

# *EasyMoney*

**Un impulso al desarrollo del negocio a través del marketing analítico.**

## Máster en Data Science

Corbella Felip, Cecília

Crespo Vázquez, Noé

Font Navarri, Raquel

Mejía Arias, Carlos Francisco

# ÍNDICE

<b>Resumen</b>	<b>3</b>
<b>Abstract</b>	<b>3</b>
<b>1. Introducción</b>	<b>4</b>
<b>2. Objetivos</b>	<b>6</b>
<b>3. Metodología</b>	<b>7</b>
3.1. Dataset Inicial	7
3.2. Preprocessing	8
3.2.1. Preparación de Datos	8
3.3. Datasets Finales	9
3.4. Visualización de los datos	10
3.4.1. Conocimiento del negocio	11
3.4.2. Reconocimiento de los clientes	11
3.4.3. Propensión de ventas	11
3.5. Asignación de probabilidades	12
3.6. Agrupación de Clientes	14
3.7. Campaña de Marketing	15
<b>4. Resultados y Conclusiones</b>	<b>16</b>
4.1. Modelo de Propensión	16
4.2. Modelo de clustering	19
4.3. Campaña de marketing	22
4.3.1. Campaña clientes internos:	22
4.3.2. Campaña clientes externos:	22
4.4. Conclusiones	24
<b>5. Bibliografía</b>	<b>25</b>

---

## Resumen

En el presente trabajo se han utilizado técnicas avanzadas de Data Science para ayudar a la empresa EasyMoney a reorientar su estrategia comercial y aprovechar al máximo la base actual de clientes de la empresa optimizando así su rentabilidad. Se han analizado los datos y se han utilizado técnicas de Machine Learning para desarrollar modelos de propensión de compra y segmentación de clientes. Los resultados obtenidos han proporcionado una comprensión detallada de los tipos de clientes de la empresa permitiendo la aplicación de estrategias comerciales más específicas y orientadas a cada cliente. Este enfoque orientado a datos podrá ayudar a la empresa a la toma de decisiones comerciales más informadas y eficaces.

---

## Abstract

In this study, advanced Data Science techniques have been employed to assist EasyMoney in redefining its commercial strategy and maximizing the utilization of its existing customer base, thus optimizing profitability. Data was analyzed and Machine Learning techniques were applied to develop propensity-to-buy models and customer segmentation. The obtained results offered a detailed understanding of the company's customers, enabling the implementation of more specific and customer-oriented commercial strategies. This data-driven approach will help the company to make more informed and effective business decisions.

---

## 1. Introducción

La transformación digital ha abierto nuevas oportunidades y desafíos para las empresas en todos los sectores. En este contexto, la empresa EasyMoney ha decidido emplear técnicas avanzadas de Data Science para reorientar su estrategia comercial.

EasyMoney, es una plataforma innovadora de comercialización de productos financieros fundada por Carol Denver, una profesional con años de experiencia en banca de inversión. EasyMoney destaca por su enfoque pionero en la creación de soluciones financieras simples y accesibles para sus clientes pero a pesar de su éxito inicial, se enfrenta actualmente a desafíos significativos. Actualmente, la empresa se encuentra en un momento decisivo debido a la falta de inversión y a la rápida expansión del catálogo de productos que ha alejado a la empresa de su visión original.

En este contexto, nuestro trabajo se enfoca en abordar esta problemática mediante el análisis de datos. Nuestro objetivo principal es aprovechar al máximo la base actual de clientes de la empresa optimizando así la rentabilidad (“Understanding the Importance of Marketing Data,” 2017). Los objetivos específicos que perseguimos en este proyecto incluyen el desarrollo de un modelo de propensión de compra, la segmentación de los clientes mediante Clustering y finalmente la implementación de campañas de marketing totalmente fundamentadas en el análisis de los datos obtenidos.

Para llevar a cabo los objetivos mencionados, se ha trabajado con un conjunto diverso de datos proporcionados por EasyMoney. Estos incluyen información sobre la actividad comercial, tipos de productos, datos sociodemográficos y detalles de ventas.

Al inicio del proyecto, empezamos con una fase crucial, el entendimiento y análisis de los datos. Este proceso se facilitó mediante la herramienta PowerBI, una potente plataforma de visualización de datos que nos permitió transformar datos complejos en información significativa y visualmente atractiva. Esta fase inicial, fue fundamental para comprender en profundidad la estructura de los datos.

A continuación, utilizando el lenguaje de programación Python y técnicas de Machine Learning hemos desarrollado un modelo de propensión a compra que nos ha permitido predecir las probabilidades de compra de los clientes para unos productos seleccionados.

Además, hemos aplicado técnicas de clustering para la segmentación de los clientes, lo que nos ha permitido agrupar a los clientes en función de sus características.

Este enfoque multidisciplinar permite traducir datos complejos en insights accionables que posteriormente serán la base para las estrategias de marketing basadas en los datos obtenidos (Vicario & Coleman, 2020).

Comprender el público objetivo es la garantía de que podremos ofrecer los productos que el cliente realmente necesita y asegurar el éxito de nuestra campaña de marketing (*What Is Data-Driven Marketing & Why Is It Important?*, 2022). Por este motivo, la estrategia de Marketing propuesta va dirigida a los clientes actuales más propensos a comprar los productos elegidos. Y adicionalmente proponemos la realización de una campaña hipersegmentada de clientes nuevos, tomando como base los clusters obtenidos y sus perfiles sociodemográficos.

---

## 2. Objetivos

El principal objetivo del proyecto es utilizar técnicas de Data Science para ayudar a la empresa EasyMoney a reorientar su estrategia comercial y aprovechar al máximo la base actual de clientes de la empresa optimizando así la rentabilidad.

Concretamente nos centraremos en tres objetivos:

- Desarrollar, implementar y evaluar un modelo de propensión a compra que mejore la eficiencia de la estrategia de ventas y aumente el rendimiento financiero de la empresa.
- Utilizar técnicas de clustering para segmentar la base de clientes con perfiles similares, con el fin de proporcionar información estratégica que oriente las decisiones de marketing y ventas de la empresa.
- Implementar una campaña de marketing basada en los clusters de clientes y las probabilidades de compra estimadas, con el propósito de calcular las ganancias potenciales para la empresa.

---

## 3. Metodología

Para llevar a cabo este estudio, se han empleado diversas herramientas y técnicas de Data Science. Se ha usado el lenguaje de programación Python 3.9.6 para la manipulación, limpieza y análisis de datos. Concretamente se han utilizado las librerías especializadas sklearn, pandas, numpy, matplotlib, seaborn, os, re, graphviz y también importamos la función SMOTE de la librería imbalanced-learn.

### 3.1. Dataset Inicial

Para el desarrollo de los modelos, se han utilizado los datos proporcionados por la empresa EasyMoney. El conjunto de datos recibido estaba dividido en 5 datasets:

- **Dataset Customer Commercial Activity:** Este conjunto de datos contiene información sobre la actividad comercial de los clientes. Incluye detalles como la identificación, actividad, segmento comercial y canal de captación. Contiene 5.962.924 filas y 6 columnas.
- **Dataset Customer Products:** Este conjunto de datos recopila información sobre los productos que los clientes han adquirido. Cada fila representa un cliente, y las columnas indican qué productos específicos tienen. Contiene 5.962.863 filas y 17 columnas.
- **Dataset Customer Sociodemographics:** Este conjunto de datos incluye datos sociodemográficos de los clientes, como información sobre su edad, género, patrimonio, país de origen y otros atributos demográficos. Contiene 5.962.924 filas y 8 columnas.
- **Dataset Sales:** Este conjunto de datos presenta detalles sobre las ventas de los productos. Las columnas incluyen información sobre la fecha de venta, el identificador del producto vendido y el margen neto de la venta. Contiene 240.773 filas y 5 columnas.
- **Dataset Product description:** Este conjunto de datos es una descripción de los productos ofrecidos por la empresa. Cada fila representa un producto, y las columnas proporcionan detalles sobre su nombre, categoría y características. Contiene 12 filas y 3 columnas.

## 3.2. Preprocessing

Para poder trabajar con los datos proporcionados por la dirección de EasyMoney, procedimos a preparar los datos para ajustarlos a las necesidades requeridas para emplear técnicas de Machine Learning.

### 3.2.1. Preparación de Datos

Para cada uno de los datasets se realizaron las siguientes técnicas de limpieza:

- **Eliminación de Duplicados:** No se encontraron datos duplicados.
- **Tratamiento de Datos Nulos:** Se identificaron y gestionaron los valores nulos en los conjuntos de datos, aplicando técnicas de imputación cuando fue necesario para garantizar la integridad de los datos.

Para las variables categóricas que contenían nulos se imputó “no dato”:

- *gender*
- *region\_code*
- *entry\_channel*
- *segment*

Para las variables numéricas, con nulos, se imputó “-1”:

- *salary*

Se eliminaron las filas cuando suponían una cantidad muy reducida de nulos:

- *pension\_plan*
- *payroll*
- **Cambios a Formato de Fecha:** Ajustamos las columnas de fecha al formato de fecha y hora para facilitar su manipulación y análisis. Este fue el caso de *month\_sale*, *pk\_partition* y *entry\_date*.
- **Unión de Datos:** Unimos los datasets de datos demográficos, de actividad de los clientes y de tenencia de productos, creando un conjunto de datos completo para poder trabajar con un único dataset.
- **Borrado de columnas:** Se eliminaron las siguientes columnas por no presentar variabilidad y por lo tanto no aportar información al modelo: *em\_account\_pp*, *em\_account\_p* y *region\_code*.
- **Agrupación de Categorías Minoritarias:** Para simplificar la representación de los datos se llevaron a cabo agrupaciones de categorías minoritarias en aquellas



columnas categóricas con un número muy elevado de etiquetas. Fueron las siguientes:

- *country\_id*: Agrupamos las categorías con frecuencias iguales o menores a 10.
- *entry\_channel*: Agrupamos las categorías con frecuencias iguales o menores a 10000.
- *region\_code*: Agrupamos las categorías con frecuencias iguales o menores a 5000.
- **Transformaciones de Datos**: Se aplicó la transformaciones One Hot Encoding a las variables categóricas: *gender*, *entry\_channel*, *segment*, *region\_code* y *country\_id* para asegurar que los datos fueran adecuados para el análisis.

### 3.3. Datasets Finales

Para calcular los productos más vendidos hemos utilizado el siguiente dataset. Hemos unido los datasets de Sales y Customer\_Products. Para la unión de ellos integramos los datos teniendo en cuenta el identificador del cliente (*cid* / *pk\_cid*) y nos quedamos con los datos de cada partición (*month\_sale*/ *pk\_partition*). Las uniones siempre se han realizado de tal forma que siempre nos hemos quedado con las filas del dataset de Sales que era el que nos daba la información deseada.

Para obtener el dataset que se han usado para generar los modelos se han unido los siguientes datasets: Customer\_Commercial\_Activity, Customer\_Sociodemographics y Customer\_Products utilizando las columnas *pk\_cid* y *pk\_partition* como claves primarias.

A continuación, se escogió únicamente la partición más reciente del dataset y se indicó si el producto se adquirió en esa última partición (1), o si se había adquirido en una partición anterior (0).

Este dataset se usó para el análisis de propensión de compra y la segmentación de clientes. Este conjunto de datos está formado por 442.995 registros y 40 columnas. A continuación, se describen las columnas y su significado:

- *pk\_partition*: Esta columna contiene la fecha y la hora en la que se registraron los datos.
- *entry\_date*: Registra la fecha en la que cada cliente se incorporó al sistema.

- *active\_customer*: Es un indicador binario que indica si un cliente se encuentra activo en un momento dado.
- *age*: Refleja la edad del cliente.
- *salary*: Contiene detalles sobre el patrimonio del cliente.
- *short\_term\_deposit*, *loans*, *mortgage*, *funds*, *securities*, *long\_term\_deposit*, *credit\_card*, *payroll*, *pension\_plan*, *payroll\_account*, *emc account*, *debit\_card*, *em\_account*: son columnas binarias que indican la tenencia o no de productos o servicios específicos por parte de los clientes.
- *gender H*, *gender No Dato*, *gender V*: Estas columnas binarias representan el género del cliente, con categorías "H" (hembra), "No Dato" y "V" (varón).
- *entry\_channel\_KAT*, *entry\_channel\_KFC*, *entry\_channel\_KHE*, *entry\_channel\_KHK*, *entry\_channel\_KHM*, *entry\_channel\_KHN*, *entry\_channel\_KHQ*, *entry\_channel\_Other*: Estas columnas binarias reflejan el canal de entrada por el cual los clientes se unieron a la empresa, incluyendo "Other" para otros canales no especificados.
- *segment\_No\_Dato*, *segment\_PARTICULARES*, *segment\_TOP*, *segment\_UNIVERSITARIO*: Estas columnas binarias definen el segmento al que pertenece cada cliente.
- *country\_id\_CH*, *country\_id\_DE*, *country\_id\_ES*, *country\_id\_FR*, *country\_id\_GB*, *country\_id\_Other*, *country\_id\_US*: Estas columnas binarias representan el país de origen de los clientes, con categorías específicas y "Other" para otros países no especificados.

### 3.4. Visualización de los datos

Para comprender los datos y presentarlos de manera efectiva a EasyMoney, hemos utilizado PowerBI. Trabajamos con diversos conjuntos de datos y creamos fórmulas internas para enriquecer nuestras visualizaciones, proporcionando así un valor adicional a los datos presentados (*Get Started with Power BI Desktop - Power BI*, 2023).

Durante este proceso, dividimos la visualización de datos en tres secciones distintas: negocio, clientes y ventas. Esto nos permitió examinar detenidamente diferentes aspectos del negocio y resaltar información clave en cada área, brindando así una visión completa y detallada.

### **3.4.1. Conocimiento del negocio**

En esta sección, nos enfocamos en entender la estructura de negocio de EasyMoney. El objetivo era obtener una comprensión global del mismo para poder evaluar la relevancia y el valor de cada conjunto de datos disponible.

Durante este análisis, hemos identificado varios aspectos cruciales del negocio, que incluyen:

- Ventas generadas
- Margen medio de la totalidad de los productos; nos ayuda a entender que productos generan más margen para la empresa, así pues, nos da información de que productos son más atractivos para la dirección de EasyMoney.
- Clientes activos
- Ratio de Conversión, nos permite ver cuantos clientes son activos sobre el total de clientes que tiene la empresa.
- País con mejores ventas

### **3.4.2. Reconocimiento de los clientes**

En la sección dedicada a la visualización de los clientes, nos centramos en los datos sociodemográficos de los clientes, exploramos sus márgenes e ingresos, y analizamos los canales de captación. Esta profundización de la información nos permite tomar mejores decisiones respecto a nuestros clientes actuales pudiendo conocer mejor sus intereses y necesidades.

### **3.4.3. Propensión de ventas**

Finalmente, nos centramos en las ventas de la empresa. Visualizamos las ventas individuales de cada producto, las ventas agrupadas por familias de productos, los márgenes netos asociados a cada producto y la evolución histórica de las ventas. Este enfoque nos proporcionó una mejor comprensión del rendimiento de las ventas, permitiéndonos identificar patrones y tendencias.

Gracias a las posibilidades que ofrece PowerBI pudimos extraer también información cruzada entre los perfiles de los clientes y los productos vendidos. Esta capacidad para extraer información interrelacionada fué básica para entender las dinámicas de los clientes.

### 3.5. Asignación de probabilidades

Después de haber hecho el análisis de la base de datos de EasyMoney, procedimos a trabajar con los datos para poder ayudar a la empresa a entender que productos les podrían hacer vender más y mejor. Para ello, nos centramos en predecir qué clientes eran más propensos a comprar los productos que seleccionamos para nuestro estudio.

El criterio que usamos para seleccionar qué productos elegir, fue seleccionar los 4 productos más vendidos (con más de 60.000 unidades vendidas). Esto nos llevó a elegir los siguientes productos: *em\_account*, *debit\_card*, *pension\_plan* y *payroll*.

Así pues, nuestro objetivo era crear un modelo de propensión a compra para los clientes ya internos de la empresa. Observamos que todos los clientes ya existentes tenían en su gran mayoría ya contratado la cuenta *em\_account*. También comprobamos que la curva del tiempo de *em\_account* coincidía con la curva del tiempo de la adquisición de nuevos clientes. Por lo tanto, adquirir el *em\_account* era una condición intrínseca de ser cliente de la empresa.

Para lograr el objetivo de la asignación de probabilidades, se realizaron las siguientes acciones:

1. **Seleccionamos los Productos Objetivo:** como previamente hemos comentado, se definió la lista de productos que se pretendía analizar: *em\_account*, *debit\_card*, *pension\_plan* y *payroll*.
2. **Pre Procesamos los Datos:** Para todos los modelos se eliminaron las columnas de fecha *pk\_partition*, *entry\_date* y *active\_customer* ya que no aportaban información relevante para nuestro estudio. Además, en cada modelo se tuvo en cuenta únicamente el producto estudiado por lo que se eliminaron las columnas pertenecientes a los demás productos financieros. De esta forma, centramos el análisis en las características demográficas y la tenencia del producto.
3. **Escalamos las Características:** Las características numéricas *age* y *salary* se escalaron utilizando el StandardScaler para asegurar una comparación adecuada.

Después de preparar los datos, procedimos a la creación del modelo y a la predicción de las probabilidades. Para cada producto objetivo, se creó un modelo de regresión logística. Elegimos este algoritmo de clasificación ya que se utiliza para predecir la probabilidad de una variable dependiente binaria. En nuestro caso la variable dependiente es binaria y representa si un cliente comprará o no un producto específico.

Además, se eligió este modelo ya que es fácilmente interpretable, pues los coeficientes nos indican cómo la variación en una variable independiente afectará la probabilidad del evento en cuestión. También es un modelo computacionalmente eficiente, lo que es crucial cuando se trabaja con grandes conjuntos de datos. Finalmente, destaca por su rendimiento ya que la Regresión Logística puede proporcionar resultados sorprendentemente buenos utilizando baja potencia computacional.

Posteriormente, se dividieron los datos en conjuntos de entrenamiento y prueba, y se aplicó la técnica de sobre muestreo (SMOTE) para abordar el desequilibrio de clases. Nuestras variables objetivo estaban muy desbalanceadas ya que la proporción de clientes que no tenían el producto era considerablemente mayor que aquellos que sí lo tenían. Al aplicar la técnica de sobre muestreo, nos generó un target balanceado. SMOTE generó ejemplos sintéticos de la clase minoritaria para equilibrar el conjunto de datos, lo que resultó en un modelo más balanceado y preciso. Con esto conseguimos mejorar el rendimiento ya que al equilibrar las clases, el modelo puede aprender de manera más efectiva. Y conseguimos reducir el sesgo hacia la clase mayoritaria.

Posteriormente, se aplicó el modelo sobre todo el conjunto de datos para predecir la propensión de todos los clientes.

Se generaron nuevas columnas con las probabilidades obtenidas y se añadieron al dataset: *prob\_em\_account*, *prob\_pension\_plan*, *prob\_debit\_card* y *prob\_payroll*.

Para evaluar la capacidad que tiene el modelo para discriminar entre aquellos clientes que adquirirán el producto y aquellos que no lo harán, utilizamos el AUC (Area Under the ROC Curve).

### 3.6. Agrupación de Clientes

Para entender mejor el comportamiento y las necesidades de cada cliente se han segmentado todos los clientes de la empresa en grupos homogéneos. Para ello, se ha utilizado la técnica de clustering utilizando el algoritmo K-Means. Escogimos el algoritmo K-means por ser ampliamente utilizado y por la abundante literatura que respalda sus resultados. Además, es un algoritmo sencillo y fácilmente aplicable a nuestros datos. El algoritmo K-Means funciona asignando cada muestra en el conjunto de datos a uno de los clusters, basándose en la distancia euclidiana entre la muestra y el centroide de cada clúster. El objetivo es minimizar la suma de las distancias cuadradas entre las muestras y los centroides de sus respectivos clusters, logrando así una agrupación compacta e internamente coherente.

Para obtener un buen resultado, el primer paso fué eliminar las columnas que consideramos no relevantes para la segmentación de los clientes. Las columnas eliminadas fueron las siguientes: *pk\_partition*, *entry\_date*, *short\_term\_deposit*, *loans*, *mortgage*, *funds*, *securities*, *long\_term\_deposit*, *credit\_card*, *payroll*, *pension\_plan*, *payroll\_account*, *emc\_account*, *debit\_card*, *em\_account* y *gender*. Además, para asegurar que todas las características tuvieran la misma importancia en el análisis, la variable numéricas *salario* y *edad* fueron escaladas utilizando StandardScaler.

Para determinar el número óptimo de clusters, se utilizó el Método del Codo (Fig. 1). Este método consiste en calcular la suma de las distancias al cuadrado entre los puntos de datos y su centroide en función del número de clusters. Estos valores se grafican, lo que permite identificar el punto en el gráfico donde la disminución en la variabilidad dentro de un mismo cluster comienza a disminuir significativamente, formando un codo en la gráfica. Este codo, nos permitió detectar el número óptimo de clusters.

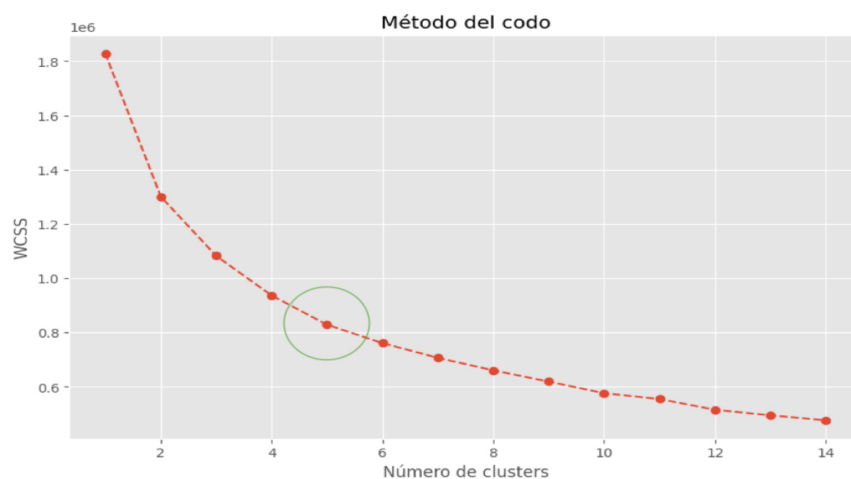


Fig 1. Método del codo para identificar el número óptimo de clusters

### 3.7. Campaña de Marketing

Se han realizado dos campañas de Marketing una para clientes internos y otra para captación de clientes nuevos.

- **Campaña clientes internos:** esta campaña consiste en la venta de tres productos (*debit\_card*, *payroll* y *pension\_plan*) mediante un pack. Así pues, los clientes tendrán que comprar el pack entero si quieren entrar en el sorteo de un Ipad y obtener un Cash Back (se les devuelve un porcentaje del dinero gastado) del 4% para todas sus compras con su tarjeta de débito EasyMoney.

Esta campaña será realizada por 6 agentes comerciales. Los clientes objetivo son aquellos que no tienen los productos (*debit\_card*, *payroll* y *pension\_plan*) y que su probabilidad de compra de los tres es alta. Mediante una llamada telefónica se les ofrecerá una promoción de los tres productos en conjunto. Dicha campaña tendrá una duración de 3 meses (del 1 de noviembre de 2023 al 31 de enero de 2024) y el coste total será de 30.000 euros. Se han seleccionado 28.472 clientes para ser sujetos a llamadas telefónicas, se espera que la ratio de respuesta a nuestras llamadas sea del 80% (22.778 personas responden nuestras llamadas). De esos, esperamos una tasa de conversión a compra del 2% para adquirir los 3 productos. Esto equivaldría a 2.278 personas.

- **Campaña clientes externos:** campaña realizada en la plataforma de facebook ads, para la generación de anuncios en las redes sociales de Facebook e Instagram.

El objetivo de la campaña es la captación de clientes potenciales. La llamada a la acción es una solicitud de llamada a un agente comercial. Se hará un conjunto de anuncios donde se realizará un anuncio para cada producto ofrecido (*pension\_plan*, *debit\_card*, *em\_account*, *payroll*), con imágenes diferentes según la sección donde se presentarán los anuncios (post cuadrado, vertical y video), además de 5 títulos y 5 descripciones diferentes lo que permitirá al algoritmo de la herramienta aprender y optimizar los anuncios para mejorar el rendimiento.

Tomamos como base el cluster Mind Senior Workers para la segmentación de la campaña (País España, 32 a 52 años, hombres y mujeres), la campaña tendrá una duración de 2 meses iniciando el 4 de diciembre 2023 y terminará el 8 de enero de 2024 con una inversión total de 15.000 €. Teniendo en cuenta los datos estadísticos de la herramienta de facebook ads, estimamos tener un alcance de 99 mil a 286 mil y de 43 a 125 solicitudes de llamadas al día. Tomando como base un escenario positivo de conversión de un 10% se estarían generando de 4 a 13 clientes diarios.

---

## 4. Resultados y Conclusiones

### 4.1. Modelo de Propensión

Al implementar el modelo de propensión para cada producto seleccionado (*pension\_plan*, *debit\_card*, *em\_account* y *payroll*), hemos podido identificar aquellos clientes con una mayor probabilidad de contratación de cada producto.

Los resultados de la métrica AUC (Area Under the ROC Curve) fueron los siguientes para cada uno de los productos seleccionados:

*pension\_plan*: 0.7195

*debit\_card*: 0.7788

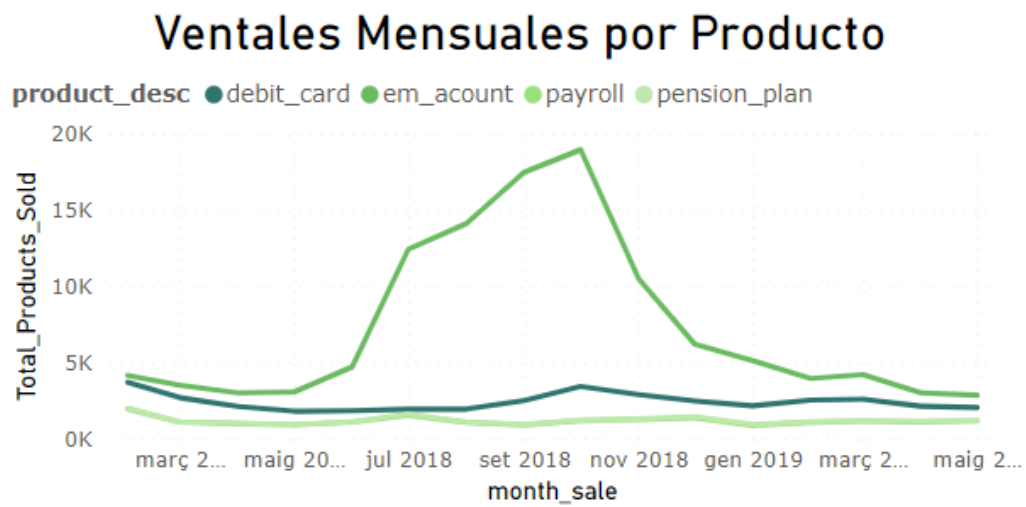
*em\_account*: 0.9425

*payroll*: 0.7162

Con estos resultados podemos concluir que el modelo tiene una capacidad moderada para predecir la adquisición de un *pension\_plan*, *debit\_card* y *payroll*. Sin embargo, el producto *em\_account* destaca notablemente con un AUC excepcionalmente alto. Después de revisar exhaustivamente los datos, descartamos que se estuviera produciendo una fuga de datos (data leakage) y llegamos a la conclusión de que tener una cuenta es algo fundamental para que cualquier cliente pueda operar. En este contexto, este AUC alto tiene sentido ya que indica que los cliente tiene una muy alta probabilidad de tener una cuenta. Este hecho se puede observar claramente en las siguientes gráficas (*Fig. 2*), donde podemos ver que el progreso que realizan las ventas mensuales de *em\_account* es muy parecido al de los clientes.



(A)



(B)

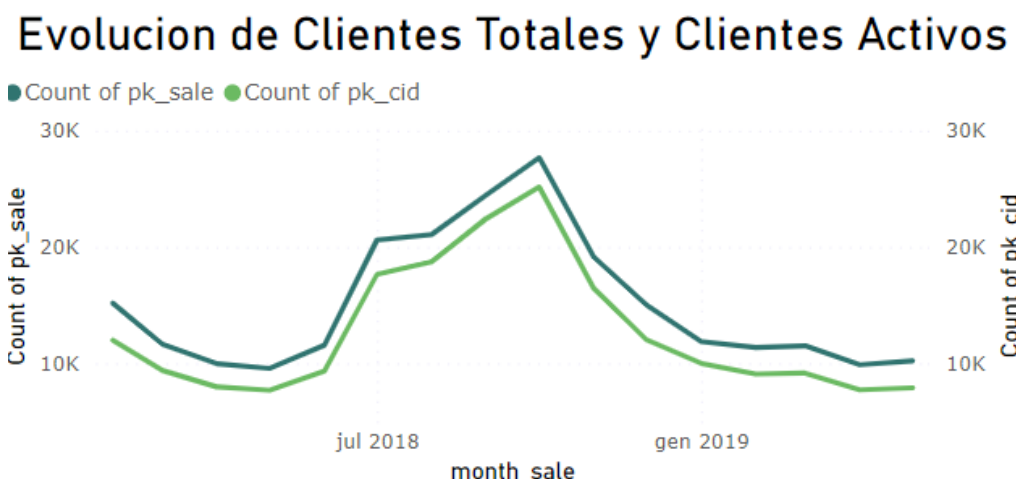


Fig 2. (A) Representación gráfica de las ventas mensuales por producto (B) Representación gráfica de la evolución de los clientes.

Utilizando nuestro modelo de propensión, hemos clasificado a los clientes en tres categorías distintas: alta, baja y media probabilidad de adquisición de cada producto. Esta clasificación se basa en las probabilidades predichas por el modelo para cada cliente en particular. Los clientes clasificados como "alta probabilidad" son aquellos que se encuentran en el percentil 90. Los clientes con "baja probabilidad" se encuentran en el percentil 10 y finalmente los clientes con "media probabilidad" se ubican entre el percentil 20 y el 80.

En la siguiente gráfica podemos observar que las variables más importantes para el modelo son la edad y el patrimonio familiar (Fig. 3):

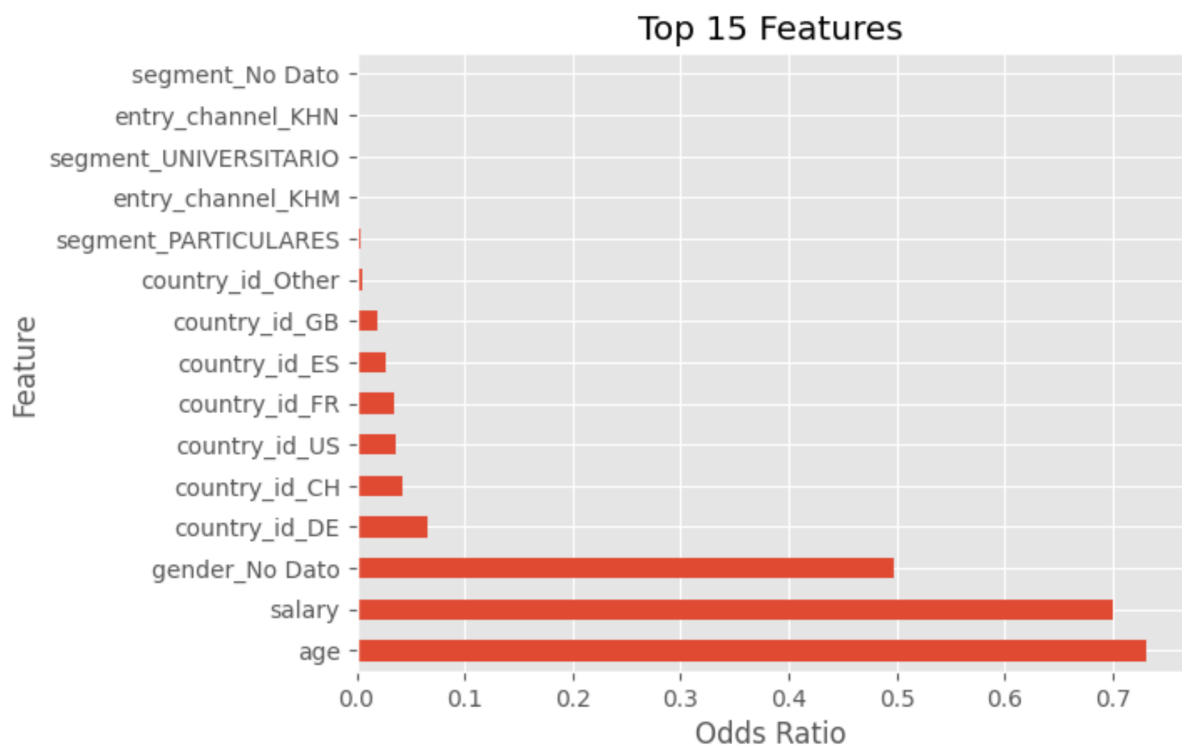
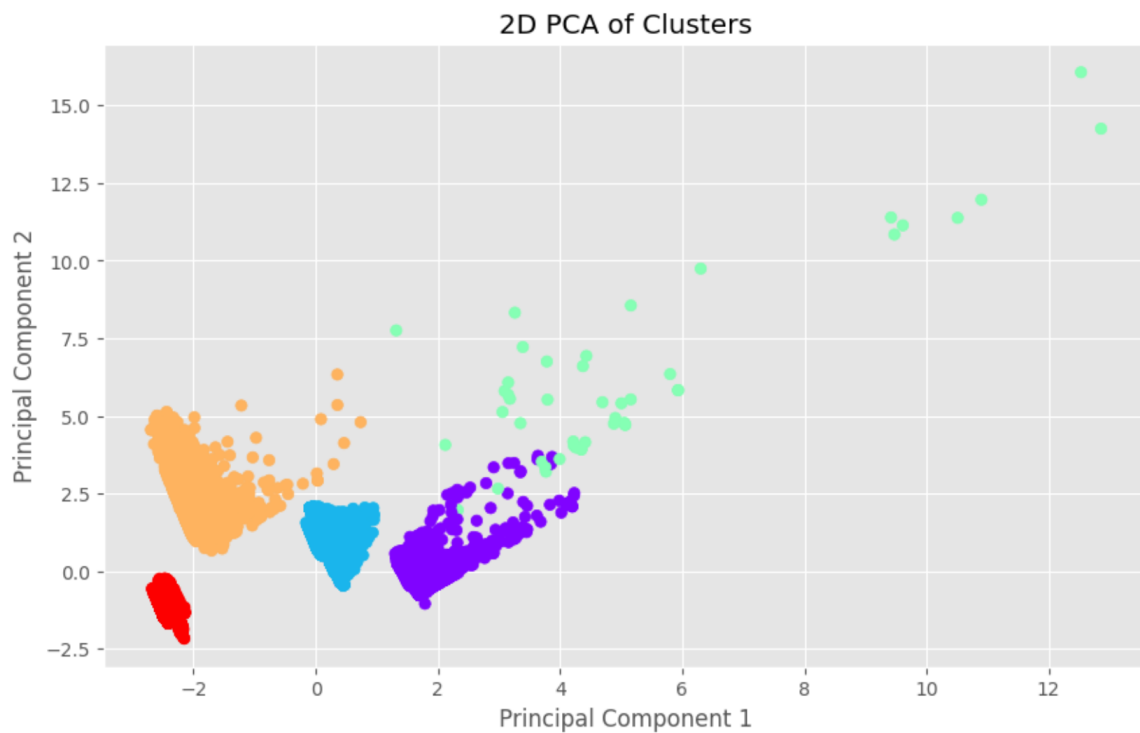


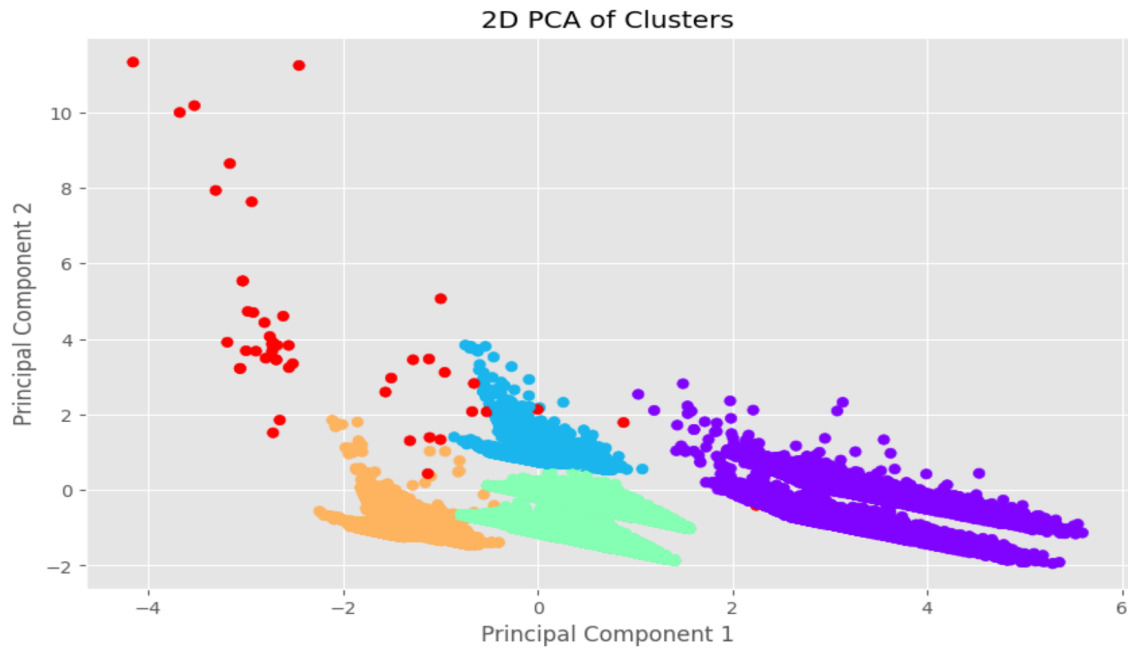
Fig 3. Representación gráfica de las feature importance de cada variable

## 4.2. Modelo de clustering

Durante el proceso de clustering, nuestro objetivo principal era identificar grupos de clientes basados en características significativas. Inicialmente, incluimos la columna de género como una de las características importantes para realizar la segmentación. Sin embargo, al analizar los datos hemos visto que teníamos dos grupos muy similares en los que únicamente cambiaba el género. Para poder verificar si debíamos dejar o quitar la variable género, realizamos una visualización de datos mediante PCA y observamos que teniendo en cuenta el género los grupos obtenidos eran más dispersos y heterogéneos (Fig.4 y 5) Por tanto vemos que el género no es una característica determinante por lo que decidimos eliminarla del modelo.



*Fig 4. Principal Component Analysis. Visualización de los grupos de clientes creados sin tener en cuenta el género de los clientes*



*Fig 5. Principal Component Analysis. Visualización de los grupos de clientes creados teniendo en cuenta el género de los clientes.*

A continuación, se presentan estos grupos con sus características detalladas:

- **Junior Premium:** 186.660 clientes (42,14%)

Este grupo está mayoritariamente integrado por jóvenes universitarios con una edad promedio de 24,22 años y un patrimonio medio de 97.009,13€. Aproximadamente el 23% de los clientes en este cluster son activos. En cuanto a los canales de captación, Google Ads es el principal medio de entrada, atrayendo al 99,03% de los clientes de este grupo. Este grupo tiene una probabilidad media en adquirir alguno de los productos seleccionados.

- **Mind Senior Workers:** 110,973 clientes (25,05%)

Este grupo está formado por clientes particulares (94.74%) con una edad promedio de 36.05 años y con un patrimonio promedio de 64.168,03€. Aproximadamente el 58,28% de los clientes son clientes activos. El canal de entrada más común es Facebook Ads con un 35.66%. En términos de probabilidad de adquisición de alguno de los productos seleccionados, este grupo muestra una alta predisposición a la contratación.

- **Unicorns:** 47 clientes (0,01%)

Este grupo se caracteriza por agrupar a clientes universitarios (55,32%) y clientes particulares (43%) con una edad promedio de 33,15 años. Es relevante destacar que este grupo muestra un patrimonio medio notablemente alto de 12.179.580€. En términos de actividad, el 29.79% de los individuos en este cluster son clientes activos. El canal de entrada más común es Google Ads con un 46.81%. Sin embargo, las probabilidades promedio de adquisición de alguno de los productos seleccionados son extremadamente bajas.

- **Senior:** 45.880 clientes (10,35%)

Este cluster se caracteriza por ser clientes particulares (86.89%) y tener una edad promedio de 58.77 años con un patrimonio medio de 86.455,04€ . El 61,28% de los individuos en este cluster están activos. En cuanto al canal de entrada más común es Facebook Ads con un 29.78%. Este grupo tiene una probabilidad media en adquirir alguno de los productos seleccionados.

- **Juniors:** 99.435 clientes (22,45%)

Este cluster se caracteriza por estar integrado por jóvenes universitarios (97.50%), con una media de edad de 22.61 años y un patrimonio promedio de 44.348,73 €. El 35.76% de los individuos en este cluster están activos. El canal de entrada más común son los podcasts con un 73.22%. Este grupo muestra una disposición media a adquirir alguno de los productos seleccionados.

### 4.3. Campaña de marketing

#### 4.3.1. Campaña clientes internos:

Los resultados esperados en nuestra campaña de marketing para clientes internos se presentan en la Tabla 1. En este resumen detallado, hemos calculado el margen medio, la conversión esperada y los ingresos esperados generados para tres productos clave: el Plan de Pensión, la Tarjeta de Débito y la Nómina. Estos datos nos han permitido estimar un ingreso total esperado de 2.778.398,35 €. Es importante destacar que este cálculo ha sido efectuado considerando un costo total de la campaña de 30.000 €, lo que nos deja con un ingreso neto esperado de 2.748.398,35 €.

PRODUCTO	MARGEN MEDIO	CONVERSIÓN ESPERADA	INGRESOS ESPERADOS GENERADOS
PENSION PLAN	5.976,44 €	455 personas	2.719.280,2€
DEBIT CARD	60,00 €	455 personas	27.300,00 €
PAYROLL	69,93 €	455 personas	31.818,15 €

*Tabla 1. Resumen de los resultados esperados en la campaña de marketing para los clientes internos.*

#### 4.3.2. Campaña clientes externos:

En la Tabla 2, presentamos los resultados diarios previstos de nuestra campaña de marketing enfocada en los clientes externos. La campaña tiene una duración de 36 días.

Para esta campaña hemos usado la plataforma Facebook Ads. Esta nos ha permitido calcular el margen medio, la conversión esperada y los ingresos esperados generados para los cuatro productos: la Cuenta Corriente, el Plan de Pensión, la Tarjeta de Débito y la Nómina. Facebook Ads nos proporciona el intervalo de conversión esperado. Estos datos nos han permitido estimar un ingreso total esperado de entre 222.348,24 € y 667.044,72 €. Es importante destacar que este cálculo ha sido efectuado considerando un costo total de la campaña de 15.000 €, lo que nos deja con un intervalo de ingresos netos esperados de 207.348,24 € a 652.044,72 €.

PRODUCTO	MARGEN MEDIO	CONVERSIÓN MÍNIMA	CONVERSIÓN MÁXIMA	MÍNIMO GENERADO	MÁXIMO GENERADO
PENSION PLAN	5.976,44 €	1 persona	3 personas	5.976,44 €	17929,32 €
EM ACCOUNT	69,97 €	1 persona	3 personas	69,97 €	209,91 €
DEBIT CARD	60,00 €	1 persona	3 personas	60,00 €	180,00 €
PAYROLL	69,93 €	1 persona	3 personas	69,93 €	209,79 €

*Tabla 2. Resumen de los resultados esperados en la campaña de marketing para los clientes externos.*

## 4.4. Conclusiones

1. Hemos clasificado a los clientes en tres categorías distintas: alta, baja y media probabilidad de adquisición de cada producto seleccionado.
2. Para el producto *em\_account* la curva temporal es igual a la de captación de clientes, lo que indica que es el producto base para poder operar en EasyMoney.
3. Se han segmentado los clientes según sus características en 5 grupos diferenciados (Junior Premium, Mind Senior Workers, Unicorns, Seniors, Juniors). Esto ha permitido una comprensión detallada del perfil de los clientes.
4. El cluster Mind Senior Workers muestra una alta predisposición a la contratación de los productos seleccionados (*pension\_plan*, *debit\_card* y *payroll*).
5. Se han diseñado campañas específicas para clientes internos y externos.
6. La segmentación de clientes proporciona oportunidades para personalizar aún más las estrategias de marketing y ventas, lo que puede aumentar la eficacia de las campañas.



---

## 5. Bibliografía

- 1) *Get started with Power BI Desktop—Power BI.* (2023).  
<https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>
- 2) Understanding the Importance of Marketing Data. (2017). *Maryville Online*.  
<https://online.maryville.edu/blog/understanding-the-importance-of-marketing-data>
- 3) Vicario, G., & Coleman, S. (2020). A review of data science in business and industry and a future view. *Applied Stochastic Models in Business and Industry*, 36(1), 6–18.  
<https://doi.org/10.1002/asmb.2488>
- 4) *What Is Data-Driven Marketing & Why Is It Important?* (2022). Semrush Blog.  
<https://www.semrush.com/blog/data-driven-marketing/>