



# Memoria Técnica

Miroslavna Kovylna Zabyako, Dácil Talón Ramírez, Javier Arderiu  
Franco, Javier González Medina, Miguel Benítez Montesinos

<b>Índice</b>	
<b>I. Introducción</b>	<b>4</b>
<b>II. Descripción técnica</b>	<b>5</b>
<b>II.1 Tarea 1: Análisis</b>	<b>5</b>
<b>II.2 Tarea 2: Clusterización</b>	<b>9</b>
<b>II.3 Tarea 3: Propensión de compra</b>	<b>10</b>
<b>II.4 Tarea 4: Personalización</b>	<b>12</b>
<b>II.5 Tarea 5: Seguimiento</b>	<b>13</b>
<b>II.6 Tarea 6: Coordinación</b>	<b>16</b>
<b>III. Conclusiones</b>	<b>17</b>
<b>IV. Referencias</b>	<b>17</b>

## **Agradecimientos**

### **Miroslavna:**

A todos los brillantes profesores de Nuclio, gracias por enseñarme, cada uno, una faceta única del mundo del Machine Learning.

Si estoy aquí, orgullosa de haber superado el intenso máster y el desafío de todos los entregables y el TFM, es porque siempre he estado apoyada por mis mágicos amigos.

A David, por ver más allá de las barreras y saber quién soy realmente.

A José, por quererme sin límites, siempre.

A Lucía, que supo sentir y darme justo lo que necesitaba en cada momento.

A Alicia, por ser la arquitecta de mi mente.

A Laura, que creyó en mí.

A Marina, Mónica y Cristina, por encontrar las palabras de apoyo y consuelo.

A José Miguel, que comparte conmigo cada una de mis aventuras intelectuales.

A Dácil, Javier, Xavi y Miguel, por ser compañeros de camino que confiaron plenamente en mí.

A Guillermo, por su pasión y valentía para desafiar mis ideas sobre lo que es importante y lo que no.

Y, sobre todo, a Juan. Por caminar a mi lado día tras día, por secar mis lágrimas y convertirlas en sonrisas, por pasar infinitas horas escuchándome y revisando conmigo los conceptos y el código mientras me transformaba en una científica de datos. Por ser mi compañero y mi sherpa en el fascinante viaje de subir juntos al Everest. Por conocer cada una de mis ideas y cómo nacían, mientras yo me enamoraba de patrones en los datos.

### **Dácil:**

A mis padres, por estar siempre a mi lado, apoyándome en todos mis proyectos, alentándome en los momentos difíciles y celebrando conmigo los éxitos. Y a mis hermanos, por creer siempre en mí. Muchas gracias también a todo el profesorado y a mis compañeros de grupo, Miroslavna, Javier, Miguel y Xavi, especialmente a Miroslavna por su gran trabajo y su apoyo incondicional.

### **Javier:**

Quiero expresar mi sincero agradecimiento a todas las personas que formaron parte de este proyecto, especialmente a Dácil, Miroslavna, Miguel y Xavi. Trabajar con un equipo tan comprometido y profesional ha sido un verdadero placer, y su esfuerzo ha sido clave para alcanzar los objetivos. En particular, agradezco a Miroslavna, cuya energía mantuvo al equipo unido. Finalmente, gracias a mi familia, siempre presentes y apoyándome. Gracias a todos por su dedicación, esfuerzo y compromiso.

### **Miguel:**

Al equipo educativo que nos ha acompañado en esta aventura, a mis compañeros de Máster y de TFM y a mi familia por apoyarme en mis proyectos personales y laborales.

## I. Introducción

Ser científico es observar y descubrir patrones. Ser científico de datos es ver patrones en los datos, traducirlos al lenguaje de las matemáticas, convertir esas matemáticas en herramientas de Machine Learning y, finalmente, crear soluciones prácticas<sup>1</sup>. Este enfoque no solo abre las puertas a la innovación, sino que también nos permite resolver desafíos complejos. En el presente TFM, hemos enfrentado uno de esos desafíos: analizar millones de registros de una plataforma financiera para diseñar una campaña de marketing por correo electrónico que maximice los ingresos.

Nuestra misión era clara: asumir el rol de un equipo de científicos de datos en la empresa easyMoney, con el objetivo de transformar datos dispersos en estrategias personalizadas que conecten con los clientes a nivel individual (Fig.1). En un entorno competitivo, una campaña genérica ya no es suficiente; para alcanzar resultados significativos, el marketing debe dirigirse directamente a cada cliente, comprender sus intereses y ofrecerle soluciones específicas y relevantes.



**Figura 1** Ciclo completo de análisis y optimización en easyMoney: desde la segmentación y personalización hasta la ejecución y seguimiento de campañas de marketing basadas en datos.

En el corazón de esta estrategia está la personalización, un elemento esencial para aumentar el impacto de cualquier acción comercial. Utilizando herramientas avanzadas de Machine Learning, hemos creado modelos que predicen qué cliente comprará, qué productos tienen mayor atractivo y qué características comparten los clientes con tendencias similares.

A continuación, hemos seleccionado 10.000 clientes que representan el mayor potencial de ingresos. Los modelos no solo han permitido identificar patrones de comportamiento, sino también segmentar a los clientes en clusters bien definidos. La segmentación es clave para diseñar campañas que optimicen la comunicación y potencien la rentabilidad. Cada cluster refleja un perfil único, desde jóvenes ahorradores hasta mujeres emprendedoras interesadas en productos financieros avanzados.

Basándose en esta selección, se han diseñado mensajes adaptados a cada grupo, asegurando que el contenido sea relevante y que las probabilidades de conversión aumenten significativamente.

La capacidad de predecir comportamientos futuros nos ha llevado a crear un marco analítico sólido. Este marco no solo identifica oportunidades, sino que también evalúa su impacto económico mediante indicadores clave de rendimiento (KPIs). Estos KPIs permiten medir desde la tasa de apertura de correos hasta el retorno de inversión (ROI) de la campaña.

Sin embargo, este proyecto no se limita al análisis y la predicción. También hemos diseñado estrategias de seguimiento que permitirán optimizar futuras campañas. Estos pasos aseguran que la empresa pueda continuar mejorando sus resultados y adaptándose a las necesidades cambiantes del mercado.

El presente TFM demuestra que la ciencia de datos no solo responde preguntas, sino que también abre nuevas posibilidades <sup>1</sup>. A través de técnicas avanzadas, creatividad y rigor, hemos transformado datos en conocimiento y conocimiento en acción. En las páginas siguientes, se detallan los procesos, herramientas y resultados que han dado vida a esta solución.

## **II. Descripción técnica**

En este proyecto se han importado librerías para optimizar el desarrollo del código y se han creado funciones reutilizables a lo largo del proceso <sup>2-4</sup>. También se han diseñado diccionarios que organizan datos geográficos, económicos y auxiliares relacionados con género, edad y salario, facilitando el análisis <sup>5,6</sup>.

Los datos provienen de tres archivos divididos en sectores: sociodemográficos, actividad comercial y productos. Estos se consolidaron en un único DataFrame de pandas que recoge toda la información sobre los clientes, optimizado para reducir el uso de memoria al conservar sólo las columnas relevantes.

Durante la preparación, se eliminaron columnas irrelevantes como 'Unnamed' y productos no adquiridos, así como las filas correspondientes a clientes fallecidos. Además, se convirtió la columna 'gender' a formato binario y se simplificaron las categorías en 'segment'.

Finalmente, se verificaron duplicados y se confirmó que ninguna fila superaba un umbral del 20% de valores nulos. Se establecieron puntos de restauración para garantizar flexibilidad durante el análisis.

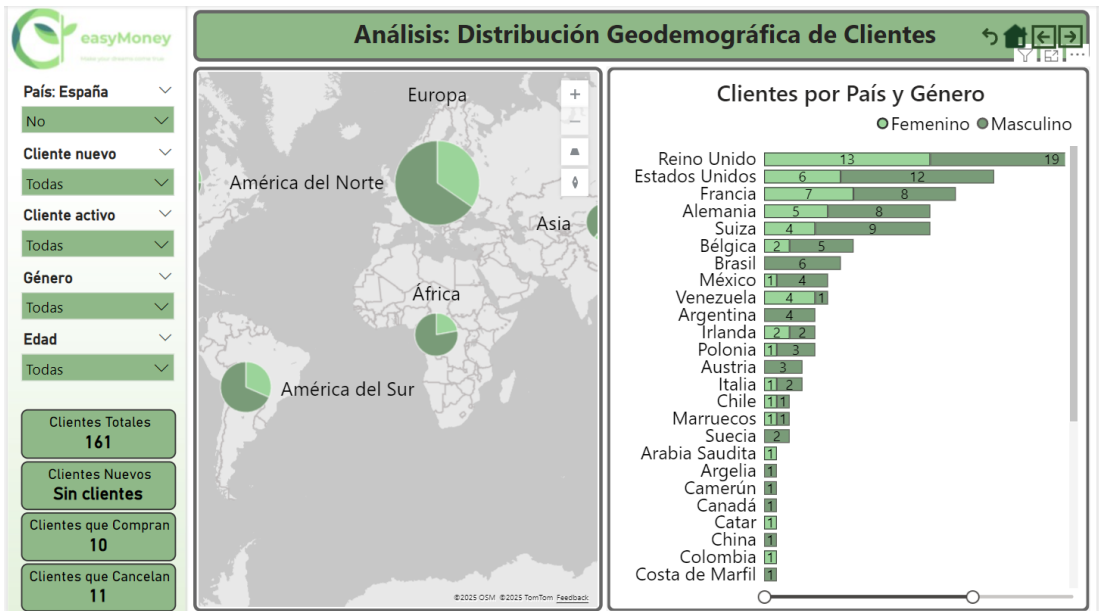
### **II.1 Tarea 1: Análisis**

El Análisis Exploratorio de Datos (EDA) es crucial para comprender la estructura y características de los datos, identificar patrones y preparar la información para un análisis más profundo. En este proyecto, el EDA se centró en evaluar la calidad de los datos, corregir inconsistencias y crear variables útiles para la modelización.

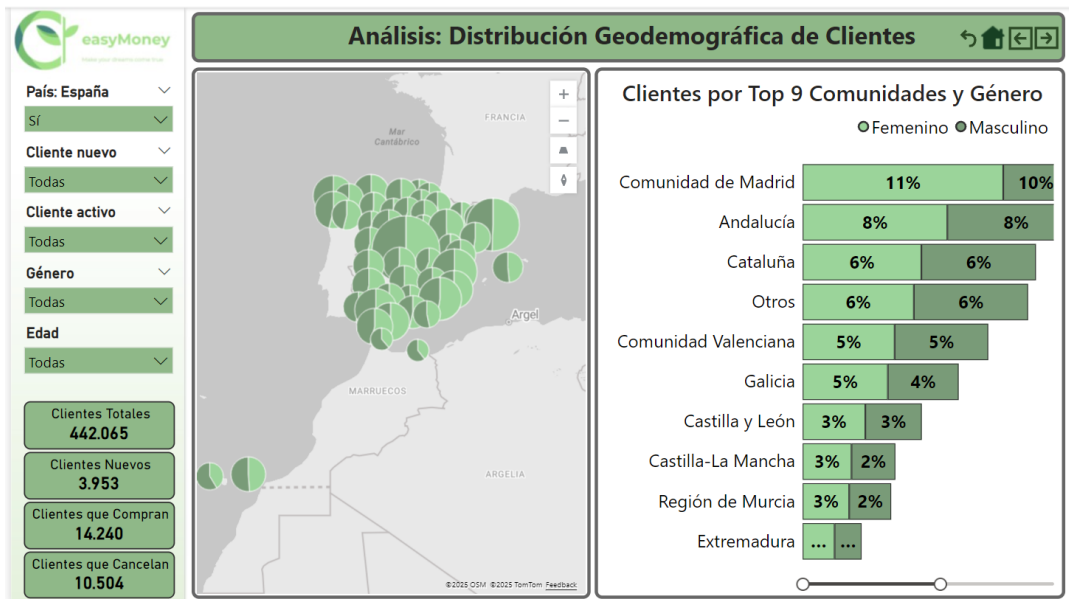
Primero, se analizaron y gestionaron los valores nulos, rellenando los datos faltantes con información de clientes en otras fechas, priorizando los registros más recientes siempre que fue posible. En la columna region\_code, los valores nulos se reemplazaron por "0", indicando clientes internacionales. En las columnas gender\_male, payroll y pension\_plan, se contabilizaron y eliminaron las filas con valores nulos, tras lo cual el DataFrame fue reindexado. Además, se aplicaron los métodos bfill y ffill para completar valores en las columnas entry\_channel y segment, asignando "Desconocido" a los casos en los que no se pudo rellenar la información.

Las fechas se revisaron para garantizar consistencia, asegurando que la fecha de contratación no fuera posterior a la de partición. Se generaron variables temporales como month\_from\_contract, last\_year y last\_6\_months para un análisis más detallado. También se analizaron patrones en las contrataciones por días, semanas y meses, destacando un promedio de alta actividad en miércoles y hacia la segunda mitad del año.

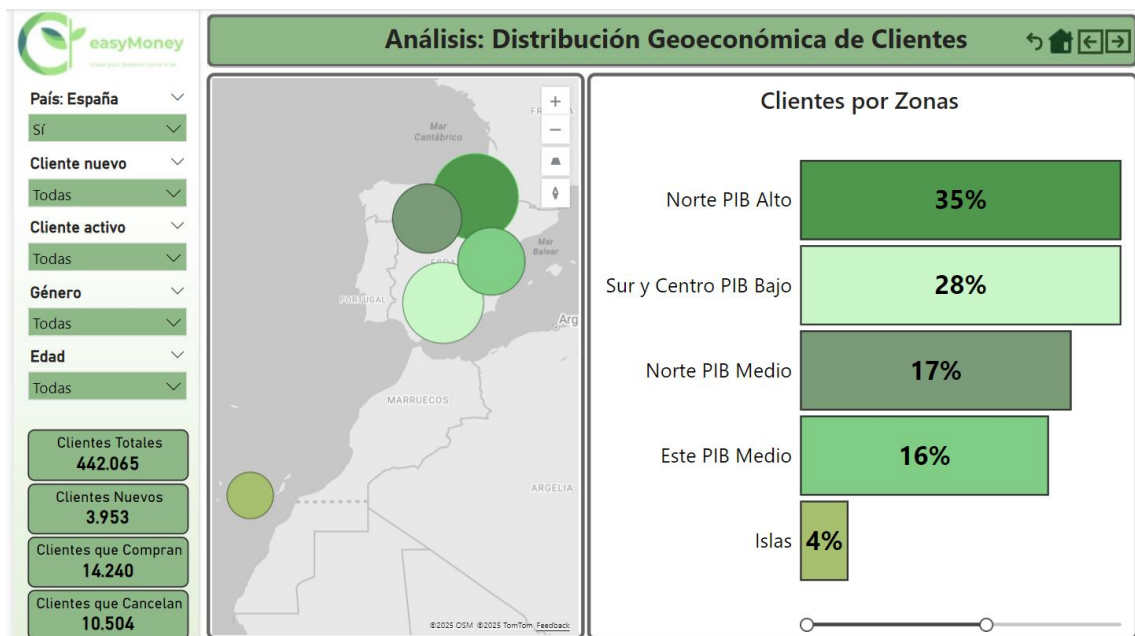
Se renombró la columna country\_id a country\_name y se asignaron nombres a los códigos de país, diferenciando entre clientes nacionales y extranjeros. Se añadieron columnas como continent, Zone, y region para agrupar provincias y países en categorías relevantes, consolidando las comunidades autónomas principales en comunidad\_top9 para simplificar la segmentación (Fig. 2-4).



**Figura 2** Análisis Geodemográfico: Clientes por Continente, País y Género.

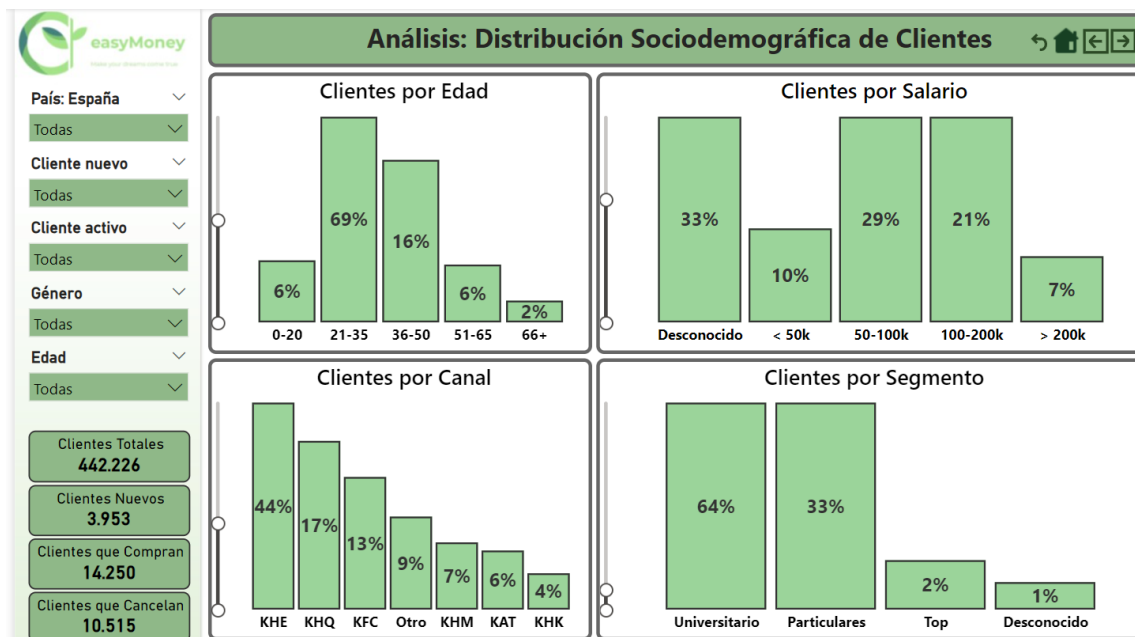


**Figura 3** Análisis Geodemográfico: Clientes en España por Región y Género.



**Figura 4** Análisis Geoeconómico: Distribución de Clientes por Zonas y Niveles de PIB.

Las variables demográficas y financieras también fueron consideradas para ver cómo se distribuyen los clientes (Fig. 5 y 6). La columna age se categorizó en rangos de edad (age\_range), y un proceso similar se aplicó a la columna salary para generar salary\_range. Asimismo, los canales de entrada se redujeron a seis categorías clave en entry\_channel\_range. Por último, se identificaron clientes nuevos en la columna cliente\_nuevo y los productos financieros se agruparon en 3 categorías principales: cuentas, ahorros e inversiones y financiación.



**Figura 5** Análisis Sociodemográfico: Clientes por Edad, Salario, Canal y Segmento.

Se nos solicitó analizar las ventas del último mes (Fig.7) para identificar los productos adquiridos, incluyendo aquellos comprados por clientes nuevos. Para ello, se realizaron las operaciones correspondientes y se preparó un archivo CSV, diseñado para su posterior visualización y análisis en Power BI.

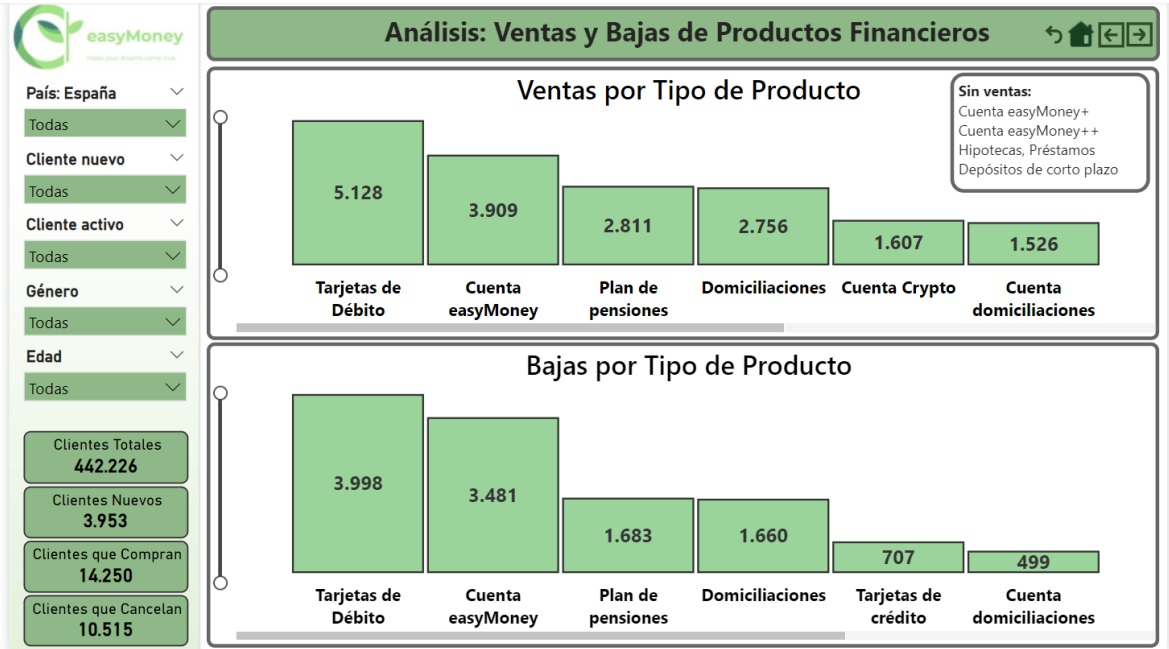


Figura 6 Análisis de Productos Financieros: Ventas y Bajas por Tipo de Producto.

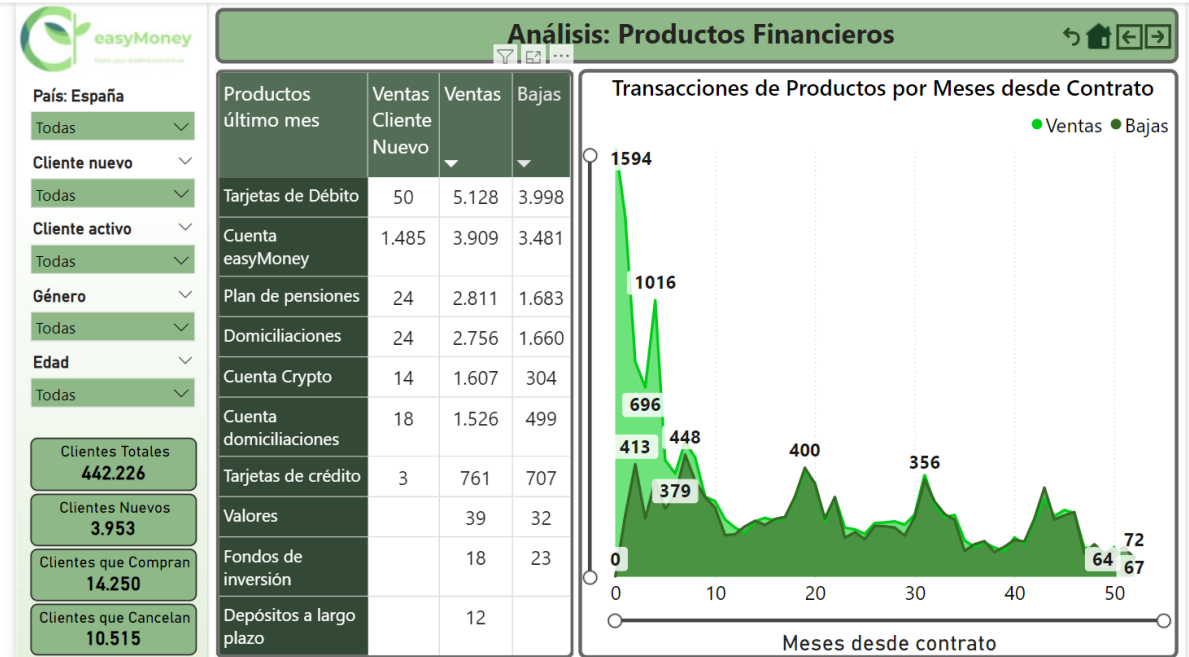


Figura 7 Análisis Temporal de Productos Financieros: Ventas y Bajas.



## II.2. Tarea 2: Clusterización

La clusterización agrupa datos en categorías homogéneas y es clave para seleccionar variables que maximicen el beneficio del cliente. En el caso de easyMoney, el análisis se enfocó en las ventas de productos financieros y la segmentación de clientes con comportamientos de compra similares, optimizando las estrategias comerciales.

El proceso comenzó con la selección de variables relevantes y la preparación del DataFrame. Se transformaron variables categóricas como segment, salary\_range y zone para adaptarlas al análisis, y se generaron histogramas para visualizar la distribución de los datos. Posteriormente, se realizó un análisis preliminar con el algoritmo KMeans <sup>7</sup>, evaluando diversas combinaciones de variables y escaladores para determinar la configuración óptima.

Las variables clave incluyeron género, edad, actividad en la plataforma, tiempo desde el contrato, segmento, nivel salarial, zonas geoeconómicas y ventas agrupadas en categorías como cuentas, ahorro e inversión, y financiación. Estas categorías se analizaron tanto para el último año como para los últimos seis meses, incluyendo datos de cancelaciones. Para garantizar una clusterización precisa basada en distancias euclidianas, se probaron diferentes escaladores (StandardScaler, RobustScaler y MinMaxScaler) <sup>8</sup>. Finalmente, StandardScaler se seleccionó como el escalador óptimo, ya que homogeneizó las escalas sin comprometer la calidad del agrupamiento.

Dado que se solicitó agrupar a los clientes en 7-8 clusters, la clusterización óptima para easyMoney consistió en 8 grupos (Tabla 1).

Estos clusters integraron variables clave como género, edad, meses desde el contrato, actividad en la plataforma y ventas anuales por tipo de producto. Aunque el PIB ofreció una leve diferenciación, su impacto fue menor, mientras que variables como zonas geoeconómicas y niveles salariales resultaron poco útiles para este análisis.

Nombre	Universitarias				Universitarios				Público Estable				Clientes Nuevos				Ahorro Joven				Top Activos				Clientes Premium				Créditos VIP			
Nivel de consumo	0				0				0				1				1				2				3				3			
Número de clientes	123.957				92.457				122.336				3.988				57.187				7.151				20.495				14.655			
Género	Mujer				Hombre				Hombres 57%				Hombres 56%				Hombres 45%				Hombres 58%				Hombres 54%				Hombres 60%			
Edad	24				24				42				40				22				54				37				36			
Actividad en plataforma (%)	20				21				51				35				42				97				93				92			
Meses desde contrato	29				29				23				0				9				20				21				28			
PIB (€)	26.330				26.388				27.743				27.625				25.853				29.994				27.885				27.218			
Segmento	Universitario				Universitario				Particulares				Desconocido				Universitario				Top				Particulares 72%				Particulares 63%			
Ventas productos	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%	mean	25%	50%	75%
Cuentas	0	0	0	0	0.01	0	0	0	0.27	0	0	0	0.38	0	0	1	1.03	1	1	1	0.8	0	1	1	1.5	1	1	2	0.55	0	0	1
Ahorro e inversión	0	0	0	0	0	0	0	0	0.02	0	0	0	0.01	0	0	0	0.01	0	0	0	0.45	0	0	1	1.4	1	1	2	0.17	0	0	0
Financiación	0.02	0	0	0	0.03	0	0	0	0.10	0	0	0	0.01	0	0	0	0	0	0	0	0.22	0	0	0	0.5	0	0	1	2.6	2	2	3
Productos totales	0.02	0	0	0	0.04	0	0	0	0.39	0	0	1	0.40	0	0	1	1.10	1	1	1	1.5	0	1	2	3.4	3	3	4	3.3	2	3	4
Productos cancelados	0.04	0	0	0	0.05	0	0	0	0.19	0	0	0	0	0	0	0	0.14	0	0	0	1	0	0	1	2.4	1	2	3	3.0	2	3	4
Comportamiento	Hacen compras limitadas, principalmente de productos de financiación, muestran baja actividad en la plataforma y mínima participación en productos de ahorro e inversión				Efectúan compras mínimas, centradas en productos de financiación, aunque mantienen baja actividad en la plataforma y no participan en productos de ahorro o inversión				Adquieren productos de forma limitada, principalmente cuentas, con una participación baja en productos de ahorro o financiación				En sus inicios, su comportamiento se centra exclusivamente en la adquisición de cuentas, sin interés en otros productos financieros o de inversión				Realizan compras moderadas, centradas en cuentas y productos de ahorro, con una participación baja en productos de financiación				Extremadamente activos, compran cuentas y participan en productos de ahorro e inversión, reflejando un perfil diversificado				Muy activos en la plataforma, adquieren cuentas y productos de ahorro e inversión, además de tener un interés creciente en financiación				Altamente enfocados en productos de financiación, con menor interés en cuentas u otros productos			

**Tabla 1** Clusterización I de clientes.

Se generaron múltiples modelos de clusterización, cuyos resultados se analizaron y visualizaron para iterar y mejorar las configuraciones. Los mejores resultados se obtuvieron en el segundo test, y los datos finales se almacenaron en un archivo CSV para su posterior análisis y presentación en Power BI.

### II.3. Tarea 3: Propensión de compra

La propensión de compra (Fig. 8) estima la probabilidad de que un cliente realice una compra en el futuro, basándose en el análisis de datos históricos, patrones de comportamiento y características del cliente, como historial de compras, interacción con la marca, preferencias y demografía. Este concepto es esencial en marketing predictivo, ya que permite identificar clientes con mayor intención de compra, optimizar recursos y aumentar la efectividad de las campañas comerciales.

El análisis comenzó con la preparación del DataFrame, creando columnas objetivo que indican si un cliente ha realizado compras en categorías clave: cuentas, ahorro e inversión, o financiación. Estas columnas binarias se generaron a partir de variables existentes, asignando un valor de 1 si el cliente realizó al menos una compra en la categoría correspondiente, y 0 en caso contrario. El enfoque escogido facilitó la identificación de patrones de comportamiento y el entrenamiento de modelos predictivos.

Para enriquecer el análisis, se generaron variables cíclicas relacionadas con el mes, lo que permitió capturar patrones temporales relevantes. Las variables numéricas se analizaron para evaluar su correlación con las columnas objetivo; como resultado, la variable `month_from_contract` fue descartada debido a su redundancia con `entry_year` y su menor correlación con las variables objetivo. Las variables categóricas también fueron evaluadas en función de su relación con las columnas objetivo, destacándose `comunidad_top9` sobre `Zone`, lo que llevó a excluir esta última por su menor relevancia.

Posteriormente, se filtraron los datos del último mes, enfocándose en los clientes activos en la plataforma. Para garantizar un análisis más preciso, se estableció un umbral de 50% de probabilidad de compra, asignando ganancias de cero a cualquier cliente cuya probabilidad estuviera por debajo de este límite. Además, solo se incluyeron clientes que no habían adquirido previamente el grupo de productos analizado, permitiendo así identificar oportunidades de adquisición de nuevos productos.

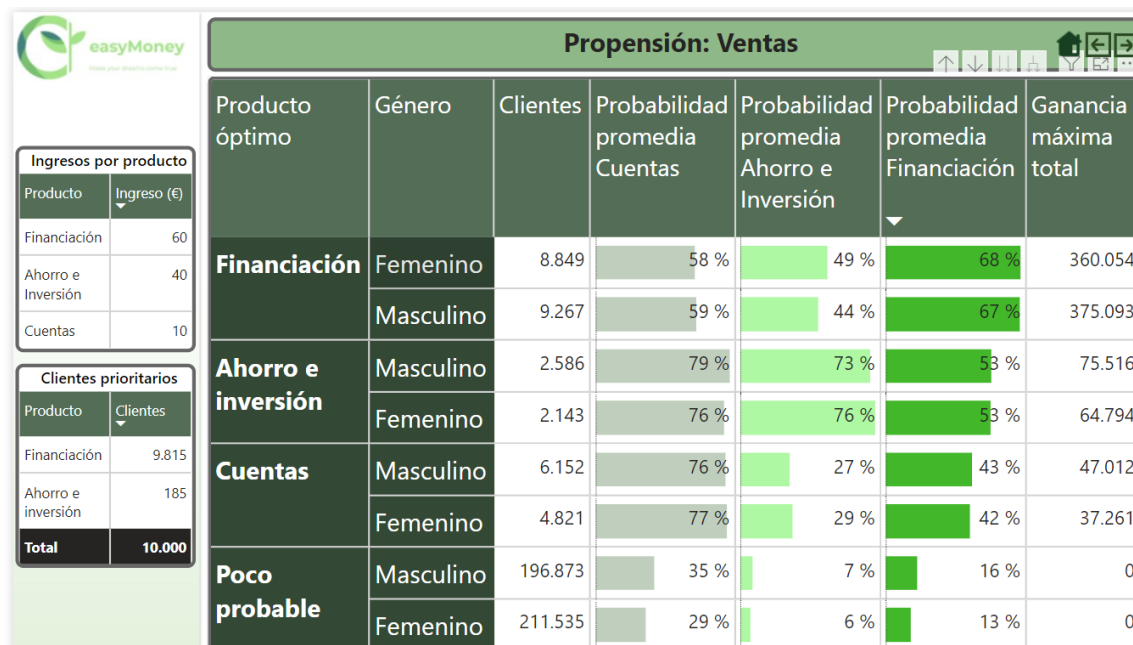
Finalmente, se aplicó One-Hot Encoding a las variables categóricas, asegurando claridad y consistencia en los nombres de las columnas, y optimizando la estructura del DataFrame para el análisis y modelado posterior.

Durante el preprocesamiento, se dividió el conjunto de datos en entrenamiento y prueba. Se detectó un desbalance significativo en las variables objetivo, especialmente en las categorías de ahorro e inversión y financiación, donde la clase positiva representaba un porcentaje reducido del total. Este desequilibrio podría impactar negativamente en el aprendizaje de los modelos, ya que tenderían a favorecer la clase mayoritaria. Para abordar este problema, se evaluaron cinco técnicas de balanceo de datos <sup>9</sup>:

1. **Submuestreo aleatorio (Undersampling):** Esta técnica reduce aleatoriamente las muestras de la clase mayoritaria para equilibrar ambas clases. Aunque eficaz para evitar el sesgo hacia la clase mayoritaria, tiene el riesgo de eliminar datos valiosos que podrían ser útiles para el aprendizaje del modelo.
2. **Sobremuestreo con SMOTE (Synthetic Minority Oversampling Technique):** SMOTE genera nuevas muestras sintéticas para la clase minoritaria mediante combinaciones lineales de ejemplos cercanos en el espacio de características. Esto aumenta la representación de la clase minoritaria sin duplicar datos existentes y reduce el riesgo de sobreajuste.

3. **Muestreo mixto:** Combina SMOTE para aumentar las muestras de la clase minoritaria con submuestreo para reducir el tamaño de la clase mayoritaria. Este enfoque equilibra ambas clases y mitiga las limitaciones individuales de cada técnica.
4. **SMOTEENN:** Esta técnica combina SMOTE con ENN (Edited Nearest Neighbors), que elimina muestras inconsistentes o ruidosas después del sobremuestreo. Este enfoque mejora la calidad y limpieza de los datos, asegurando que el modelo aprenda de ejemplos representativos.
5. **SMOTETomek:** Combina SMOTE con Tomek Links, que identifica y elimina pares de instancias ambiguas cercanas en el espacio de características. Esto permite al modelo aprender límites más claros entre las clases, mejorando su capacidad de generalización.

Cada técnica fue evaluada utilizando métricas clave como accuracy, precision, recall, f1\_score y ROC AUC, priorizando especialmente el recall, dado que el objetivo principal es captar el mayor número posible de clientes interesados. Sin embargo, también se consideró la precisión, para garantizar que una alta proporción de las predicciones positivas correspondiera efectivamente a clientes potenciales, equilibrando así la eficacia y el uso de recursos.



**Figura 8** Propensión de Compra.

Se probaron diversos modelos predictivos, priorizando el recall, ROC AUC y precisión como métricas fundamentales. El modelo CatBoostClassifier con undersampling destacó en la categoría de cuentas, alcanzando un recall de 0.8856, un ROC AUC de 0.9643 y una precisión de 0.8216, logrando un balance ideal entre estas métricas. Para ahorro e inversión, el modelo CatBoostClassifier con SMOTEENN obtuvo un recall de 0.8137, un ROC AUC de 0.9773 y una precisión de 0.7717. En la categoría de financiación, el LGBMClassifier con SMOTEENN fue la mejor opción, logrando un recall de 0.6863, un ROC AUC de 0.9339 y una precisión de 0.6737.

Posteriormente, se llevó a cabo la hiperparametrización de los modelos para optimizar su rendimiento en las tres categorías. Se seleccionaron los 10.000 clientes con mayor

potencial de ingresos, calculados como el producto de la probabilidad de compra por el ingreso estimado para cada categoría: 10 euros para cuentas, 40 euros para ahorro e inversión, y 60 euros para financiación.

A partir de estos resultados, se procedió al cálculo del ROI (Retorno de la Inversión), considerando variables clave como el porcentaje de rebote en correos, tasas de apertura y el coste por envío. Este análisis se realizó para tres escenarios de ROI (base, optimista y pesimista), evaluando cómo las variaciones en las tasas de rebote, apertura y costes afectan los resultados. Este enfoque proporciona una visión integral del rendimiento de la campaña, permitiendo ajustar estrategias y optimizar recursos para maximizar el impacto económico.

## II.4. Tarea 4: Personalización

Para personalizar la comunicación y hacerla más efectiva, es fundamental segmentar a los 10.000 clientes seleccionados en 4 o 5 grupos (Tabla 2), incrementando así el potencial de generar ganancias. La segmentación permite adaptar los mensajes y estrategias a las características específicas de cada grupo, optimizando la relevancia de las campañas, mejorando la experiencia del cliente y maximizando el retorno de inversión.

Nombre de cluster	Emprendedores	Emprendedoras	Universitarios	Ahorradores
Número de clientes	3.492	3.320	3.003	185
Género	Hombre	Mujer	Hombres 49%	Hombres 37%
Edad	40	40	28	30
Segmento	Particulares	Particulares	Universitario	Universitario 64%
Compra de producto financiero en el último año	No	No	No	25% al menos una compra
Producto óptimo	Financiación	Financiación	Financiación	Ahorro e inversión
Descripción clientes	Clientes hombres de 40 años en promedio del segmento particulares, sin compras recientes de productos financieros	Clientes mujeres de 40 años en promedio del segmento particulares, sin compras recientes de productos financieros	Clientes jóvenes universitarios de 28 años en promedio, sin compras recientes de productos financieros	Clientes jóvenes, en su mayoría universitarios, con un promedio de 30 años, algunos de los cuales ya han adquirido productos financieros
Ideas para atraer atención	Financiar proyectos personales (remodelación, coche, casa), seguros asociados, tasas preferenciales y herramientas digitales para simulación de préstamos	Apoyo a negocios liderados por mujeres, financiación familiar, transición laboral, y productos sostenibles (vehículos eléctricos, paneles solares)	Préstamos educativos, apoyo a startups, intercambios internacionales y recompensas por contratación de productos financieros	Planes de ahorro con propósito, bonificaciones por objetivos, ahorro flexible y herramientas de automatización de ahorro

**Tabla 2** Clusterización II de clientes.

El proceso de clusterización comienza con la selección de variables relevantes, un paso clave para comprender las necesidades y comportamientos de los clientes. Estas variables incluyen factores sociodemográficos como la edad, género, nivel educativo y lugar de residencia, que proporcionan una visión detallada de las características de los clientes. Además, se considera el sector al que pertenece cada cliente, lo que ayuda a anticipar sus intereses y preferencias. La capacidad económica, medida a través de los ingresos, influye directamente en los productos financieros que pueden adquirir. El análisis de los productos ya contratados permite comprender el comportamiento actual

del cliente y prever posibles necesidades futuras. Otra variable clave son los productos óptimos, identificados mediante modelos de propensión de compra, que reflejan las opciones más adecuadas para cada cliente.

Por último, la ganancia máxima, calculada como la probabilidad de que un cliente adquiera un producto específico multiplicada por el beneficio esperado según el tipo de producto, optimiza la selección del producto ideal. Este indicador, basado en modelos predictivos de Machine Learning, asegura que las campañas se alineen con los intereses y necesidades de los clientes, maximizando tanto su impacto como la rentabilidad.

Se han diseñado cuatro plantillas de correo electrónico personalizadas según los clusters de clientes identificados, integrando propuestas atractivas y alineadas con las necesidades específicas de cada grupo. Estas plantillas tienen como objetivo maximizar la efectividad de la comunicación, destacando soluciones relevantes y acordes con las aspiraciones de los clientes.

La **primera plantilla**, dirigida al cluster de emprendedores, presenta soluciones financieras innovadoras diseñadas para impulsar el inicio de nuevos negocios y apoyar el progreso profesional. También incluye opciones para financiar proyectos personales, como la remodelación del hogar o la adquisición de vehículos, respaldadas por herramientas digitales avanzadas que inspiran decisiones inteligentes y estratégicas.

La **segunda plantilla**, pensada para mujeres emprendedoras, se enfoca en fomentar el inicio o expansión de negocios, así como en respaldar proyectos sostenibles, como la adquisición de vehículos eléctricos o paneles solares. Además, incluye opciones de apoyo en momentos clave de la vida, como tratamientos médicos, la construcción de una familia o la educación de los hijos, ayudando a convertir cada reto en un paso hacia el éxito.

La **tercera plantilla**, diseñada para universitarios, ofrece opciones de financiación enfocadas en su desarrollo académico y profesional. Esto incluye apoyo para su educación, intercambios internacionales y el lanzamiento de proyectos innovadores o startups. Estas propuestas están orientadas a ayudarles a construir un futuro sólido y convertir sus ideas en realidades transformadoras.

Finalmente, la **cuarta plantilla**, dirigida a ahorradores estratégicos, se centra en proporcionar beneficios exclusivos, acceso flexible a sus fondos y herramientas automatizadas que simplifican el proceso de ahorro. Estas ofertas están diseñadas para ayudarles a alcanzar sus metas personales, educativas o profesionales de manera eficiente y sin complicaciones.

## II.5. Tarea 5: Seguimiento

Tras el envío masivo de correos, es fundamental realizar un seguimiento detallado de la campaña de marketing para evaluar su efectividad. En el caso de easyMoney, los indicadores clave de desempeño (KPIs) se dividen en dos categorías principales <sup>10,11</sup>: interacción y beneficios económicos, proporcionando una visión integral del impacto de la campaña.

Los KPIs de interacción (Tabla 3) miden la respuesta de los usuarios a los correos enviados y ofrecen información crucial para evaluar el alcance de la campaña. Incluyen la tasa de rebote, que refleja el porcentaje de correos que no llegaron a la bandeja de entrada, señalando posibles problemas en la calidad de las direcciones o el sistema de envío. La tasa de apertura mide el porcentaje de destinatarios que abrieron los correos, un indicador clave para evaluar el atractivo del asunto y el contenido inicial. La tasa de clics calcula la proporción de usuarios que hicieron clic en los enlaces, lo que indica el interés en las ofertas presentadas. También se incluyen la tasa de interacción con simuladores financieros y la tasa de apertura de simuladores de ahorro, que evalúan el

uso de herramientas específicas proporcionadas en los correos, ofreciendo una visión detallada de la participación del cliente.

KPI	Fórmula	Importancia	Escenario		
			Pesimista	Base	Optimista
Tasa de rebote	$\frac{\text{Correos rebotados}}{\text{Correos enviados}} * 100$	Impacta la efectividad de la campaña al influir en la cantidad de correos que llegan a los destinatarios	1,5%	0,70%	0,3%
Tasa de apertura de correos	$\frac{\text{Correos abiertos}}{\text{Correos enviados}} * 100$	Evalúa el interés inicial por la campaña y la efectividad del asunto del correo	15%	22,5%	30%
Tasa de clics	$\frac{\text{Clics en el enlace}}{\text{Correos abiertos}} * 100$	Mide la interacción del cliente con el contenido del correo	2%	3%	5%
Tasa de interacción con simuladores financieros	$\frac{\text{Usuarios de simulador financiero}}{\text{Usuarios totales alcanzados}} * 100$	Mide el nivel de interés en herramientas digitales útiles para la decisión financiera post-campaña	5%	10%	20%
Tasa de apertura de simuladores de ahorro	$\frac{\text{Usuarios de simulador ahorro/inversión}}{\text{Usuarios totales alcanzados}} * 100$	Evalúa la atracción que generan los simuladores de ahorro e inversión post-campaña	10%	15%	25%

**Tabla 3** KPIs de interacción.

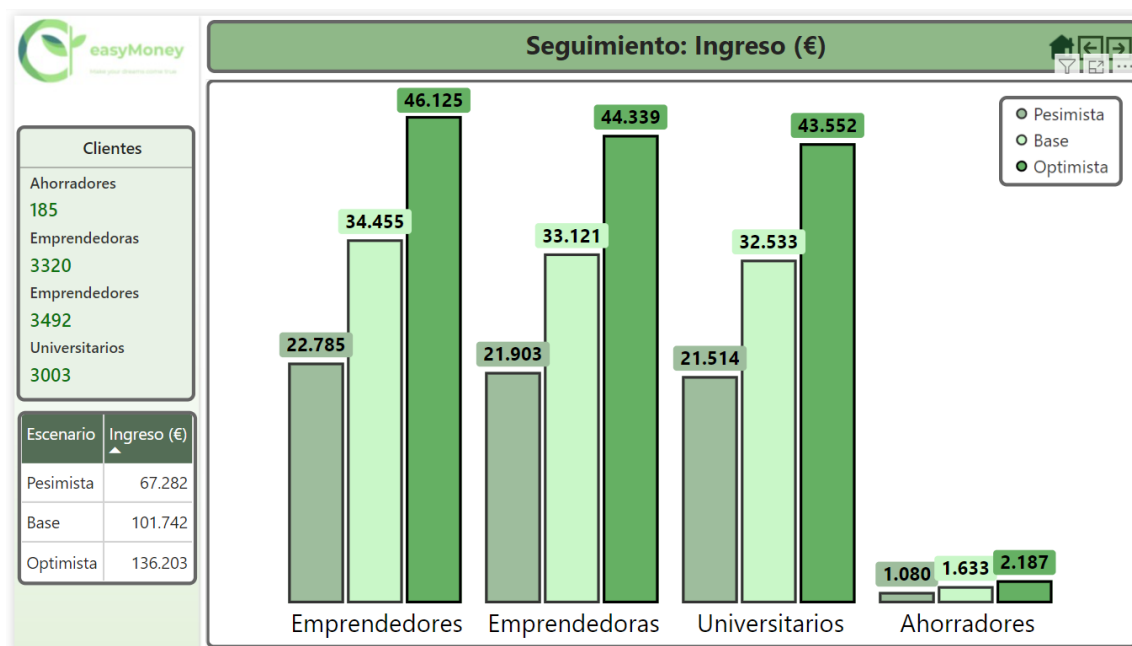
KPI	Fórmula	Importancia	Escenario		
			Pesimista	Base	Optimista
Tasa de adquisición de productos financieros (según propensión de compra)	$\frac{\text{Compras de productos financieros}}{\text{Correos enviados}} * 100$	Mide la eficacia general de la campaña en la venta de productos financieros	76.08%		
Tasa de adquisición de productos de ahorro e inversión (según propensión de compra)	$\frac{\text{Compras de ahorro e inversión}}{\text{Correos enviados}} * 100$	Indica qué tan efectiva es la campaña para captar clientes interesados en productos de ahorro e inversión	98.79%		
Ingresos generados por productos financieros	Productos vendidos <sub>F</sub> * Beneficio por producto <sub>F</sub>	Mide la efectividad global de la campaña para vender productos financieros	66.202 €	100.109 €	134.016 €
Ingresos generados por productos de ahorro e inversión	Productos vendidos <sub>A/I</sub> * Beneficio por producto <sub>A/I</sub>	Evalúa la efectividad económica de la campaña en la promoción de productos de ahorro e inversión	1.080 €	1.633 €	2.187 €
ROI por cluster (ej. Ahorradores)	$\frac{\text{Ingresos por cluster} - \text{Costes por cluster}}{\text{Costes por cluster}} * 100$	Mide la rentabilidad de los correos personalizados por segmento para optimizar las futuras campañas	58286%	882805%	1313177%
Ingresos totales generados	Productos vendidos * Beneficio por producto	Determina el impacto financiero directo de la campaña en términos de ingresos	67.282 €	101.742 €	136.203 €
ROI de la campaña	$\frac{\text{Ingresos totales} - \text{Costes totales}}{\text{Costes totales}} * 100$	Mide la rentabilidad general de la campaña en comparación con su coste	67182%	1017320%	1513267%

**Tabla 4** KPIs de beneficios económicos.

Los KPIs de beneficios económicos (Tabla 4), por su parte, se centran en medir el impacto financiero de la campaña. La tasa de adquisición de productos financieros

calcula el porcentaje de clientes que compraron productos tras la campaña, mientras que la tasa de adquisición de productos de ahorro e inversión analiza específicamente la efectividad en esta categoría. Los ingresos generados por productos financieros cuantifican los beneficios totales obtenidos a través de las compras, proporcionando una métrica clara del impacto directo de la campaña. Además, el ROI por cluster evalúa el retorno de inversión a nivel de cada segmento de clientes, lo que permite identificar los grupos más rentables. Finalmente, el ROI global de la campaña integra el impacto financiero total considerando todos los segmentos, ofreciendo una visión completa del desempeño económico.

Los indicadores clave de rendimiento (KPIs) se analizan bajo tres escenarios: pesimista, base y optimista. Estos escenarios reflejan tanto los gastos como las interacciones de los clientes, subrayando cómo la segmentación y la personalización potencian los resultados. En el escenario optimista, el segmento de emprendedores lidera con ingresos proyectados de hasta 46.125 €, seguido de los segmentos de emprendedoras y universitarios, quienes también generan contribuciones significativas. Por otro lado, aunque el segmento de ahorradores muestra ingresos más modestos, alcanzando un máximo de 2.187 € en el escenario optimista, sigue siendo rentable. Esto se debe a que, a pesar de su menor aporte absoluto, la probabilidad de compra dentro de este grupo es significativamente mayor en comparación con otros clientes que no están incluidos en los 10.000 seleccionados (Fig. 9).



**Figura 9** Seguimiento de Ingresos: Comparación de Escenarios (Pesimista, Base, Optimista) por Segmento de Clientes.

Este análisis pone de relieve la importancia de identificar y priorizar segmentos específicos para maximizar la rentabilidad. La segmentación y personalización no solo incrementan la efectividad de las campañas, sino que también optimizan el uso de recursos, asegurando que los esfuerzos estén dirigidos hacia los clientes con mayor potencial de conversión y beneficio económico.



## II.6. Tarea 6: Coordinación

Al final, se establece la coordinación de los próximos pasos y la planificación a futuro para optimizar las campañas de e-marketing, estructurando un flujo de trabajo en varias etapas clave para garantizar el éxito de las estrategias implementadas (Fig. 10).

El primer paso consiste en implementar un proceso ETL (Extracción, Transformación y Carga) para recopilar y procesar datos, entrenar modelos predictivos, calcular la propensión de compra y segmentar a los clientes en clusters relevantes. Esto asegura que las bases de datos estén limpias, actualizadas y preparadas para el análisis.

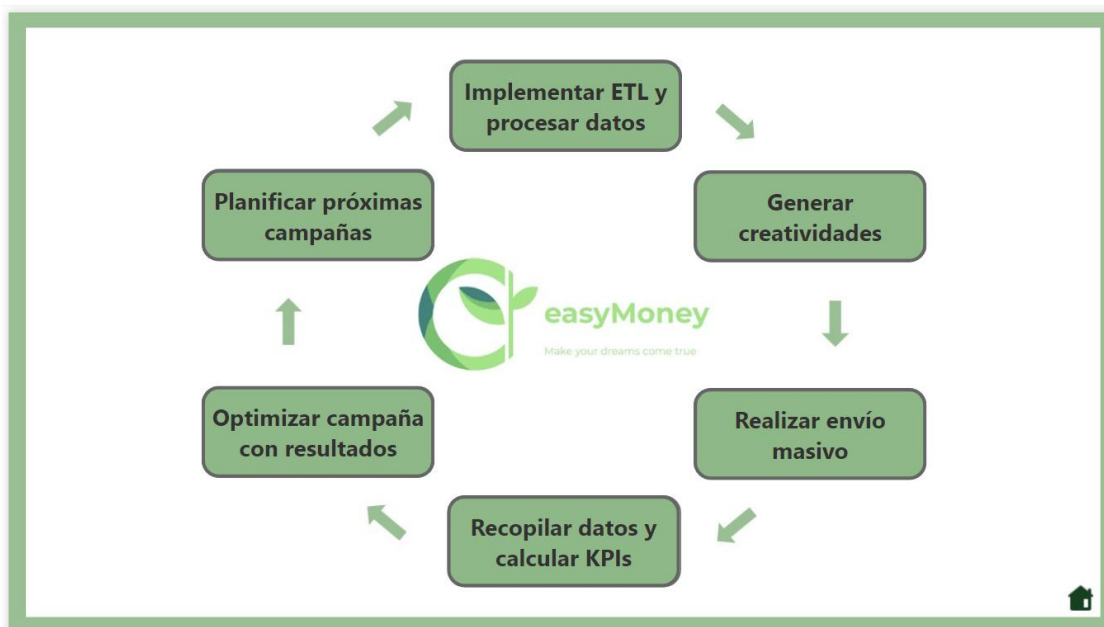
A partir de esta segmentación, se generan creatividades personalizadas para cada grupo de clientes, diseñando contenido atractivo que se alinee con las necesidades, intereses y preferencias de cada segmento. Este enfoque personalizado aumenta las posibilidades de captar la atención de los clientes y mejorar las tasas de conversión.

Posteriormente, se realiza el envío masivo de correos electrónicos segmentados, programados estratégicamente para optimizar la interacción y las conversiones. El tiempo y la frecuencia de estos envíos se calculan cuidadosamente para maximizar su impacto.

Tras el envío, se recopilan datos detallados para calcular los principales KPIs de la campaña, como tasas de apertura, clics, conversiones y otros indicadores clave. Estos resultados permiten medir el rendimiento de la campaña y sirven de base para identificar áreas de mejora.

Con los datos recopilados, se analizan los resultados en tiempo real para ajustar las estrategias y garantizar una mejora continua en el impacto y el retorno de inversión (ROI). Esto incluye optimizar tanto los mensajes como la segmentación para futuros envíos.

Finalmente, los aprendizajes obtenidos se incorporan a la planificación de campañas futuras, perfeccionando las estrategias a partir de los resultados anteriores y asegurando una mayor efectividad en las acciones futuras. Este enfoque iterativo permite que las campañas de e-marketing evolucionen y generen un impacto creciente en la rentabilidad de la empresa.



**Figura 10** Flujo del Proceso de Coordinación en Campañas de Marketing.



### III. Conclusiones

El presente trabajo final de máster demuestra cómo la ciencia de datos puede transformar datos dispersos en estrategias prácticas y efectivas, logrando resolver desafíos complejos en el ámbito financiero <sup>12</sup>. A través de técnicas avanzadas de análisis, clusterización y modelado predictivo, se diseñaron campañas de marketing personalizadas que no solo optimizan el retorno de inversión, sino que también conectan con los clientes a un nivel más personal. La segmentación y personalización han demostrado ser herramientas clave para maximizar la efectividad y eficiencia de los recursos, generando un impacto positivo tanto en los ingresos como en la experiencia del cliente.

Además, la metodología implementada destaca la importancia de una estrategia integrada que combina la preparación y limpieza de datos, el uso de herramientas de Machine Learning y el análisis continuo de resultados. Este enfoque iterativo permite no solo responder a los objetivos inmediatos, sino también construir una base sólida para el futuro. En un entorno competitivo y en constante evolución, este marco analítico y estratégico ofrece a las empresas una ventaja significativa al abordar las necesidades de sus clientes de manera más precisa y efectiva.

### IV. Referencias

1. Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
2. Müller, A., & Guido, S. (2018). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media.
3. Harrison, M. (2021). *Treading on Python: Illustrated Guide to Python 3*. Treading Publications.
4. Harrison, M. (2021). *Effective Pandas: Patterns for Data Manipulation*. Treading Publications.
5. Banco Mundial. Producto Interno Bruto (PIB). <https://data.worldbank.org>
6. Instituto Nacional de Estadística (INE). Producto Interior Bruto (PIB) por comunidades autónomas y sectores. <https://www.ine.es>
7. Rokach, L., & Maimon, O. (2005). *Clustering Methods*. Springer.
8. Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
9. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
10. Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15th ed.). Pearson.
11. Marr, B. (2016). *Key Performance Indicators (KPI): The 75 Measures Every Manager Needs to Know*. Pearson Education.
12. Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Harvard Business Review Press.