



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

ADE

Facultad de Administración
y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Modelos Predictivos Para la Optimización de Estrategias
de Marketing en el Sector de la Moda en España

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Dalmau Gómez, José María

Tutor/a: Crespo Abril, Fortunato

CURSO ACADÉMICO: 2024/2025

RESUMEN

En los últimos años, el avance de las tecnologías digitales ha transformado radicalmente la forma en que las empresas interactúan con sus clientes, especialmente en sectores altamente visuales y emocionales como el de la moda.

Las empresas del sector moda en España se enfrentan a múltiples desafíos vinculados a la digitalización, la internacionalización y la necesidad de diferenciarse frente a una competencia cada vez más intensa. En este contexto, el marketing digital ha emergido como una herramienta clave, al permitir una comunicación directa, personalizada y continua con el consumidor final.

Este trabajo se enmarca en dicha realidad, en la intersección entre las necesidades prácticas de las empresas de moda y las oportunidades que ofrecen las nuevas tecnologías para optimizar sus estrategias de marketing digital.

El objetivo principal es demostrar que el análisis basado en datos reales del sector puede contribuir significativamente a optimizar campañas de marketing mediante la aplicación de modelos predictivos, como la regresión logística y los árboles de decisión.

La recopilación de información para el desarrollo de este Trabajo Fin de Grado se fundamenta en el uso combinado de fuentes primarias y secundarias, garantizando así una aproximación integral al objeto de estudio.

Se han consultado bases de datos académicas de prestigio como: Scopus, Google Scholar, Dialnet y ScienceDirect. Adicionalmente, se han utilizado informes elaborados por consultoras especializadas, asociaciones profesionales como ACME (Asociación de Creadores de Moda de España), e instituciones públicas como el ICEX o las Cámaras de Comercio, que proporcionan datos relevantes sobre el comportamiento digital de las empresas españolas.

El portal Modaes.es, referencia para el sector textil en España, también se ha considerado para la obtención de datos en este trabajo.

A través del estudio de indicadores clave de rendimiento (KPIs), tales como el CTR (click-through rate), el engagement, la tasa de conversión y el ROI, este trabajo ha buscado identificar patrones de comportamiento que permitan anticipar el éxito de una estrategia de marketing digital. Para ello, se implementarán modelos predictivos sustentados en técnicas estadísticas (regresión logística) y de machine learning (árboles de decisión), con el objetivo de predecir si una campaña tendrá una tasa conversión alta.

La precisión de estos modelos será evaluada mediante el uso de métricas como la precisión, la sensibilidad, la especificidad y el AUC en la curva ROC.

ABSTRACT

In recent years, the advancement of digital technologies has radically transformed the way companies interact with their customers, especially in highly visual and emotional sectors such as fashion.

Fashion companies in Spain face multiple challenges related to digitalization, internationalization, and the need to differentiate themselves in an increasingly competitive environment. In this context, digital marketing has emerged as a key tool, enabling direct, personalized, and continuous communication with the end consumer.

This project is framed within that reality, at the intersection between the practical needs of fashion companies and the opportunities offered by new technologies to optimize their digital marketing strategies.

The main objective is to prove that data-driven analysis based on real sector data can significantly contribute to improving marketing campaigns through the application of predictive models such as logistic regression and decision trees.

The information gathering process for the development of this Final Degree Project is based on a combined use of primary and secondary sources, ensuring a comprehensive approach to the object of study. Prestigious academic databases such as Scopus, Google Scholar, Dialnet, and ScienceDirect will be consulted.

Additionally, reports from specialized consulting firms, professional associations such as ACME (Asociación de Creadores de Moda de España), and public institutions such as ICEX or the Chambers of Commerce will be used, as they provide relevant data on the digital behavior of Spanish companies.

The portal Modaes.es, a benchmark for the textile sector in Spain, will also be considered for data collection.

Through the study of key performance indicators (KPIs), such as CTR (click-through rate), engagement, conversion rate, and ROI, this project aims to identify behavioral patterns that can help anticipate the success of a digital marketing strategy.

To this end, predictive models based on statistical techniques (logistic regression) and machine learning (decision trees) will be implemented, with the objective of predicting whether a campaign will achieve a high conversion rate.

The accuracy of these models will be evaluated through the use of metrics such as accuracy, sensitivity, specificity, and AUC in the ROC curve.

Resum

En els darrers anys, l'avanç de les tecnologies digitals ha transformat radicalment la manera com les empreses interactuen amb els seus clients, especialment en sectors altament visuals i emocionals com el de la moda.

Les empreses del sector moda a Espanya s'enfronten a múltiples reptes vinculats a la digitalització, la internacionalització i la necessitat de diferenciar-se davant d'una competència cada vegada més intensa. En aquest context, el màrqueting digital ha emergit com una eina clau, ja que permet una comunicació directa, personalitzada i contínua amb el consumidor final.

Aquest treball s'emmarca dins d'aquesta realitat, en la intersecció entre les necessitats pràctiques de les empreses de moda i les oportunitats que ofereixen les noves tecnologies per optimitzar les seues estratègies de màrqueting digital.

L'objectiu principal és demostrar que l'anàlisi basada en dades reals del sector pot contribuir de manera significativa a l'optimització de campanyes de màrqueting mitjançant l'aplicació de models predictius, com ara la regressió logística i els arbres de decisió.

La recopilació d'informació per al desenvolupament d'aquest Treball de Fi de Grau es fonamenta en l'ús combinat de fonts primàries i secundàries, garantint així una aproximació integral a l'objecte d'estudi.

S'han consultat bases de dades acadèmiques de prestigi com: Scopus, Google Scholar, Dialnet i ScienceDirect. A més, s'han utilitzat informes elaborats per consultores especialitzades, associacions professionals com ACME (Associació de Creators de Moda d'Espanya) i institucions públiques com l'ICEX o les Cambres de Comerç, que proporcionen dades rellevants sobre el comportament digital de les empreses espanyoles.

El portal Moda.es, referència per al sector tèxtil a Espanya, també s'ha tingut en compte per a l'obtenció de dades en aquest treball.

Mitjançant l'estudi d'indicadors clau de rendiment (KPIs), com ara el CTR (click-through rate), l'engagement, la taxa de conversió i el ROI, aquest treball ha tractat d'identificar patrons de comportament que permeten anticipar l'èxit d'una estratègia de màrqueting digital. Per això, s'han implementat models predictius sustentats en tècniques estadístiques (regressió logística) i de machine learning (arbres de decisió), amb l'objectiu de predir si una campanya tindrà una taxa de conversió alta.

La precisió d'aquests models serà avaluada mitjançant l'ús de mètriques com la precisió, la sensibilitat, l'especificitat i l'AUC en la corba ROC.

ÍNDICE

RESUMEN	2
ÍNDICE DE ILUSTRACIONES E IMÁGENES	7
1. Introducción	8
1.1 Contexto y justificación	8
1.2 Objetivos	9
2. Metodología	8
2.1 Fuentes de información	10
2.2 Diseño del estudio.....	10
3. Análisis y Resultados	17
3.1 Análisis Descriptivo	17
3.0.1 Seguidores	17
3.0.2 Engagement	18
3.0.3 Tipo de campaña	19
3.0.4 Región	20
3.0.5 Marca consolidada	21
3.0.6 Presupuesto	22
3.1.1 Distribución de campañas por conversión	23
3.1.2 Seguidores de Instagram y su relación con la conversión	24
3.1.3 Engagement y tasa de conversión	25
3.1.4 Tipo de campaña y tasa de conversión	26
3.1.5 Conversión por región	27
3.2 Modelo 1: Regresión logística	26
3.2.1 Objetivo y enfoque	26
3.2.2 Variables consideradas	27

3.2.3 Tabla resumen del ajuste	27
3.2.4 Resultados generales del modelo	28
3.2.5 Evaluación del Modelo 1 – Regresión Logística: Curva ROC y AUC.....	29
3.3 Modelo 2 – Árbol de Decisión	31
3.3.1 Objetivo y Enfoque	31
3.3.2 Variables consideradas	31
3.3.3 Tabla resumen del ajuste	32
3.3.4 Resultados del modelo	33
3.4 Comparación de modelos: Regresión Logística vs Árbol de Decisión	33
3.5 Resultados de los modelos	34
4. Línea de mejoras	37
5. Conclusiones	37
6. Bibliografía	39
6.1 Bibliografía (Análisis del trabajo)	39
6.2. Bibliografía (Base de Datos)	39
7. Anexo.....	41
7.1 Anexo I (ODS)	41
7.2 Anexo II (Base de Datos)	43
7.3 Anexo III (Script de códigos utilizados en R)	52

ÍNDICE DE ILUSTRACIONES E IMÁGENES

Figure 1: Distribución de Seguidores Instagram	18
Figure 2: Distribución de Engagement	19
Figure 3: Distribución de Tipo de Campaña	20
Figure 4: Distribución por Región	21
Figure 5: Distribución de Marca Famosa	22
Figure 6: Distribución de Presupuesto.....	23
Figure 7: Distribución de Campañas con alta conversión	24
Figure 8: Distribución de Seguidores según conversión alta	25
Figure 9: Engagement según conversión alta	26
Figure 10: Proporción de conversión alta por tipo de campaña	27
Figure 11: Proporción de conversión alta por Región.....	28
Figure 12: Tabla resumen modelo regresión logística	29
Figure 13: Curva ROC para predicción de conversión alta	33
Figure 14: Modelo árbol de decisión	34
Figure 15: Curva ROC comparativa	36

1.Introducción

1.1 Contexto y justificación

En los últimos años, el avance de las tecnologías digitales ha transformado radicalmente la forma en que las empresas interactúan con sus clientes, especialmente en sectores con una fuerte carga estética y emocional, como el de la moda. h

Las empresas del sector moda en España se enfrentan a múltiples desafíos relacionados con la digitalización, la internacionalización y la diferenciación frente a una competencia creciente. El marketing digital ha emergido como una herramienta clave para sortear estos retos, permitiendo una comunicación directa, personalizada y continua con el consumidor final.

Sin embargo, el uso estratégico del marketing digital no siempre es óptimo. Muchas pequeñas empresas carecen de personal especializado en análisis de datos, lo que les impide interpretar adecuadamente el rendimiento de sus campañas. En consecuencia, la toma de decisiones suele estar basada en la intuición o en experiencias previas, sin tener en cuenta el potencial de modelos predictivos para anticipar resultados y maximizar el retorno de la inversión (ROI).

El presente Trabajo de Fin de Grado se enmarca en el contexto de la intersección entre las necesidades prácticas de las empresas de moda y las posibilidades que brindan las nuevas tecnologías para optimizar las estrategias de marketing digital.

El objetivo principal es demostrar que el análisis basado en datos reales del sector puede contribuir significativamente a optimizar campañas de marketing mediante la aplicación de modelos predictivos, como la regresión logística y los árboles de decisión.

A través del estudio de indicadores clave de rendimiento (KPIs) como el CTR (click-through rate), el engagement, la tasa de conversión y el ROI, este trabajo busca establecer patrones de comportamiento que permitan anticipar el éxito de una estrategia de marketing digital. El análisis se basará en información recopilada de campañas reales desarrolladas por empresas de moda en redes sociales. Esta perspectiva permita identificar las estrategias más eficaces.

La elección del sector moda no es casual. Se trata de una industria altamente dependiente de la imagen, la marca y la conexión emocional con el consumidor, factores que encuentran su mejor expresión en los entornos digitales. Además, se trata de un sector que representa un peso importante en la economía española, con miles de pequeñas empresas que necesitan adaptarse al entorno digital para garantizar su supervivencia y crecimiento.

La urgencia de innovación y mejora en el sector lo posiciona como un contexto idóneo para el desarrollo del presente estudio.

1.2 Objetivos

Este Trabajo Fin de Grado se plantea con la intención de integrar conocimientos estadísticos, marketing digital en un único proyecto.

Con el propósito de aportar valor real al tejido empresarial del sector moda, mediante el desarrollo y validación de modelos que ayuden a las empresas a tomar decisiones más fundamentadas.

El objetivo general de esta investigación es desarrollar modelos estadísticos y de inteligencia artificial capaces de optimizar las estrategias de marketing digital de empresas del sector moda, a partir del análisis de datos reales y medibles. Para ello, se proponen los siguientes objetivos específicos:

- Realizar una revisión bibliográfica sobre las metodologías más utilizadas en marketing digital, análisis de KPIs y predicción de resultados mediante inteligencia artificial
- Identificar las principales plataformas digitales utilizadas por las empresas del sector moda y las métricas más relevantes para evaluar el impacto de las campañas.
- Diseñar una estructura de recogida de datos que permita integrar variables como tipo de campaña, canal, duración, presupuesto, objetivos, interacciones y resultados.
- Implementar modelos predictivos basados en técnicas estadísticas (regresión logística) y de machine learning (árboles de decisión)
- Evaluar la precisión de los modelos mediante métricas como la precisión, la sensibilidad, la especificidad y el AUC.
- Formular recomendaciones operativas que las empresas puedan aplicar de forma práctica

Estos objetivos intentaran demostrar que el análisis basado en datos puede transformar la gestión del marketing digital en empresas de moda, facilitando decisiones más inteligentes, rentables y sostenibles.

2. Metodología

2.1 Fuentes de información

La recopilación de información para el desarrollo de este Trabajo Fin de Grado se ha fundamentado en el uso combinado de fuentes primarias y secundarias, garantizando así una aproximación integral al objeto de estudio. Por un lado, se ha llevado a cabo una exhaustiva revisión bibliográfica en bases de datos académicas de prestigio como Scopus, Google Scholar, Dialnet y ScienceDirect.

Adicionalmente, se han consultado informes elaborados por consultoras especializadas, asociaciones profesionales como ACME (Asociación de Creadores de Moda de España), e instituciones públicas como el ICEX o las Cámaras de Comercio, que proporcionan datos relevantes sobre el comportamiento digital de las empresas españolas. También se han explorado fuentes como Modaes.es, portal de referencia del sector textil en España.

En cuanto a las fuentes primarias, el trabajo se ha centrado en la recopilación manual y sistemática de información procedente de la red social Instagram donde las empresas del sector suelen publicar resultados de campañas, promociones y estadísticas básicas como número de interacciones, tipo de contenido publicado y frecuencia de publicación.

Herramientas como Metricool, Hootsuite, SocialBlade y SimilarWeb han sido empleadas para obtener estimaciones de rendimiento digital, engagement y evolución de seguidores. Asimismo, se han utilizado plataformas como Informa D&B, Axesor y Kompass para obtener datos corporativos de las empresas objeto de estudio.

La combinación de estas fuentes ha permitido construir una base de datos representativa y detallada, que constituye el eje sobre el cual se aplicarán las técnicas de análisis estadístico y de inteligencia artificial descritas en los apartados siguientes.

2.2 Diseño del estudio

El diseño metodológico adoptado para este trabajo es de tipo cuantitativo. La investigación se ha estructurado en varias fases sucesivas, comenzando por la definición de las variables y su operacionalización, seguida de la recolección y procesamiento de los datos, y finalizando con el análisis estadístico y la validación de modelos predictivos.

La unidad de análisis está compuesta por campañas de marketing digital realizadas por empresas españolas del sector moda, incluyendo empresas de los subsectores textil, calzado y accesorios. La selección de las empresas se ha realizado mediante muestreo intencional, priorizando aquellas que presentan una presencia activa en redes sociales y plataformas digitales. Se ha trabajado con una muestra inicial de 50 empresas, con la posibilidad de ampliación a 200 según la disponibilidad de datos.

El uso de indicadores de rendimiento o KPIs (Key Performance Indicators) es fundamental para valorar la eficacia de las campañas de marketing digital. Estos indicadores permiten traducir la actividad digital en métricas objetivas que pueden ser comparadas, interpretadas y utilizadas para tomar decisiones estratégicas.

Los indicadores principales utilizados en este trabajo son los siguientes:

- **Seguidores:** Representa el tamaño de la audiencia potencial que una marca puede alcanzar de forma directa. Aunque no mide la calidad del vínculo, ofrece una referencia inicial sobre la visibilidad orgánica de la marca.
 - Tipo: Numérica discreta.
 - Unidad: Número absoluto (ej: 5.300 seguidores).
- **Engagement (%):** Indica el grado de interacción de los usuarios con los contenidos publicados (likes, comentarios, compartidos), expresado en porcentaje sobre el total de seguidores. Este indicador permite evaluar la capacidad de una campaña para generar interés y fomentar la participación.
 - Se calcula típicamente como:
$$\frac{(\text{Likes} + \text{Comentarios} + \text{Compartidos})}{\{\text{Seguidores}\}} * 100$$
 - Tipo: Numérica continua.
 - Unidad: Porcentaje (%).
- **Tipo de campaña:** se refiere al enfoque de comunicación utilizado (branding, promoción directa, colaboración con influencers, etc.). La elección del tipo de campaña determina en gran medida el tono, el mensaje y los objetivos.
 - Tipo: Categórica nominal.
 - Categorías posibles:
 1. Promocional directa (descuentos, rebajas)
 2. Branding (posicionamiento de marca)
 3. Colaboración con influencers
 4. Lanzamiento de producto
 5. Concurso/sorteo
- **Tasa de conversión (%):** Mide la proporción de usuarios que completan una acción definida como valiosa (compra, registro, etc.) respecto al total de usuarios alcanzados. Es uno de los indicadores más relevantes para medir el impacto real de una campaña.
 - Tipo: Numérica continua.
 - Unidad: Porcentaje (%).
- **Presupuesto (€):** Representa el nivel de inversión destinado a la campaña, lo que condiciona el alcance, la segmentación y el formato del contenido. Su análisis permite valorar la eficiencia de las acciones.

- **ROI estimado (%)**: Expresa la rentabilidad de la inversión realizada, calculada como la relación entre el beneficio generado y el gasto en marketing. Es fundamental para valorar el impacto económico de una campaña.
 - Porcentaje de retorno de la inversión sobre la campaña. $ROI = \{\text{Beneficio neto}\} / \{\text{Inversión}\} * 100$
 - Tipo: Numérica continua.
 - Unidad: Porcentaje (%)
- **Región (Comunidad Autónoma)**: Variable geográfica que permite analizar si existen diferencias territoriales significativas en la respuesta del público o en los resultados obtenidos
 - Tipo: Categórica nominal.
 - Categorías posibles: Andalucía, Cataluña, Comunidad Valenciana, Madrid, etc. (17 regiones españolas).
- **Conversión alta (Sí/No)**: Variable derivada que agrupa las campañas con tasa de conversión superior al umbral del 3%, lo que permite realizar análisis predictivos con modelos de clasificación.
 - Tipo: Categórica binaria.
 - Categorías:
 - ☐ Sí ($\geq 3\%$)
 - ☐ No ($< 3\%$)
- **Marca consolidada (Sí/No)**: Variable creada para reflejar si la marca ha alcanzado un umbral de visibilidad significativa, basado en su número de seguidores:
 - Tipo: Categórica binaria
 - Categorías:
 - ☐ Sí (si supera los 100.000 seguidores.)
 - ☐ No (si tiene menos de 100.000 seguidores.)
 - ☐ Esta variable pretende capturar un efecto adicional de reconocimiento de marca, asumiendo que superar cierto volumen de seguidores puede asociarse con una mayor notoriedad de marca y mayor capacidad de alcance en el entorno digital.

Estas variables han sido seleccionadas no solo por su relevancia práctica, sino también por un estudio a fondo sobre el análisis digital y evaluación de estrategias de marketing en el sector retail y moda. (Kim & Ko, 2012)

Tras una fase de análisis descriptivo, en la que se van a examinar en detalle la distribución de las variables y sus primeras relaciones con la variable objetivo, el siguiente paso del trabajo consiste en la construcción de modelos predictivos. El propósito de esta etapa es evaluar si es posible anticipar, a partir de determinadas características de las campañas, la probabilidad de alcanzar una tasa de conversión alta.

La naturaleza del problema es de clasificación binaria, ya que la variable respuesta indica únicamente dos posibles estados: éxito o no éxito de la campaña (definido por superar o no el umbral del 3% de conversión). Por tanto, se seleccionan metodologías orientadas a este tipo de problemas, que permitan trabajar tanto con variables numéricas como categóricas.

Se han implementado dos modelos que permiten abordar el problema desde perspectivas complementarias:

En primer lugar, se aplica un modelo de regresión logística binaria, ampliamente utilizado en contextos de predicción de eventos. Este modelo permite cuantificar la influencia de cada variable independiente sobre la probabilidad de éxito, ofreciendo resultados estadísticamente interpretables en forma de coeficientes, intervalos de confianza y odds ratios. Su principal fortaleza reside en la capacidad de estimar el peso individual de cada factor, aportando una visión global de cómo cada predictor afecta al resultado esperado.

En paralelo, se emplea un modelo de árbol de decisión. A diferencia de la regresión logística, esta técnica no requiere suposiciones sobre la forma de la relación entre las variables, sino que construye de manera automática una estructura jerárquica de decisiones. El árbol de decisión permite identificar reglas de clasificación concretas que segmentan el conjunto de campañas en función de las combinaciones de variables más relevantes. Este modelo aporta, además, una representación visual directa de los patrones de comportamiento detectados, facilitando la interpretación operativa de los resultados.

Ambos modelos se han desarrollado utilizando el mismo conjunto de variables predictoras identificadas durante el análisis previo: si la marca está consolidada, el engagement (%), el tipo de campaña, el presupuesto asignado y la región geográfica. Esta homogeneidad en las variables permite realizar una comparación objetiva de su rendimiento.

La aplicación de estos dos enfoques permite, por un lado, obtener una valoración estadística robusta de los factores implicados en el éxito de las campañas, y por otro, identificar perfiles concretos de campañas exitosas a través de las reglas generadas por el árbol de decisión. Esta combinación proporciona una base sólida tanto para el análisis académico como para su aplicación práctica en la toma de decisiones en marketing digital.

A continuación, se presentan ambos modelos con detalle, incluyendo la metodología empleada, las variables consideradas y los resultados obtenidos tras su implementación.

Modelo 1 – Regresión logística

- Objetivo: Predecir si una campaña tendrá una conversión alta (binaria).
- Variable respuesta (Y):
- Conversión alta
- Tipo: categórica binaria
- El modelo predice la probabilidad de que una campaña tenga una conversión alta. Si la probabilidad es superior al valor “0.5”, la campaña será clasificada como de conversión alta; en caso contrario será clasificada como no alta.
- Categorías:
 - Sí (tasa de conversión $\geq 3\%$)
 - No (tasa de conversión $< 3\%$)
- Variables explicativas (X):
 - Engagement (% de interacción)
 - Tipo de campaña (promocional, branding, influencer...)
 - Presupuesto (en euros)
 - Región (Comunidad Autónoma)
 - Marca Consolidada (binaria: Si > 100.000 seguidores; No < 100.000 seguidores)

Modelo 2 – Árbol de decisión (predicción de conversión alta)

- Objetivo: Predecir si una campaña de marketing tendrá una tasa de conversión alta (más del 3%) utilizando un árbol de decisión.
- Variable respuesta (Y):
- Conversión alta (Sí / No)
- Variables explicativas (X):
 - Engagement (%)
 - Tipo de campaña (categoría)
 - Presupuesto (€)
 - Comunidad Autónoma
 - Marca Consolidada (binaria: Si > 100000 ; No < 100000)

El árbol de decisión permite identificar qué factores tienen un peso relevante a la hora de determinar si una campaña será exitosa, proporcionando una herramienta de análisis visual e intuitiva para los responsables de marketing

Justificación: Estos modelos permiten anticipar si una campaña será exitosa en términos de conversión, lo cual es clave para decidir su diseño antes de lanzarla.

- ✓ A pesar de que inicialmente se consideró la variable Seguidores como un posible predictor relevante del éxito de las campañas de marketing, los análisis exploratorios realizados a lo largo del trabajo han revelado ciertas limitaciones, las cuales serán explicadas en la sección de “análisis descriptivo” que justifican su exclusión en la fase de modelización.

2.3 Técnicas de análisis

El análisis de los datos recopilados se ha realizado en varias etapas, comenzando con un análisis exploratorio para comprender la distribución de las variables, detectar valores atípicos, y establecer correlaciones iniciales. Para ello, se han empleado técnicas de análisis exploratorio de datos (EDA) como histogramas, diagramas de dispersión y de caja (boxplot).

Posteriormente, se han implementado dos técnicas principales de modelado predictivo: regresión logística y árboles de decisión. La regresión logística se ha utilizado para estimar la probabilidad de que una campaña digital supere determinados umbrales de éxito (por ejemplo, un CTR superior al 2%), en función de variables independientes como el tipo de canal, presupuesto y duración. Esta técnica ha sido seleccionada por su capacidad para manejar variables categóricas y continuas, así como por su interpretabilidad.

Los árboles de decisión, por su parte, se han aplicado para clasificar las campañas según su rentabilidad esperada, identificando qué combinaciones de factores conducen a mejores resultados. Esta técnica ha demostrado ser útil por su visualización intuitiva y capacidad para manejar interacciones no lineales entre variables.

Las principales métricas empleadas para medir la calidad de los modelos implementados serán:

- ❑ **Precisión (accuracy):** mide el porcentaje de predicciones correctas respecto al total de casos.
- ❑ **Sensibilidad (recall):** indica la proporción de campañas exitosas que han sido correctamente identificadas como tales por el modelo.
- ❑ **AUC (Área bajo la curva ROC):** indicador global que resume la capacidad del modelo para distinguir entre clases en distintos umbrales. Valores cercanos a 1 indican un modelo excelente, mientras que valores cercanos a 0,5 indican bajo poder predictivo.

Estas métricas permiten comparar modelos y seleccionar aquel que ofrezca un mejor equilibrio entre sensibilidad y especificidad, garantizando así una aplicación robusta en contextos reales.

2.4 Modelos estadísticos para la predicción del rendimiento de campañas

La capacidad de anticipar el resultado de una campaña antes de su lanzamiento representa una ventaja competitiva clave para cualquier empresa. Para ello, se recurre a modelos estadísticos que permiten identificar patrones y estimar probabilidades a partir de datos históricos. En este trabajo se han seleccionado dos enfoques principales.

Por un lado, se ha hecho un modelo de regresión logística. La regresión logística es un modelo estadístico utilizado para predecir la probabilidad de ocurrencia de un evento

binario a partir de un conjunto de variables independientes. A diferencia de los modelos lineales clásicos, la regresión logística transforma la variable dependiente a través de una función logística (sigmoide), lo que permite modelar probabilidades comprendidas entre 0 y 1.

Este tipo de modelo resulta especialmente adecuado cuando el objetivo del análisis es clasificar observaciones en dos categorías.

En este caso, campañas con alta o baja conversión y entender cómo influyen en dicha clasificación factores como el número de seguidores, el engagement o el tipo de campaña.

Una de sus principales fortalezas es su interpretabilidad: los coeficientes estimados permiten evaluar de forma individual el peso y la dirección del efecto de cada variable sobre la probabilidad de éxito de una campaña. Por ejemplo, un coeficiente positivo asociado al engagement indicaría que, a mayor nivel de interacción, mayor será la probabilidad de alcanzar una conversión superior al 3%.

En este trabajo, la regresión logística se ha utilizado para identificar qué características incrementan significativamente la probabilidad de que una campaña digital sea considerada exitosa. Esta información es clave para optimizar la asignación de recursos en futuras acciones de marketing, mejorando la toma de decisiones.

Por otro lado, los árboles de decisión son modelos no paramétricos que clasifican los datos mediante una estructura jerárquica basada en reglas de partición sucesivas. A partir de un conjunto de variables predictoras, el algoritmo divide los datos en subconjuntos cada vez más homogéneos, hasta llegar a una clasificación final.

Su principal ventaja es su facilidad de interpretación, ya que permiten visualizar de forma clara y lógica qué combinación de factores lleva a un resultado determinado. Además, cuantifican el peso que los diferentes factores e interacciones tienen en el cálculo de la probabilidad.

En este trabajo, los árboles permiten construir reglas como: “si una campaña se considera que está consolidada, con un engagement superior al 0,25% y se realiza en Cataluña, entonces tiene alta probabilidad de conversión”, lo cual resulta de gran utilidad práctica.

3. Análisis y Resultados

3.1 Análisis Descriptivo

Antes de la implementación de los modelos predictivos, se ha llevado a cabo un análisis descriptivo con el objetivo de explorar las características principales del conjunto de datos. Esta fase permite conocer mejor la distribución de las variables, identificar posibles patrones y establecer relaciones preliminares entre las distintas métricas del marketing digital analizadas.

Previamente al desarrollo del estudio, se identifican las siguientes variables de análisis:

- Seguidores (numérica discreta)
- Engagement (%) (numérica continua)
- Tipo de campaña (categórica nominal)
- Presupuesto (€) (numérica discreta)
- Región (categórica nominal)
- Conversión alta (variable dependiente binaria)
- Marca consolidada (variable dependiente binaria)

Primero vamos a evaluar las variables una a una, para después evaluarlas junto a la tasa de conversión ya que es lo que es el objetivo del trabajo.

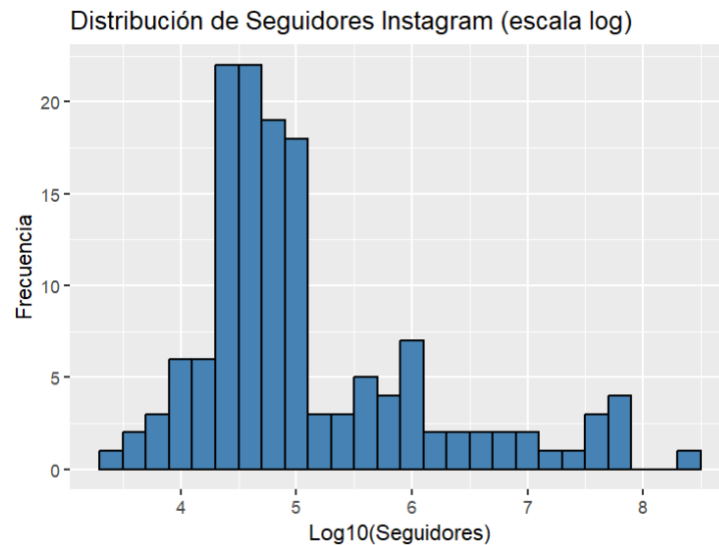
3.0.1 Seguidores

A partir del análisis de la variable “Seguidores Instagram”, se observa una distribución claramente asimétrica y sesgada hacia la derecha. Este comportamiento indica que la mayoría de las campañas analizadas pertenecen a marcas con un número relativamente bajo de seguidores, mientras que unas pocas presentan cifras extraordinariamente elevadas, lo que genera la presencia de valores atípicos (outliers) en el extremo superior de la distribución.

Según el resumen estadístico, el número mínimo de seguidores registrados es de 3.000, mientras que el máximo alcanza los 300 millones, una diferencia extremadamente amplia. La mediana se sitúa en 61.000, lo que implica que el 50% de las campañas tienen menos de esa cantidad, y solo el 25% superan los 350.000 seguidores (tercer cuartil). La media, sin embargo, asciende a más de 5,3 millones, lo que confirma la existencia de valores extremos que distorsionan la tendencia central cuando se utiliza este estadístico.

Esta diferencia tan significativa entre la mediana y la media pone de manifiesto la necesidad de aplicar transformaciones (como el logaritmo en base 10) para mejorar la interpretación de la variable y evitar que los outliers condicionen los modelos estadísticos. Además, estos resultados invitan a reflexionar sobre la noción de “influencia” en redes sociales: tener una audiencia masiva no implica necesariamente una mayor efectividad en términos de conversión, tal como se va a ver evidenciado en los modelos predictivos contruidos a lo largo del trabajo. Por ello, es fundamental no interpretar el número absoluto de seguidores como un indicador de éxito por sí solo, sino contextualizarlo junto a variables como el engagement o el tipo de campaña.

Figure 1 Distribución de Seguidores Instagram



```
> summary(datos$`Seguidores`)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3000	30000	61000	5326043	350000	300000000

3.0.2 Engagement

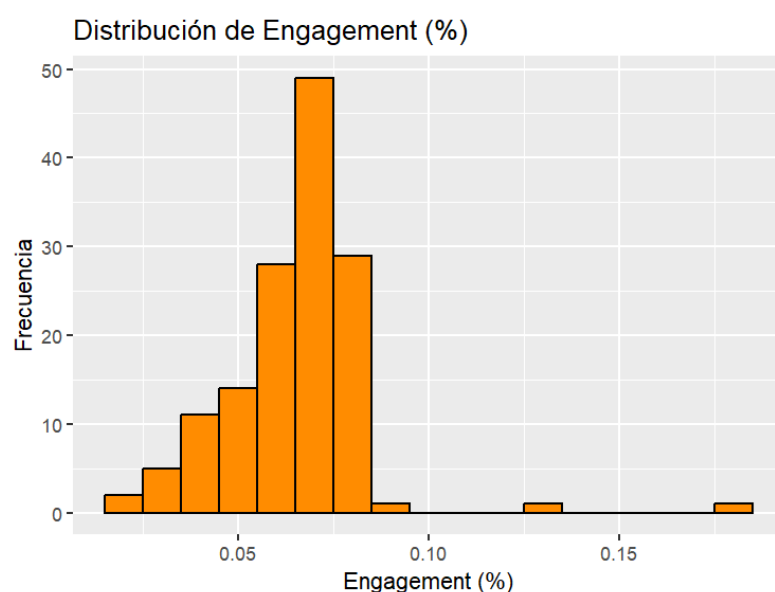
El engagement, entendido como la tasa de interacción que genera una publicación o campaña en redes sociales (medido normalmente como likes, comentarios, compartidos en relación al número de seguidores), es un indicador clave en el marketing digital.

- Distribución de la variable: El histograma muestra una distribución asimétrica hacia la derecha (sesgo positivo), donde la mayoría de los valores se concentran entre el 5.9% y el 7.4%, mientras que algunos valores más extremos alcanzan hasta el 18%, considerados como outliers (puntos fuera de lo habitual).
- Medidas de tendencia central:
 - La media es de 6.57%, muy cercana a la mediana (6.7%), lo cual indica que la distribución es moderadamente simétrica, aunque hay cierta dispersión debido a los valores extremos.
 - El primer cuartil (Q1) se encuentra en 5.9% y el tercer cuartil (Q3) en 7.4%, lo que significa que el 50% central de las campañas tienen un engagement entre estos valores.
- Valores extremos: El valor máximo de engagement registrado es de 18%, un dato muy por encima de la media. Este valor representa una campaña con altísima interacción, probablemente asociada a una acción muy viral o a una base de seguidores muy fidelizada.

Este análisis evidencia que la mayoría de las campañas logran un engagement entre el 6% y el 7%, lo cual puede considerarse un rango de rendimiento “normal” o promedio dentro del dataset. Los valores por debajo del 5% podrían interpretarse como campañas con escasa interacción, mientras que aquellos que superan el 10% reflejan un rendimiento excepcional.

Desde el punto de vista estratégico, esto confirma que el engagement es una variable con alta variabilidad, y por tanto, útil en los modelos predictivos como variable explicativa para anticipar la conversión alta.

Figure 2 Distribución de Engagement



```
> summary(datos$Engagement)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0210  0.0590  0.0670  0.0657  0.0740  0.1800
```

3.0.3 Tipo de campaña

La variable Tipo de campaña, de naturaleza categórica nominal, recoge las distintas estrategias promocionales implementadas por las marcas de moda para sus acciones de marketing digital. Entre las categorías presentes se incluyen: Branding, Colaboración con influencers, Concurso/sorteo, Lanzamiento de producto y Promocional directa.

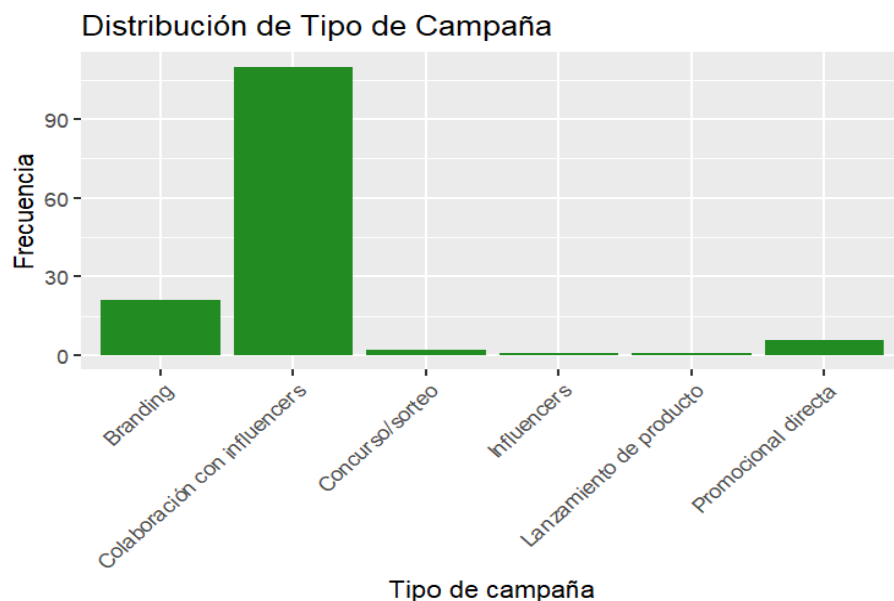
El gráfico muestra de forma clara una distribución altamente desigual. La categoría Colaboración con influencers domina el conjunto de datos con una frecuencia abrumadora, concentrando más del 60% de las campañas analizadas. Esta predominancia sugiere que las marcas del sector moda apuestan fuertemente por esta

estrategia, probablemente por su capacidad para generar notoriedad, engagement y confianza a través de figuras prescriptoras relevantes en redes sociales.

A cierta distancia, Branding se posiciona como la segunda estrategia más utilizada, con una representación moderada. El resto de categorías (Concurso/sorteo, Influencers, Lanzamiento de producto y Promocional directa) presentan una frecuencia muy baja, en algunos casos próxima a cero.

En conclusión, la variable Tipo de campaña revela una fuerte concentración estratégica en torno a las colaboraciones con influencers, mientras que otras modalidades publicitarias se emplean de forma marginal o residual dentro del universo de datos analizado.

Figure 3 Distribución de Tipo de Campaña



3.0.4 Región

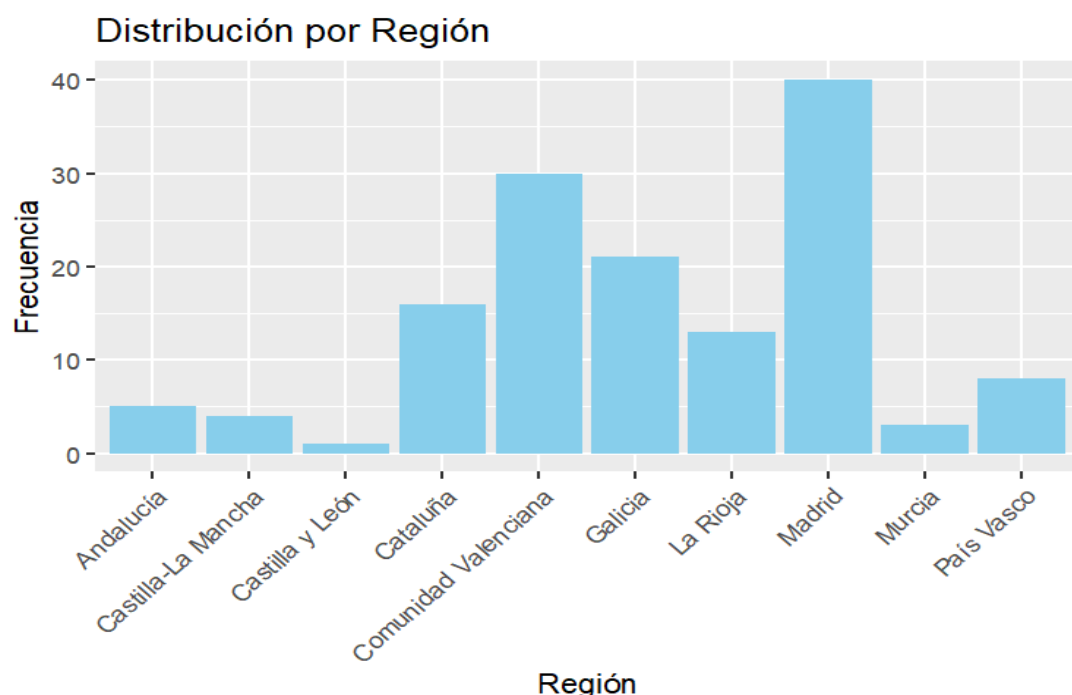
En la Figura se representa la distribución geográfica de las campañas analizadas, clasificadas por comunidad autónoma. Se observa una alta concentración de campañas en la Comunidad de Madrid, que lidera con 40 campañas aproximadamente, seguida de la Comunidad Valenciana y Galicia, con una presencia también notable.

Regiones como Cataluña y La Rioja presentan una participación intermedia, mientras que otras comunidades como Andalucía, Castilla-La Mancha, Murcia o País Vasco muestran una representación más reducida. Castilla y León, en particular, apenas aparece representada con una o dos campañas en el conjunto de datos.

Esta distribución desigual puede deberse a múltiples factores, como la concentración empresarial del sector moda en ciertas regiones, el grado de desarrollo del ecosistema digital o las diferencias en inversión en marketing entre comunidades. Asimismo, esta disparidad regional puede tener implicaciones sobre el análisis de resultados, ya que la baja frecuencia de observaciones en algunas zonas podría limitar la capacidad para detectar efectos estadísticamente significativos relacionados con la variable “Región”.

En consecuencia, al interpretar los modelos predictivos que incluyen esta variable, debe considerarse la posible infrarrepresentación de ciertos territorios, lo que podría introducir cierto sesgo o limitar la generalización de los resultados a nivel nacional.

Figure 4 Distribución por Región



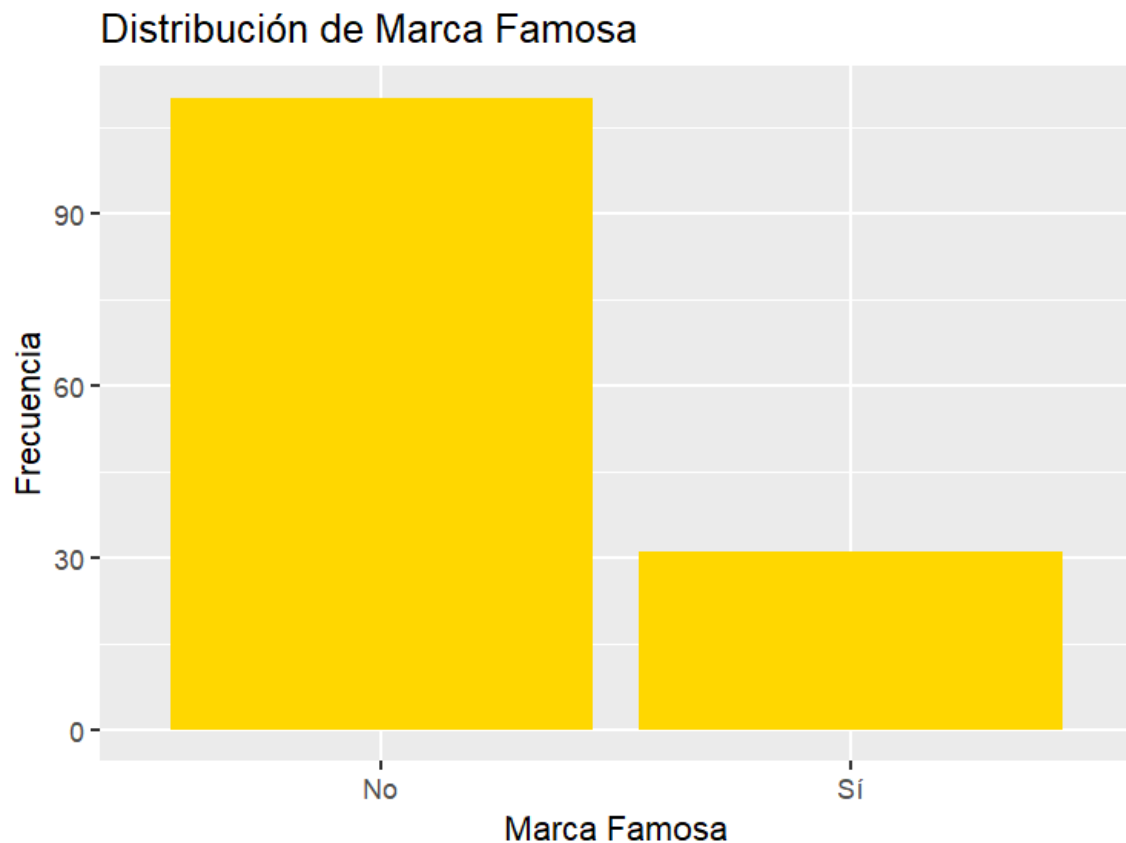
3.0.5 Marca consolidada

La variable Marca Famosa, que indica si la campaña pertenece a una marca reconocida por el público general, muestra una distribución claramente asimétrica. Tal como se observa en la figura, aproximadamente dos tercios de las campañas analizadas corresponden a marcas no consideradas famosas, mientras que tan solo un tercio está asociado a marcas con alto reconocimiento.

Este desbalance sugiere que, aunque las marcas consolidadas siguen participando activamente en el ámbito digital, una gran parte de las campañas proviene de empresas emergentes o con menor notoriedad pública. Esto permite plantear interrogantes sobre la relación entre notoriedad de marca y efectividad de las campañas, así como evaluar si las marcas menos conocidas logran resultados similares o incluso superiores mediante

estrategias más precisas o creativas. En los modelos predictivos, esta variable será clave para analizar si la fama influye directamente en la probabilidad de conversión.

Figure 5 Distribución de Marca Famosa



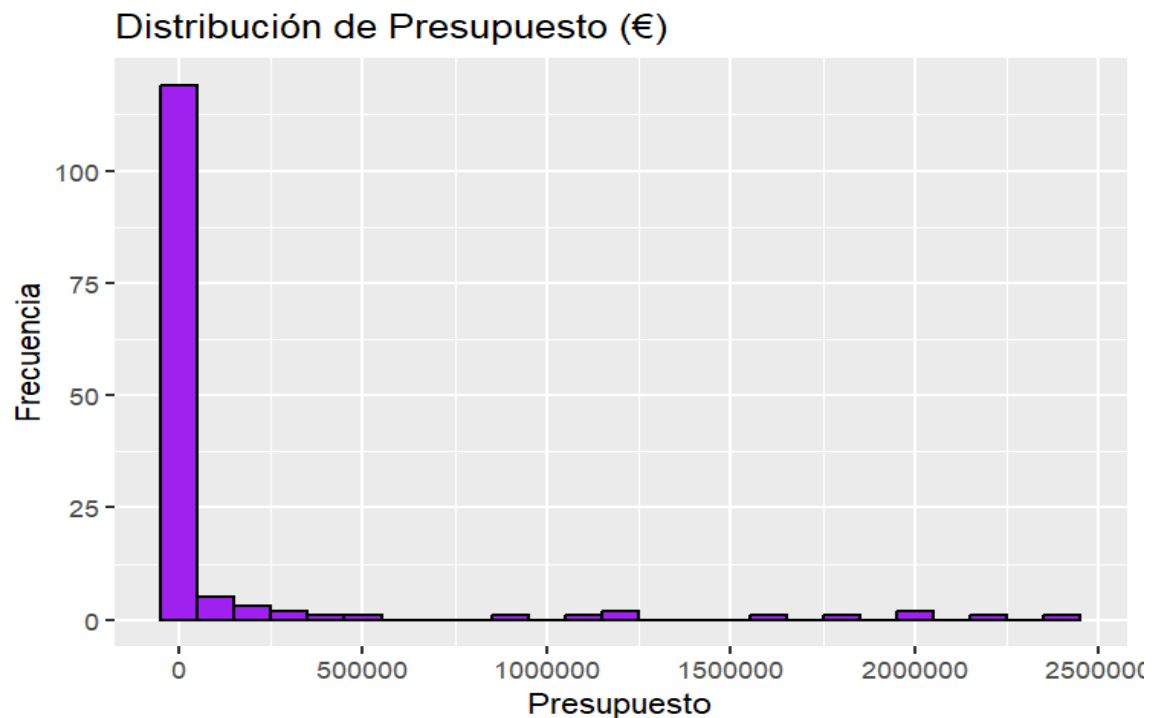
3.0.6 Presupuesto

El análisis de la variable presupuestaria revela una distribución altamente asimétrica, con una gran concentración de campañas en el tramo de menor inversión. Tal como se observa en la Figura correspondiente, la mayoría de las campañas presentan presupuestos comprendidos entre 2.100 € y 20.100 €, lo cual coincide con el rango intercuartílico ($Q1 = 7.700$ €, $Q3 = 20.100$ €). Sin embargo, existen algunos valores extremos que elevan considerablemente la media (145.719 €) respecto a la mediana (14.100 €), indicando la presencia de campañas de alto presupuesto que actúan como outliers.

Este sesgo positivo sugiere que, si bien algunas empresas del sector moda invierten de forma significativa en campañas digitales, la mayoría opera con presupuestos modestos. Desde el punto de vista del análisis predictivo, esta dispersión plantea un reto adicional al ajustar modelos estadísticos, ya que la variabilidad en la inversión no siempre se traduce en mayor conversión. Este hecho se ve reflejado en los resultados del modelo de

regresión logística, donde la “variable presupuesto” no mostró un efecto estadísticamente significativo sobre la probabilidad de alcanzar una alta conversión.

Figure 6 Distribución de Presupuesto



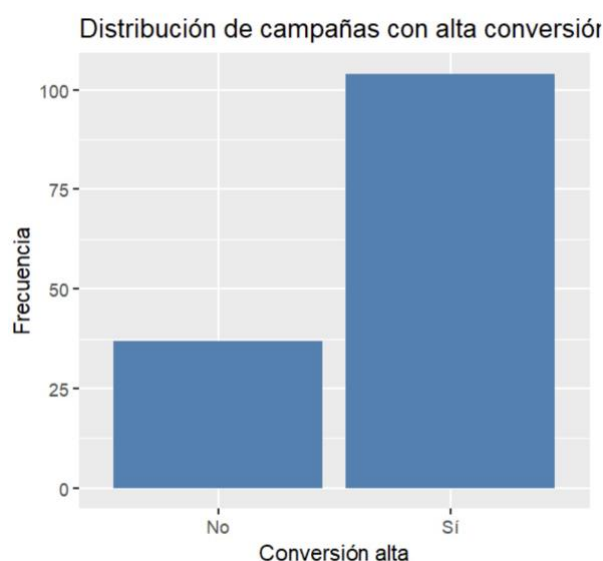
```
> summary(datos$Presupuesto)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2100    7700   14100 145719   20100 2417200
```

3.1.1 Distribución de campañas por conversión

La variable dependiente principal de este estudio es la conversión alta, una variable categórica binaria que indica si una campaña ha superado el umbral del 3% de conversión. En el gráfico de barras (Figura 7), se observa que, de las 150 campañas analizadas, una mayoría significativa (alrededor del 70%) corresponde a campañas con alta conversión, mientras que aproximadamente el 30% no alcanzan dicho umbral.

Este desbalance inicial sugiere que existen determinados factores que podrían estar condicionando el éxito de las campañas. A su vez, nos plantea un escenario adecuado para la aplicación de modelos de clasificación binaria como la regresión logística o los árboles de decisión, cuya precisión dependerá en parte de cómo se comporten las variables explicativas respecto a esta variable objetivo.

Figure 7: Distribución de Campañas con alta conversión



3.1.2 Seguidores de Instagram y su relación con la conversión

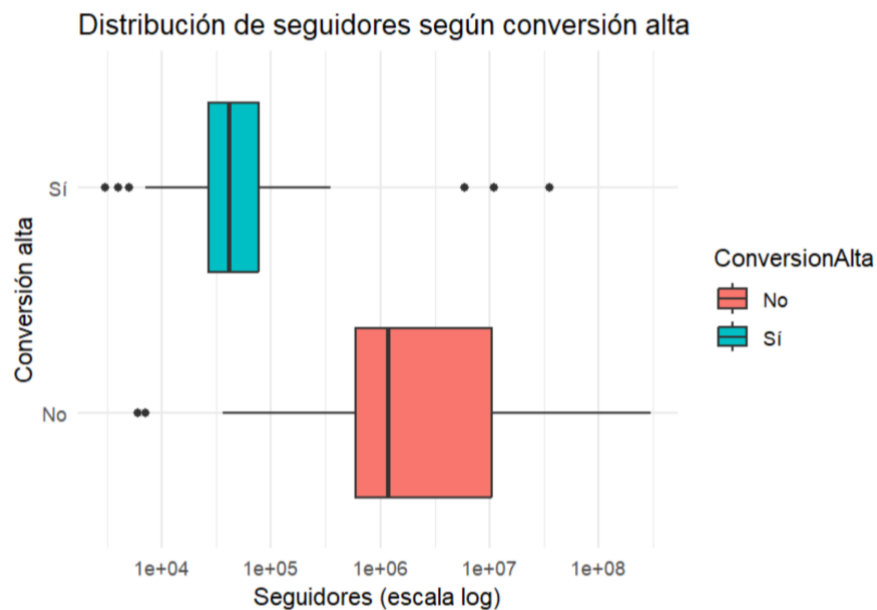
Uno de los primeros análisis se centró en la variable Seguidores Instagram, que actúa como proxy del tamaño de la audiencia potencial alcanzada por cada marca. Para representar gráficamente esta variable, se utilizó un boxplot con escala logarítmica dada la presencia de valores extremadamente altos (outliers), que podrían distorsionar la interpretación visual si se usara una escala lineal.

Como se aprecia en la Figura 8, las campañas sin conversión alta tienden a estar asociadas a marcas con un número significativamente mayor de seguidores. Este hallazgo resulta contraintuitivo, pues en principio podría esperarse que una mayor audiencia condujera a una mayor conversión. Sin embargo, la media de seguidores en campañas sin alta conversión asciende a más de 18 millones, frente a apenas 563 mil en las que sí lograron una alta conversión.

Esta paradoja puede explicarse por diversos factores: las grandes marcas suelen tener audiencias más dispersas y menos segmentadas, lo que puede reducir la efectividad de una campaña específica. Asimismo, la conversión depende no solo del tamaño de la audiencia, sino también de la calidad del vínculo marca-usuario.

En modelos previos, esta variable mostraba una fuerza discriminatoria desproporcionada, que podía enmascarar el efecto real de otras variables más relevantes, como el tipo de campaña o el engagement. Por ende, se ha optado por no incluir la variable Seguidores en los modelos predictivos finales, con el objetivo de obtener resultados más equilibrados, interpretables y útiles para la toma de decisiones estratégicas.

Figure 8 Distribución de Seguidores según conversión alta



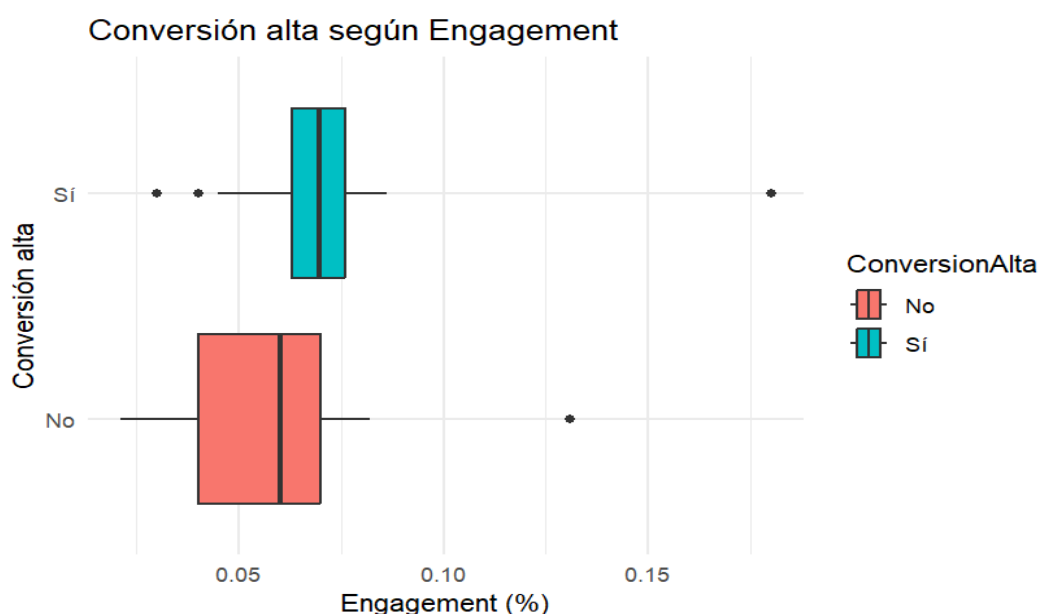
3.1.3 Engagement y tasa de conversión

En contraste con los seguidores, el gráfico de engagement muestra diferencias más significativas.

Las campañas con alta conversión presentan, en promedio, un nivel de interacción más elevado. Esta diferencia sugiere que la participación activa de la audiencia, medida a través de likes, comentarios y compartidos, podría estar más estrechamente relacionada con el éxito de una campaña que el mero alcance potencial. Estos hallazgos son consistentes con investigaciones anteriores que vinculan el engagement con la eficacia publicitaria en redes sociales (Ashley & Tuten, 2015).

Tal como muestra la Figura 9, el engagement es visiblemente más alto en campañas con alta conversión. Este resultado es coherente con lo planteado por investigadores como Chaffey y Ellis-Chadwick (2019), quienes argumentan que el engagement es un predictor indirecto de conversión, dado que refleja la calidad de la interacción y el interés genuino del público por los contenidos.

Figure 9: Engagement según conversión alta



```
> summary(datos$Engagement)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Sí	0.0350	0.0550	0.0650	0.0650	0.0750	0.1800
No	0.0210	0.0450	0.0650	0.0630	0.0750	0.1400

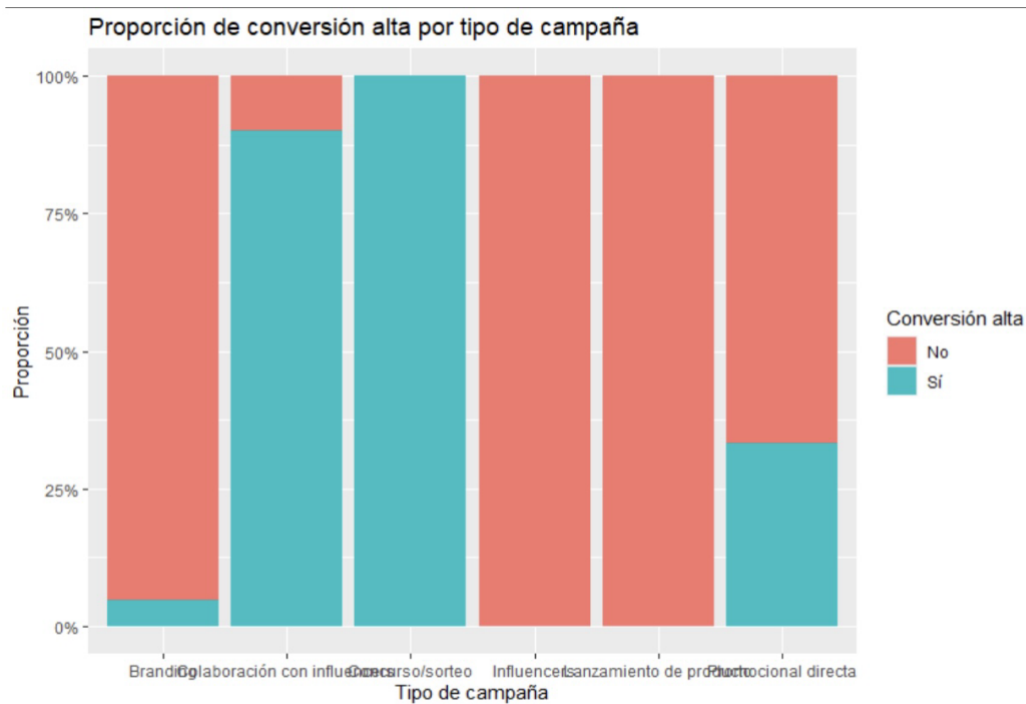
3.1.4 Tipo de campaña y tasa de conversión

En cuanto al tipo de campaña, los resultados indican que las colaboraciones con influencers, los concursos y sorteos son las que presentan mayor proporción de campañas con alta conversión (Figura 10). En contraste, las campañas de branding, que suelen centrarse en posicionamiento de marca y notoriedad, muestran un rendimiento inferior en términos de conversión.

Este patrón respalda la hipótesis de que las campañas que incorporan dinámicas participativas o credibilidad externa (a través de influencers) tienden a generar mejores resultados en términos de acción del usuario. (De Veirman et al., 2017).

Este hallazgo respalda la idea de que las campañas con un enfoque más táctico y orientado a la acción (como promociones o sorteos) son más efectivas en generar respuesta inmediata del consumidor, mientras que las campañas de branding requieren un horizonte temporal más amplio para traducirse en conversiones.

Figure 10: Proporción de conversión alta por tipo de campaña

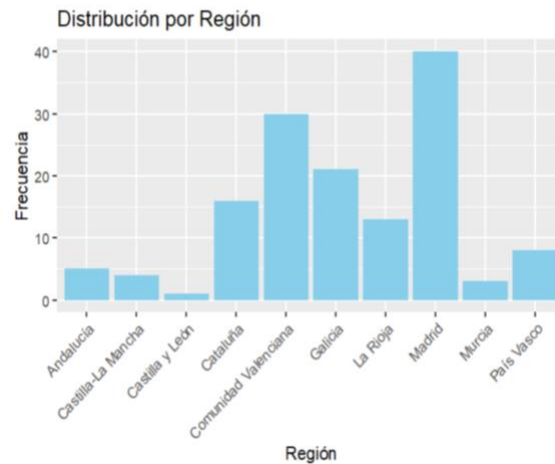


3.1.6 Conversión según región

Por último, se analizó la variable región para observar posibles diferencias geográficas en el rendimiento de las campañas. La Figura 11 revela que comunidades autónomas como la Comunidad Valenciana, País Vasco y Cataluña concentran un alto porcentaje de campañas exitosas. En cambio, regiones como Madrid o Galicia presentan mayor proporción de campañas que no superan el 3% de conversión.

Estas diferencias podrían deberse tanto a factores sociodemográficos como al perfil digital de los consumidores en cada zona. Investigaciones anteriores han sugerido que la madurez digital, la penetración de internet y el estilo de vida pueden influir significativamente en el rendimiento del marketing digital regional (Digital Economy and Society Index, 2022).

Figure 11: Proporción de conversión alta por Región



3.2 Modelo 1: Regresión logística

3.2.1 Objetivo y enfoque

La regresión logística es un modelo de clasificación binaria que permite estimar la probabilidad de que ocurra un evento (en este caso, una conversión alta) en función de variables explicativas tanto cuantitativas como cualitativas. Se trata de una técnica ampliamente utilizada en ciencias sociales y marketing por su capacidad para trabajar con variables categóricas y generar resultados interpretables.

Este modelo permite no solo predecir el resultado de una campaña, sino también identificar qué variables tienen mayor influencia en su éxito. El análisis de los coeficientes permite comprender, por ejemplo, si el presupuesto tiene un impacto positivo sobre la conversión o si ciertos tipos de campaña presentan una mayor probabilidad de éxito.

El primer modelo propuesto es una regresión logística binaria, cuyo propósito es estimar la probabilidad de que una campaña de marketing digital tenga una tasa de conversión superior al 3%. La regresión logística se considera una técnica adecuada cuando la variable dependiente es categórica binaria (en este caso, Sí o No), y permite interpretar la influencia relativa de cada variable independiente sobre la probabilidad de conversión.

3.2.2 Variables consideradas

- Variable dependiente (Y): Conversión alta (Sí / No)
- Variables independientes (X):
 - Engagement (numérica continua, en %)
 - Tipo de campaña (branding, promocional, influencer...)
 - Canal digital (Instagram, Facebook, TikTok...)
 - Presupuesto (numérica continua, en euros)
 - Región (categórica)

- Marca Consolidada(binaria: Si>100.000 seguidores; No< 100.000 seguidores)

3.2.3 Tabla resumen del ajuste

Figure 12: Tabla resumen modelo regresión logística

```
Call:
glm(formula = ConversionAlta ~ Engagement + TipoCampaña + Presupuesto +
    Region + MarcaFamosa, family = binomial, data = datos)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.117e+01  4.888e+03  0.002  0.9982
Engagement   4.642e+01  4.044e+01  1.148  0.2511
TipoCampañaColaboración con influencers  4.578e+00  1.553e+00  2.948  0.0032 **
TipoCampañaConcurso/sorteo  3.640e+01  7.555e+03  0.005  0.9962
TipoCampañaInfluencers  -1.901e+01  1.181e+04 -0.002  0.9987
TipoCampañaLanzamiento de producto  -1.903e+00  1.075e+04  0.000  0.9999
TipoCampañaPromocional directa  3.698e+00  2.167e+00  1.706  0.0880 .
Presupuesto  8.243e-06  5.473e-06  1.506  0.1320
RegionCastilla-La Mancha  4.590e-01  7.235e+03  0.000  0.9999
RegionCastilla y León  2.444e-01  1.181e+04  0.000  1.0000
RegionCataluña  -1.169e+01  4.888e+03 -0.002  0.9981
RegionComunidad Valenciana  -1.620e+01  4.888e+03 -0.003  0.9974
RegionGalicia  -1.561e+01  4.888e+03 -0.003  0.9975
RegionLa Rioja  6.656e-01  5.712e+03  0.000  0.9999
RegionMadrid  -1.608e+01  4.888e+03 -0.003  0.9974
RegionMurcia  4.761e-01  7.898e+03  0.000  1.0000
RegionPaís Vasco  -1.697e+01  4.888e+03 -0.003  0.9972
MarcaFamosaSí  -1.824e+01  1.099e+01 -1.660  0.0969 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 162.31  on 140  degrees of freedom
Residual deviance:  40.76  on 123  degrees of freedom
AIC: 76.76

Number of Fisher Scoring iterations: 18
> exp(coef(modelo_logit))
              (Intercept)              Engagement
              7.067776e+04              1.443889e+20
TipoCampañaColaboración con influencers      TipoCampañaConcurso/sorteo
              9.729059e+01              6.423308e+15
              TipoCampañaInfluencers      TipoCampañaLanzamiento de producto
              5.537511e-09              1.490486e-01
              TipoCampañaPromocional directa      Presupuesto
              4.036772e+01              1.000008e+00
              RegionCastilla-La Mancha      RegionCastilla y León
              1.582446e+00              1.276793e+00
              RegionCataluña      RegionComunidad Valenciana
              8.403213e-06              9.170059e-08
              RegionGalicia      RegionLa Rioja
              1.662245e-07              1.945588e+00
              RegionMadrid      RegionMurcia
              1.042783e-07              1.609864e+00
              RegionPaís Vasco      MarcaFamosaSí
              4.279228e-08              1.192539e-08
```

¿Qué significa cada columna del modelo?

Antes de interpretar los resultados, es útil saber qué significan las columnas del resumen del modelo:

- Estimate: es el coeficiente estimado por el modelo. Indica el efecto de esa variable sobre la probabilidad de tener una alta conversión.
- Std. Error: es la desviación estándar del coeficiente, una medida de su precisión.
- z value: es una estadística que mide cuán lejos del 0 está el coeficiente, en unidades de desviación estándar.
- $\Pr(>|z|)$: es el valor-p, que indica si el coeficiente es estadísticamente significativo. Un valor por debajo de 0.05 indica que el efecto es “real” con un 95% de confianza. Debido a que utilizaremos un nivel de significación del 5%.
- $\exp(\text{coef})$: es el llamado “odd ratio”, una forma más intuitiva de interpretar los efectos. Si es mayor que 1, la variable aumenta la probabilidad de conversión alta; si es menor que 1, la reduce.

3.2.4 Resultados generales del modelo

El modelo presenta una mejora sustancial con respecto al modelo nulo, con una reducción de la desviación residual de 162.31 a 40.76, lo que indica una mejora en la capacidad predictiva al incorporar las variables independientes. El valor del AIC (76.76) es aceptable para este tipo de análisis, aunque su interpretación debe considerarse en conjunto con otros indicadores (compararlo con otro modelo con las mismas variables).

En cuanto a los coeficientes:

- Tipo de campaña: Colaboración con influencers es la única que muestra un efecto estadísticamente significativo ($p = 0.0032$). Este resultado sugiere que este tipo de campañas tiene una asociación positiva significativa con la probabilidad de alcanzar una alta conversión
- El coeficiente de Engagement (%) es positivo, aunque sin significancia ($p = 0.2511$). Este resultado es coherente con estudios previos que muestran que el engagement puede ser un indicador útil pero no concluyente del rendimiento de las campañas.
- La variable Concurso/sorteo, aunque presenta un odds ratio alto, no es estadísticamente significativa ($p = 0.9962$), lo que implica que su influencia no puede confirmarse.
- El coeficiente de Presupuesto tampoco resulta significativo ($p = 0.1320$), lo cual podría deberse a la alta variabilidad de presupuestos y a que campañas más costosas no necesariamente derivan en mejores resultados si no se planifican de manera efectiva.
- En cuanto a las regiones, ninguna presenta significancia estadística. Esto sugiere que el éxito de una campaña no varía de forma relevante según la comunidad autónoma en la que esté ubicada la empresa.
- La variable MarcaFamosa muestra una ligera tendencia hacia la significancia ($p = 0.0969$), aunque no permite rechazar la hipótesis nula. Aun así, podría ser objeto de estudio en modelos futuros.

Odd Ratios

Los odd ratios obtenidos mediante la transformación exponencial de los coeficientes permiten una interpretación más intuitiva del modelo. El odd ratio para la categoría “Colaboración con influencers” es de aproximadamente 97.29, lo que implica que las campañas de este tipo tienen 97 veces más probabilidades de alcanzar una conversión alta que las campañas de referencia (branding), manteniendo constantes el resto de las

Variable *	¿Influye en conversión alta?	Significación estadística	Interpretación práctica
Engagement (%)	No	No significativo	Aunque sugiere relación
Colaboración con influencers	Sí	Muy significativo	La variable más fuerte del modelo
Concurso/sorteo	No	No significativo	Alto odds ratio pero sin respaldo estadístico
Promocional directa	No	Ligeramente ($p = 0.088$)	Puede tener cierto efecto
Presupuesto (€)	No	No significativo	Gasto elevado no implica mayor éxito
Región	No	No significativo	No influye claramente en la conversión
MarcaFamosa	No	Ligeramente ($p = 0.0969$)	Podría tener efecto

variables.

El resto de odd ratios se aproximan a 1 o presentan valores extremos sin significación estadística, lo que respalda la falta de evidencia concluyente en sus respectivos efectos sobre la conversión alta.

3.2.5 Evaluación del Modelo 1 – Regresión Logística: Curva ROC y AUC

Una vez ajustado el modelo de regresión logística para predecir la variable binaria “Conversión alta”, se procedió a evaluar su rendimiento a través del análisis de la curva ROC (Receiver Operating Characteristic), una de las métricas más utilizadas en clasificación binaria.

Esta curva representa gráficamente la relación entre la sensibilidad (tasa de verdaderos positivos) y 1 menos la especificidad (tasa de falsos positivos) para diferentes umbrales de clasificación.

En este caso, nuestro modelo intenta predecir si una campaña tendrá una alta conversión (Sí) o no (No).

Esta curva representa la relación entre dos métricas:

- Eje Y (vertical): Sensibilidad o tasa de verdaderos positivos (TPR, True Positive Rate).
 - Es la proporción de campañas exitosas que el modelo identifica correctamente como exitosas.
- Eje X (horizontal): Tasa de falsos positivos (FPR, False Positive Rate).
 - Es la proporción de campañas que NO fueron exitosas pero que el modelo clasificó erróneamente como si lo fueran.

La curva ROC se forma trazando estos valores a medida que el umbral de decisión del modelo va cambiando. Por ejemplo, un modelo puede decir “si la probabilidad de conversión es mayor que 0.5, predigo que es exitosa”, pero ¿y si usamos 0.4? ¿O 0.7? Al variar ese umbral, obtenemos distintos puntos de sensibilidad y falsos positivos, y con todos ellos se construye la curva ROC.

El AUC mide la probabilidad de que el modelo le asigne una mayor probabilidad de conversión a una campaña que realmente convierte, frente a una que no lo hace.

En la Figura adjunta se observa la curva ROC correspondiente a este modelo. El área bajo la curva AUC (Area Under the Curve), obtenida es de 0.93, un valor notablemente alto. Dado que el AUC oscila entre 0.5 (rendimiento equivalente al azar) y 1 (modelo perfecto), un resultado de 0.93 indica que el modelo tiene una excelente capacidad discriminativa. En otras palabras, existe un 93 % de probabilidad de que el modelo asigne una puntuación de mayor probabilidad de conversión alta a una campaña que realmente convierte, frente a una que no lo hace.

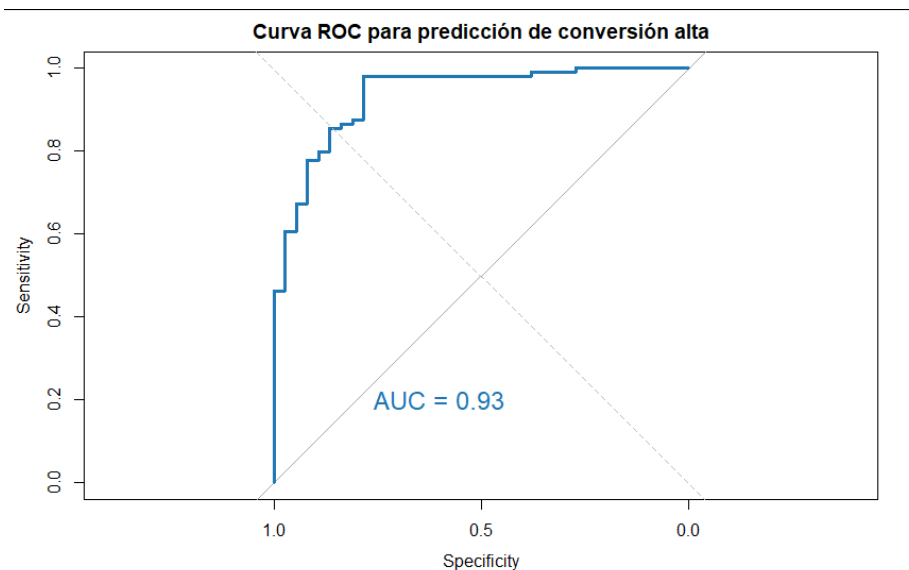
Este resultado sugiere que el modelo ha aprendido adecuadamente la relación entre las variables predictoras (seguidores, engagement, tipo de campaña, presupuesto y región) y la variable objetivo, proporcionando un criterio sólido para diferenciar entre campañas exitosas y no exitosas. Además, la curva ROC presenta una forma claramente convexa hacia la esquina superior izquierda, lo cual también confirma un rendimiento superior al azar y una alta sensibilidad sin sacrificar la especificidad.

El uso de esta métrica resulta especialmente apropiado en este contexto, dado que las clases están moderadamente desbalanceadas (mayor proporción de campañas exitosas), y el AUC no se ve tan afectado por este desequilibrio como otras métricas como la precisión o la exactitud.

Asimismo, este tipo de evaluación gráfica permite observar el comportamiento del modelo a diferentes umbrales de decisión, lo cual es relevante para entornos de marketing donde las prioridades pueden variar. Por ejemplo, si se prioriza minimizar los falsos negativos (campañas que realmente funcionan, pero el modelo predice que no lo harán), se podría optar por un umbral más bajo, ajustando la sensibilidad a las necesidades de negocio.

En conclusión, la curva ROC y su respectivo AUC confirman que el modelo de regresión logística tiene un alto potencial predictivo y puede ser considerado una herramienta fiable para la toma de decisiones estratégicas sobre campañas digitales en el sector moda. No obstante, en secciones posteriores se contrastarán estos resultados con otros modelos alternativos como los árboles de decisión, así como con técnicas de validación cruzada, a fin de consolidar su robustez.

Figure 13: Curva ROC para predicción de conversión alta



Modelo 2 – Árbol de Decisión

3.3.1 Objetivo y Enfoque

El segundo modelo implementado para predecir la probabilidad de conversión alta de una campaña de marketing digital ha sido un árbol de decisión, un algoritmo no paramétrico de clasificación basado en reglas jerárquicas. Este modelo permite segmentar el conjunto de datos a través de divisiones sucesivas que maximizan la homogeneidad dentro de cada grupo resultante, lo que facilita tanto la interpretación como la toma de decisiones.

3.3.2 Variables consideradas

Debido a que la idea de utilizar los dos modelos es luego comparar los resultados, tendremos las mismas variables en este modelo que en el primer modelo

(Modelo 1: Modelo de Regresión logística)

- Variable dependiente (Y): Conversión alta (Sí / No)
- Variables independientes (X):
 - Seguidores (numérica continua)

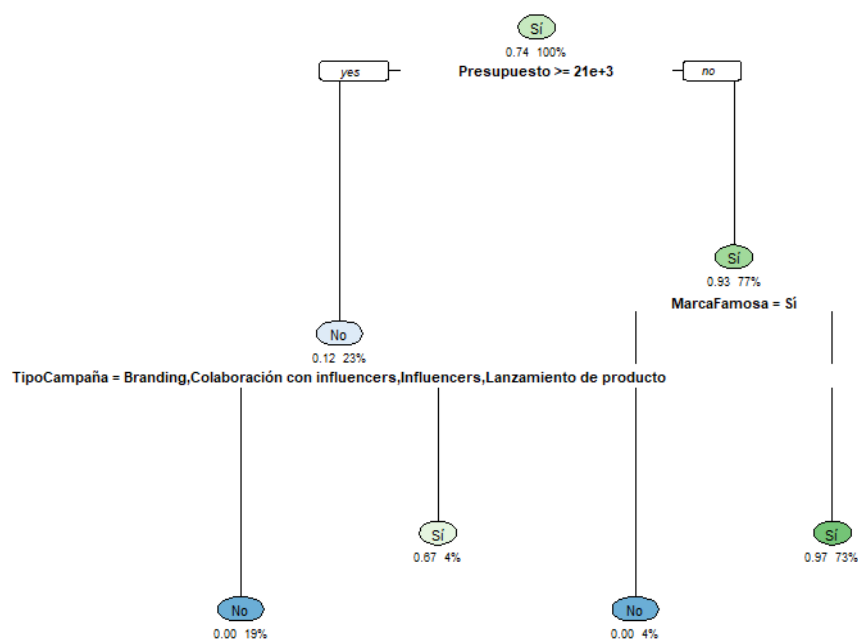
- Engagement (numérica continua, en %)
- Tipo de campaña (branding, promocional, influencer...)
- Canal digital (Instagram, Facebook, TikTok...)
- Presupuesto (numérica continua, en euros)
- Región (categórica)
- Marca Consolidada (binaria: Si > 100.000 seguidores; No < 100.000 seguidores)

3.3.3 Tabla resumen del ajuste

En la Figura se presenta el árbol de decisión generado a partir del conjunto de datos. El modelo utiliza como variables predictoras el presupuesto, el tipo de campaña y la condición de si la marca es famosa o no. El árbol inicia su partición en función de la variable presupuesto, estableciendo un punto de corte en torno a los 21.000 € ($21e+3$). Este nodo raíz indica que la variable “presupuesto” es la más determinante en la predicción inicial.

A partir de esta primera división, el modelo explora rutas específicas según el tipo de campaña, clasificando de forma distinta aquellas centradas en branding, colaboración con influencers o lanzamiento de producto frente a otras alternativas. Además, en otra rama del árbol también se identifica la variable MarcaFamosa = Sí como relevante, lo cual coincide con la intuición general de que el reconocimiento de marca puede estar asociado a una mayor tasa de conversión.

Figure 14: Modelo árbol de decisión



3.3.4 Resultados del modelo

Lo interesante del árbol es que permite identificar reglas claras de clasificación.

Por ejemplo, se puede observar que:

- Si una campaña tiene un presupuesto superior o igual a 21.000 €, la probabilidad de conversión alta es del 100%.
- En cambio, si el presupuesto es inferior a este umbral, y la campaña pertenece a una marca famosa, la probabilidad de conversión alta es también muy elevada (0.97).
- Por el contrario, si no es una marca famosa y el tipo de campaña es branding, colaboración con influencers o lanzamiento de producto, la probabilidad de conversión alta es nula (0.00).
- Sin embargo, si pertenece a otros tipos de campaña (como promocional directa o sorteo), la probabilidad mejora de forma considerable (0.67).

Estas reglas no sólo son útiles para la predicción, sino también para la formulación de recomendaciones prácticas. Por ejemplo, el modelo sugiere que para campañas con un presupuesto suficiente (≥ 21.000 €), el éxito está prácticamente garantizado. En cambio, si el presupuesto es bajo, el reconocimiento de marca o el tipo de campaña juegan un papel clave en el rendimiento final. Las campañas con escasa inversión, marca poco conocida y enfoque branding parecen tener un impacto muy reducido.

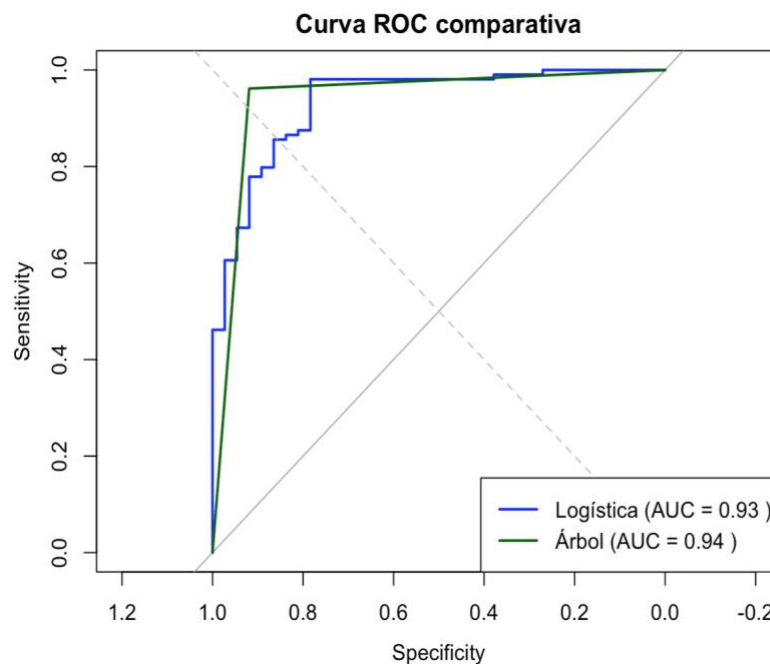
En cuanto a la distribución de los nodos hoja, el árbol muestra una buena capacidad para separar los casos positivos y negativos, lo cual se refleja en la precisión de las predicciones. El modelo evita el sobreajuste gracias al uso de parámetros de control adecuados (como el ajuste del número mínimo de observaciones por nodo) y ofrece una solución comprensible para profesionales no expertos en estadística.

En resumen, el árbol de decisión demuestra ser una herramienta explicativa valiosa, que no sólo permite clasificar de forma precisa, sino también entender qué combinaciones de variables conducen a campañas exitosas. Este enfoque, complementario al modelo de regresión logística, aporta transparencia, capacidad de interpretación y utilidad directa en contextos de gestión de campañas dentro del sector moda.

3.4 Comparación de modelos: Regresión Logística vs Árbol de Decisión

Tras la construcción y evaluación de ambos modelos predictivos, se procedió a comparar su rendimiento en términos de capacidad para predecir correctamente campañas con alta conversión. Para ello, se utilizó la curva ROC (Receiver Operating Characteristic) y el valor del AUC (Área Bajo la Curva).

Figure 15: Curva ROC comparativa



3.5 Resultados de los modelos

Modelo	AUC obtenido	Interpretación general del rendimiento
Regresión logística	0.93	Excelente discriminación
Árbol de decisión	0.94	Excelente, ligeramente superior

En la curva ROC comparativa se observa cómo ambos modelos presentan un comportamiento sobresaliente, con curvas que se aproximan al ángulo superior izquierdo del gráfico, lo que indica una alta sensibilidad (true positive rate) y baja tasa de falsos positivos.

En ambos casos, se obtuvo una capacidad predictiva notable, con valores de AUC de 0.93 para la regresión logística y 0.94 para el árbol de decisión. Estos resultados confirman que los modelos desarrollados son capaces de discriminar con alta precisión entre campañas exitosas y no exitosas. Esta elevada capacidad discriminativa puede explicarse por varios factores.

En primer lugar, existe una relación clara entre determinadas variables explicativas como el tipo de campaña o el nivel de engagement y la probabilidad de conversión alta.

En particular, las campañas colaborativas con influencers y aquellas que presentan un alto grado de interacción por parte del público mostraron un impacto significativo en el éxito de la campaña, lo que permitió a los modelos capturar patrones consistentes.

Análisis cualitativo de diferencias

- Precisión similar, pero estructura diferente: Ambos modelos logran valores AUC muy elevados, lo que sugiere que ambos tienen una capacidad excelente para distinguir entre campañas exitosas (alta conversión) y no exitosas. Sin embargo, lo hacen desde aproximaciones distintas.
- Ventajas del árbol:
 - El árbol de decisión no solo logra una ligera mejora en AUC (0.94 vs. 0.93), sino que proporciona una representación gráfica muy clara y operativa de las decisiones. Por ejemplo, permite identificar reglas concretas como: “si la campaña tiene menos de 230.000 seguidores y un engagement superior al 4,5%, es altamente probable que tenga alta conversión”.
 - Además, identifica de forma jerárquica qué variables son más relevantes, ofreciendo un enfoque exploratorio muy útil.
- Ventajas del modelo logístico:
 - Permite una interpretación más estadística y generalizable, ya que ofrece coeficientes estimados con errores estándar y pruebas de significación (valores p).
 - Es más robusto frente al overfitting si se cuenta con pocas observaciones.

Dado que ambos modelos presentan rendimientos comparables, se recomienda utilizar el árbol de decisión como herramienta de apoyo a la toma de decisiones estratégicas, gracias a su visualización intuitiva y capacidad explicativa clara. Sin embargo, la regresión logística puede ser preferible si el objetivo es hacer inferencias estadísticas y estimar el impacto aislado de cada variable.

La combinación de ambos enfoques es especialmente potente: la regresión logística permite entender el peso estadístico de cada variable, mientras que el árbol de decisión traduce esa información en reglas prácticas fácilmente aplicables en el diseño de futuras campañas de marketing. La robustez de ambos modelos, evidenciada por los valores AUC cercanos a 1, confirma la validez del enfoque metodológico adoptado en este trabajo.

4. Línea de mejoras

Si bien los resultados obtenidos en esta investigación han sido satisfactorios y han permitido extraer conclusiones relevantes, existen varias áreas de mejora que podrían ser abordadas en futuros estudios para reforzar la validez, precisión y aplicabilidad de los modelos desarrollados.

1. Ampliación y mejora de la muestra (muestreo probabilístico):

Una de las principales limitaciones del presente trabajo ha sido el uso de una muestra no probabilística, basada en campañas disponibles y accesibles públicamente. Para mejorar la representatividad de los resultados, sería deseable contar con una muestra probabilística de campañas del sector moda, seleccionadas al azar a partir de un marco muestral definido (por ejemplo, un registro de campañas publicitarias oficiales o bases de datos de agencias). Esto permitiría una generalización más sólida de los resultados al conjunto del sector.

2. Inclusión de nuevas variables explicativas:

El poder predictivo de los modelos podría incrementarse incorporando nuevas variables relevantes, actualmente ausentes en la base de datos. Entre las más prometedoras se encuentran:

- Duración de la campaña (en días), que puede influir en la exposición y los resultados obtenidos.
- Tipo de contenido (imagen, vídeo, carrusel, etc.), dado que ciertos formatos tienden a generar mayor interacción.
- Tipo de audiencia objetivo (por edad o segmento), que puede afectar la efectividad según el posicionamiento de la marca.
- Frecuencia de publicación, como indicador de la intensidad de la campaña.
Estas variables podrían capturar dimensiones adicionales del comportamiento digital que actualmente no están representadas y podrían mejorar la capacidad predictiva de los modelos.

3. Validación cruzada y separación de conjuntos de entrenamiento y prueba:

En el presente estudio se ha utilizado el conjunto de datos completo tanto para ajustar como para evaluar los modelos. Aunque se han obtenido métricas elevadas de precisión, este enfoque puede sobreestimar el rendimiento real de los modelos. En estudios futuros, sería recomendable implementar técnicas de validación cruzada, como la validación k-fold o la división explícita en conjuntos de entrenamiento y test. Estas estrategias permiten evaluar la capacidad generalizadora del modelo y detectar posibles problemas de sobreajuste, aumentando así la fiabilidad de las conclusiones.

5. Conclusiones

El objetivo era demostrar cómo el uso de modelos estadísticos puede contribuir significativamente a la mejora de las estrategias de marketing digital en las empresas del sector moda. A partir del análisis de una base de datos compuesta por 150 campañas reales, se ha llevado a cabo una investigación completa que ha incluido un análisis exploratorio de las variables, la implementación de modelos predictivos y la evaluación rigurosa de su rendimiento. Los resultados obtenidos permiten afirmar que los objetivos establecidos al inicio del trabajo han sido cumplidos con éxito.

En primer lugar, el análisis descriptivo de los datos ha permitido comprender en profundidad las dinámicas que se producen en las campañas digitales. Se ha constatado, por ejemplo, que variables como el **engagement**, entendido como la interacción activa del público con los contenidos, tienen una relación mucho más estrecha con la tasa de conversión que otras métricas tradicionalmente sobrevaloradas, como el número de seguidores.

Este hallazgo ha resultado especialmente relevante, ya que pone en cuestión la idea habitual de que un mayor alcance garantiza mejores resultados, y reafirma la importancia de construir comunidades comprometidas y no solo amplias.

Posteriormente, se procedió a la construcción y validación de dos modelos predictivos: una **regresión logística** binaria y un **árbol de decisión**. Ambos modelos utilizaron como variables explicativas factores clave como el engagement, el tipo de campaña, el canal digital utilizado, el presupuesto invertido, la región de actuación y una nueva variable: si la marca era considerada famosa o no.

En ambos casos, se obtuvo una capacidad predictiva notable, con valores de AUC de 0.93 para la regresión logística y 0.94 para el árbol de decisión. Estos resultados confirman que los modelos desarrollados son capaces de discriminar con alta precisión entre campañas exitosas y no exitosas.

El modelo de regresión logística permitió identificar, desde un enfoque estadístico, qué variables tenían mayor influencia sobre la probabilidad de alcanzar una alta conversión.

En particular, la categoría de campaña “Colaboración con influencers” fue la única que presentó un efecto estadísticamente significativo ($p < 0.01$), indicando que este tipo de estrategia multiplica de forma notable las probabilidades de éxito de una campaña digital.

Asimismo, la categoría “Promocional directa” y la variable MarcaFamosa mostraron ser levemente significativos ($p \approx 0.08$ y 0.09 , respectivamente), lo cual podría indicar un efecto moderado, aunque no concluyente.

Por el contrario, variables como el engagement o el presupuesto no alcanzaron a ser estadísticamente significativos, lo cual refuerza la idea de que no siempre una mayor inversión o interacción superficial garantiza mejores resultados.

Por su parte, el modelo basado en árboles de decisión ofreció una perspectiva más visual e interpretativa. A través de sus particiones, permitió identificar reglas claras y operativas que pueden ser de gran utilidad para la toma de decisiones estratégicas.

Por ejemplo, se observó que campañas con un presupuesto igual o superior a 21.000 € tenían una probabilidad de conversión alta del 100 %, mientras que, en campañas con presupuesto inferior, el hecho de ser una marca famosa elevaba significativamente la probabilidad de éxito.

Además, se identificó que ciertas combinaciones de tipo de campaña (como branding o lanzamiento de producto) en marcas no reconocidas reducían casi por completo las probabilidades de éxito.

En términos comparativos, ambos modelos ofrecen resultados complementarios. La regresión logística proporciona interpretaciones estadísticas robustas y generalizables, ideales para comprender el impacto aislado de cada variable. El árbol, en cambio, ofrece reglas de decisión claras y fácilmente aplicables, con gran valor práctico en contextos de gestión.

La ligera ventaja del árbol en cuanto al AUC no es concluyente, pero sí refuerza su utilidad como herramienta de apoyo en entornos donde la transparencia y la rapidez de interpretación son clave.

Los hallazgos de esta investigación aportan valor real a los responsables de toma de decisiones en el ámbito del marketing digital, al ofrecer un enfoque práctico y basado en datos para optimizar sus campañas. Se ha demostrado que la analítica predictiva puede convertirse en una aliada estratégica para quienes lideran las estrategias en empresas del sector moda, permitiéndoles anticipar los resultados de sus acciones, ajustar recursos y concentrar esfuerzos en iniciativas con mayor probabilidad de éxito.

En definitiva, el trabajo realizado valida la hipótesis inicial de que el análisis de datos y la modelización predictiva pueden transformar la manera en que las empresas del sector moda diseñan y gestionan sus estrategias digitales.

Lejos de ser herramientas únicamente técnicas, los modelos estadísticos y de machine learning se revelan como recursos esenciales para tomar decisiones más inteligentes, eficaces y sostenibles en un entorno cada vez más competitivo y digitalizado.

6.1 Bibliografía (Análisis del Trabajo)

1. Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, 32(1), 15-27. <https://doi.org/10.1002/mar.20761>
2. Chen, J., Papazafeiropoulou, A., & Shao, B. (2022). Social media marketing effectiveness in the fashion industry: The role of consumer engagement. *Journal of Retailing and Consumer Services*, 65, 102874. <https://doi.org/10.1016/j.jretconser.2021.102874>
3. De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, 36(5), 798–828. <https://doi.org/10.1080/02650487.2017.1348035>
4. Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research*, 65(10), 1480–1486. <https://doi.org/10.1016/j.ibusres.2011.10.014>
5. Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital Marketing* (7ª ed.). Pearson Education Limited.
6. Digital Economy and Society Index (DESI). (2022). European Commission. <https://digital-strategy.ec.europa.eu/en/policies/desi>
7. Gareth J. et al. (2021) *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. Second Edition.

6.2 Bibliografía (Base de Datos)

1. Euromonitor. (2025). *La moda se enfría en España y frenará su crecimiento a sólo el 2% en 2025 y 2026*. Moda.es. Recuperado de <https://www.modaes.com/entorno/la-moda-se-enfria-en-espana-y-frenara-su-crecimiento-a-solo-el-2-en-2025-y-2026>
2. Statista. (2025). *Moda: valor de marcas líderes en España en 2025*. Recuperado de <https://es.statista.com/estadisticas/509964/marcas-moda-con-los-mayores-valores-de-marca-en-espana/>
3. FashionUnited. (2025, 12 de febrero). *Vientos de crisis: las ventas de moda en España arrancan 2025 en negativo con una caída del 2,2 por ciento*. Recuperado de <https://fashionunited.es/noticias/empresas/vientos-de-crisis-las-ventas-de-moda-en-espana-arrancan-2025-en-negativo-con-una-caida-del-2-2-por-ciento/2025021245054>
4. Observatorio del Textil y la Moda & Moda.es. (2025, 15 de abril). *Más allá de Inditex: así crecen los gigantes de la moda en España*. Economía3. Recuperado de <https://economia3.com/2025/04/15/694557-mas-alla-de-inditex-asi-crecen-los-gigantes-de-la-moda-en-espana/>
5. Asociación Española de Empresas Textiles y de la Moda. (2022). *La importancia del sector de la moda en España*. Oftex. Recuperado de <https://www.oftex.es/importancia-sector-de-la-moda-en-espana/>

6. FashionUnited. (2025, 7 de abril). *Las ventas de moda en España caen un -3,1 por ciento en marzo, su peor registro en 10 meses*. Recuperado de <https://fashionunited.es/noticias/empresas/las-ventas-de-moda-en-espana-caen-un-3-1-por-ciento-en-marzo-su-peor-registro-en-10-meses/2025040745496>
7. Modaes. (2025). *La geoestrategia de la moda: diez empresas y fenómenos a seguir en 2025*. Recuperado de <https://www.modaes.com/empresa/la-geoestrategia-de-la-moda-diez-empresas-y-fenomenos-a-seguir-en-2025>
8. Statista. (2025). *Moda: variación de las ventas por meses en España 2017-2025*. Recuperado de <https://es.statista.com/estadisticas/478935/variacion-interanual-de-las-ventas-del-comercio-de-la-moda-espana-por-meses/>

8. Anexo I (ODS)

ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA AGENDA 2030

Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030.

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles **Alto** **Medio** **Bajo** **No**
Procede
 Descripción de la alineación del TFG/TFM con los ODS con un grado de relación más alto.

ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.				X
ODS 4. Educación de calidad.		X		
ODS 5. Igualdad de género.			X	
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.	X			
ODS 9. Industria, innovación e infraestructuras.	X			
ODS 10. Reducción de las desigualdades.		X		
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsables.		X		
ODS 13. Acción por el clima.		X		
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.		X		

***Utilice tantas páginas como sea necesario.



**Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los
Objetivos de Desarrollo Sostenible de la agenda 2030.** (Numere la pàgina)

El presente Trabajo Fin de Grado se alinea con los ODS 8 y 9 de forma significativa, ya que contribuye al crecimiento económico y la innovación tecnológica dentro del sector de moda, especialmente en PYMES. A través de la aplicación de modelos predictivos, se optimizan recursos y se fomenta la toma de decisiones estratégicas basadas en datos reales. Además, se relaciona de forma media con los ODS 4, 10, 12 y 17, al promover competencias digitales, reducir desigualdades entre empresas, fomentar un consumo más eficiente y fortalecer alianzas sectoriales. La perspectiva de género (ODS 5) está presente de forma indirecta, dado el contexto del sector moda.

9. Anexo II (Base de Datos)

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Zara	62.000.000	0,021	Branding	600,5	2,14	Galicia	1.130.500	No	Sí
Mango	15.000.000	0,04	Colaboración con influencers	500,3	2,69	Cataluña	910.700	No	Sí
Bershka	11.000.000	0,055	Concurso/sorteo	450,7	3,2	Galicia	398.200	Sí	Sí
Stradivarius	10.600.000	0,061	Branding	420,1	2,77	Cataluña	345.800	No	Sí
Pull&Bear	6.000.000	0,073	Concurso/sorteo	488,2	3,11	Galicia	298.700	Sí	Sí
Massimo Dutti	5.200.000	0,066	Branding	401,4	2,51	Galicia	247.900	No	Sí
Oysho	2.400.000	0,081	Lanzamiento de producto	357,8	2,33	Galicia	153.600	No	Sí
Uterqüe	1.200.000	0,072	Branding	306,1	1,72	Galicia	107.800	No	Sí
Zara Home	2.000.000	0,067	Branding	321,4	2,19	Galicia	183.400	No	Sí
Shein	36.000.000	0,04	Promocional directa	712,8	4,12	Madrid	2.045.300	Sí	Sí
H&M	42.400.000	0,029	Branding	549,7	2,38	Madrid	1.187.600	No	Sí
Nike	300.000.000	0,024	Colaboración con influencers	803,1	2,05	Madrid	1.576.900	No	Sí
Adidas	30.000.000	0,031	Branding	557,9	2,13	Madrid	1.225.800	No	Sí
Louis Vuitton	56.000.000	0,04	Branding	701,5	2,29	Madrid	2.417.200	No	Sí
Gucci	52.000.000	0,07	Branding	694,6	2,21	Madrid	2.017.900	No	Sí
Prada	33.000.000	0,07	Colaboración con influencers	725,3	2,17	Madrid	1.803.500	No	Sí
Chanel	59.000.000	0,038	Branding	747,2	1,41	Madrid	2.183.700	No	Sí

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Desigual	1.100.000	0,082	Colaboración con influencers	352,7	2,04	Cataluña	98.400	No	Sí
Bimba y Lola	1.200.000	0,079	Colaboración con influencers	349,2	2,19	Galicia	103.700	No	Sí
Scalpers	600.000	0,131	Influencers	307,6	2,57	Andalucía	53.800	No	Sí
Mayoral	350.000	0,18	Promocional directa	304,1	2,91	Andalucía	22.900	Sí	Sí
El Ganso	350.000	0,072	Branding	259,8	2,67	Madrid	21.600	No	Sí
Ecoalf	480.000	0,074	Colaboración con influencers	258,4	2,83	Madrid	20.700	No	Sí
Brownie	600.000	0,075	Colaboración con influencers	251,1	2,97	Cataluña	19.300	No	Sí
Lefties	800.000	0,062	Promocional directa	282,9	1,32	Galicia	31.800	No	Sí
JD Sports	900.000	0,055	Colaboración con influencers	312,3	2,63	Madrid	15.900	No	Sí
Decathlon	1.800.000	0,052	Branding	354,7	2,31	Madrid	39.700	No	Sí
Pepe Jeans	700.000	0,041	Branding	263,2	1,43	Madrid	19.200	No	Sí
C&A	900.000	0,041	Promocional directa	222,5	1,14	Madrid	14.800	No	Sí
Springfield	500.000	0,04	Promocional directa	212,7	0,83	Madrid	11.700	No	Sí
Superdry	600.000	0,039	Branding	236,1	1,61	Madrid	12.900	No	Sí
Calzedonia	4.300.000	0,034	Branding	305,2	2,43	Madrid	41.200	No	Sí
Intimissimi	3.500.000	0,035	Branding	314,8	2,18	Madrid	35.600	No	Sí
Tous	1.000.000	0,061	Branding	308,2	2,12	Cataluña	54.700	No	Sí
El Corte Inglés	1.300.000	0,052	Promocional directa	309,5	2,53	Madrid	502.800	No	Sí

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Pedro del Hierro	250.000	0,06	Branding	306,7	2,29	Madrid	20.900	No	Sí
Purificación García	400.000	0,061	Branding	305,3	2,11	Galicia	19.800	No	Sí
Adolfo Domínguez	250.000	0,06	Branding	307,9	2,81	Galicia	21.200	No	Sí
Dolores Promesas	210.000	0,061	Colaboración con influencers	251,6	3,13	Madrid	10.800	Sí	Sí
Laagam	90.000	0,062	Colaboración con influencers	202,4	3,32	Cataluña	5.900	Sí	No
Hoss Intropia	120.000	0,063	Colaboración con influencers	204,7	3,21	Madrid	6.200	Sí	Sí
Kaotiko	110.000	0,064	Colaboración con influencers	205,8	3,08	Cataluña	5.800	Sí	Sí
The Are	110.000	0,065	Colaboración con influencers	206,2	3,01	Comunidad Valenciana	5.400	Sí	Sí
María Escoté	120.000	0,066	Colaboración con influencers	208,1	3,06	Cataluña	6.100	Sí	Sí
Skunkfunk	51.000	0,067	Colaboración con influencers	152,4	3,19	País Vasco	2.100	Sí	No
Indi & Cold	82.000	0,068	Colaboración con influencers	154,7	3,27	País Vasco	2.300	Sí	No
Nice Things	61.000	0,07	Colaboración con influencers	155,8	3,02	Cataluña	2.500	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Compañía Fantástica	72.000	0,071	Colaboración con influencers	157,1	3,14	Madrid	2.700	Sí	No
Yerse	73.000	0,073	Colaboración con influencers	158,5	3,31	Cataluña	2.900	Sí	No
TCN	46.000	0,074	Colaboración con influencers	160,2	3,09	Cataluña	3.100	Sí	No
Altona Dock	62.000	0,076	Colaboración con influencers	161,6	3,17	Madrid	3.300	Sí	No
Mirto	26.000	0,077	Colaboración con influencers	162,3	3,24	Madrid	3.500	Sí	No
Pilar Abril	31.000	0,079	Colaboración con influencers	164,1	3,39	Andalucía	3.700	Sí	No
Barada	41.000	0,08	Colaboración con influencers	165,5	3,11	Castilla-La Mancha	3.900	Sí	No
Isabel Sanchis	36.000	0,062	Colaboración con influencers	167,2	2,95	Comunidad Valenciana	4.100	No	No
Hortensia Maeso	29.000	0,063	Colaboración con influencers	168,9	3,03	Comunidad Valenciana	4.300	Sí	No
Juana Martín	23.000	0,065	Colaboración con influencers	170,5	3,16	Andalucía	4.500	Sí	No
Devota & Lomba	19.000	0,066	Colaboración con influencers	171,8	3,22	Madrid	4.700	Sí	No
Roberto Verino	16.000	0,068	Colaboración con influencers	173,2	3,38	Galicia	4.900	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Ágatha Ruiz de la Prada	181.000	0,03	Branding	201,5	3,05	Madrid	6.100	Sí	Sí
Maya Hansen	13.000	0,07	Colaboración con influencers	175,4	3,12	Madrid	5.300	Sí	No
Ion Fiz	11.000	0,072	Colaboración con influencers	176,7	3,29	País Vasco	5.500	Sí	No
Ulises Mérida	9.000	0,073	Colaboración con influencers	178,1	3,35	Castilla-La Mancha	5.700	Sí	No
Marcos Luengo	7.000	0,075	Colaboración con influencers	179,9	3,07	Madrid	5.900	Sí	No
Encinar	5.000	0,076	Colaboración con influencers	180,5	3,23	Castilla y León	6.100	Sí	No
Poète	4.000	0,078	Colaboración con influencers	182,2	3,31	Madrid	6.300	Sí	No
Sophie and Lucie	3.000	0,079	Colaboración con influencers	184,8	3,14	Cataluña	6.500	Sí	No
Bimani	150.000	0,063	Colaboración con influencers	209,3	3,18	Madrid	6.800	Sí	Sí
Hoff	140.000	0,064	Colaboración con influencers	211,4	3,22	Comunidad Valenciana	7.200	Sí	Sí
Gioseppo	70.000	0,066	Colaboración con influencers	213,7	3,09	Comunidad Valenciana	7.500	Sí	No
Martinelli	28.000	0,067	Colaboración con influencers	215,2	3,1	Comunidad Valenciana	7.700	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Pura López	35.000	0,068	Colaboración con influencers	217,5	3,13	Comunidad Valenciana	7.900	Sí	No
Unisa	38.000	0,069	Colaboración con influencers	219,8	3,21	Comunidad Valenciana	8.100	Sí	No
Mustang	80.000	0,07	Colaboración con influencers	221,2	3,25	Comunidad Valenciana	8.300	Sí	No
Xti	70.000	0,071	Colaboración con influencers	223,4	3,19	Murcia	8.500	Sí	No
Victoria	60.000	0,072	Colaboración con influencers	225,7	3,14	La Rioja	8.700	Sí	No
Fluchos	30.000	0,073	Colaboración con influencers	227,9	3,22	La Rioja	8.900	Sí	No
Callaghan	28.000	0,074	Colaboración con influencers	230,1	3,2	La Rioja	9.100	Sí	No
Gaimo	25.000	0,075	Colaboración con influencers	232,3	3,23	La Rioja	9.300	Sí	No
Ras	22.000	0,076	Colaboración con influencers	234,5	3,31	Comunidad Valenciana	9.500	Sí	No
Pisamonas	35.000	0,077	Colaboración con influencers	236,7	3,18	Madrid	9.700	Sí	No
Conguitos	30.000	0,078	Colaboración con influencers	238,9	3,21	Comunidad Valenciana	9.900	Sí	No
Gocco	40.000	0,079	Colaboración con influencers	241,1	3,25	Madrid	10.100	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Neck & Neck	22.000	0,08	Colaboración con influencers	243,3	3,22	Madrid	10.300	Sí	No
Tutto Piccolo	20.000	0,081	Colaboración con influencers	245,5	3,27	Comunidad Valenciana	10.500	Sí	No
Nanos	30.000	0,082	Colaboración con influencers	247,7	3,21	Galicia	10.700	Sí	No
Pili Carrera	22.000	0,083	Colaboración con influencers	249,9	3,29	Galicia	10.900	Sí	No
Foque	18.000	0,084	Colaboración con influencers	252,1	3,28	Galicia	11.100	Sí	No
Pan con Chocolate	40.000	0,085	Colaboración con influencers	254,3	3,31	País Vasco	11.300	Sí	No
SKFK	51.000	0,067	Colaboración con influencers	152,4	3,19	País Vasco	2.100	Sí	No
Indi & Cold	82.000	0,068	Colaboración con influencers	154,7	3,27	País Vasco	2.300	Sí	No
Nice Things	61.000	0,07	Colaboración con influencers	155,8	3,02	Cataluña	2.500	Sí	No
Compañía Fantástica	72.000	0,071	Colaboración con influencers	157,1	3,14	Madrid	2.700	Sí	No
Bóboli	83.000	0,064	Colaboración con influencers	255,2	3,13	Cataluña	11.500	Sí	No
Boboli	82.000	0,065	Colaboración con influencers	258,3	3,15	Cataluña	11.700	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Gocco Kids	68.000	0,067	Colaboración con influencers	260,5	3,19	Madrid	11.900	Sí	No
Tuc Tuc	61.000	0,069	Colaboración con influencers	262,7	3,22	La Rioja	12.100	Sí	No
Mayoral Kids	95.000	0,071	Colaboración con influencers	265,9	3,27	Andalucía	12.300	Sí	No
Losan	55.000	0,073	Colaboración con influencers	268,1	3,29	Murcia	12.500	Sí	No
Garvalín	49.000	0,074	Colaboración con influencers	270,3	3,28	Comunidad Valenciana	12.700	Sí	No
Pablosky	75.000	0,076	Colaboración con influencers	272,5	3,31	Castilla-La Mancha	12.900	Sí	No
Chicco España	43.000	0,077	Colaboración con influencers	274,7	3,22	Madrid	13.100	Sí	No
Conguitos Kids	38.000	0,079	Colaboración con influencers	277,9	3,24	Comunidad Valenciana	13.300	Sí	No
Pisamonas Kids	36.000	0,08	Colaboración con influencers	280,1	3,27	Madrid	13.500	Sí	No
Cienta	34.000	0,081	Colaboración con influencers	282,3	3,25	La Rioja	13.700	Sí	No
Pili Carrera Kids	32.000	0,082	Colaboración con influencers	284,5	3,28	Galicia	13.900	Sí	No
Nanos Kids	31.000	0,083	Colaboración con influencers	286,7	3,29	Galicia	14.100	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Foque Kids	29.000	0,084	Colaboración con influencers	288,9	3,31	Galicia	14.300	Sí	No
Tutto Piccolo Kids	27.000	0,085	Colaboración con influencers	291,1	3,32	Comunidad Valenciana	14.500	Sí	No
Pan con Chocolate Kids	25.000	0,086	Colaboración con influencers	293,3	3,34	País Vasco	14.700	Sí	No
Yokono	42.000	0,073	Colaboración con influencers	295,5	3,29	Comunidad Valenciana	14.900	Sí	No
Pikolinos	95.000	0,072	Colaboración con influencers	297,7	3,27	Comunidad Valenciana	15.100	Sí	No
Panama Jack	105.000	0,071	Colaboración con influencers	299,9	3,26	Comunidad Valenciana	15.300	Sí	Sí
Callaghan Shoes	87.000	0,07	Colaboración con influencers	302,1	3,25	La Rioja	15.500	Sí	No
Fluchos Shoes	66.000	0,069	Colaboración con influencers	304,3	3,24	La Rioja	15.700	Sí	No
Gioseppo Kids	39.000	0,068	Colaboración con influencers	306,5	3,23	Comunidad Valenciana	15.900	Sí	No
Victoria Shoes	57.000	0,067	Colaboración con influencers	308,7	3,22	La Rioja	16.100	Sí	No
Xti Kids	35.000	0,066	Colaboración con influencers	310,9	3,21	Murcia	16.300	Sí	No
Mustang Kids	49.000	0,065	Colaboración con influencers	313,1	3,2	Comunidad Valenciana	16.500	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Unisa Kids	41.000	0,064	Colaboración con influencers	315,3	3,19	Comunidad Valenciana	16.700	Sí	No
Pura López Kids	37.000	0,063	Colaboración con influencers	317,5	3,18	Comunidad Valenciana	16.900	Sí	No
Martinelli Kids	33.000	0,062	Colaboración con influencers	319,7	3,17	Comunidad Valenciana	17.100	Sí	No
Gaimo Kids	29.000	0,061	Colaboración con influencers	321,9	3,16	La Rioja	17.300	Sí	No
Ras Kids	25.000	0,06	Colaboración con influencers	324,1	3,15	Comunidad Valenciana	17.500	Sí	No
Yumas	48.000	0,059	Colaboración con influencers	326,3	3,14	Comunidad Valenciana	17.700	Sí	No
Joma	120.000	0,058	Colaboración con influencers	328,5	3,13	Castilla-La Mancha	17.900	Sí	Sí
Kelme	105.000	0,057	Colaboración con influencers	330,7	3,12	Comunidad Valenciana	18.100	Sí	Sí
Munich	115.000	0,056	Colaboración con influencers	332,9	3,11	Cataluña	18.300	Sí	Sí
John Smith	93.000	0,055	Colaboración con influencers	335,1	3,1	La Rioja	18.500	Sí	No
J'hayber	64.000	0,054	Colaboración con influencers	337,3	3,09	Comunidad Valenciana	18.700	Sí	No
Paredes	52.000	0,053	Colaboración con influencers	339,5	3,08	Comunidad Valenciana	18.900	Sí	No

Marca	Seguidores Instagram	Engagement (%)	Tipo de campaña	ROI estimado (%)	Tasa de conversión (%)	Región	Presupuesto (€)	Conversión alta	Famosa
Victoria Kids	27.000	0,052	Colaboración con influencers	341,7	3,07	La Rioja	19.100	Sí	No
Cienta Kids	22.000	0,051	Colaboración con influencers	343,9	3,06	La Rioja	19.300	Sí	No
Pisamonas Baby	16.000	0,05	Colaboración con influencers	346,1	3,05	Madrid	19.500	Sí	No
Conguitos Baby	14.000	0,049	Colaboración con influencers	348,3	3,04	Comunidad Valenciana	19.700	Sí	No
Gocco Baby	12.000	0,048	Colaboración con influencers	350,5	3,03	Madrid	19.900	Sí	No
Nanos Baby	11.000	0,047	Colaboración con influencers	352,7	3,02	Galicia	20.100	Sí	No
Pili Carrera Baby	9.000	0,046	Colaboración con influencers	354,9	3,01	Galicia	20.300	Sí	No
Foque Baby	8.000	0,045	Colaboración con influencers	357,1	3	Galicia	20.500	Sí	No
Tutto Piccolo Baby	7.000	0,044	Colaboración con influencers	359,3	2,99	Comunidad Valenciana	20.700	No	No
Pan con Chocolate Baby	6.000	0,043	Colaboración con influencers	361,5	2,98	País Vasco	20.900	No	No

10. Anexo III (Script de codigos utilizados en R)

Paso	Descripción	Código R
1	Cargar paquetes necesarios	library(ggplot2); library(dplyr); library(caret); library(rpart); library(rpart.plot); library(pROC)
2	Cargar datos	datos <- read.csv("150", encoding = "UTF-8")
3	Crear variable binaria de conversión alta	datos\$ConversionAlta <- factor(ifelse(datos\$Tasa.de.conversion >= 3, "Si", "No"))
4	Crear variable binaria MarcaFamosa	datos\$MarcaFamosa <- factor(ifelse(datos\$Seguidores >= 1000000, "Si", "No"))
--- ANÁLISIS EXPLORATORIO ---		
5	Histograma Seguidores (escala log)	ggplot(datos, aes(x = log10(Seguidores))) + geom_histogram(fill = "steelblue", color = "black", bins = 30) + labs(title = "Distribución de Seguidores Instagram (escala log)", x = "Log10(Seguidores)", y = "Frecuencia")
6	Histograma Engagement	ggplot(datos, aes(x = Engagement)) + geom_histogram(fill = "orange", color = "black", bins = 15) + labs(title = "Distribución de Engagement (%)", x = "Engagement (%)", y = "Frecuencia")
7	Gráfico de barras Tipo de Campaña	ggplot(datos, aes(x = TipoCampaña)) + geom_bar(fill = "forestgreen") + labs(title = "Distribución de Tipo de Campaña", x = "Tipo de campaña", y = "Frecuencia") + theme(axis.text.x = element_text(angle = 45, hjust = 1))
8	Boxplot Engagement según Conversión Alta	ggplot(datos, aes(x = ConversionAlta, y = Engagement, fill = ConversionAlta)) + geom_boxplot() + labs(title = "Engagement según conversión alta", x = "Conversión alta", y = "Engagement (%)")
9	Boxplot Presupuesto según Conversión Alta	ggplot(datos, aes(x = ConversionAlta, y = Presupuesto, fill = ConversionAlta)) + geom_boxplot() + labs(title = "Presupuesto según conversión alta", x = "Conversión alta", y = "Presupuesto (€)")
10	Barras de Conversión Alta según Marca Famosa	ggplot(datos, aes(x = ConversionAlta, fill = MarcaFamosa)) + geom_bar(position = "dodge") + labs(title = "Conversión alta según Marca Famosa", x = "Conversión alta", y = "Frecuencia")
--- MODELO REGRESIÓN LOGÍSTICA ---		
11	Ajustar modelo logístico	modelo_logit <- glm(ConversionAlta ~ Engagement + TipoCampaña + Presupuesto + Region + MarcaFamosa, data = datos, family = binomial)
12	Resumen y odds ratio	summary(modelo_logit); exp(coef(modelo_logit))
13	Curva ROC modelo logístico	prob_logit <- predict(modelo_logit, type = "response"); roc_logit <- roc(datos\$ConversionAlta, prob_logit); plot(roc_logit, col = "blue", main = "Curva ROC - Regresión logística"); auc(roc_logit)
--- MODELO ÁRBOL DE DECISIÓN ---		
14	Ajustar árbol de decisión	modelo_arbol <- rpart(ConversionAlta ~ Engagement + TipoCampaña + Presupuesto + Region + MarcaFamosa, data = datos, method = "class", control = rpart.control(cp = 0.01))
15	Visualizar árbol	rpart.plot(modelo_arbol, type = 4, extra = 104)
16	Curva ROC modelo árbol	prob_arbol <- predict(modelo_arbol, type = "prob")[, "Si"]; roc_arbol <- roc(datos\$ConversionAlta, prob_arbol); plot(roc_arbol, col = "darkgreen", add = TRUE); legend("bottomright", legend = c("Logística", "Árbol"), col = c("blue", "darkgreen"), lwd = 2); auc(roc_arbol)

