

PROYECTO FINAL DATA SCIENCE

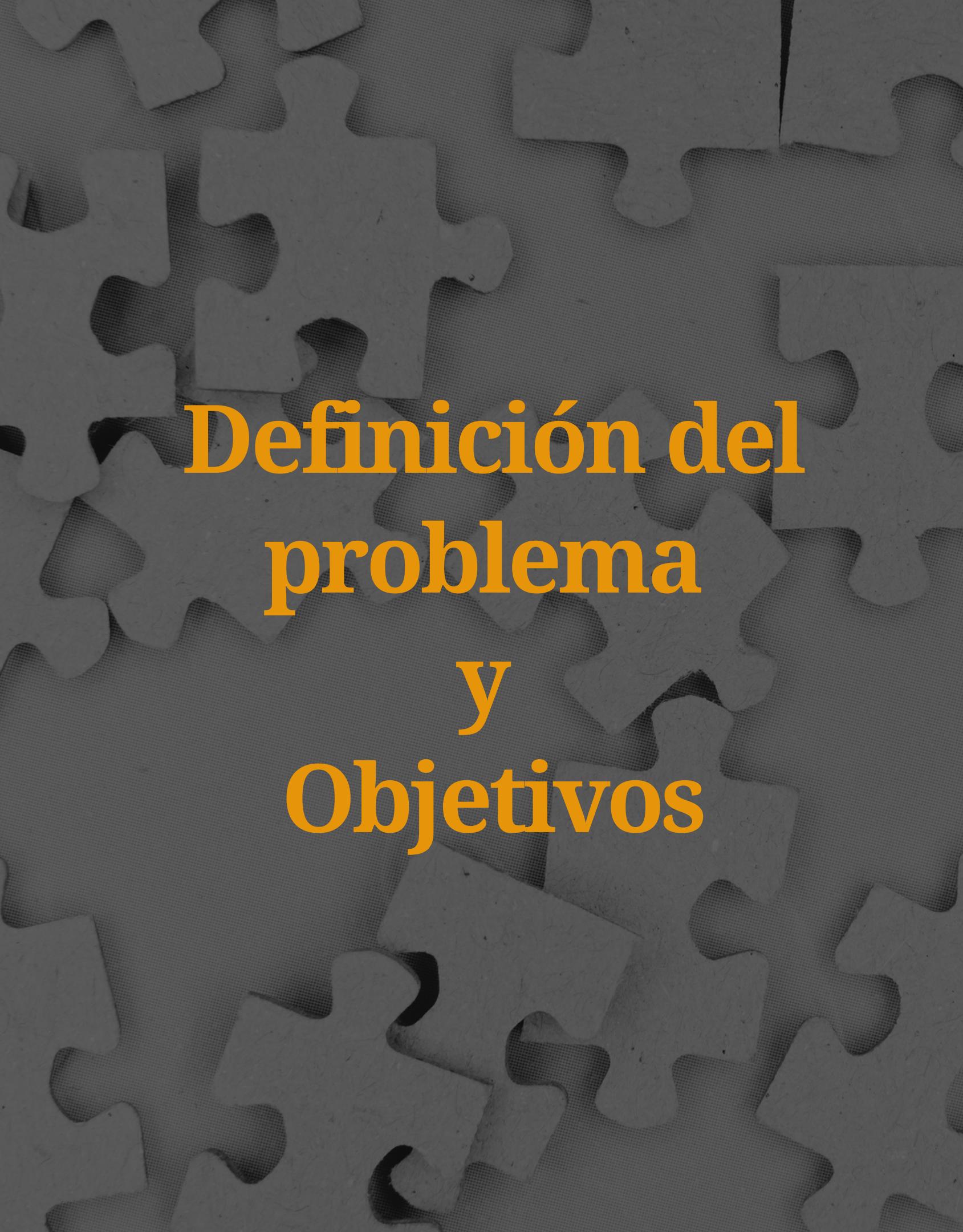
FUGA DE CLIENTES

Guido Cagliani Rincon



Tabla de contenidos

-
- 01 Definición del problema y objetivos
 - 02 Preguntas de interés
 - 03 Diccionario de datos
 - 04 Exploratory Data Análisis (EDA)
 - 05 Modelos
 - 06 Boosting models
 - 07 Conclusiones
 - 08 Futuras líneas
-



Definición del problema y Objetivos

Unos de los problemas mas comunes en los bancos son los ‘clientes fugados’, estos son aquellos que tienen una cuenta en el banco pero por alguna razón dejan de realizar operaciones o dejan de utilizar las tarjetas del mismo.

Por eso el gerente del banco Provincia se puso en contacto con nosotros, para que creemos un modelo que pueda lograr una estimación de si un cliente se va a fugar o no para así poder generar campañas para evitar que el cliente se fuge.

Nuestros objetivos a cumplir son:

1. Identificar y visualizar cuales son los factores que contribuyen al abandono de los clientes.
2. Construir un modelo de predicción que asigne una probabilidad de abandono a cada cliente para así poder abordarlo y tratar de que cambie de opinión, con promociones o con una mejora del servicio.



Preguntas de Interés

- ¿Cuántas cuentas hay activas? ¿Cuántas cerradas?
- ¿Cuál es el monto máximo de crédito otorgado? ¿El mínimo?
- ¿Qué rango de edad tienen los clientes?
- ¿Hay alguna relación entre los pedidos de crédito y los ingresos anuales de los clientes?
- ¿Hay alguna relación entre el tipo de tarjeta que poseen y el tiempo que llevan en el banco?
- Hay datos marcados como 'Unknown' en los ingresos de los clientes ¿A qué se debe?

Diccionario de datos

- CLIENTNUM: Numero unico identificador del cliente.
- Attrition_Flag: Actividad del cliente, si la cuenta esta cerrada vale 1 sino 0.
- Customer_Age: Edad del cliente.
- Gender: Genero del cliente.
- Dependent_count: Variable demografica, numero de dependencias.
- Education_Level: Nivel de educacion maxima alcanzada del cliente.
- Marital_Status: Estado civil del cliente.
- Income_Category: Ingresos anuales del cliente
- Card_Category: Tipo de tarjeta que tiene el cliente
- Monthsonbook: Periodo de relacion con el banco.
- TotalRelationshipcount: Numero total de productos mantenidos por el cliente
- MonthsInactive12_mon: N° de meses inactivos en los ultimos 12 meses.
- ContactsCount12_mon: N° de contactos en los ultimos 12 meses.
- Credit_Limit: Limite de la tarjeta de credito.
- TotalRevolvingBal: Saldo rotatorio total en la tarjeta de credito
- AvgOpenTo_Buy: Disponibilidad de comprar una linea de credito (porcentaje de los ultimos 12 meses)
- TotalAmtChngQ4Q1: Cambio en el monto de la transaccion (Q4 over Q1)
- TotalTransAmt: Importe total de las transacciones (Ultimos 12 meses)

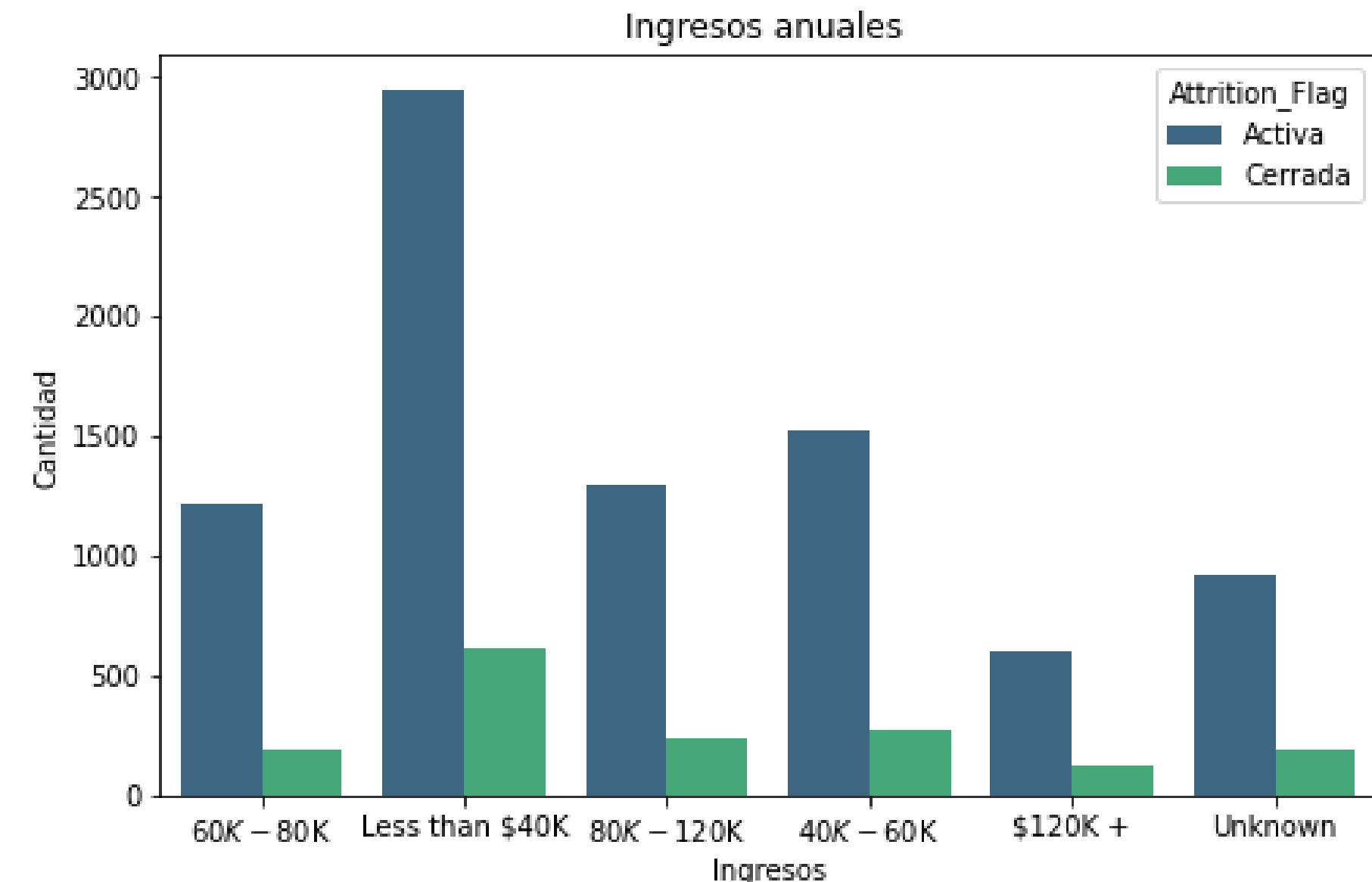
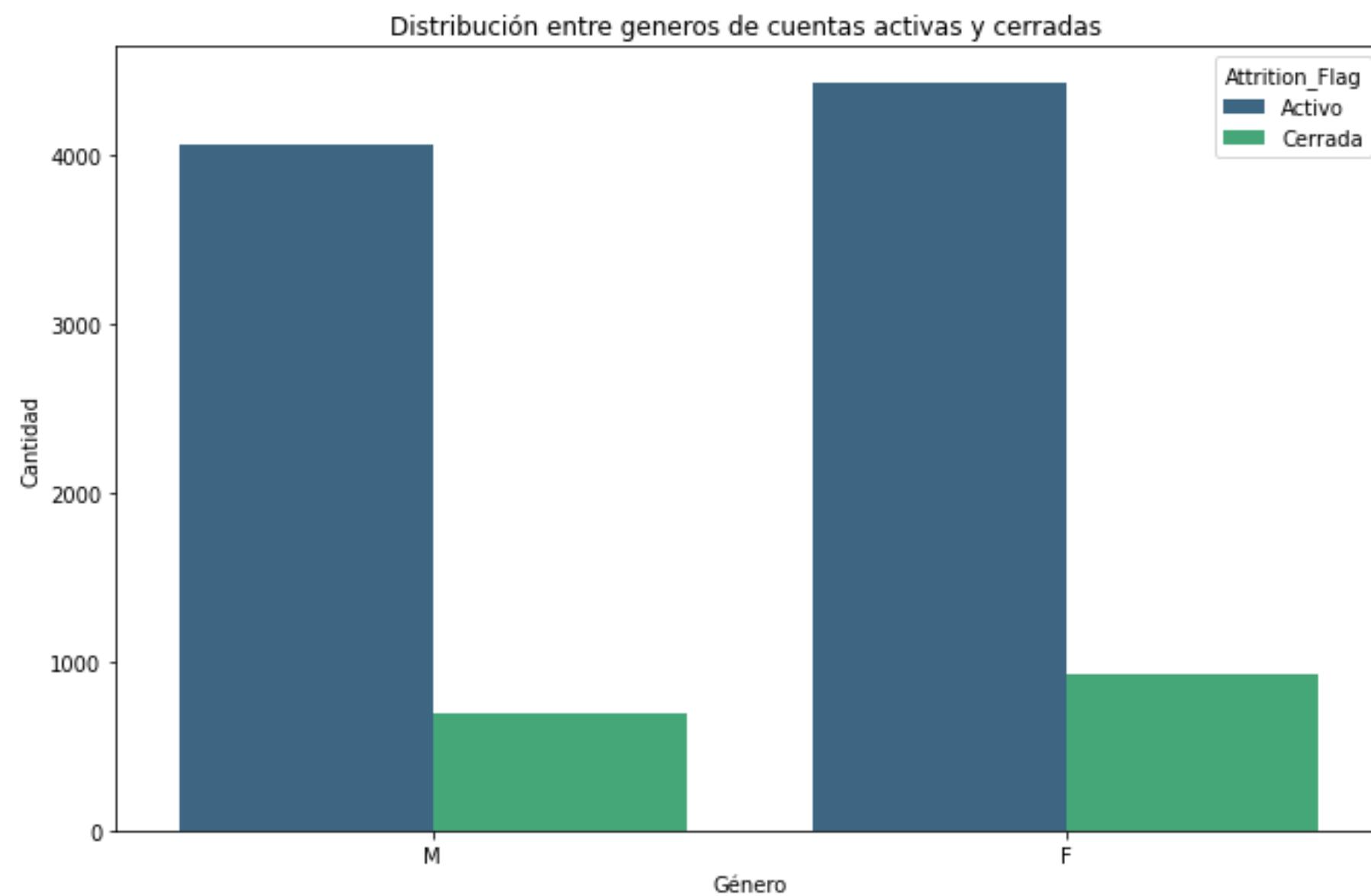
Análisis Exploratorio



Datos generales clientes

La mayoria de clientes en el banco son mujeres y tambien son las que mas cuentas cerradas tienen. Hay que mejorar el servicio al cliente para las mujeres para asi evitar que cierren su cuenta.

La mayoria de clientes tienen un ingreso anual bajo. Hay una gran cantidad de clientes con el ingreso como 'Unknown'. ¿A que se debe? ¿Error humano o de maquina?
Esto es un problema a solucionar inmediatamente.



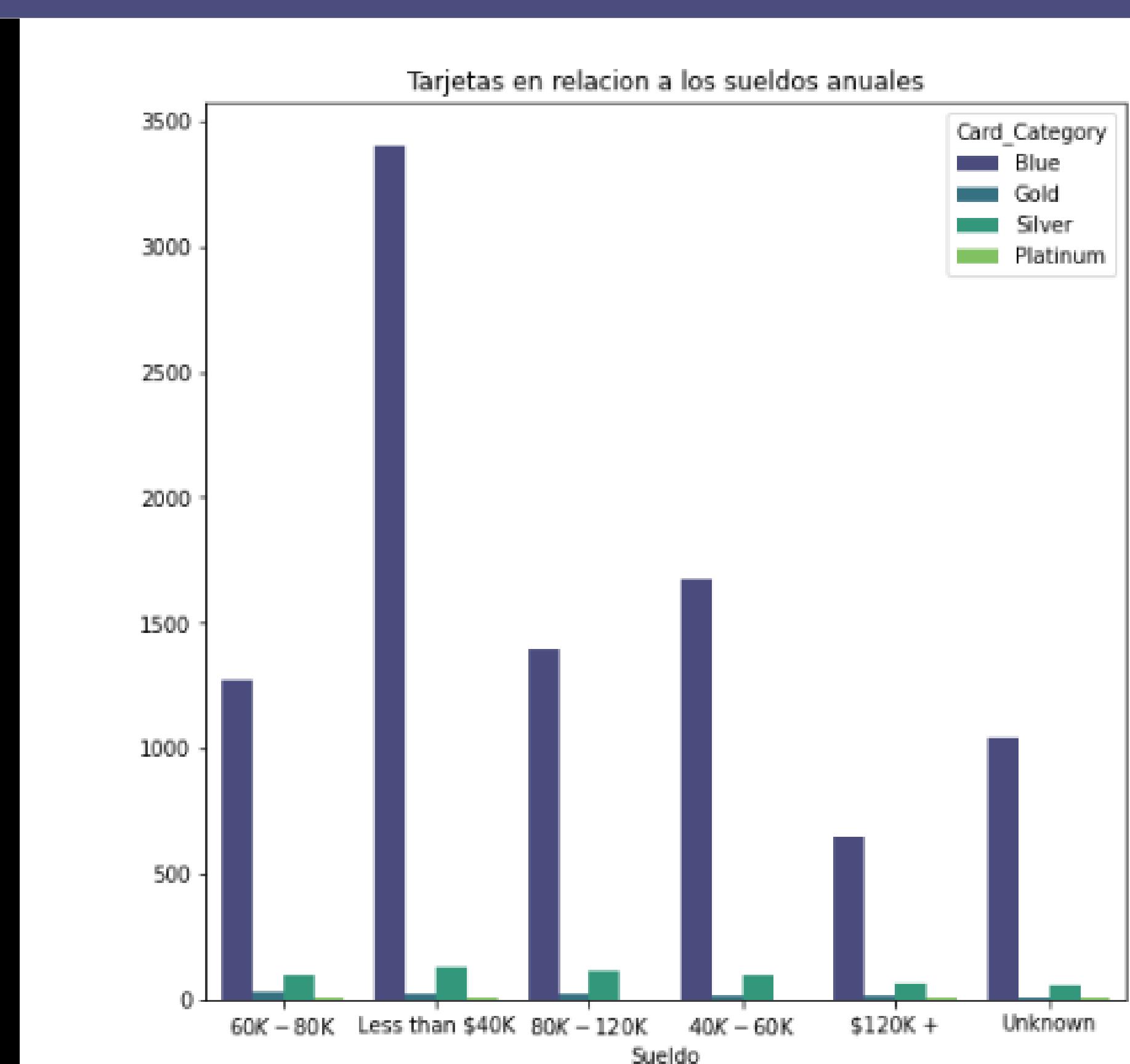
Como se relaciona la tarjeta que tiene un cliente con sus ingresos

La mayoría de clientes tienen ingresos bajos y la mayoría de estos tienen una tarjeta de tipo 'Blue'.

¿Esto influira en los pedidos de credito?

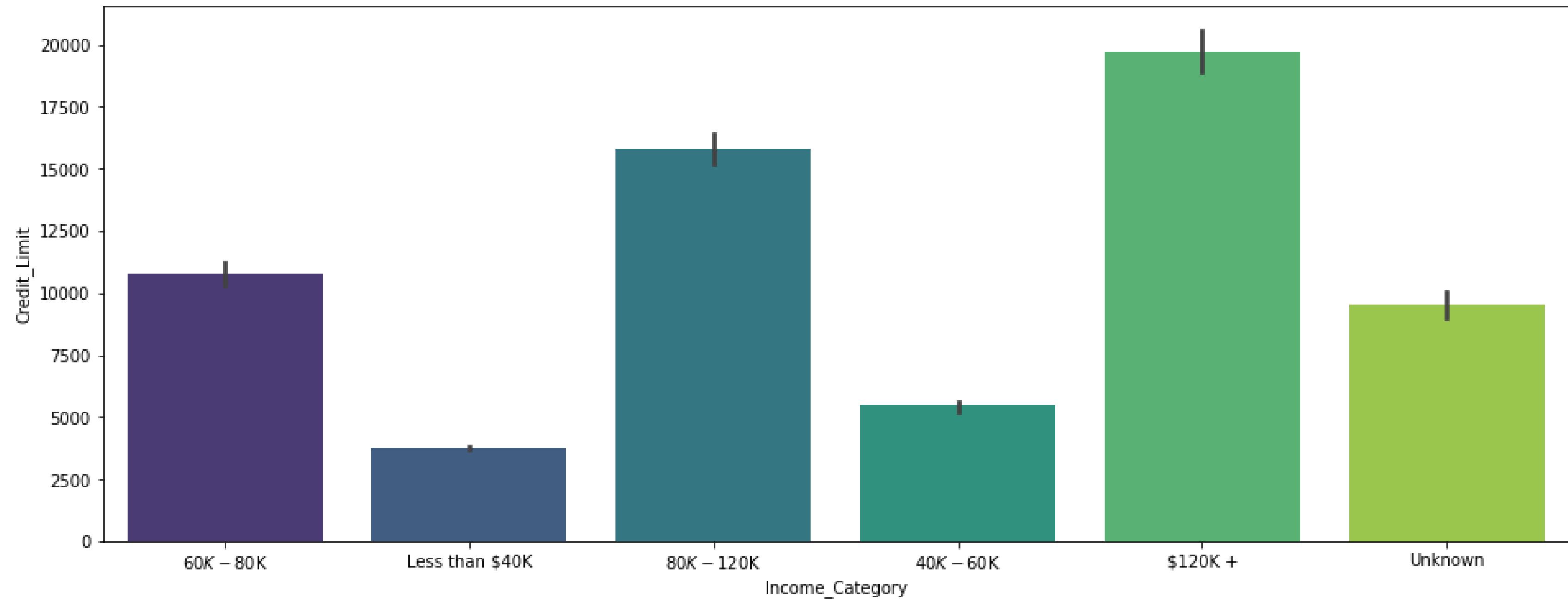
Los clientes con mayores ingresos son los que mas tienen la tarjeta de tipo 'Silver'.

Hay pocos clientes con la tarjeta de tipo 'Platinum' y la mayoría no tienen registrado sus ingresos. Aconsejo cambiar esto lo antes posible y tratar de que mas clientes tengan este tipo de tarjeta.

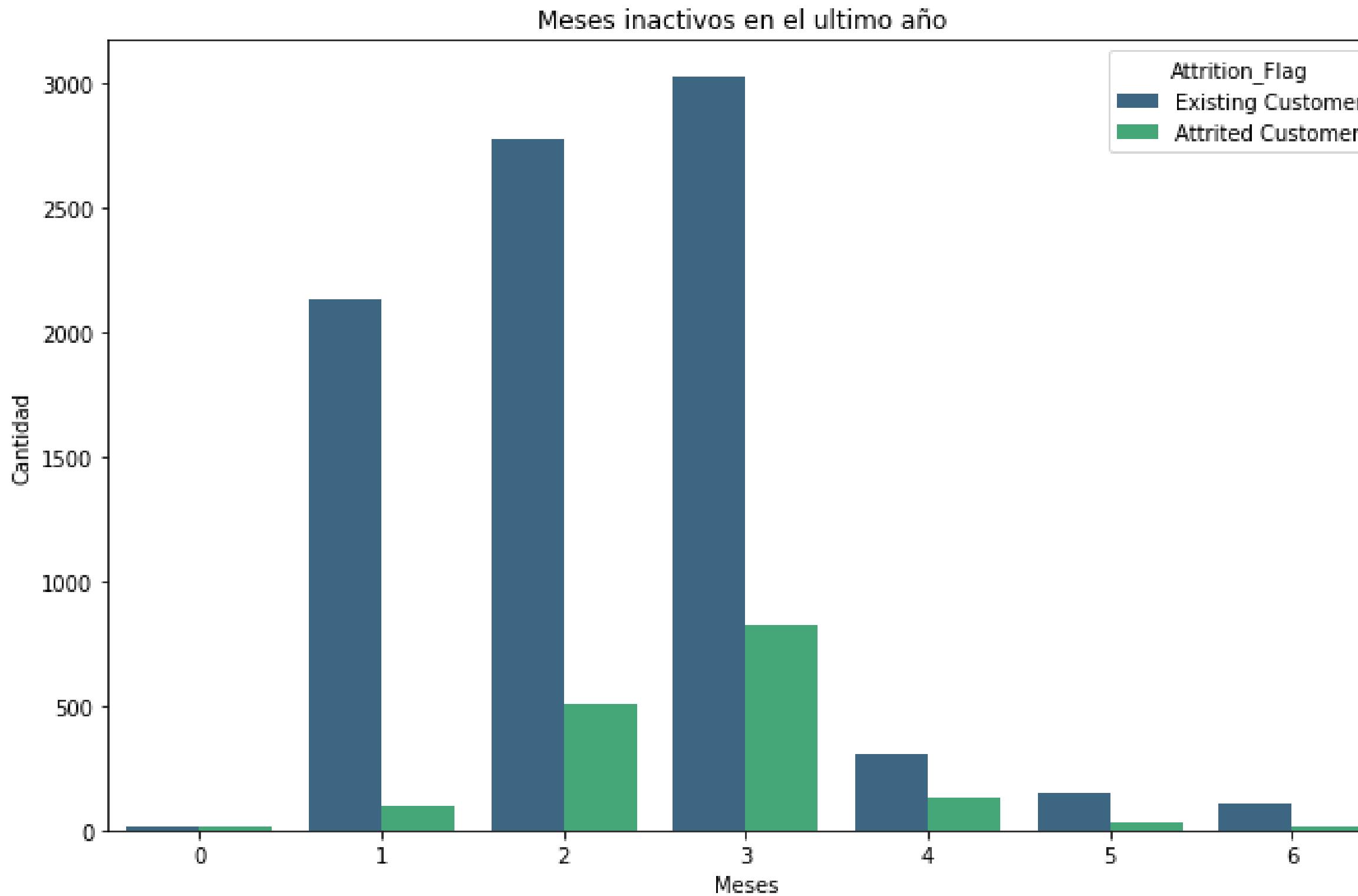


Relación entre los sueldos y el límite de credito otorgado

- Podemos observar que los clientes con menores ingresos, tienen un límite de credito menor mientras que los cliente con mayores ingresos son los que tienen un límite mayor.



¿Cuánto tiempo llevan inactivas las cuentas de los clientes? ¿Por qué?



- Podemos observar que la mayoria de clientes estan inactivos hace 3 meses, tambien se puede ver hace cuanto cerraron su cuenta.

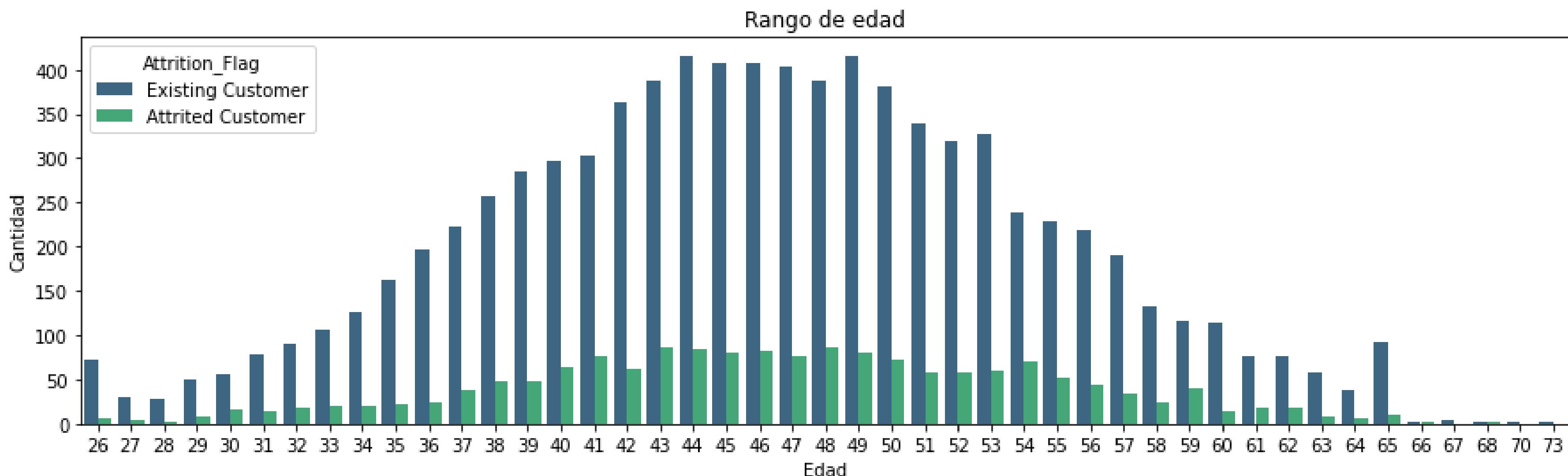
Propongo la siguiente solución:

- Abordar al cliente en la etapa de 3 meses de inactividad para saber porque esta inactivo y si piensa en cerrar su cuenta para asi poder ofrecerle una mejora en el servicio.

¿Cuál es el rango de edad de los clientes?

Podemos ver que en el rango de 37 y 55 años son los que mas abandonan el banco, tal vez reciben mejores ofertas de otros bancos.

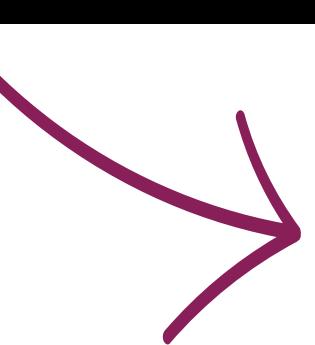
Hay pocos clientes mayores de 73 años, tal vez podriamos tener un paquete de ofertas para las personas mayores.



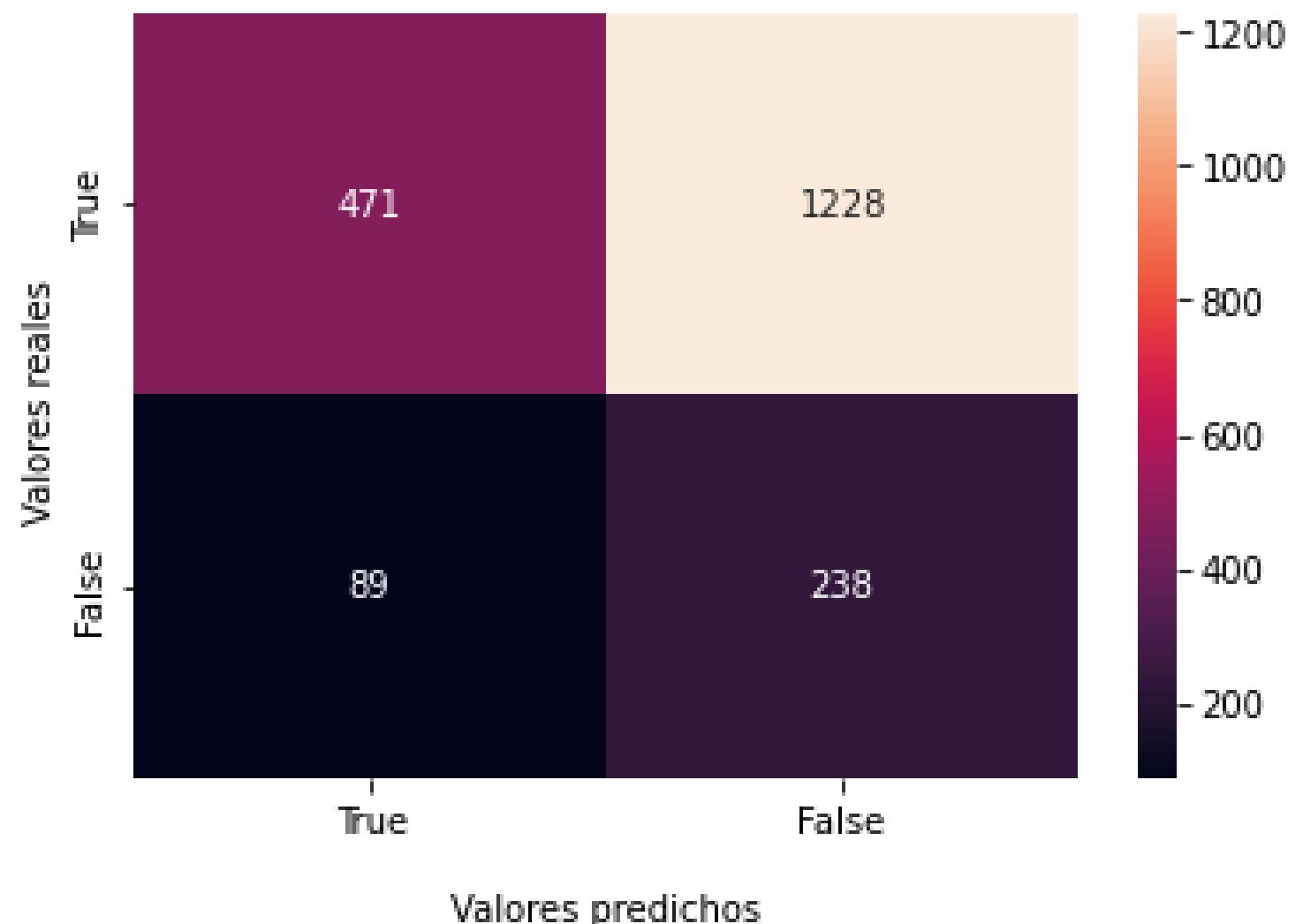
Modelos

* Para todos los modelos se tomo un 80% para entrenar y un 20% para testear

Regresión Logística

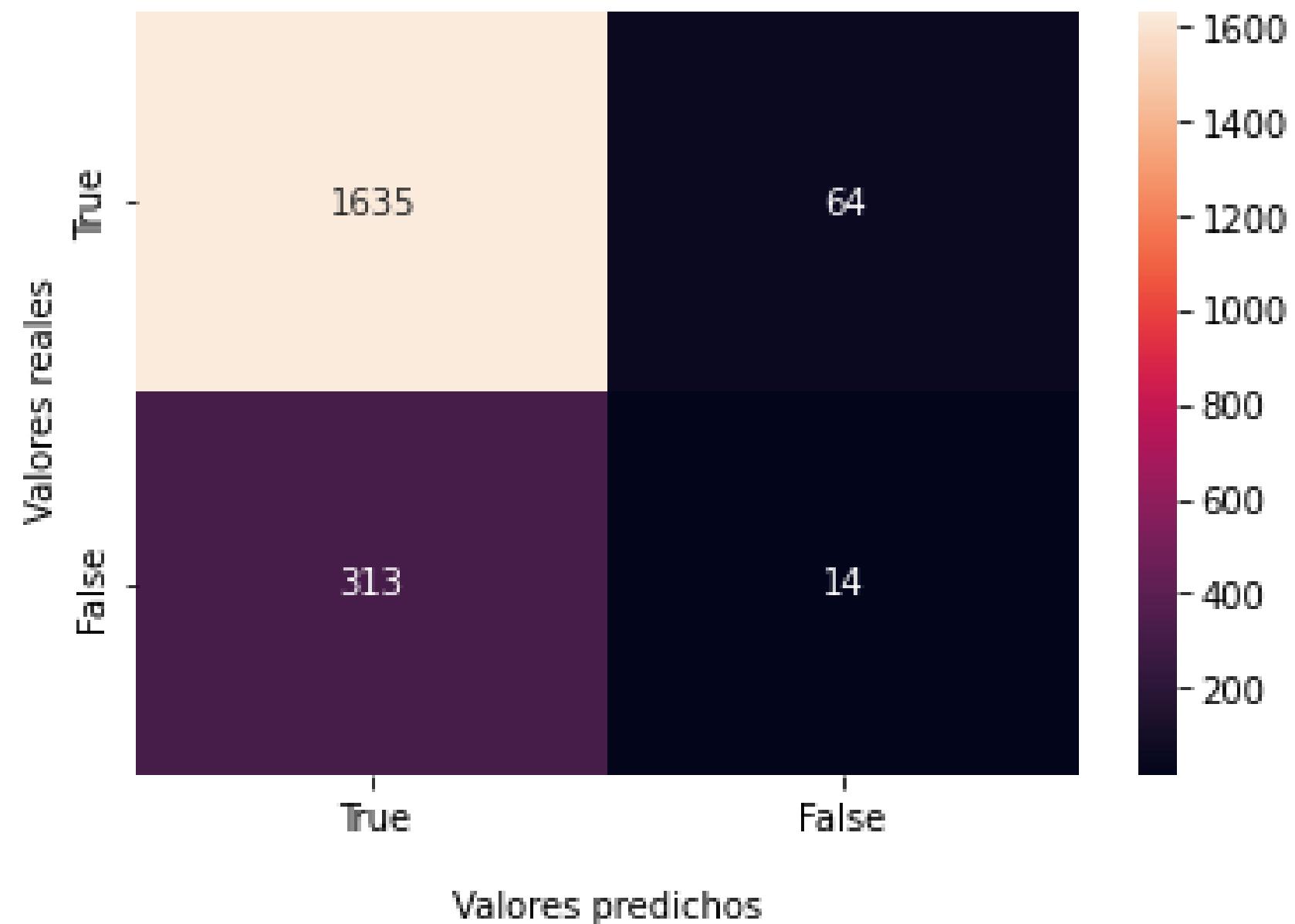


	precision	recall	f1-score	support
0	0.84	0.28	0.42	1699
1	0.16	0.73	0.27	327
accuracy			0.35	2026
macro avg	0.50	0.50	0.34	2026
weighted avg	0.73	0.35	0.39	2026



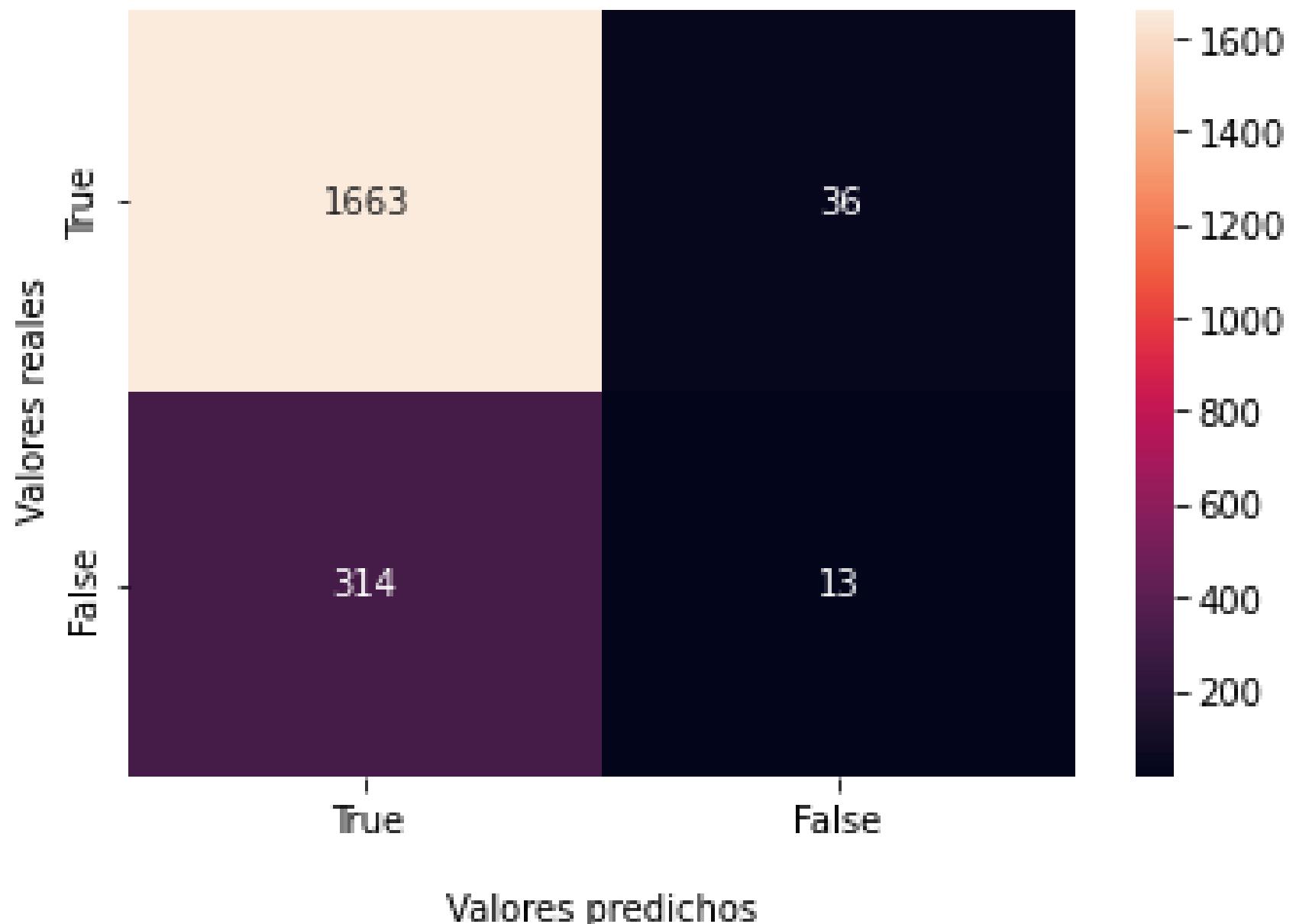
Random Forest

	precision	recall	f1-score	support
0	0.84	0.96	0.90	1699
1	0.18	0.04	0.07	327
accuracy			0.81	2026
macro avg	0.51	0.50	0.48	2026
weighted avg	0.73	0.81	0.76	2026



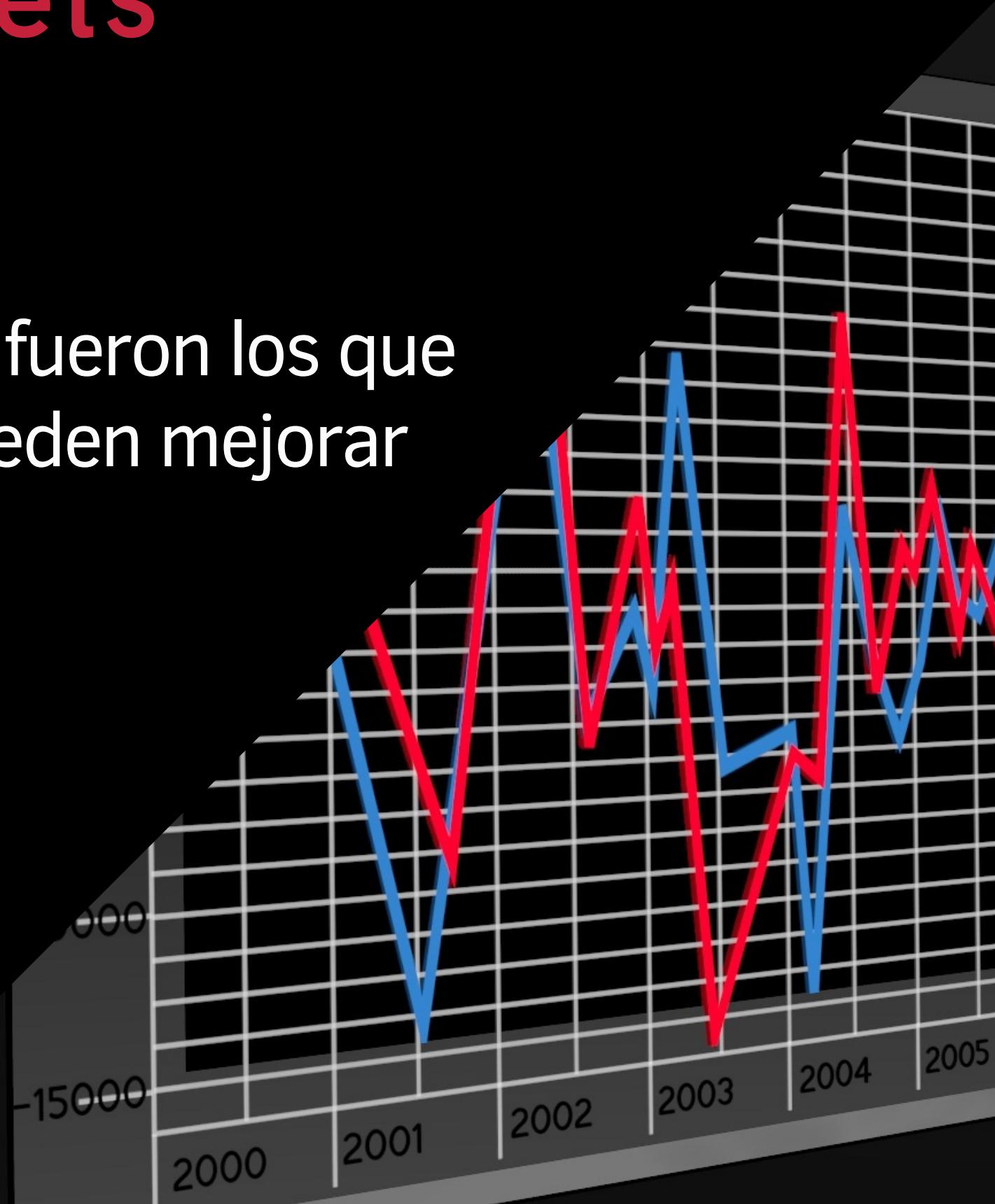
XGBoost

	precision	recall	f1-score	support
0	0.84	0.98	0.90	1699
1	0.27	0.04	0.07	327
accuracy			0.83	2026
macro avg	0.55	0.51	0.49	2026
weighted avg	0.75	0.83	0.77	2026



Boosting Models

Elegimos ‘Random Forest’ y ‘XGBoost’, ya que fueron los que mejor accuracy obtuvieron, para ver si se pueden mejorar



Stratified- K fold - Random Forest

Accuracy: 82%

Randomized Search CV - XGBoost

Accuracy: 84%



Conclusiones

Calculando el accuracy vemos que XGBoost y Random Forest son los que ofrecen los mejores resultados. Elegimos XGBoost porque tiene el accuracy más alto (84%) brindándonos una mejor predicción en las personas que abandonaran o no.

Con los datos obtenidos en el presente notebook nos van a servir para hacer diferentes campañas de marketing dependiendo del perfil del usuario como tambien mejorar las ya ofrecidas a clientes actuales o a clientes nuevos.

Podemos concluir también que al implementar la solución ofrecida, el banco se beneficiaría en una reducción de costos en las campañas que no se orientan al tipo de usuario adecuado, sino que también sería redituable a nivel de ingresos, ya que al mejorar la eficiencia de las campañas también puede orientarse a captar nuevos clientes.

Futuras Lineas

- Para seguir mejorando el proyecto se pueden usar métodos en Feature selection, un método exhaustivo por ejemplo. Para comparar la relevancia de los features y de allí elegir los mejores según los métodos empleados.
- Se podría generar categorización de perfiles de forma automática para clientes existentes o nuevos.
- Para que el modelo sea más robusto y no esté sesgado solo a un caso de negocio, podríamos utilizar datasets de otros bancos para ver como impactaron en los ingresos.
- Implementar una estrategia de Marketing para lograr que mas clientes tengan la tarjeta de tipo 'Platinum'.
- Solucionar el error en la base de datos de clientes con el ingreso 'Unknown'.