es exactamente lo mismo que un KPI, aunque ambos conceptos están profundamente relacionados. Por un lado, las métricas, son mediciones de cualquier tipo de actividad, sin embargo, los KPIs son un subconjunto de métricas directamente relacionadas con los objetivos estratégicos de la empresa u organización. Mientras que las métricas son datos brutos que registran eventos o resultados, los indicadores clave de desempeño permiten evaluar el progreso que lleva una compañía hacia sus objetivos. A diferencia de las métricas, los KPIs son específicos, medibles, alcanzables, relevantes y con plazos definidos (SMART). Mientras que una métrica puede ser aumentar el número de visualizaciones, un KPI sería aumentar el número de visualizaciones en una campaña específica en un porcentaje establecido durante un periodo predeterminado.

En marketing digital, existe una amplia variedad tanto de métricas como de KPIs que ayudan a las empresas a poder analizar y optimizar las acciones realizadas. Entre las más relevantes podemos destacar:

- **Número de impresiones (*Impressions*):** representa el número de veces que se muestra un contenido o anuncio, sin tener en cuenta si un mismo usuario lo ve más de una vez. Mide la exposición y el volumen total de visibilidad que una campaña está generando. Una impresión se contabiliza cada vez que el anuncio aparece en pantalla, independientemente de que el usuario lleve a cabo una interacción o no.

- **Número de visualizaciones (*Views*):** refleja el número de veces que un contenido audiovisual ha sido visto o consumido por los usuarios. A diferencia de las impresiones, normalmente exige que se cumpla un tiempo mínimo de reproducción para que la visualización sea válida, lo que permite medir un nivel de atención más real. Este KPI es especialmente relevante en campañas de vídeo, ya que indica consumo efectivo del contenido.
- **Alcance (*Reach*):** mide el número de usuarios únicos que han visto un contenido o anuncio durante un periodo determinado. No contabiliza múltiples exposiciones de un mismo usuario, al contrario que el número de impresiones. Representa el número de personas que han sido impactadas por la campaña, siendo un indicador clave para evaluar la amplitud de difusión.
- **Engagement (*Engagement*):** indica el número total de interacciones que tiene un anuncio o campaña. Incluye acciones como comentarios, me gusta, guardados, compartidos, clics u otras formas de interacción. Puede analizarse tanto en volumen absoluto como en tasa respecto al alcance o a las impresiones y refleja el grado de interés, conexión y relevancia que ha generado la campaña en la audiencia.
- **Número de visitas (*Website Visits*):** representa la cantidad total de veces que los usuarios han accedido a un sitio web en un tiempo determinado, sin tener en cuenta que un mismo usuario haya podido entrar varias veces. Indica la cantidad de tráfico que recibe la página web.
- **Páginas vistas (*Page Views*):** refleja el número total de páginas que han sido visualizadas dentro de un sitio web. Contabiliza cada página individual que se consulta durante una sesión. Permite analizar el nivel de navegación interna y el interés por distintos contenidos o productos.
- **Número de clics (*Clicks*):** mide el número de veces que los usuarios han pinchado en un anuncio, enlace o contenido.
- **Tasa de clics (*Click-Through Rate, CTR*):** representa el porcentaje de usuarios que pincharon en un anuncio, enlace o contenido sobre el total de impresiones. Se calcula dividiendo el total de clics entre el total de impresiones.
- **Coste por clic (*Cost Per Click, CPC*):** indica el coste que paga una empresa por cada clic obtenido en una campaña publicitaria. Se calcula dividiendo el coste total de la campaña entre el número de clics obtenidos. Permite analizar la eficiencia económica en la generación de tráfico y comparar el rendimiento entre distintas campañas o plataformas. Este precio se fija, en muchas ocasiones, mediante un sistema de puja.

- **Coste por adquisición (*Cost Per Acquisition, CPA*):** representa el coste que paga una empresa por cada conversión conseguida. Se calcula dividiendo el total del coste de la campaña entre el número total de conversiones.
- **Coste por mil impresiones (*Cost Per Mille, CPM*):** es el precio que paga la empresa por recibir 1.000 impresiones. Se calcula dividiendo el coste total de la campaña entre el número de impresiones, todo ello multiplicado por 1000.
- **Coste por Lead (*Cost Per Lead, CPL*):** es el precio que paga una marca por conseguir cada lead. Un lead es cada persona que deja los datos de contacto como el nombre, correo electrónico o teléfono porque está interesado en el producto o servicio que la empresa ofrece. Se calcula dividiendo la inversión publicitaria entre el número de leads obtenidos.
- **Tasa de conversión (*Conversion Rate*):** Mide la proporción de usuarios que realizan una acción deseada. Esta acción puede ser efectuar una compra, suscribirse a la página web, solicitar información, completar un formulario, entre otras. Se calcula dividiendo el número total de conversiones entre el número total de usuarios o de clics, según el tipo de análisis que se quiera realizar. Este KPI es clave para cualquier campaña publicitaria, ya que permite evaluar la eficacia de la campaña a la hora de conseguir que un usuario realice la acción deseada.
- **Retorno de la inversión publicitaria (*Return on Ad Spend, ROAS*):** representa los ingresos generados por cada unidad monetaria invertida en publicidad. Si tengo un ROAS de 20€, significa que, por cada 1€ invertido en publicidad, se han generado 20€ de ingresos. Se calcula dividiendo los ingresos generados por la campaña entre la inversión publicitaria. Permite evaluar de forma directa la efectividad de la inversión en anuncios.
- **Retorno de la inversión (*Return on Investment, ROI*):** se calcula dividiendo el beneficio neto entre la inversión total. Mide la rentabilidad de una inversión teniendo en cuenta tanto los ingresos como los costes asociados. Al contrario que el ROAS, el ROI tiene en cuenta todos los costes implicados, no solo la inversión publicitaria, lo que proporciona una visión más completa de la rentabilidad. Como evolución de esta métrica, se utiliza el MROI (*Marketing Return on Investment*), que se centra específicamente en la rentabilidad derivada de las actividades de marketing, lo que facilita una evaluación y análisis más preciso del impacto de las campañas y recursos dedicados.
- **Tasa de rebote (*Bounce Rate*):** representa el porcentaje de usuarios que abandonan un sitio web sin realizar ninguna interacción ni acceder a otras páginas del sitio. Se calcula dividiendo el número de sesiones de una sola página entre el total de sesiones.

- **Valor del ciclo de vida del cliente (*Customer Lifetime Value, CLV o LTV*):** es una estimación de los ingresos totales que va a generar un cliente a la empresa durante toda la relación con la marca. Para calcularlo se tienen en cuenta distintos factores como la frecuencia de compra, el valor medio de cada pedido o la duración de la relación comercial. Este indicador resulta bastante útil a la hora de definir estrategias de fidelización.
- **Valor medio del pedido (*Average Order Value, AOV*):** es el importe promedio gastado por los clientes en cada compra o transacción. Se calcula dividiendo los ingresos totales entre el número total de compras o pedidos.
- **Tasa de recompra (*Repeat Purchase Rate*):** representa el porcentaje de clientes que vuelven a realizar una compra durante un periodo determinado. Es un indicador clave para analizar el grado de fidelización, lealtad y satisfacción de los clientes hacia la marca. Se calcula dividiendo el número de clientes recurrentes entre el total de clientes.
- **Referencias (*Referrals*):** número de nuevos clientes que llegan a la marca a través de recomendaciones de otros clientes ya existentes mediante programas de referidos o enlaces compartidos. Refleja el grado de satisfacción y confianza en la marca, además de ser una fuente de adquisición con alto potencial de conversión.
- **Menciones sociales (*Social Mentions*):** número de veces que una marca es mencionada en redes sociales. Resulta bastante útil para analizar la reputación de la marca.

Es muy importante tener en cuenta que, dependiendo de la etapa del *Marketing Funnel* que se encuentre la estrategia (*Awareness, Consideration, Conversion, Loyalty y Advocacy*), los objetivos que queramos conseguir pueden ser distintos y, por ello, analizaremos unos KPIs u otros. Cada fase del *funnel* exige un análisis específico alineado con el objetivo estratégico correspondiente.

En la fase de *awareness*, tal y como ya hemos explicado, el objetivo principal es alcanzar al mayor número posible de personas generando reconocimiento de la marca, por lo que, cobrarán especial relevancia métricas como el número de impresiones o alcance. Por otro lado, en la etapa de *consideration*, lo que se busca es generar tanto interés como tráfico hacia nuestra marca, por lo que se analizarán indicadores como número de visitas, *engagement*, o el *click-through rate*. En la fase de *conversion*, el objetivo es generar ventas o leads, por lo que los KPIs clave pasan a ser el número de compras o suscripciones, el *cost per acquisition*, la tasa de conversión o el ROAS. Finalmente, en la etapa de *Loyalty*, el objetivo es retener clientes y

maximizar su valor a largo plazo, por lo que se priorizan métricas como el CLV y la tasa de recompra.

Tabla 1: KPIs por etapa del Marketing Funnel

KPIs por etapa del Marketing Funnel	
<i>Awareness</i>	Impresiones y alcance
<i>Consideration</i>	Visitas, <i>engagement</i> , CTR
<i>Conversion</i>	Compras, suscripciones, CPA, tasa de conversión y ROAS
<i>Loyalty</i>	CLV y tasa de recompra

Fuente: elaboración propia

Como ya hemos explicado, el *funnel* puede dividirse en tres partes: *Top of the Funnel* (TOFU), *Middle of the Funnel* (MOFU) y *Bottom of the Funnel* (BOFU). Es fundamental tener en cuenta que el objetivo que persigue la empresa y sus campañas publicitarias varía en cada una de estas etapas, por lo que también deben ser distintos los KPIs e indicadores que se analicen o monitoreen para evaluar de forma efectiva el progreso, impacto y desempeño de la campaña

3. Caso Práctico:

3.1. Lectura y limpieza de los datos

Una vez desarrollado todo el marco teórico y habiendo profundizado en los conceptos clave del marketing digital, se ha llevado a cabo un análisis exploratorio de distintas variables, un modelo explicativo y un modelo predictivo a partir de la base de datos “*Digital Marketing Campaign Dataset*” que cuenta con 8.000 observaciones y 20 variables distintas. El objetivo será, por un lado, mediante el análisis exploratorio y el modelo explicativo, entender las posibles correlaciones que pueda existir entre las distintas variables y que factores pueden resultar más o menos significativo a la hora de determinar nuestra variable objetivo. Por otro lado, mediante el modelo predictivo se intentará desarrollar un modelo que sea capaz de establecer la probabilidad de conversión de un cliente a partir de los valores de determinadas variables. Este conjunto de datos proporciona una visión completa de las interacciones de los clientes con diferentes campañas de marketing digital. Incluye tanto datos demográficos de cada individuo, como métricas específicas de marketing, indicadores de comportamiento y datos históricos de compras, lo que hace que dicha base de datos sea adecuada para llevar a cabo este análisis. Es un conjunto de datos generado con fines educativos, propiedad de Sr.

Rabie El Kharoua, lo que lo hace ideal para proyectos de ciencia de datos y aprendizaje automático.

Como ya hemos visto, el marketing digital juega hoy en día un papel fundamental en el mundo empresarial, ya que permite la segmentación de audiencias de manera precisa, facilitando así la optimización de recursos y la toma de decisiones basada en datos reales. Es por ello por lo que, como ya hemos mencionado, mediante este caso práctico, no solo se pretende predecir la probabilidad de conversión de un cliente, sino también identificar cuáles son las variables significativas que influyen en dicha conversión, así como la posible correlación entre ellas. Asimismo, este análisis se relaciona con los modelos de atribución en marketing digital, que sirven para entender qué puntos o canales influyen en que un usuario termine tomando la decisión final de compra o conversión. Tradicionalmente, estos modelos asignaban el mérito de forma muy simplificada, por ejemplo, al primer contacto, al último o repartiendo todo por igual entre los distintos pasos del proceso. Sin embargo, esto no refleja de forma correcta cómo compran realmente las personas hoy en día, ya que los usuarios suelen interactuar varias veces con la marca y antes de tomar la decisión final. Por ello, con la ayuda de la inteligencia artificial, y de los algoritmos de aprendizaje de *machine learning*, se están desarrollando modelos más avanzados que permiten analizar mejor estos recorridos y entender de forma más precisa qué factores influyen en la decisión de conversión del consumidor.

En primer lugar, se ha llevado a cabo la instalación de todas las librerías necesarias que vamos a necesitar a lo largo de todo el análisis. Una vez instaladas y descargadas las librerías necesarias, se ha procedido a la lectura del *dataset*.

Ilustración 3: Lectura del dataset

```
df = pd.read_csv('digital_marketing_campaign_dataset.csv', sep=';', index_col=False)
df = df.loc[~df.columns.str.contains('^unnamed')]
df.head()
```

	CustomerID	Age	Gender	Income	CampaignChannel	CampaignType	AdSpend	ClickThroughRate	ConversionRate	WebsiteVisits	PagesPerVisit	TimeOnSite	SocialShares	EmailOpens	EmailClicks	PreviousPurchases	LoyaltyPoints	AdvertisingPlatform	AdvertisingTool	Conversion
0	8000	55	Female	135912	Social Media	Awareness	6.497.870.008.417.780	0.043919	0.088031	0	2.399.019.527.783.840	73.988.025.807.060.000	19	8	9	4	888	IsConfId	ToolConfId	1
1	8001	60	Male	41760	Email	Retention	38.688.088.088.007.200	0.155725	0.182725	42	29.171.377.546.985.100	6.352.548.051.242.100	5	2	7	2	3459	IsConfId	ToolConfId	1
2	8002	45	Female	88465	PPC	Awareness	15.454.295.987.900.600	0.277480	0.079423	2	8.223.919.007.895.170	13.794.901.485.538.000	0	11	2	8	2337	IsConfId	ToolConfId	1
3	8003	32	Female	44085	PPC	Conversion	5.395.295.187.793.690	0.137611	0.088004	47	4.540.938.073.132.840	14.688.352.621.260.000	89	2	2	0	2463	IsConfId	ToolConfId	1
4	8004	60	Female	83864	PPC	Conversion	16.790.438.729.670.600	0.252851	0.109940	0	2.048.947.118.073.750	13.993.370.402.828.000	6	8	8	8	4345	IsConfId	ToolConfId	1

Fuente: elaboración propia a partir de los datos del dataset

A continuación, se ha extraído el número de variables y de observaciones de la base de datos y se ha realizado un análisis de la distribución de la variable objetivo “*Conversión*”.

Ilustración 4: Análisis descriptivo del conjunto de datos y variable objetivo

```

num_variables = df.shape[1]
num_observaciones = df.shape[0]
conversion_porcentajes = df['Conversion'].value_counts(normalize=True) * 100
print(f"Número de variables (columnas): {num_variables}")
print(f"Número de observaciones (filas): {num_observaciones}")
print(f"Porcentaje de 0s y 1s en la variable 'Conversion':")
print(conversion_porcentajes)

```

```

Número de variables (columnas): 20
Número de observaciones (filas): 8000

Porcentaje de 0s y 1s en la variable 'Conversion':
Conversion
1    87.65
0    12.35
Name: proportion, dtype: float64

```

Fuente: elaboración propia a partir de los datos del dataset

Podemos observar cómo nuestra base de datos consta de 20 variables distintas, tanto numéricas como categóricas. Contamos además con un total de 8.000 observaciones distintas, que, en este caso, son diferentes usuarios. En cuanto a la variable objetivo, conversión, se trata de una variable dicotómica, que contiene valores de 1 o 0 para cada individuo dependiendo de si este ha o no convertido, (ha convertido = 1, no ha convertido = 0). Se ha realizado un análisis de la distribución de la misma, y, según los resultados, podemos observar cómo contamos con un 87,65% de individuos que si convirtieron frente a un 12,35% que no lo hicieron. Para evitar que nuestro modelo aprenda un sesgo hacia la clase mayoritaria ignorando la minoritaria, vamos a corregir este equilibrio manualmente mediante las siguientes líneas de código:

Ilustración 5: Ajuste del dataset para equilibrar las clases

```

from sklearn.utils import resample
df_majority = df[df['Conversion'] == 1]
df_minority = df[df['Conversion'] == 0]
df_majority_undersampled = resample(df_majority,
                                   replace=False,
                                   n_samples=len(df_minority),
                                   random_state=42)
df_balanced = pd.concat([df_majority_undersampled, df_minority])
print("Nueva distribución de la variable 'Conversion' en df_balanced después del submuestreo:")
print(df_balanced['Conversion'].value_counts())

```

```

Nueva distribución de la variable 'Conversion' en df_balanced después del submuestreo:
Conversion
1    988
0    988
Name: count, dtype: int64

```

Fuente: elaboración propia a partir de los datos del dataset

Previo a la limpieza de datos, se ha llevado a cabo un análisis preliminar de las diferentes variables que van a formar parte del análisis, con el objetivo de entender su utilidad y finalidad dentro del estudio, así como identificar posibles inconsistencias o información irrelevante para los modelos.

Análisis de las variables:

Para llevar a cabo el análisis de las variables, estas se han dividido en cuatro bloques diferentes dependiendo de la información que aportan. Estas categorías son: variables sociodemográficas,

variables de datos históricos de la interacción de los clientes con la marca, variables sobre las características de cada campaña publicitaria y variables de engagement. Adicionalmente se ha analizado también por separado la variable objetivo.

Información sociodemográfica de los clientes

- Customer ID: *Customer ID* es una variable categórica nominal. Representa un identificador único de cada cliente.
- Age: *Age* es una variable numérica nominal. Representa la edad de cada uno de los usuarios.
- Gender: Variable dicotómica (hombre/mujer). Indica el género de cada cliente.
- Income: la variable *income* representa los ingresos anuales de cada consumidor.

Datos históricos sobre los clientes con la marca

- Previous Purchases: variable numérica nominal. Representa el número de compras realizadas anteriormente por cada uno de los clientes de la base de datos.
- Loyalty Points: variable numérica nominal. Indica la cantidad de puntos de fidelización que tiene un cliente.

Variables sobre las características de la campaña de marketing

- Campaign Channel: variable categórica nominal. Indica a través de qué canal se ha distribuido cada campaña. (*Email, Social, Media, SEO, PPC y Referral*).
- Campaign Type: variable categórica nominal. Representa el tipo de campaña de la que se trataba (*Awareness, Consideration, Conversion o Retention*)
- Advertising Platform: variable categórica nominal. Nos proporciona información acerca de en qué plataforma se ejecutó la campaña.
- Advertising Tool: variable categórica nominal. Representa la herramienta específica usada para gestionar la campaña.

Variables de engagement de cada campaña

- Click Through Rate: variable numérica nominal. Representa el porcentaje de personas que hacen clic. Se calcula dividiendo el número total de clics entre el número total de impresiones.

- Conversion Rate: variable numérica continua. Representa la proporción de clics que se convierten en una acción deseada.
- Website Visits: variable numérica nominal. Representa el número de visitas que obtuvo la web.
- Pages Per Visit: variable numérica nominal. Indica la media de páginas vistas por sesión.
- Time On Site: variable numérica nominal. Representa el tiempo medio que un usuario pasa en la web (en minutos).
- Social Shares: variable numérica nominal. Representa el número de veces que se compartió la campaña.
- Email Opens: variable numérica nominal. Representa el número de veces que se abrieron los mails.
- Email Clicks: variable numérica nominal. Representa el número de veces que se pinchó en los enlaces que contenían los mails.

Variable objetivo

- Conversion: variable dicotómica (1/0) que representa si un cliente realizó la acción objetivo o no.

Una vez realizada la lectura de la base y analizadas las diferentes variables, se ha llevado a cabo la limpieza de datos.

En primer lugar, se ha decidido prescindir de algunas variables que se ha considerado que no aportaban valor o información relevante a nuestro análisis. Estas han sido:

- *Customer ID*
- *Advertising Tool*
- *Advertising Platform*

La variable de *Customer ID* se eliminó ya que, al tratarse de un identificador único de cada cliente, no proporcionaba ningún tipo de información útil ni al modelo predictivo ni al modelo explicativo, y no iba a tener ningún tipo de correlación con la variable objetivo. En segundo lugar, tanto *Advertising Tool* como *Advertising Platform* fueron descartadas debido a que

ambas presentaban un único valor en la base de datos, por lo que tampoco ofrecían ningún tipo de información significativa al análisis.

En segundo lugar, se llevó a cabo un proceso de dicotomización de todas aquellas variables categóricas, transformándolas en numéricas mediante *One-Hot Encoding*, facilitando así su tratamiento. Este procedimiento consiste en crear una nueva columna binaria por cada categoría de la variable original, dejando fuera una categoría de referencia para evitar la multicolinealidad en el modelo. Cada observación recibe un 1 si pertenece a la categoría correspondiente o un 0 en caso contrario.

Las variables categóricas que se transformaron en variables numéricas dicotómicas mediante este proceso fueron las siguientes:

- *Campaign Channel*
- *Campaign Type*
- *Gender*

En tercer lugar, se han convertido todas aquellas variables numéricas continuas como *ad spend*, *pages per visit* o *time on site* en formato decimal con una coma, unificando así el tratamiento de los datos.

En cuarto lugar, se han estandarizado todas las variables numéricas. Este proceso consiste en transformarlas para que tengan una media de 0 y una desviación estándar de 1, es decir, en llevar todos los valores a una misma escala. Esto resulta fundamental, ya que las variables originales pueden presentar rangos muy distintos, lo que podría provocar que el modelo otorgue mayor peso a aquellas con valores más elevados simplemente por su magnitud y no por su relevancia real. Al estandarizar las variables, se garantiza que todas contribuyan de manera equilibrada, mejorando así el rendimiento y la estabilidad del modelo.

Una vez estandarizadas las variables y realizada la limpieza de datos, se ha comprobado que todas las variables estén en formato numérico a través de las siguientes líneas de código:

Ilustración 6: Tipo de variables

```

variables=pd.DataFrame(columns=['Nombre','Tipo'])
for column in df_balanced.columns:
    data_type = df_balanced[column].dtype
    variables = pd.concat([variables,pd.DataFrame({'Nombre':[column],'Tipo':[str(data_type)]},
                                                ignore_index = True)],ignore_index = True)
print(variables)

```

Nombre	Tipo
Age	float64
Income	float64
AdSpend	float64
ClickThroughRate	float64
ConversionRate	float64
WebsiteVisits	float64
PagesPerVisit	float64
TimeOnSite	float64
SocialShares	float64
EmailOpens	float64
EmailClicks	float64
PreviousPurchases	float64
LoyaltyPoints	float64
Conversion	int64
CampaignChannel_PPC	int64
CampaignChannel_Referral	int64
CampaignChannel_SEO	int64
CampaignChannel_Social Media	int64
Gender_Male	int64
CampaignType_Consideration	int64
CampaignType_Conversion	int64
CampaignType_Retention	int64

Fuente: elaboración propia a partir de los datos del dataset

Podemos observar que todas las variables son numéricas, ya sean continuas o discretas, (*float64* o *int64*). Para desarrollar tanto el modelo explicativo como el predictivo, es necesario que todas las variables sean numéricas, ya que los modelos requieren procesarlas mediante operaciones matemáticas. Esto permite identificar correlaciones y patrones presentes en la base de datos, facilitando así el aprendizaje, la predicción y la interpretación de los resultados.

Como último paso en el análisis y limpieza de la base de datos y de sus variables, se va a analizar la posible presencia de valores vacíos en nuestra base de datos. Afortunadamente, no existen valores nulos en ninguna de nuestras variables, por lo que no ha sido necesario hacer nada más respecto de este aspecto.

3.2. Análisis exploratorio por segmentos

Después de haber llevado a cabo la lectura y limpieza de los datos, se va a proceder a un análisis exploratorio por segmentos con el objetivo de identificar posibles correlaciones que pueda haber entre las variables explicativas y nuestra variable objetivo.

Mediante este análisis se busca identificar posibles patrones preliminares que puedan ser relevantes, así como conocer la estructura y distribución de las variables. Esto nos ayudará a tener una base sólida para la construcción posterior de tanto el modelo explicativo como el predictivo, así como la interpretación de los mismos. El análisis se va a dividir en tres subapartados dependiendo de las variables de las que se esté examinando la correlación:

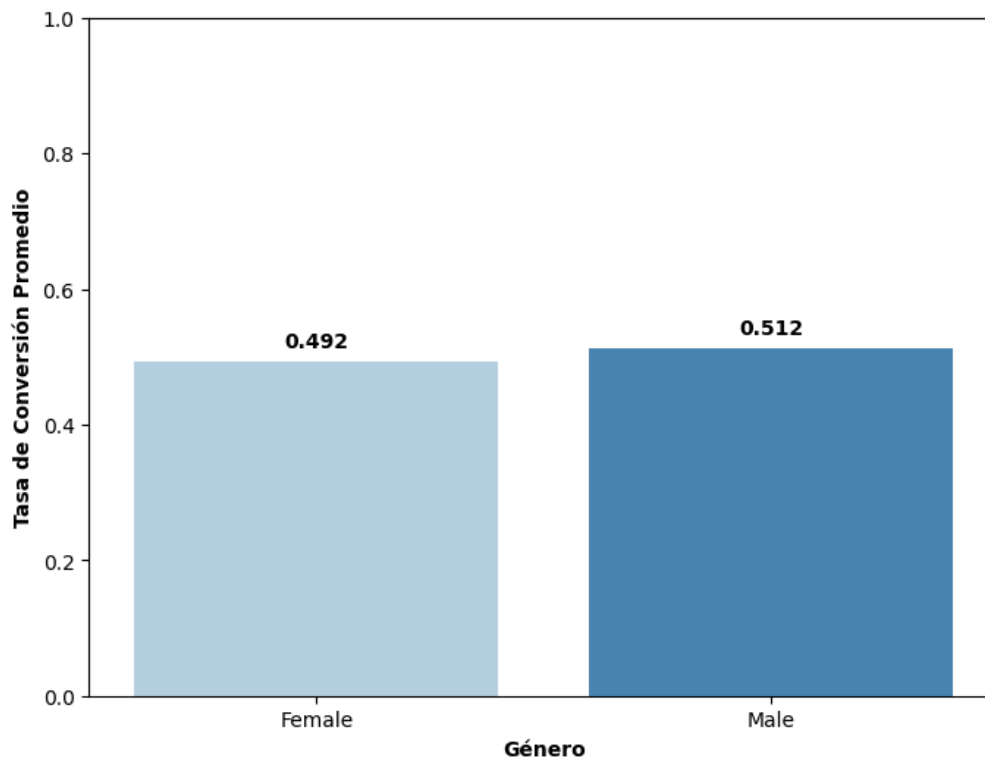
aspectos sociodemográficos, comportamiento digital del consumidor o cliente con las campañas y características de la campaña.

3.2.1. Aspectos sociodemográficos

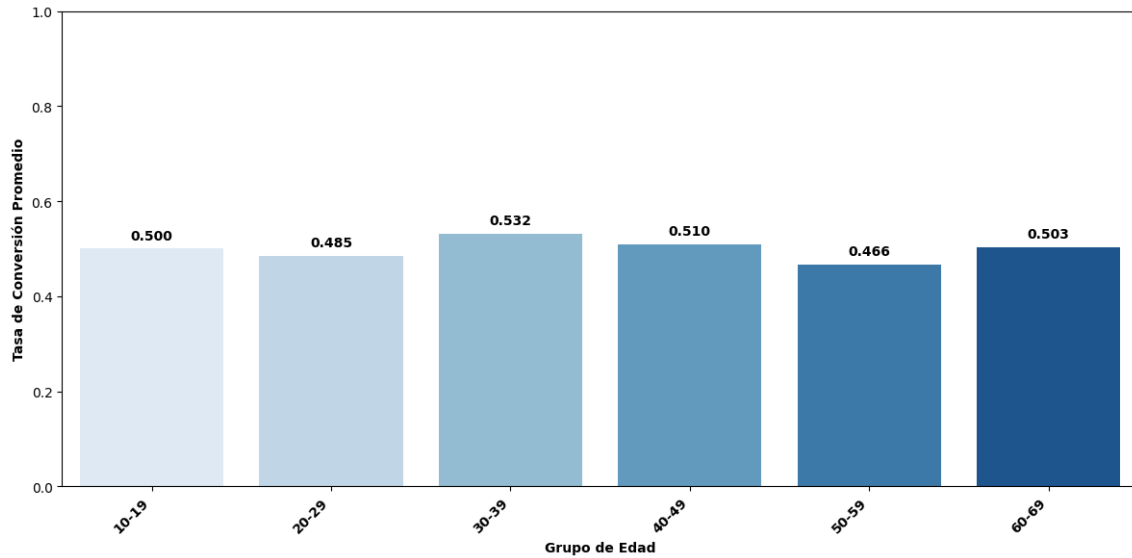
Tal y como acabamos de mencionar, en primer lugar, se va a realizar un análisis de las posibles relaciones que pueda haber entre las variables sociodemográficas de nuestra base de datos (género, edad e ingresos) y la variable objetivo de conversión. El objetivo es identificar si, alguna de estas características propias de cada cliente puede estar asociada a una mayor o menor tasa de conversión.

A continuación, se presentarán algunas visualizaciones segmentadas por características sociodemográficas clave de los clientes:

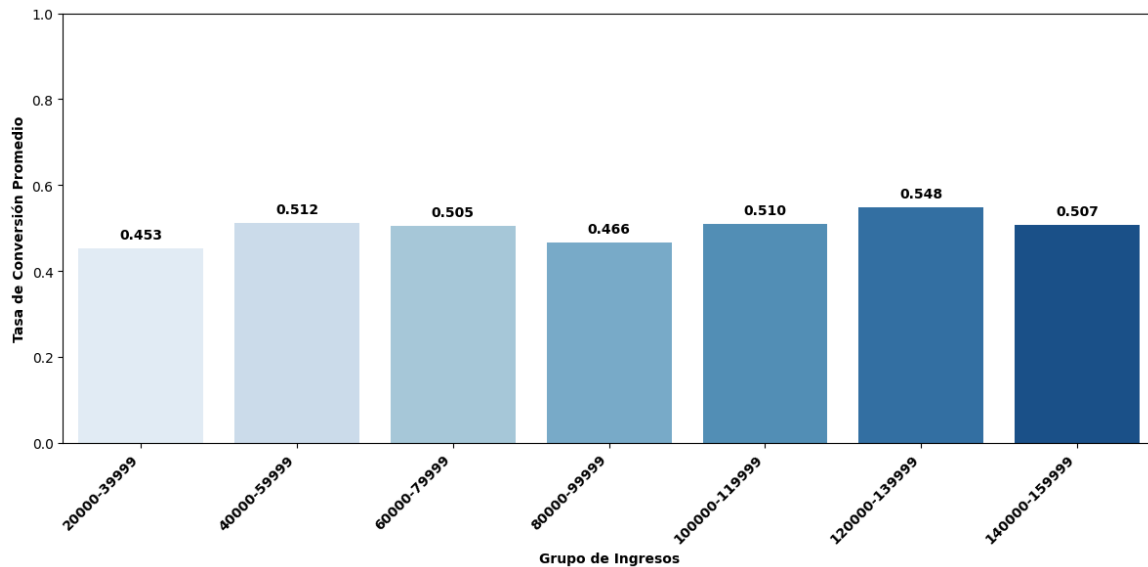
Ilustración 7: Tasa de Conversión Promedio por Género



Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 8: Tasa de conversión promedio por edad

Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 9: tasa de conversión promedio por rango salarial

Fuente: elaboración propia a partir de los resultados del dataset

A partir del análisis de los tres gráficos, podemos observar que, en primer lugar, en cuanto al análisis de la tasa de conversión por género, parece que, aunque la tasa de conversión es ligeramente más alta en hombres (51,2%) que en mujeres (49,2%), la diferencia es mínima, por lo que no parece que la variable género resulte significativa a la hora de explicar la conversión de los clientes.

En segundo lugar, respecto al análisis de la tasa de conversión por edad, podemos ver cómo, el rango de edad que presenta una mayor tasa de conversión es el de 30-39 años (53,2%), seguido de los grupos de edad de 40-49 y 60-69 años, con unas tasas de conversión promedio del 51,0% y 50,3% respectivamente. Por otro lado, el rango de 50-59 años es el grupo de edad que menor tasa de conversión, un 46,6%. Aunque existen ciertas variaciones entre los grupos de edad, las diferencias no parecen ser muy significativas, lo que podría sugerir que la edad no parece ser un determinante importante en la conversión de los clientes.

Por último, en lo referente a la tasa de conversión por rango salarial, el análisis muestra una diferencia algo más marcada: el rango de ingresos de 120.000-139.999 € presenta la tasa de conversión más alta (54,8%), mientras que el rango más bajo, de 20.000-39.999 €, alcanza solo un 45,3%. Esta diferencia de casi el 10% podría indicar que el nivel de ingresos podría ser una variable relativamente significativa a la hora de explicar la probabilidad de conversión de los clientes, en contraste con el género o la edad.

Aunque los resultados del análisis muestren que, en términos generales, la tasa de conversión parece mantenerse relativamente consistente entre los diferentes grupos sociodemográficos, es importante destacar la importancia de la segmentación de clientes como herramienta de optimización de las estrategias de marketing digital. Entender cómo la tasa de conversión puede variar entre los distintos grupos de edad, género o nivel de ingresos permite a las empresas identificar a aquellos grupos con mayor o menor propensión a convertir, pudiendo así adaptar las campañas de manera más efectiva.

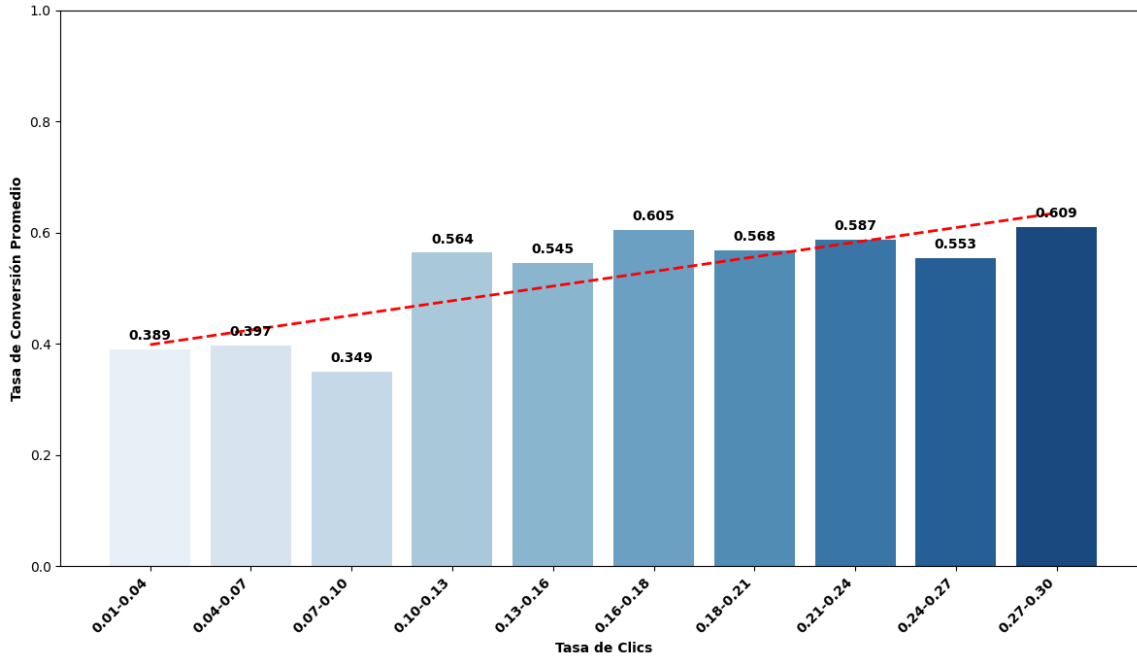
3.2.2. Comportamiento del consumidor

En segundo lugar, se va a analizar la posible relación que pueda haber entre la variable objetivo y diferentes métricas de comportamiento digital e interacción con la campaña publicitaria de cada uno de los clientes. Entre las variables analizadas vamos a considerar: *Click-Through Rate*, *Conversion Rate*, *Website Visits*, *Pages per Visit*, *Time On Site*, *Social Shares*, *Email Opens*, *Email Clicks*, *Previous Purchases* y *Loyalty Points*.

El propósito de este análisis es identificar patrones y asociaciones entre el comportamiento del cliente y su probabilidad de conversión, proporcionando una visión preliminar sobre qué métricas pueden tener mayor influencia en la decisión final del consumidor.

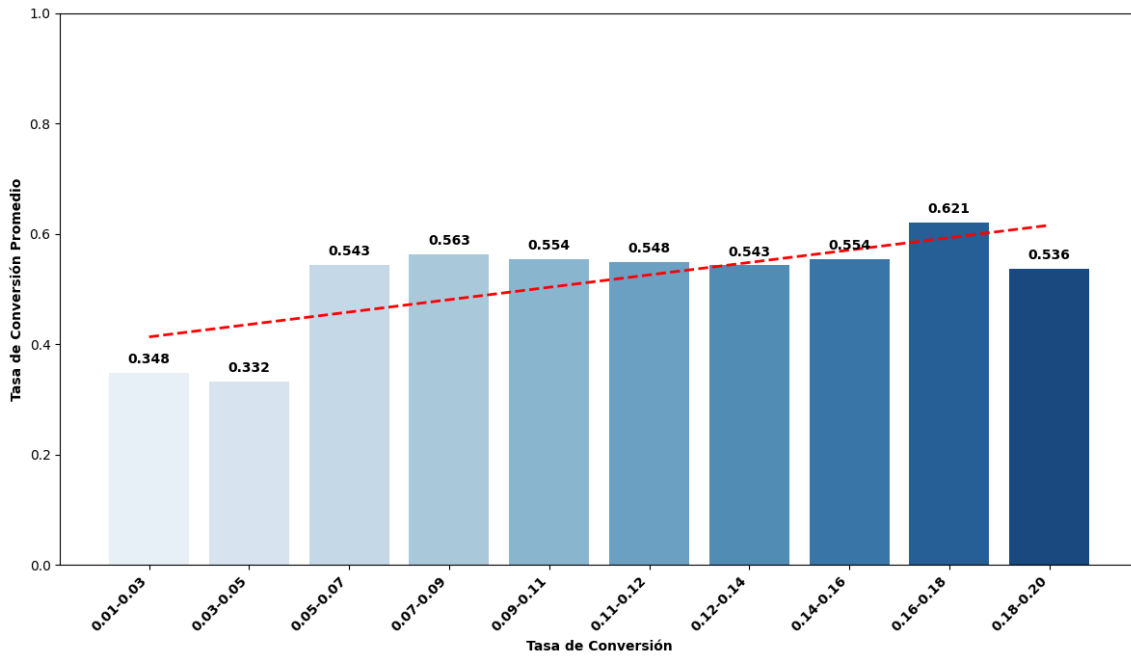
A continuación, se presentarán diversas visualizaciones segmentadas en función del comportamiento del consumidor:

Ilustración 10: tasa de conversión promedio por CTR



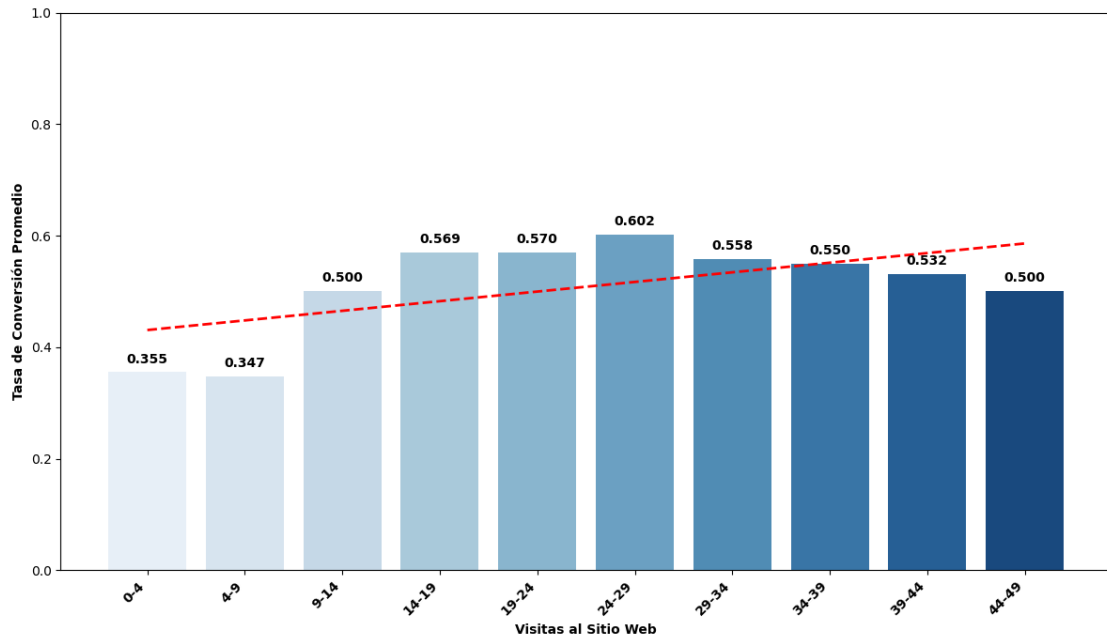
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 11: Tasa de conversión promedio por Conversion Rate



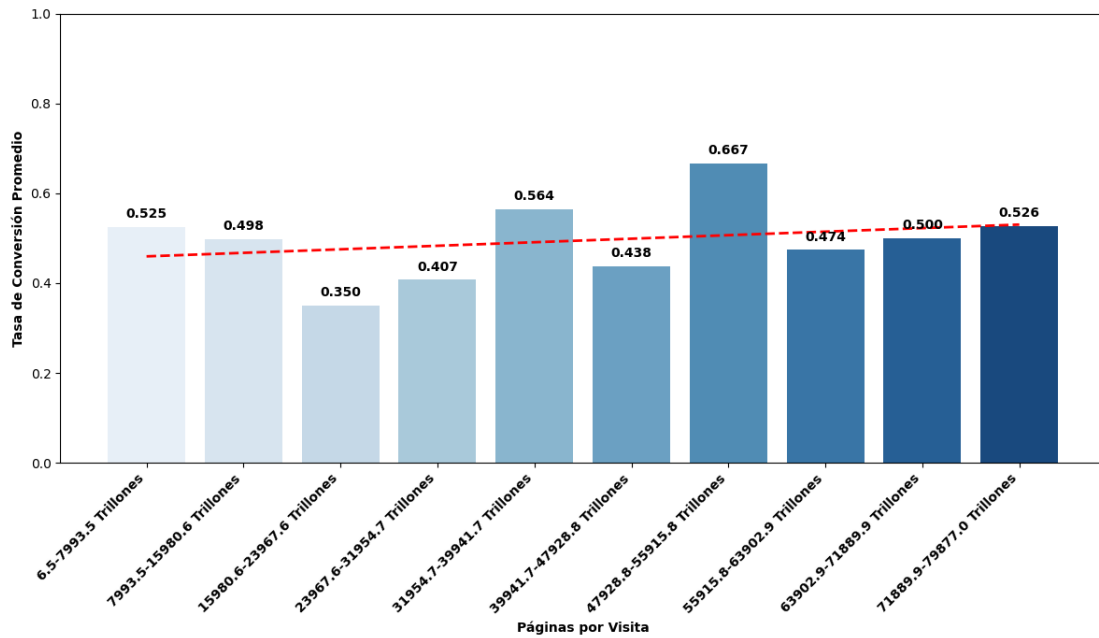
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 12: Tasa de conversión promedio por website visits



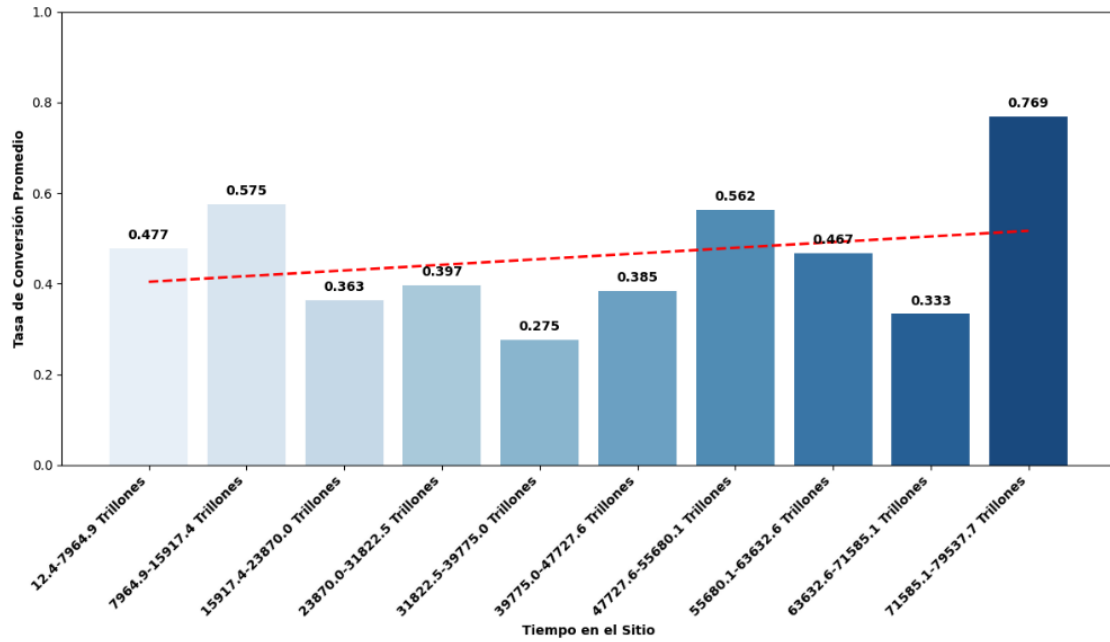
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 13: tasa de conversión por medio por páginas por visita



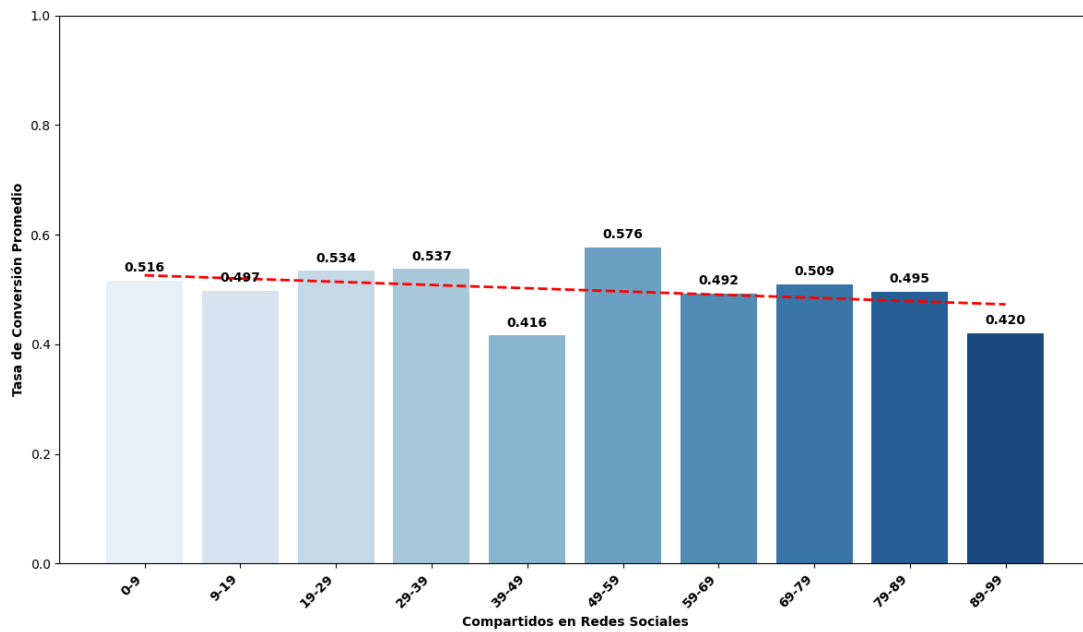
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 14: tasa de conversión promedio por time on site



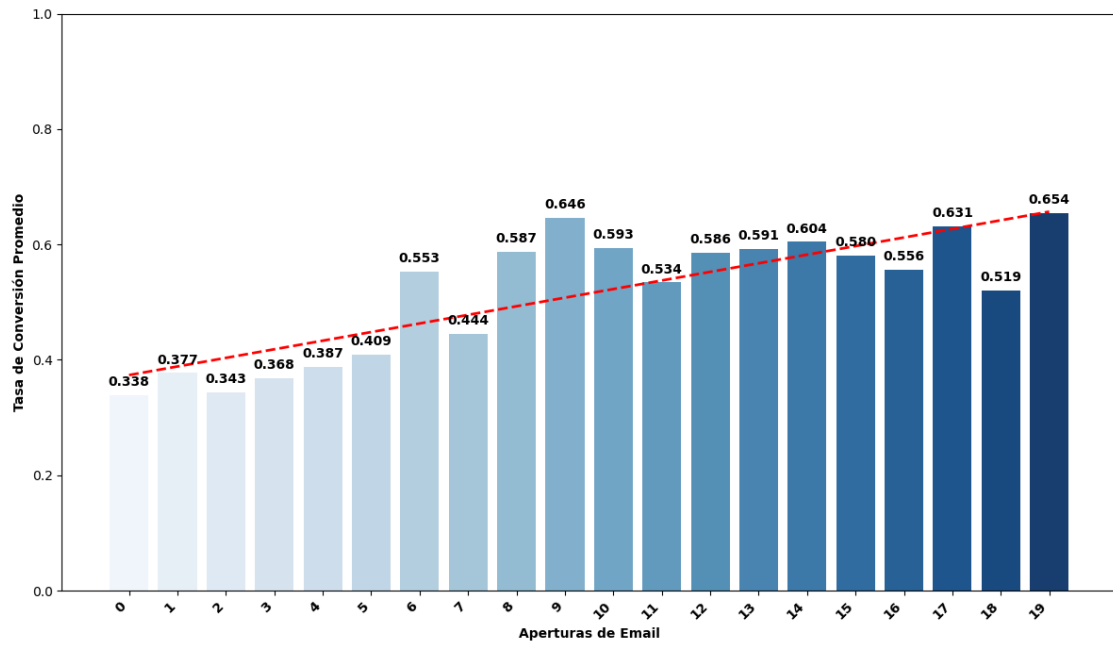
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 15: tasa de conversión promedio por social shares



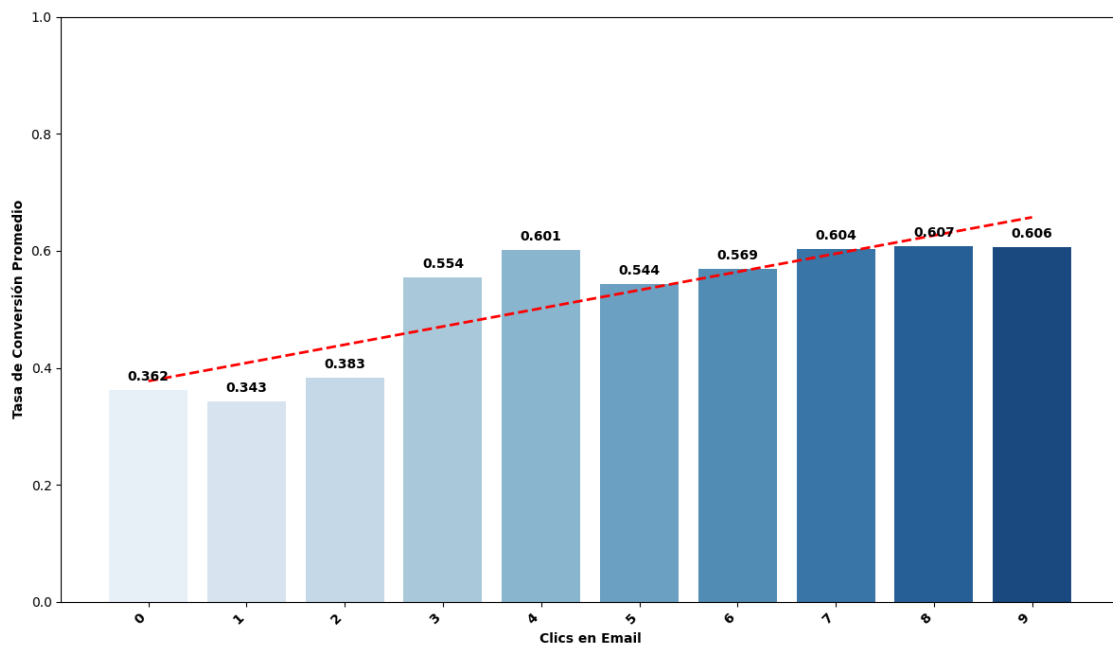
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 16: tasa de conversión promedio por email opens



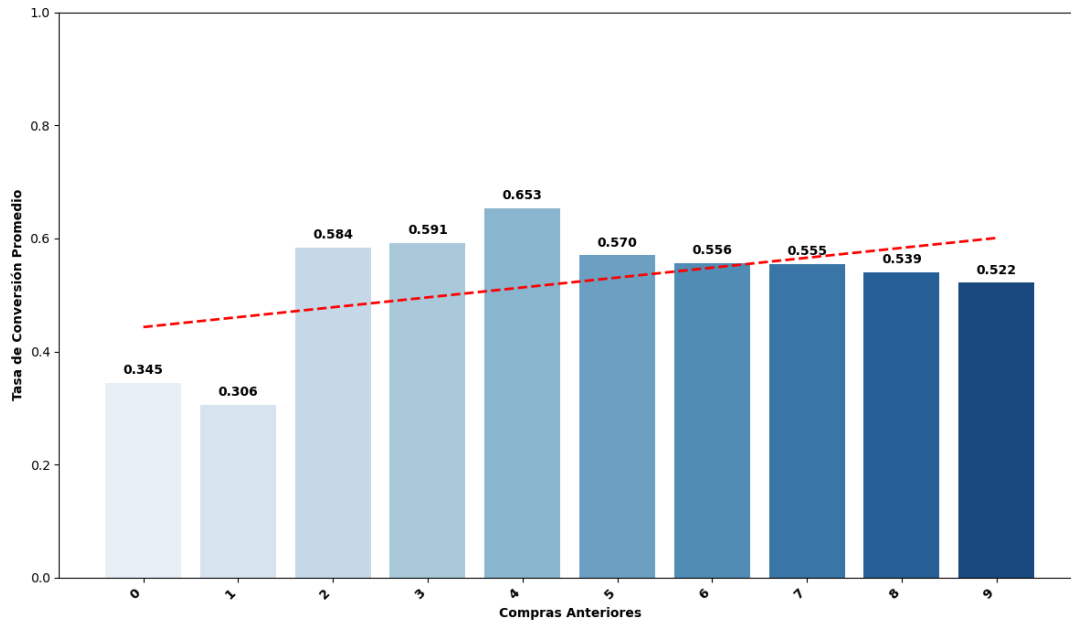
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 17: tasa de conversión promedio por email clicks



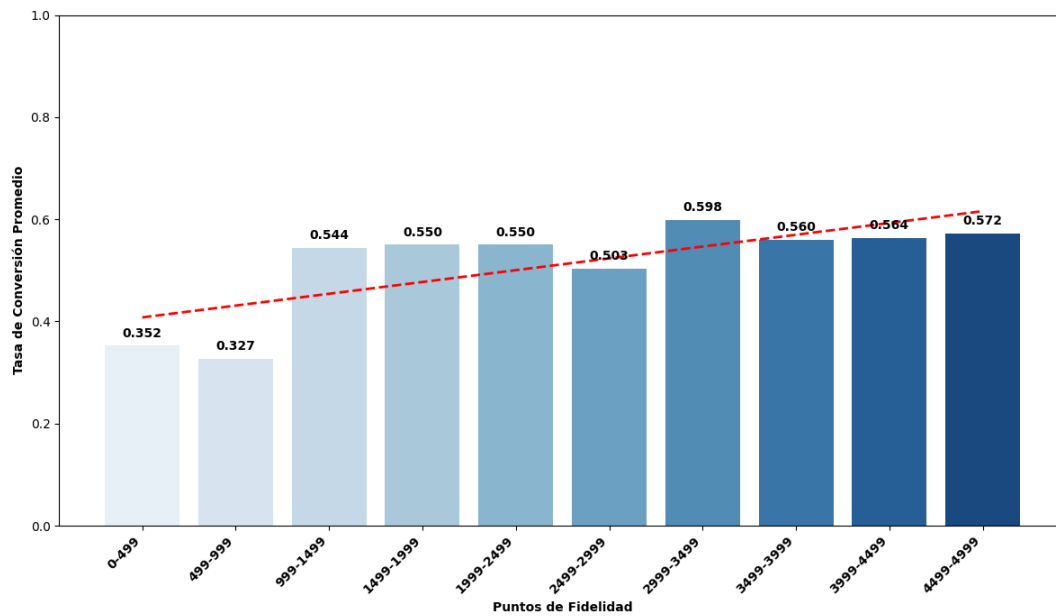
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 18: Tasa de Conversión Promedio por Previous Purchases



Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 19: Tasa de conversión promedio por loyalty points



Fuente: elaboración propia a partir de los resultados del dataset

Después de analizar los resultados obtenidos, podemos observar cómo algunas métricas sí parecen ser significativas en cuanto a la decisión final del cliente, mientras que otras parecen que no influyen o son irrelevantes respecto a determinar si existe una mayor o menor tasa de conversión promedio. Vamos a analizar los resultados de todas ellas:

En primer lugar, en el análisis de la tasa de conversión promedio en relación con el *click-through rate* (CTR), se percibe una posible correlación positiva entre ambas variables. Podemos observar cómo, a partir de un CTR del 10%, la tasa de conversión comienza a incrementarse de manera notable, alcanzando valores especialmente elevados, alcanzando hasta un 60,9% con un CTR dentro del rango entre el 27% al 30%. Estos resultados sugieren que la tasa de clics podría ser una variable determinante de la probabilidad de conversión de los clientes.

En segundo lugar, al analizar la tasa de conversión promedio en relación con la tasa de conversión histórica (*conversion rate*), también se observa una correlación positiva entre ambas variables. Este resultado es coherente, ya que indica que, cuanto mayor haya sido la tasa de conversión previa de un individuo, mayor es la probabilidad de que vuelva a convertir. A partir de una tasa de conversión histórica del 5%, se aprecian incrementos significativos en la tasa de conversión promedio, alcanzando su valor máximo en el intervalo comprendido entre el 16% y el 18%.

En tercer lugar, al analizar la correlación entre la tasa de conversión promedio y el número de visitas al sitio web (*website visits*), se observa nuevamente una tendencia general positiva. Esto sugiere que, cuanto mayor sea el número de visitas realizadas por un usuario, mayor será su probabilidad de conversión.

En cuarto lugar, en relación con la posible correlación entre las páginas por visita (*pages per visit*) y la tasa de conversión promedio, se aprecia una ligera tendencia positiva, aunque esta es bastante débil y presenta ciertas fluctuaciones. Por lo tanto, aunque exista cierta correlación entre ambas variables, el número de páginas visitadas no parece constituir un factor determinante para explicar la probabilidad de conversión de un cliente.

De manera similar, ocurre con la relación entre el tiempo en el sitio (*time on site*) y la conversión. Aunque se observa una correlación positiva y la tasa de conversión promedio tiende a ser más alta cuando un usuario pasa más tiempo en la web, esto tiene sentido, ya que quienes finalmente realizan una acción suelen permanecer más tiempo en la página. Sin embargo, la correlación no es especialmente significativa: se observan notables fluctuaciones,

lo que indica que la duración de la visita no constituye un factor decisivo para predecir la probabilidad de conversión de un cliente.

En cuanto al análisis de la correlación entre la tasa de conversión promedio y el número de compartidos en redes sociales, algunos intervalos, como los rangos de 29-39 ó 49-59 compartidos, muestran tasas de conversión ligeramente más altas, pero, en general, los valores fluctúan y no parece que se siga una tendencia consistente. Esto sugiere que el número de compartidos por sí solo no es un factor determinante en la probabilidad de conversión. Esto podría deberse a que otros elementos como la relevancia y calidad del contenido, o el momento de la publicación probablemente influyen de manera más significativa.

En sexto lugar, al analizar la correlación entre la tasa de conversión promedio y las variables *Email Opens* y *Email Clicks*, se observa que ambas muestran una relación positiva con la tasa de conversión. Esto indica que, a mayor interacción con los correos electrónicos, ya sea abriéndolos o haciendo clic en ellos, aumenta la probabilidad de que los usuarios realicen una conversión.

Por último, al analizar la relación entre las variables *Previous Purchases* y *Loyalty Points*, se observa que ambas presentan una correlación positiva con la tasa de conversión promedio. Esto significa que los clientes que han realizado más compras anteriormente o que acumulan más puntos de fidelidad tienden, en general, a convertir con mayor frecuencia, lo que resalta la importancia de fomentar la lealtad del cliente.

A partir de todo este análisis, podemos establecer que algunas variables muestran una relación más significativa con la tasa de conversión promedio mientras que otras parecen tener un impacto más limitado o inconsistente. Las variables que, parecen ser más significativas en cuanto a la probabilidad de conversión de un cliente son: *Click-Through Rate*, *Conversion Rate*, *Website Visits*, *Email Clicks*, *Previous Purchases* y *Loyalty Points*. Se observa como todas estas métricas presentan una correlación positiva con la tasa de conversión promedio, lo que sugiere que un mayor desempeño o interacción en cada una de ellas tiende a asociarse a una mayor probabilidad de que los clientes realicen una conversión.

Este análisis es especialmente útil ya que permite a las empresas identificar qué métricas o indicadores de comportamiento del cliente son más relevantes para el éxito comercial. Esto permite diseñar estrategias más efectivas, así como optimizar las campañas y esfuerzos. En resumen, conocer y ser capaz de identificar qué variables generan un mayor impacto en la

conversión ayuda a priorizar recursos y acciones maximizando así el retorno sobre la inversión y mejorando el rendimiento global del negocio.

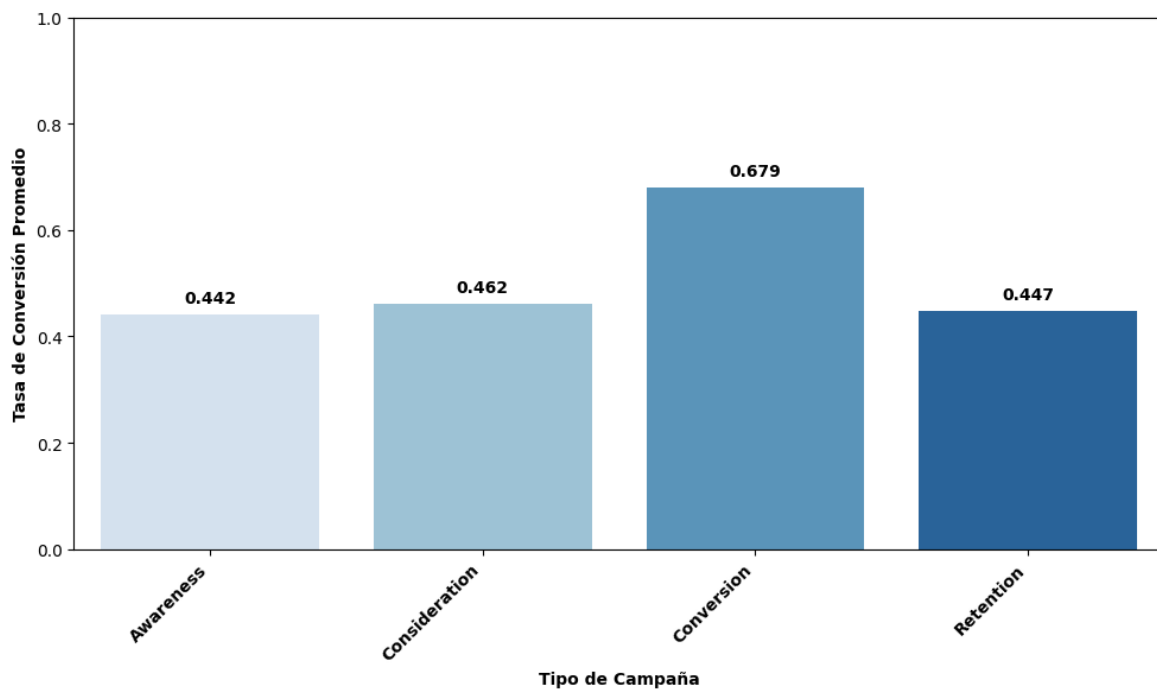
3.2.3. Características de la campaña

En este último apartado de análisis por segmentación, se va a estudiar las posibles relaciones que puedan existir entre las características de la campaña publicitaria (tipo de campaña, canal y dinero destinado), y la tasa promedio de conversión de un cliente.

El objetivo de este apartado es determinar si existe alguna correlación entre estas variables explicativas y la variable objetivo, así como ser capaces de determinar qué tipos de campaña, canales o presupuesto están siendo más efectivos.

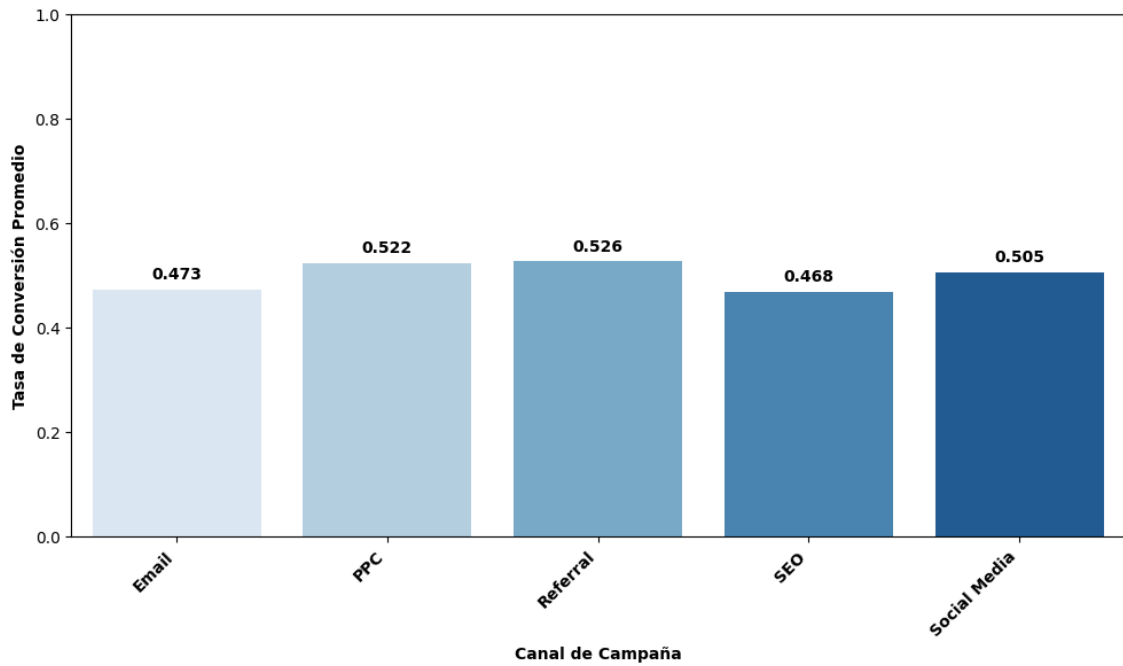
A continuación, se muestran las visualizaciones de los resultados del análisis de segmentación según las características de cada campaña publicitaria:

Ilustración 20: tasa de conversión promedio por campaign type



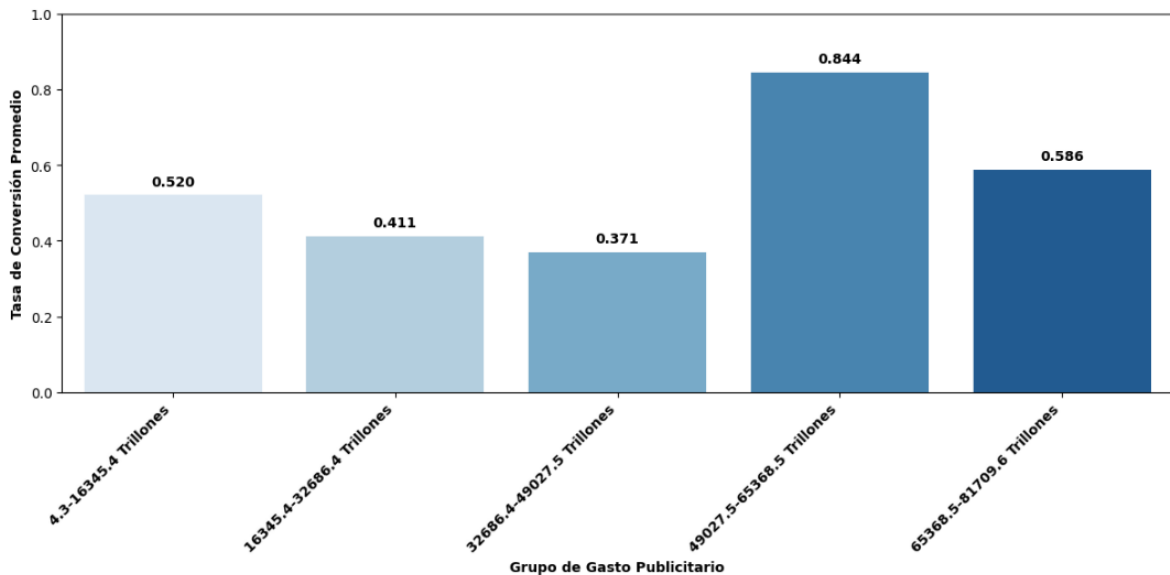
Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 21: tasa de conversión promedio por campaign channel



Fuente: elaboración propia a partir de los resultados del dataset

Ilustración 22: tasa de conversión promedio por ad spend



Fuente: elaboración propia a partir de los resultados del dataset

En primer lugar, según el análisis de correlación entre el tipo de campaña publicitaria y la tasa promedio de conversión, se puede observar cómo aquellas campañas de tipo *Conversion*

presentan una tasa significativamente más alta, alrededor del 67,9%, en comparación con el resto de las campañas como *Awareness*, *Consideration* o *Retention*, que rondan unas tasas de conversión entre el 44% y el 46%. Esto sugiere que, las campañas diseñadas explícitamente para la conversión del cliente son más efectivas a la hora de conseguir que un cliente realice una acción deseada.

En segundo lugar, respecto al análisis de la tasa de conversión promedio por el canal de la campaña, parece que los canales de *PPC* y *Referral* muestran tasas de conversión ligeramente superiores en comparación con *Email*, *SEO* o *Social Media*. Aunque no parece que haya una diferencia muy significativa, los resultados sugieren que aquellas campañas de *PPC* y *Referral* parecen ser marginalmente más efectivas.

Por último, en cuanto al análisis de la correlación entre la tasa de conversión y la variable *Ad Spend*, podemos observar cómo, la tasa promedio de conversión es notablemente más alta en el cuarto grupo de gasto publicitario, con una tasa de conversión del 88,4% frente al resto de grupos, que mantienen una tasa de conversión promedio entre el 37% y el 58%. Esto nos podría indicar que, aunque no exista una correlación lineal positiva o negativa entre ambas variables, sí pudiera existir un rango óptimo de inversiones publicitarias en el que se maximice la tasa de conversión y, por tanto, los beneficios.

Este tipo de análisis es fundamental para una empresa porque permite identificar qué estrategias, canales y niveles de inversión generan un mayor retorno, lo que permite la optimización y planificación de campañas publicitarias, asignar presupuestos de manera más eficiente y diseñar acciones más efectivas para aumentar la conversión de clientes, reduciendo desperdicio de recursos y potenciando los beneficios.

En conclusión, el análisis realizado sobre las características sociodemográficas de los clientes, las métricas de comportamiento y los tipos de campañas publicitarias ha permitido identificar patrones clave que influyen en la probabilidad de conversión. Las segmentaciones sociodemográficas ayudan a entender qué grupos de clientes tienen mayor predisposición a interactuar o comprar, mientras que el análisis de métricas de comportamiento, como visitas al sitio, interacciones con emails o historial de compras, revela qué acciones de los usuarios se asocian más estrechamente con la conversión. Asimismo, el estudio de los distintos tipos de campañas y canales permite determinar qué estrategias de marketing generan un mayor retorno sobre la inversión.

3.3. Modelo explicativo:

Después de haber realizado la lectura y limpieza de base de datos y haber estudiado los resultados obtenidos tras el análisis de segmentación se va a proceder al desarrollo de un modelo explicativo.

Dado que el objetivo de este caso práctico es analizar la probabilidad de conversión de los clientes, se va a emplear un modelo de clasificación. Este tipo de modelo es el más adecuado tanto para explicar qué factores incluyen en la probabilidad de que un cliente convierta mediante un modelo explicativo, el cual se va a desarrollar a lo largo de este apartado, como para predecir dicha probabilidad mediante un modelo predictivo, el cual se va a llevar a cabo más adelante.

A continuación, se dispondrán los resultados obtenidos del modelo explicativo:

Tabla 2: Resultados modelo explicativo de clasificación

```

*** Optimization terminated successfully.
    Current function value: 0.592435
    Iterations 5
  
```

Logit Regression Results						
Dep. Variable:	Conversion	No. Observations:	1976			
Model:	Logit	Df Residuals:	1954			
Method:	MLE	Df Model:	21			
Date:	Sun, 29 Mar 2026	Pseudo R-squ.:	0.1453			
Time:	20:27:03	Log-Likelihood:	-1170.7			
converged:	True	LL-Null:	-1369.7			
Covariance Type:	nonrobust	LLR p-value:	2.388e-71			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-0.4245	0.146	-2.898	0.004	-0.711	-0.137
Age	0.0068	0.050	0.136	0.892	-0.092	0.105
Income	0.0878	0.050	1.759	0.079	-0.010	0.186
AdSpend	-0.0249	0.050	-0.496	0.620	-0.123	0.073
ClickThroughRate	0.3587	0.051	7.074	0.000	0.259	0.458
ConversionRate	0.3176	0.051	6.258	0.000	0.218	0.417
WebsiteVisits	0.2294	0.050	4.546	0.000	0.130	0.328
PagesPerVisit	0.0013	0.050	0.027	0.979	-0.097	0.100
TimeOnSite	-0.0520	0.050	-1.043	0.297	-0.150	0.046
SocialShares	-0.0559	0.050	-1.110	0.267	-0.154	0.043
EmailOpens	0.3663	0.051	7.181	0.000	0.266	0.466
EmailClicks	0.3882	0.051	7.630	0.000	0.288	0.488
PreviousPurchases	0.2792	0.050	5.539	0.000	0.180	0.378
LoyaltyPoints	0.3274	0.051	6.468	0.000	0.228	0.427
CampaignChannel_PPC	0.2678	0.159	1.683	0.092	-0.044	0.580
CampaignChannel_Referral	0.2545	0.158	1.613	0.107	-0.055	0.564
CampaignChannel_SEO	0.0521	0.164	0.317	0.751	-0.270	0.374
Q("CampaignChannel_Social Media")	0.1685	0.159	1.058	0.290	-0.144	0.481
Gender_Male	0.0774	0.102	0.757	0.449	-0.123	0.278
CampaignType_Consideration	0.1038	0.136	0.765	0.444	-0.162	0.370
CampaignType_Conversion	0.9703	0.149	6.525	0.000	0.679	1.262
CampaignType_Retention	0.0542	0.139	0.391	0.696	-0.218	0.326

Fuente: elaboración propia a partir de los resultados del dataset

Antes de analizar los resultados obtenidos, es importante explicar brevemente el significado de algunos de los principales indicadores del modelo.

- Pseudo R²: es el indicador que evalúa la capacidad que tiene un modelo de explicar la variabilidad de la variable objetivo. Es la medida utilizada para evaluar la bondad de ajuste de modelos, cuanto más alto es, mejor capacidad de predicción tiene el modelo.
- P-valor: muestra la probabilidad de que el efecto observado sea producto del azar, es decir, si una variable presenta un p-valor del 15%, con una probabilidad del 15%, esto significa que hay un 15% de probabilidad de que el efecto observado no sea real, sino que se deba a fluctuaciones aleatorias en los datos. Vamos a considerar que aquellas variables con un p-valor menor al 10% serán significativas.
- Coefficiente: indica la dirección del efecto de cada variable. Es decir, si el coeficiente de una variable significativa es positivo, significa que existe una correlación positiva entre dicha variable y la probabilidad de que la variable objetivo sea igual a 1, si, por el contrario, el coeficiente es negativo, cuando aumenta dicha variable, disminuye la probabilidad de que la variable objetivo sea igual a 1.

Una vez comprendido qué representa cada indicador, vamos a proceder al análisis de los resultados del modelo.

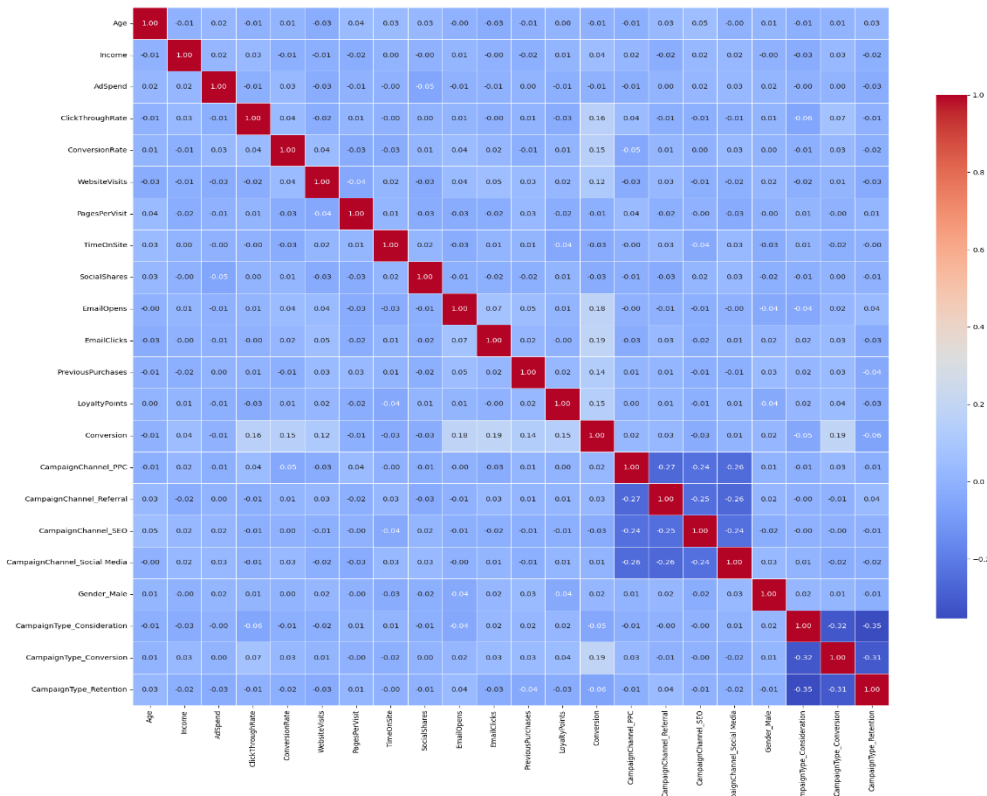
Podemos observar que nuestro modelo explicativo presenta un Pseudo R² inicial del 14,3%, lo que significa que, nuestro modelo de predicción funciona moderadamente mejor que un modelo de azar. Cabe destacar que, en términos generales, un Pseudo R² entre 0,2 y 0,4 para modelos de clasificación, es considerado como un buen nivel de ajuste, (Lee, 2025). Es importante tener en cuenta que, en el ámbito del marketing, la decisión final del cliente suele estar influenciada por múltiples factores externos difíciles de capturar completamente en un modelo, como las recomendaciones de familiares o amigos o factores emocionales o psicológicos.

Es importante también tener en cuenta que la relación entre las variables independientes y la probabilidad de conversión puede no ser estrictamente lineal, y que algunas variables podrían ejercer un efecto conjunto que no queda reflejado al incluirlas de forma aislada.

Es por ello por lo que se va a realizar una matriz de correlación, con el objetivo de identificar posibles relaciones entre las variables y determinar qué interacciones podrían incluirse en el modelo para mejorar su capacidad explicativa.

A continuación, se muestran los resultados obtenidos de la matriz de correlación:

Ilustración 23: Resultados Matriz de correlación



Fuente: elaboración propia a partir de los resultados del dataset

Se observa que algunas variables presentan ligeras correlaciones entre sí, siendo las más destacables las derivadas de la dicotomización mediante *one-hot Encoding* de las variables *Campaign Type* y *Campaign Channel*. Esto se debe a que, al ser categorías mutuamente excluyentes, cuando una variable toma el valor de 1, las demás necesariamente toman 0, generando así correlaciones negativas entre ellas.

Por otro lado, observamos algunas correlaciones entre el 10 y el 20% de otras variables con la variable objetivo como: *email opens*, *email clicks*, *previous purchases*, *loyalty points* y *campaign type conversion*. Esto refuerza los resultados que habíamos obtenido del primer análisis de segmentación, sugiriendo que estas son algunas de las variables que podrían ser significativas.

Además de esto, no se observan correlaciones fuertes entre las variables. Aun así, se ha decidido incorporar algunas interacciones específicas para intentar capturar posibles relaciones conjuntas que podrían influir en la conversión de manera más precisa, mejorando así el funcionamiento del modelo. Se han incluido interacciones entre distintas variables. En primer

lugar, se han interrelacionado *emailopens* y *emailclicks*, ya que ambas variables están relacionadas con el comportamiento del usuario frente a campañas de *email marketing*. La combinación de ambas variables, no solo abrir el correo sino también hacer clic en el enlace dentro del mensaje, podría reflejar un nivel más alto de compromiso, y, por lo tanto, una mayor probabilidad de conversión. En segundo lugar, *previous purchases* y *loyalty points*. Un cliente con muchas compras anteriores y, además, muchos puntos de fidelidad acumulados pueden tener una propensión mucho mayor a convertir gracias a su lealtad y familiaridad con la marca. En tercer lugar, se ha realizado una interacción entre *time on site* y *website visits*. Ambas variables miden la interacción del usuario con el sitio web. La interacción entre ambas podría ayudarnos a diferenciar patrones de comportamiento que puedan influir en la probabilidad de conversión. Por último, se ha incluido la interacción entre *campaign type conversion e email clicks*, intentando combinar una variable de características de la campaña con la interacción del usuario.

A continuación, se muestran los resultados del modelo explicativo tras las interacciones incorporadas:

Tabla 3: Resultados modelo explicativo de clasificación II

```

... Optimization terminated successfully.
      Current function value: 0.591529
      Iterations 5
      Logit Regression Results
-----
Dep. Variable:      Conversion      No. Observations:      1976
Model:              Logit          Df Residuals:          1950
Method:             MLE           Df Model:              25
Date:               Sun, 29 Mar 2026    Pseudo R-squ.:        0.1466
Time:               20:37:56          Log-Likelihood:       -1168.9
converged:          True           LL-Null:               -1369.7
covariance Type:   nonrobust      LLR p-value:          1.466e-69
-----
                    coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept           -0.4244      0.147      -2.884      0.004      -0.713      -0.136
Age                  0.0069      0.050      0.137      0.891      -0.092      0.105
Income               0.0837      0.050      1.675      0.094      -0.014      0.182
AdSpend              -0.0263      0.050      -0.523      0.601      -0.125      0.072
ClickThroughRate     0.3568      0.051      7.022      0.000      0.257      0.456
conversionRate       0.3173      0.051      6.243      0.000      0.218      0.417
WebsiteVisits        0.2323      0.051      4.591      0.000      0.133      0.332
PagesPerVisit        0.0005      0.050      0.010      0.992      -0.090      0.099
TimeOnSite           -0.0507      0.050      -1.017      0.309      -0.148      0.047
SocialShares         -0.0599      0.050      -1.188      0.235      -0.159      0.039
EmailOpens           0.3666      0.051      7.171      0.000      0.266      0.467
EmailClicks          0.4146      0.057      7.246      0.000      0.302      0.527
PreviousPurchases    0.2792      0.050      5.532      0.000      0.180      0.378
LoyaltyPoints        0.3264      0.051      6.440      0.000      0.227      0.426
CampaignChannel_PPC  0.2722      0.159      1.708      0.088      -0.040      0.585
CampaignChannel_Referral
0.2664      0.158      1.682      0.093      -0.044      0.577
CampaignChannel_SEO  0.0519      0.165      0.315      0.753      -0.271      0.375
Q("CampaignChannel_Social Media")
0.1755      0.160      1.098      0.272      -0.138      0.489
Gender_Male          0.0700      0.102      0.683      0.494      -0.131      0.271
CampaignType_Consideration
0.0963      0.136      0.707      0.480      -0.171      0.363
CampaignType_Conversion
0.9694      0.149      6.526      0.000      0.678      1.260
CampaignType_Retention
0.0528      0.139      0.380      0.704      -0.220      0.325
EmailOpens:EmailClicks
0.0312      0.051      0.612      0.541      -0.069      0.131
PreviousPurchases:LoyaltyPoints
-0.0302      0.051      -0.592      0.554      -0.130      0.070
CampaignType_Conversion:EmailClicks
-0.1160      0.126      -0.921      0.357      -0.363      0.131
WebsiteVisits:TimeOnSite
-0.0668      0.050      -1.346      0.178      -0.164      0.030
-----

```

Fuente: elaboración propia a partir de los resultados del dataset

En este segundo modelo se ha logrado un Pseudo R² del 14,66%, ligeramente superior al obtenido en el primer modelo. Tras esta mejora, vamos a analizar cada una de las variables para identificar cuáles resultan estadísticamente significativas y cuál es su efecto sobre la

conversión. Para ello, consideraremos como significativas aquellas variables cuyo p-valor sea inferior al 10%.

VARIABLES SIGNIFICATIVAS E INTERPRETACIÓN DE SU COEFICIENTE:

- Income: presenta un p-valor del 0,094, lo que la convierte en una variable significativa al 10%. Su coeficiente es positivo, lo que indica que, a mayores ingresos del individuo, mayor probabilidad hay de que este convierta.
- Click-Through Rate: presenta un p-valor próximo a 0, lo que la convierte en una variable altamente significativa. Su coeficiente es positivo, indicando que, a mayor tasa de clics, mayor es la probabilidad de que el usuario convierta.
- Conversion Rate: con un p-valor cercano a 0, esta variable es altamente significativa. Su coeficiente positivo sugiere que a medida que aumenta la tasa de conversión histórica, también aumenta la probabilidad de conversión del individuo.
- Website Visits: muestra un p-valor cercano a 0, lo que la convierte en altamente significativa. Su coeficiente positivo indica que un mayor número de visitas al sitio se asocia con una mayor probabilidad de conversión.
- Email Opens: tiene un p-valor en torno a 0 por lo que es altamente significativa. El coeficiente positivo refleja que cuanto más abre emails un usuario, mayor es su probabilidad de conversión.
- Email Clicks: presenta un p-valor cercano a 0, siendo altamente significativa. Su coeficiente positivo indica que los clics en emails aumentan la probabilidad de conversión.
- Previous Purchases: con un p-valor próximo a 0, es altamente significativa. Su coeficiente positivo sugiere que los usuarios con compras previas tienen más probabilidades de convertir nuevamente.
- Loyalty Points: muestra un p-valor en torno a 0, siendo altamente significativa. Su coeficiente positivo indica que los usuarios con más puntos de lealtad tienen mayor probabilidad de conversión.
- Campaign Channel PPC: presenta un p-valor de 0,088, lo que la convierte en significativa al 10%. Su coeficiente positivo indica que los usuarios alcanzados a través de campañas PPC tienen mayor probabilidad de conversión.
- Campaign Channel Referral: con un p-valor de 0,093, es significativa al 10%. Su coeficiente positivo refleja que los usuarios que llegan mediante referidos tienen una mayor probabilidad de convertir.

- Campaign Type Conversion: presenta un p-valor próximo a 0 siendo altamente significativa. Su coeficiente positivo indica que las campañas orientadas a la conversión aumentan la probabilidad de que un usuario convierta.

En conclusión, podemos ver cómo, el modelo explicativo desarrollado, nos permite identificar de manera clara algunos de los principales factores que influyen de manera significativa en la probabilidad de conversión de los clientes. El modelo destaca algunas variables clave relacionadas con la interacción del usuario con las campañas, como los clics, aperturas de emails o visitas a la web, así como con su comportamiento histórico, como el número de compras previas o puntos de fidelización, todas ellas con un efecto positivo sobre la conversión.

Podemos observar cómo los resultados obtenidos del modelo explicativo coinciden con las conclusiones que habíamos obtenido previamente tras el análisis de segmentación. De esta forma, podemos confirmar la relevancia de variables vinculadas tanto al *engagement* y comportamiento del usuario como a su relación previa con la marca como factores determinantes en el proceso de conversión.

Una vez comprendidos los principales factores que parecen determinantes de la conversión desde un enfoque explicativo, vamos a desarrollar un modelo predictivo que permita estimar de forma más precisa la probabilidad de conversión de cada cliente, con el objetivo de facilitar la toma de decisiones y optimizar las estrategias de marketing.

3.4. Modelo predictivo

Antes de desarrollar el modelo predictivo, es importante destacar la diferencia entre los modelos supervisados y no supervisados en aprendizaje automático. Por un lado, los modelos de aprendizaje supervisado son aquellos que utilizan datos previamente etiquetados para entrenar al modelo. Por otro lado, los modelos de aprendizaje no supervisado son aquellos que trabajan con datos sin etiquetar para identificar patrones o estructuras.

Los modelos predictivos, que pueden ser tanto de aprendizaje supervisado como no supervisado, son aquellos que, a partir de datos históricos, son capaces de identificar patrones y relaciones entre variables para predecir el comportamiento futuro de un fenómeno o individuo. Al igual que los modelos explicativos, estos pueden emplearse en problemas de clasificación o de regresión, dependiendo del tipo de variable. En nuestro caso, de la misma forma que en el apartado anterior, se va a trabajar con un modelo de clasificación, ya que la variable objetivo sigue siendo la misma: estimar la probabilidad de conversión de un usuario.

El modelo va a pretender determinar qué probabilidad hay de que un individuo realice una acción deseada, a partir de una serie de características tanto sociodemográficas, de comportamiento digital, como del tipo de campaña y características de la misma. Este tipo de modelos, al permitir calcular la probabilidad de que ocurra un evento específico, facilita la toma de decisiones basadas en datos y, consecuentemente, favorece la implementación de estrategias de marketing más efectivas.

Existen distintos tipos de modelos de aprendizaje automático; para este problema se ha decidido emplear un *Random Forest*.

El *Random Forest* es un algoritmo de aprendizaje automático supervisado no paramétrico que combina múltiples árboles de decisión para generar una predicción única y estable. Se utiliza tanto en problemas de clasificación como de regresión.

Cada árbol de decisión, también conocido por su nombre en inglés, *Decision Tree*, es también un algoritmo de aprendizaje supervisado no paramétrico que, al igual que *Random Forest*, puede aplicarse tanto en problemas de clasificación como de regresión. Los árboles de decisión clasifican los datos dividiendo repetidamente el conjunto de estos según los valores que tengan en determinadas variables de entrada. En cada paso el árbol evalúa todas las variables disponibles de la base de datos y elige la que mejor “separa” el conjunto. El objetivo es obtener una separación de los datos en una serie de subconjuntos lo más homogéneos posibles en términos de la variable objetivo.

Los árboles de decisión presentan una estructura jerárquica, similar a la de un árbol invertido, tal y como indica su nombre. Está formado por un nodo raíz, ramas, nodos internos y nodos hoja.

El nodo raíz es el primer nodo del árbol y representa la muestra completa de entrenamiento. En este punto se elige la variable que mejor separa las clases, utilizando los criterios de *Gini Impurity*, *Entropy* o *Information Gain*. Por otro lado, las ramas son aquellas líneas que conectan los nodos y representan las decisiones que se toman según los valores de las variables. Cada rama divide los datos en diferentes subgrupos. Los nodos internos son aquellos nodos de decisión situados entre el nodo raíz y los nodos hoja, que siguen dividiendo los datos según variables relevantes. Por último, los nodos hoja son los nodos finales del árbol, donde se establece la predicción final, ya sea la clase a la que pertenece el individuo en problemas de clasificación o el valor estimado de la variable objetivo en casos de regresión.

Como se ha mencionado al principio, el algoritmo de *Random Forest* es la combinación de múltiples árboles de decisión. Sin embargo, es importante tener en cuenta que este no consiste únicamente en construir varios árboles y juntarlos, sino que introduce dos mecanismos clave que mejoran de forma significativa el modelo: el *bagging* y la selección aleatoria de variables.

El *bagging* pertenece a los denominados métodos *ensemble*, que consisten en combinar múltiples modelos simples o débiles para crear un modelo predictivo más fuerte y robusto. A su vez, el término de *bagging* proviene de *Bootstrap aggregating*, que implica que, con el objetivo de contrarrestar la alta varianza que puede haber en un único árbol de decisión además de aumentar la precisión de la predicción, cada árbol del bosque se entrena a partir de una muestra diferente del conjunto de datos original. Esto implica que cada árbol analiza una versión ligeramente distinta de los datos, introduciendo de esta forma variabilidad entre ellos.

En segundo lugar, en cada nodo de cada árbol no se consideran todas las variables disponibles para realizar la división, sino únicamente un subconjunto aleatorio de ellas. Este proceso reduce la correlación entre los árboles y favorece que cada uno aprenda patrones diferentes. Una vez entrenados todos los árboles, el modelo combina sus predicciones para obtener un resultado final más robusto.

Se ha seleccionado el algoritmo de *Random Forest* para el desarrollo del modelo por diversos motivos. En primer lugar, este algoritmo es capaz de captar relaciones no lineales y complejas entre las variables, lo cual resulta especialmente relevante en el ámbito del marketing digital, donde la conversión de un usuario no depende de un único factor.

En segundo lugar, el *Random Forest*, al combinar múltiples árboles de decisión, reduce el error del modelo y mejora su precisión, ofreciendo en la mayoría de los casos un rendimiento superior al de un modelo individual.

Por otro lado, se trata de un algoritmo robusto frente al ruido que pueda haber en la base de datos, como clics accidentales, *scrolls* rápidos sin interacción real, tiempo en la página inflado porque el usuario haya dejado la pestaña abierta o sesiones de corta duración, habituales en entornos digitales. Asimismo, es capaz de manejar un gran número de variables sin necesidad de un preprocesamiento exhaustivo y permite además establecer la importancia de las variables, lo que facilita la identificación de los factores que influyen en la conversión y contribuye a la mejora de las estrategias de marketing.

El algoritmo de *Random Forest* cuenta con tres hiperparámetros principales que deben configurarse antes del entrenamiento: el tamaño del nodo (número mínimo de muestras por nodo hoja), la cantidad de árboles que compondrán el bosque y la cantidad de características muestreadas en cada división de los árboles.

Para establecer estos hiperparámetros, se ha utilizado la herramienta *GridSearchCV* de la librería *scikit-learn*, que, mediante validación cruzada, evalúa las distintas combinaciones de hiperparámetros estableciendo los valores de estos con los que se llegará a un mejor resultado. Se ha probado diferentes valores para el número de árboles (*n_estimators*), la profundidad máxima de los árboles (*max_depth*) y el número mínimo de muestras por nodo hoja (*min_samples_leaf*).

De esta forma, se han obtenido los hiperparámetros que maximizan la capacidad predictiva del modelo, que más adelante utilizaremos para entrenar el modelo. La mejor combinación de hiperparámetros que se encontró incluye un bosque de 100 árboles, cada uno con una profundidad máxima de 10 niveles y un mínimo de 10 muestras requeridas para formar un nodo hoja.

Una vez establecidos los hiperparámetros, se procedió al desarrollo del modelo predictivo, del cual obtuvimos los siguientes resultados:

Tabla 4: Resultados modelo predictivo Random Forest

```

Accuracy: 0.7133
ROC AUC: 0.7715

Classification Report:
      precision    recall  f1-score   support

     0       0.73      0.68      0.70      297
     1       0.70      0.75      0.72      296

 accuracy          0.71      0.71      0.71      593
 macro avg         0.71      0.71      0.71      593
 weighted avg      0.71      0.71      0.71      593

```

Fuente: elaboración propia a partir de los resultados del dataset

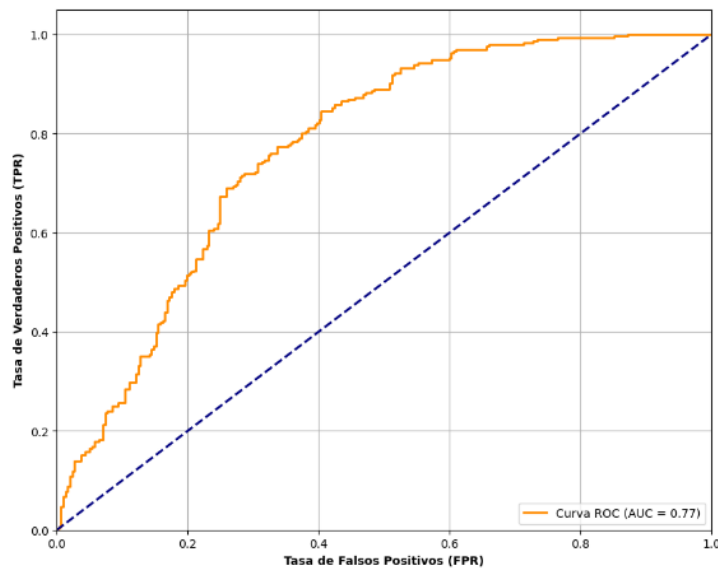
Antes de interpretar los resultados obtenidos, es importante comprender el significado de cada uno de los indicadores del modelo:

- **Accuracy:** Este indicador refleja la proporción de predicciones correctas de nuestro modelo. Se calcula sumando todas las predicciones que resultaron acertadas entre el total de predicciones realizadas.

- ROC AUC: Mide la capacidad del modelo para diferenciar entre las clases. Un valor más alto indica mejor capacidad para distinguir correctamente entre quienes convierten y quienes no. Valores cercanos a 0,5 indican que el modelo no parece mejor que el azar, es decir, que este tiene aproximadamente un 50% de probabilidades de acertar. Sin embargo, valores cercanos a 1 reflejan una alta capacidad del modelo para distinguir de forma adecuada entre quienes pertenecen a un grupo u otro.
- Precision (Precisión): La precisión para cada clase representa el porcentaje de predicciones de esa clase que fueron correctas. Es decir, de todos los usuarios que el modelo había predicho como que convertirían cuántos realmente convirtieron, y de todos los predichos como no convertidores, cuántos, efectivamente, no convirtieron.
- Recall (Sensibilidad): Indica la tasa de verdaderos positivos, es decir, qué proporción de los casos reales positivos fueron correctamente identificados por el modelo.
- F1-score: Combina precisión y sensibilidad, indicando cómo de bueno es el modelo haciendo predicciones teniendo en cuenta que este no se equivoque mucho diciendo que algo es positivo cuando no lo es (precisión) o que el modelo no sea capaz de predecir la mayoría de los casos positivos (sensibilidad). Cuanto más cerca de 1 esté este indicador, mejor será el modelo.

Una vez entendidos todos los indicadores, vamos a proceder a la interpretación de los resultados.

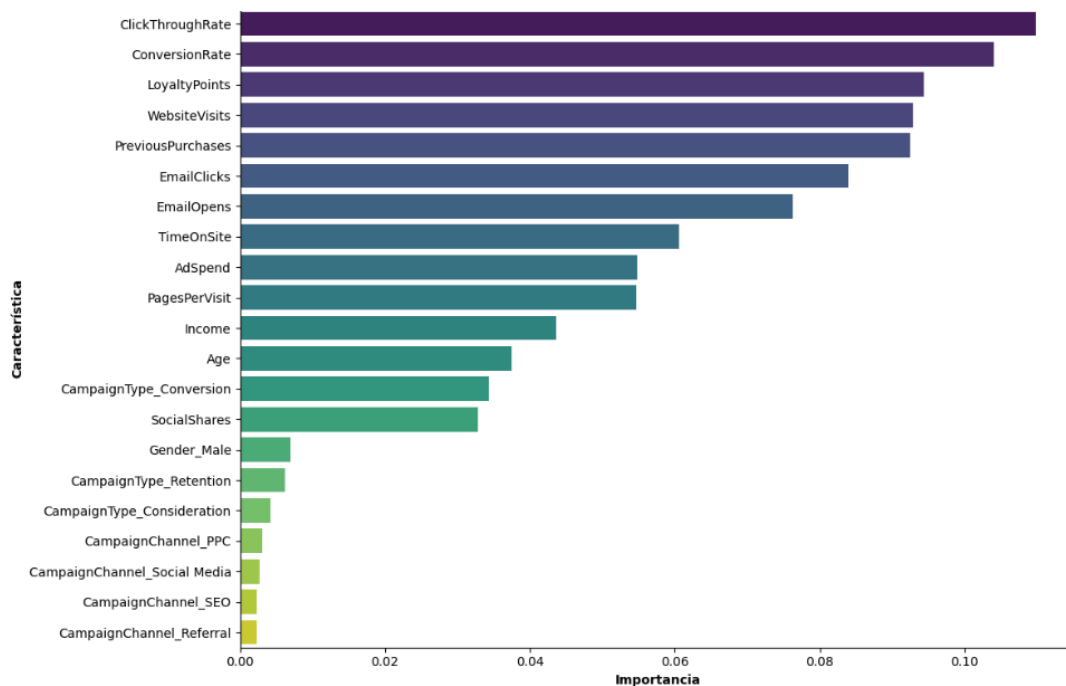
Podemos observar cómo nuestro modelo ha obtenido un *Accuracy* del 71,33%. Tal y como ya hemos explicado, esto significa que el modelo ha sido capaz de acertar más del 71% de los casos del conjunto de prueba. Podemos ver también cómo el modelo ha obtenido un ROC AUC de 77,15%. Este resultado indica que nuestro modelo tiene más de un 77% de probabilidades de acertar si un usuario convirtió o no. A continuación, se muestra la visualización de dicho indicador:

Ilustración 24: Curva Característica Operativa del Receptor (ROC)

Fuente: elaboración propia a partir de los resultados del dataset

La curva ROC (*Receiver Operating Characteristics*) representa la relación entre la tasa de verdaderos positivos (sensibilidad), y la tasa de falsos positivos. El AUC (*Area Under the Curve*) es el área que queda bajo esa curva. Por otro lado, la línea diagonal azul representa la capacidad predictiva de un modelo aleatorio. Cuanto más se acerque la curva hacia la esquina superior izquierda, significará que mejor es el modelo, ya supondría que está logrando tanto una alta tasa de verdaderos positivos como una baja tasa de falsos positivos.

Como hemos mencionado al principio del apartado, el modelo de Random Forest nos permite identificar la importancia de las diferentes variables a la hora de poder ir clasificando a cada uno de los individuos. A continuación, se muestra un gráfico de barras en el aparecen ordenadas de mayor a menor relevancia las distintas variables.

Ilustración 25: Importancia de las variables en el modelo predictivo Random Forest

Fuente: elaboración propia a partir de los resultados del dataset

El gráfico representa la importancia relativa de cada variable utilizada por el modelo para predecir la variable objetivo. Se observa que las variables más relevantes están relacionadas tanto con el comportamiento digital del usuario frente a la campaña como con el compromiso del cliente con la marca, así como con el historial y la actividad previa del individuo. Entre ellas, destacan *Click-Through Rate*, *Conversion Rate*, *Loyalty Points*, *Website Visits* y *Previous Purchases*.

A continuación, aparecen variables como *Email Clicks*, *Email Opens* y *Time On Site*, que también aportan información valiosa sobre la interacción digital y el nivel de compromiso del usuario. Además, dentro de las diez variables más importantes se encuentran *Ad Spend* y *Pages Per Visit*.

Por otro lado, las variables sociodemográficas, como *Income*, *Age* o *Gender_Male*, junto con aquellas relacionadas con el tipo de campaña y el canal utilizado, presentan una menor relevancia para el modelo. Esto indica que, en este caso y con este conjunto de datos, el comportamiento del usuario tiene un peso mucho mayor para la predicción que sus características demográficas o las especificidades de la campaña.

Una vez analizados los resultados del algoritmo y la importancia de cada variable dentro del modelo, vamos a desarrollar una matriz de confusión y a interpretar sus resultados.

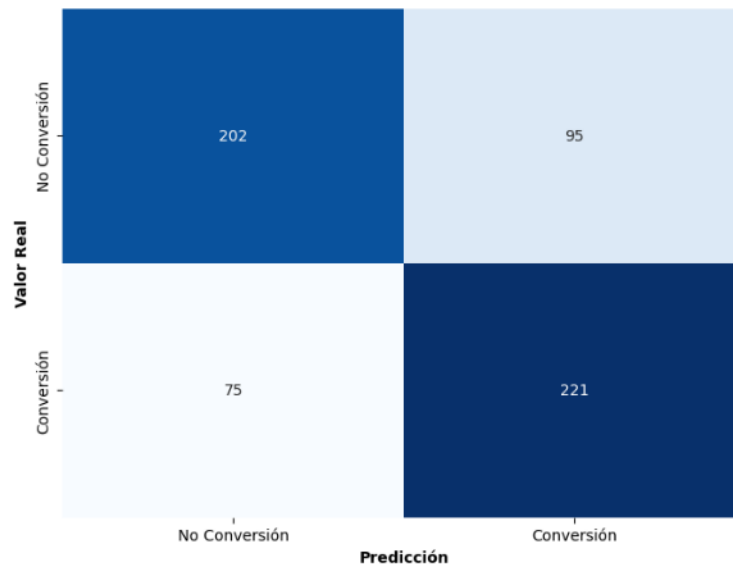
La matriz de confusión es una tabla que permite evaluar el rendimiento de un modelo de clasificación, mostrando cómo de acertadas o no han sido las predicciones que ha realizado el modelo, separadas por cada clase. En nuestro caso, recordamos que las clases son: 1 = convierte y 0 = no convierte.

La matriz refleja el número de casos en cada categoría:

- *True Negative (TN)*: predicciones de 0 que realmente eran 0. Representa a aquellos usuarios que fueron correctamente identificados como no conversores.
- *True Positive (TP)*: predicciones de 1 que realmente eran 1. Indica el número de individuos que fueron correctamente identificados como conversores.
- *False Negative (FN)*: predicciones de 0 que en realidad eran 1. Nos muestra la cantidad de usuarios que el modelo predijo como no conversores de forma incorrecta, es decir, que si habían convertido.
- *False Positive (FP)*: predicciones de 1 que en realidad eran 0. Representa el número de individuos que el modelo había predicho como conversores de forma errónea, ya que si que realmente no habían convertido.

Cuanto más altos sean los valores de True Negative y True Positive frente a los valores de false Negative y false Positive, significará que mejor capacidad de predicción está teniendo el modelo.

A continuación, se presenta la matriz de confusión de nuestro modelo

Ilustración 26: Resultados matriz de confusión Random Forest

Fuente: elaboración propia a partir de los resultados del dataset

En base a los resultados obtenidos en la matriz de confusión, podemos analizar cómo el modelo ha sido capaz de predecir correctamente a 221 usuarios como conversores, que efectivamente convirtieron, así como a 202 usuarios como no conversores, que realmente no convirtieron. Por otro lado, el modelo predijo incorrectamente a 75 usuarios como no conversores, cuando en realidad sí convirtieron, y a 95 usuarios como conversores, que en realidad no convirtieron.

Una vez desarrollado el apartado y analizados los resultados obtenidos, se puede concluir que se ha construido con éxito un modelo predictivo, en este caso, un Random Forest, que muestra una capacidad sólida para estimar la probabilidad de conversión de los usuarios. Con un *accuracy* superior al 71% y un ROC AUC cercano al 77%, el modelo no solo supera claramente el rendimiento esperado por azar, sino que también presenta una buena capacidad tanto de predicción como de discriminación entre usuarios que convertirán y aquellos que no.

El algoritmo Random Forest, permite capturar relaciones no lineales entre las variables, lo que resulta especialmente útil en un entorno de marketing digital, donde el comportamiento del usuario es multifactorial. Además, ofrece la ventaja de ser capaz de identificar qué variables son más relevantes en la predicción. En este caso, el hecho de que variables de comportamiento del usuario con la campaña como la interacción con la web, el historial de compras o el *engagement* con campañas, tengan mayor peso que las sociodemográficas sugiere que, si la base de datos fuese un registro real de una empresa y sus clientes, las estrategias de marketing

de dicha compañía deberían centrarse más en el comportamiento digital del usuario y sus interacciones con la campaña que en su perfil sociodemográfico.

En términos prácticos, este tipo de modelos constituye una herramienta clave para la toma de decisiones, ya que permite segmentar audiencias, optimizar campañas y orientar de forma más eficiente la inversión publicitaria.

3.5. Conclusiones y recomendaciones según los resultados obtenidos

Después de haber desarrollado todo el caso práctico y realizado un análisis exploratorio de la variable objetivo por segmentos, junto con un modelo explicativo que nos ha permitido identificar las variables determinantes para predecir la probabilidad de conversión de un usuario, y un modelo predictivo capaz de estimar la probabilidad de que un usuario convierta o no, podemos establecer una serie de conclusiones respaldadas por los resultados obtenidos.

A lo largo de todo el estudio y según los resultados de cada uno de los modelos, se ha observado que las variables relacionadas con el comportamiento digital y la interacción del usuario con la marca, tales como los clics, el número de visitas, la tasa de apertura de correos electrónicos, la cantidad de compras previas o los puntos de fidelización de cada cliente, son los factores que tienen un mayor peso a la hora de determinar la probabilidad de conversión de un individuo. Por otro lado, los resultados también indican que las características sociodemográficas son variables considerablemente menos relevantes, mostrando un impacto menor en estimar la probabilidad de que un cliente realice una conversión.

Es importante destacar que estas conclusiones se basan en el tratamiento de la base de datos utilizada, la cual, como se mencionó al inicio del trabajo, fue creada con fines educativos y no contiene información real de una empresa. Sin embargo, si se tratase de una base de datos real que recogiese la información de los clientes de una empresa, según las conclusiones obtenidas se podrían realizar una serie de sugerencias a la empresa.

En este sentido, las principales recomendaciones para la marca serían, en primer lugar, priorizar la recopilación y análisis de datos relacionados con el comportamiento digital, tales como las acciones dentro de las campañas y las visitas al sitio web, con el fin de enfocar mejor las estrategias de marketing. En segundo lugar, diseñar campañas personalizadas basadas en el nivel de compromiso y los patrones de interacción de los usuarios, en lugar de centrarse exclusivamente en variables sociodemográficas. Sería importante también que la empresa fuese capaz de emplear modelos predictivos, como el *Random Forest* realizado, para identificar anticipadamente a los usuarios con mayor probabilidad de convertir y dirigir hacia ellos los

esfuerzos y recursos. Por último, es fundamental que la empresa sea capaz de mantener un monitoreo constante que permita detectar cambios en el comportamiento del usuario y ajustar las estrategias de forma ágil y dinámica.

El marketing digital no consiste únicamente en monitorizar el comportamiento de los clientes y ampliar el alcance o impacto de las campañas. Su verdadera utilidad radica en ser capaces de identificar tanto las oportunidades como los retos a los que se enfrenta la marca, para minimizar sus posibles efectos negativos y maximizar las ventajas que el entorno digital ofrece. Solo a través de esta visión integral, las empresas pueden mantenerse competitivas y construir relaciones sólidas y duraderas con sus clientes.

4. Reflexión crítica:

Para finalizar el trabajo, me gustaría concluir con una reflexión crítica que incluye las principales conclusiones obtenidas de todo el estudio remarcando las ventajas y oportunidades que ofrece el marketing digital.

Como ya hemos visto, nos encontramos en una era profundamente marcada por la digitalización, que ha revolucionado tanto los medios de comunicación como el entorno empresarial. Este cambio ha transformado de forma radical los hábitos de consumo del mercado y, consecuentemente, la manera en que las empresas y marcas se relacionan y comunican con sus clientes y potenciales consumidores. La inmediatez de acceso a la información que disponemos sumado a la gran variedad de productos y servicios disponibles han generado un mercado mucho más dinámico y competitivo, que exige a las empresas desarrollar una alta capacidad de adaptación para no quedarse atrás.

En este nuevo escenario, tal y como hemos visto, el marketing digital ha surgido como una evolución natural del marketing tradicional, adaptándose a estos nuevos retos y desafíos y trayendo consigo importantes ventajas y oportunidades para las empresas. Gracias a esta evolución del marketing clásico, las marcas tienen la posibilidad y capacidad de segmentar con muchísima mayor precisión a la audiencia a la que se dirigen, además de ser capaces de monitorizar prácticamente en tiempo real la respuesta de los usuarios a las diferentes campañas publicitarias y estrategias comerciales. Esto permite a las marcas poder reaccionar con mucha mayor rapidez, ajustando sus estrategias según las demandas y necesidades del mercado. Todo

esto se traduce en campañas mucho más efectivas, desarrolladas y respaldadas por decisiones tomadas sobre datos reales.

Los formatos publicitarios también han evolucionado y se han adaptado a este nuevo entorno. Hemos pasado de mensajes unidireccionales y estáticos, propios medios tradicionales como televisión, radio o prensa, a formatos mucho más interactivos y personalizados como programas de fidelización, colaboraciones con *influencers* o anuncios en redes sociales. Estos formatos digitales no solo modifican la manera de conectar con el consumidor, sino que también facilitan una medición precisa del impacto de cada acción gracias a las diversas métricas disponibles. Medir con precisión cuántas personas ven un anuncio en una marquesina o prestan atención a uno transmitido por radio, aunque es posible, puede llegar a ser una tarea muy compleja. Sin embargo, con los nuevos formatos publicitarios digitales que permiten la interacción del usuario, es posible evaluar de manera exacta cuántos clics ha recibido una publicación, cuántas personas acceden a las webs a través de esos anuncios y qué tipo de interacciones generan.

Adicionalmente se ha destacado como la combinación de distintos tipos de formatos, *owned*, *paid* y *earned media* sigue siendo esencial para maximizar el alcance y la eficacia de las campañas. No basta con contratar a un tercero para que gestione la publicidad, ni con depender únicamente de potenciar tus propias redes y medios, ni confiar solo en el boca a boca o recomendaciones de usuarios. Las estrategias publicitarias más efectivas combinan estos tres enfoques para lograr los mejores resultados.

Para valorar el éxito de estas acciones, gracias al marketing digital, las empresas pueden llevar a cabo un seguimiento de los Indicadores Clave de Desempeño, mejor conocido como KPIs por sus siglas en inglés. Estos funcionan como herramientas clave que permiten a las compañías medir con exactitud el progreso, evolución y desempeño de las diferentes campañas que se estén llevando a cabo. Es importante seleccionar indicadores adecuados según el objetivo específico de la campaña, ya sea generar conciencia, fomentar la consideración del cliente, impulsar la conversión o promover la fidelización, etapas claramente definidas, como hemos visto en el *Customer Marketing Funnel*. Sin embargo, es importante tener en cuenta, tal y como refleja el modelo *flywheel*, que el recorrido del cliente no es lineal, y es fundamental ser consciente de que los consumidores navegan entre diferentes etapas y opciones, evaluando continuamente cuál satisface mejor sus necesidades.

Podemos observar cómo, aunque el marketing tradicional y digital comparten un mismo objetivo, guiar al cliente a lo largo de todo el proceso de venta y conseguir que este tome la decisión final de compra, el marketing digital se distingue por ofrecer una precisión mayor en la medición y segmentación, lo que permite optimizar recursos y obtener resultados superiores. Asimismo, proporciona una base sólida para tomar decisiones inteligentes, apoyadas en datos y modelos analíticos, y facilita la implementación de estrategias integrales que acompañan al cliente durante todo su recorrido, desde el primer contacto hasta la fidelización.

Entre las principales ventajas que presenta el marketing digital frente al marketing tradicional podemos destacar:

En primer lugar, el marketing digital ofrece a las empresas la posibilidad de segmentar sus audiencias de manera mucho más precisa y personalizada, dirigiendo mensajes específicos basados en datos sociodemográficos o comportamentales. En segundo lugar, permite a las marcas realizar mediciones y análisis en tiempo real, lo que facilita la adaptación y optimización continua de sus estrategias para mejorar el rendimiento de las campañas. En tercer lugar, el marketing digital brinda una mayor flexibilidad en la gestión de presupuestos, ya que es posible ajustar las inversiones y pagar generalmente solo por resultados efectivos, como clics o conversiones. Además, facilita una interacción directa y bidireccional con el consumidor a través de canales como redes sociales, correos electrónicos o páginas web, las empresas pueden mantener un diálogo constante con sus clientes, fomentando la fidelidad y mejorando su experiencia.

Estos beneficios hacen que el marketing digital sea indispensable para las empresas que desean mantenerse competitivas en un mercado, que, como ya hemos visto, es cada vez más digital, dinámico y exigente.

Estos avances han revolucionado la relación entre empresas y clientes, posibilitando conexiones más auténticas, personalizadas y efectivas, esenciales para prosperar en un mundo cada vez más digital y cambiante. Sin embargo, es muy importante tener en cuenta que el entorno digital está en constante evolución y exige a las empresas mantenerse ágiles y adaptarse a las necesidades y demandas del mercado para no quedar obsoletas. La irrupción de la inteligencia artificial, como nuevo motor de cambio, es un claro ejemplo de esta transformación continua. Ser conscientes de este contexto y saber aprovechar sus ventajas, así como enfrentar

los retos y desafíos que pueda suponer, es fundamental de cara a optimizar esfuerzos tanto empresariales como económicos y asegurar la relevancia y éxito en el futuro de la marca.

En definitiva, el marketing digital no es solo una herramienta más, sino un elemento estratégico imprescindible que permite a las empresas comprender mejor a sus clientes, ofrecer experiencias más relevantes y tomar decisiones fundamentadas en datos reales. Mantenerse al día con las innovaciones y adaptarse a las nuevas realidades del mercado es un aspecto clave para construir relaciones sólidas y sostenibles de las marcas con los consumidores, garantizando así un crecimiento continuo y competitivo en la era digital.

Declaración uso responsable de herramientas de IA

Por la presente, yo, Paula García Gonzalez, estudiante de Relaciones Internacionales & Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Marketing digital: optimización de campañas mediante analítica avanzada", 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:

Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.

1. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
2. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
4. Interpretador de código: Para realizar análisis de datos preliminares.
5. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
6. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
7. Generador previo de diagramas de flujo y contenido: Para esbozar diagramas iniciales.
8. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
9. Generador de datos sintéticos de prueba: Para la creación de conjuntos de datos ficticios.
10. Generador de problemas de ejemplo: Para ilustrar conceptos y técnicas.
11. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
12. 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: Abril 2026

Bibliografía:

AIML.com. (2023, 3 de octubre). *Why is Random Forest a non-linear model? Why does it result in non-linear decision boundaries?* Recuperado de: <https://aiml.com/why-is-random-forest-a-non-linear-model-why-does-it-result-in-non-linear-decision-boundaries/>

Arturo. (2025, 14 de marzo). *KPI vs métricas: ¿Cuál es la clave del éxito?* AprendeIndustrial.com. Recuperado de: <https://aprendeindustrial.com/kpi-vs-metricas/>

El Kharoua, R. (s. f.). *Predict Conversion in Digital Marketing Dataset.* Kaggle. <https://www.kaggle.com/datasets/rabieelkharoua/predict-conversion-in-digital-marketing-dataset>

Eurostat. (2024). *E-business integration.* In *Statistics Explained.* Recuperado de: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-business_integration

Flywheel en Inbound Marketing
SB Digital. (s. f.). *Flywheel.* Recuperado de <https://www.sb.digital/diccionario-inbound-marketing/flywheel>

GeeksforGeeks. (2025, 23 de julio). *What are the advantages and disadvantages of Random Forest?* GeeksforGeeks. Recuperado de: <https://www.geeksforgeeks.org/machine-learning/what-are-the-advantages-and-disadvantages-of-random-forest/>

Globant. (n.d.). *Full-funnel marketing.* Recuperado de: <https://www.globant.com/es/tech-terms/full-funnel-marketing>

Great Learning Editorial Team. (2025, 28 de abril). *Random Forest algorithm in machine learning.* Great Learning. Recuperado de: <https://www.mygreatlearning.com/blog/random-forest-algorithm/>

Han, E. (2024, 15 de febrero). *Paid vs. owned vs. earned media: What's the difference?* Harvard Business School Online. Recuperado de: <https://online.hbs.edu/blog/post/earned-vs-paid-media>

Kemp, S. (2024). *Digital 2024: Global overview report.* DataReportal. Recuperado de: <https://datareportal.com/reports/digital-2024-global-overview-report>

Kotler, P., Kartajaya, H., & Setiawan, I. (2017). *Marketing 4.0: Moving from traditional to digital.* Recuperado de: <https://archive.org/details/marketing40movin0000kotl/page/n2/mode/1up>

Llana, C. (s.f). Data in Communication. Universidad Pontificia de Comillas ICADE. <file:///C:/Users/paula/Downloads/lesson%203%20the%20funnel%20as%20the%20basis%20of%20the%20strategy.pdf>

Llana, C. (s.f). Data in Communication. Universidad Pontificia de Comillas ICADE. <file:///C:/Users/paula/Downloads/Lesson%205%20paid%20media%20.pdf>

Llana, C. (s.f). Data in Communication. Universidad Pontificia de Comillas ICADE. <file:///C:/Users/paula/Downloads/Topic%206%20Strategic%20management%20of%20search%20marketing%20.pdf>

Llana, C. (s.f). Data in Communication. Universidad Pontificia de Comillas ICADE. <file:///C:/Users/paula/Downloads/Topic%207%20Strategic%20management%20of%20social%20media.pdf>

Llana, C. (s.f). Data in Communication. Universidad Pontificia de Comillas ICADE. <file:///C:/Users/paula/Downloads/Lesson%202%20Digital%20media%20in%20corporate%20communication%20strategy.pdf>

McFadden's R-squared (logistic regression)
 Lee, S. (2025, 13 de marzo). *A comprehensive guide to McFadden's R-squared in logistic regression*. Number Analytics. Recuperado de <https://www.numberanalytics.com/blog/comprehensive-guide-mcfaddens-r-squared-logistic-regression>

Nash, B. (2025, 22 de noviembre). *Social media marketing statistics 2025: 91+ stats & insights*. Marketing LTB. Recuperado de: <https://marketingltb.com/blog/statistics/social-media-marketing-statistics/#:~:text=92%25%20of%20businesses%20use%20social,budgets%20in%20the%201st%20year>

Newman, N., Fletcher, R., Schulz, A., Andl, S., & Kleis Nielsen, R. (2025). *Digital News Report 2025*. Reuters Institute for the Study of Journalism. Recuperado de: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2025-06/Digital_News-Report_2025.pdf

Salesforce. (s. f.). *Marking 40 years of email marketing with Salesforce's Joanna Milliken*. Recuperado de: <https://www.salesforce.com/news/stories/marketing-40-years-of-email-marketing-with-salesforces-joanna-milliken/>

Social proof / prueba social en marketing
Polanco, K. (2025, 21 de julio). *¿Qué es social proof o prueba social en marketing? + Ejemplos*. Tiendanube Blog. Recuperado de <https://www.tiendanube.com/blog/social-proof/>

TOFU, MOFU y BOFU (embudo de marketing)
Santos, D. (2023, 7 de septiembre). *TOFU, MOFU y BOFU: qué son, ejemplos e implementación*. HubSpot Blog. Recuperado de <https://blog.hubspot.es/marketing/que-significan-tofu-mofu-y-bofu>

Tosas, G. (2014, 27 de octubre). *Los hitos en la historia del banner*. *La Vanguardia*. Recuperado de: <https://www.lavanguardia.com/tecnologia/internet/20141027/54417589877/hitos-historia-banner.html>

Zunzunegui, A. (2025, 14 de marzo). *Estadísticas y datos de email marketing en España*. Acumbamail. Recuperado de: <https://acumbamail.com/blog/estadisticas-email-marketing/#:~:text=%E2%80%9CDescarga%20directa%E2%80%9D.-,Nivel%20de%20satisfacci%C3%B3n%20con%20el%20correo%20electr%C3%B3nico%20como%20canal%20de,en%20su%20eficacia%20y%20rentabilidad>

Anexo

Anexo 01. Archivo código caso práctico

<https://drive.google.com/drive/folders/1mmOW0un1vWfFZv1WYjq-DEsvvM151TCE?usp=sharing>

Anexo 02: Implementación de One-Hot-Encoding para la conversión de variables categóricas

Ilustración 27: Implementación de One-Hot-Encoding para la conversión de variables categóricas

```
df_balanced = pd.get_dummies(df_balanced, columns=['CampaignChannel', 'Gender', 'CampaignType'], drop_first=True, dtype=int)
df_balanced.columns

... Index(['Age', 'Income', 'AdSpend', 'ClickThroughRate', 'ConversionRate',
         'WebsiteVisits', 'PagesPerVisit', 'TimeOnSite', 'SocialShares',
         'EmailOpens', 'EmailClicks', 'PreviousPurchases', 'LoyaltyPoints',
         'Conversion', 'CampaignChannel_PPC', 'CampaignChannel_Referral',
         'CampaignChannel_SEO', 'CampaignChannel_Social Media', 'Gender_Male',
         'CampaignType_Consideration', 'CampaignType_Conversion',
         'CampaignType_Retention'],
        dtype='object')
```

Fuente: elaboración propia a partir del dataset

Anexo 03: Eliminación de variables

Ilustración 28: Eliminación de variables

```
df_balanced = df_balanced.drop('AdvertisingPlatform', axis=1)
df_balanced = df_balanced.drop('AdvertisingTool', axis=1)
df_balanced = df_balanced.drop('CustomerID', axis=1)
df_balanced.columns

... Index(['Age', 'Gender', 'Income', 'CampaignChannel', 'CampaignType', 'AdSpend',
         'ClickThroughRate', 'ConversionRate', 'WebsiteVisits', 'PagesPerVisit',
         'TimeOnSite', 'SocialShares', 'EmailOpens', 'EmailClicks',
         'PreviousPurchases', 'LoyaltyPoints', 'Conversion'],
        dtype='object')
```

Fuente: elaboración propia a partir del dataset

Anexo 04: Unificación formato numérico

Ilustración 29: Unificación formato numérico

```
for col in ['AdSpend', 'PagesPerVisit', 'TimeOnSite']:
    df_balanced[col] = df_balanced[col].astype(str).str.replace('.', '', regex=False)
    df_balanced[col] = df_balanced[col].str.replace(',', '.', regex=False)
    df_balanced[col] = pd.to_numeric(df_balanced[col], errors='coerce')
```

Fuente: elaboración propia a partir del dataset

Anexo 05: estandarización de variables

Ilustración 30: estandarización de variables

```
from sklearn.preprocessing import StandardScaler
numeric_cols = [
    'Age', 'Income', 'AdSpend', 'ClickThroughRate', 'ConversionRate',
    'WebsiteVisits', 'PagesPerVisit', 'TimeOnSite', 'SocialShares',
    'EmailOpens', 'EmailClicks', 'PreviousPurchases', 'LoyaltyPoints'
]
scaler = StandardScaler()
df_balanced[numeric_cols] = scaler.fit_transform(df_balanced[numeric_cols])
```

Fuente: elaboración propia a partir del dataset

Anexo 06: Valores nulos

Ilustración 31: Búsqueda de valores nulos en las variables

```
print("Valores nulos en df_balanced:")  
print(df_balanced.isnull().sum())
```

```
... Valores nulos en df_balanced:  
Age                0  
Income             0  
AdSpend            0  
ClickThroughRate  0  
ConversionRate     0  
WebsiteVisits     0  
PagesPerVisit     0  
TimeOnSite        0  
SocialShares      0  
EmailOpens        0  
EmailClicks       0  
PreviousPurchases 0  
LoyaltyPoints     0  
Conversion         0  
CampaignChannel_PPC 0  
CampaignChannel_Referral 0  
CampaignChannel_SEO 0  
CampaignChannel_Social Media 0  
Gender_Male       0  
CampaignType_Consideration 0  
CampaignType_Conversion 0  
CampaignType_Retention 0  
dtype: int64
```

Fuente: elaboración propia a partir del dataset