



# Proyecto Data Science II

Análisis sobre el abandono de  
clientes de tarjetas de crédito

# Audiencia:

---

- El proyecto está orientado al área de ventas y clientes bancarios para poder determinar cuáles son los factores que propician el abandono de los clientes al servicio de tarjetas de crédito.
  - Estos análisis están sustentados por una base de datos en donde se recopilan datos de los clientes basados en sus ingresos, estado civil, nivel académico, genero, etc.
- 





# Contexto:

- Esta presentación está basada en un dataset que contiene la tasa de abandono de clientes usuarios de tarjetas de crédito. Con esto se procederá a analizar cuáles son los motivos (de haberlos) por los que los clientes abandonan el servicio de tarjetas de crédito, y qué relación existe entre los datos proporcionados y el abandono del servicio.



# Hipótesis

---

Se abarcarán varias hipótesis para poder llegar a una conclusión para así poder retener clientes antes de que estos abandonen el servicio. Nos centraremos en analizar las siguientes variables:

- Históricas: Frecuencia de uso, monto de las transacciones, tipo de gastos, antigüedad de la cuenta, etc.
- Interacciones: Combinaciones de variables que pueden revelar patrones más profundos (por ejemplo, clientes con alta frecuencia de uso, pero bajo monto promedio).
- Demográficas: Edad, género, ubicación geográfica, nivel de ingresos, etc.

Con esto se intentará responder a la pregunta: ¿Existe alguna relación entre el abandono del servicio de tarjetas de crédito que este sustentado por alguna de las variables mencionadas?

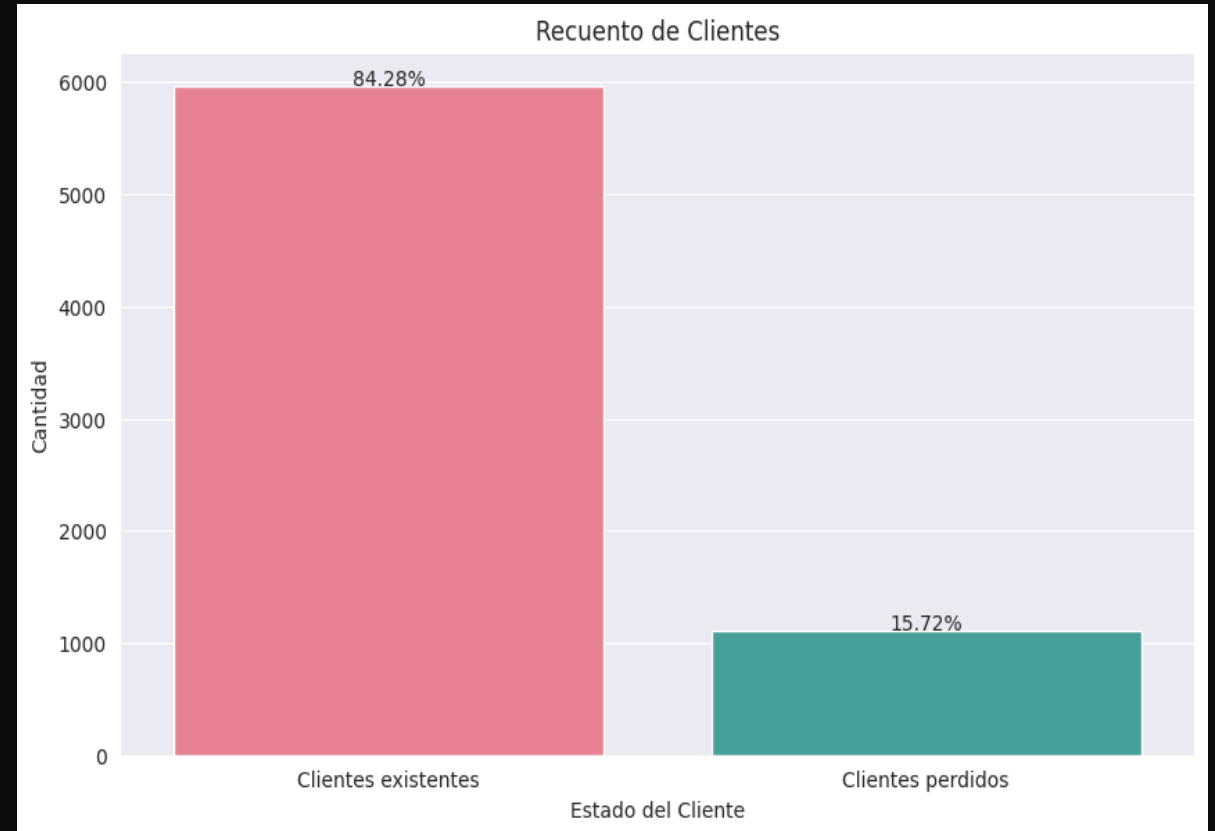
# Sobre el dataset

---

- Resumen de los datos:
  - Casos de estudio: 10127
  - Cantidad de variables examinadas: 20
- Tipo de variables:
  - Categóricas: 6
  - Numéricas: 14

Como primer análisis se muestran los porcentajes de clientes existentes y perdidos. Nosotros nos centraremos en el segundo caso.

---



# ANALISIS EXPLORATORIO DE DATOS - EDA

# Categorización de los clientes

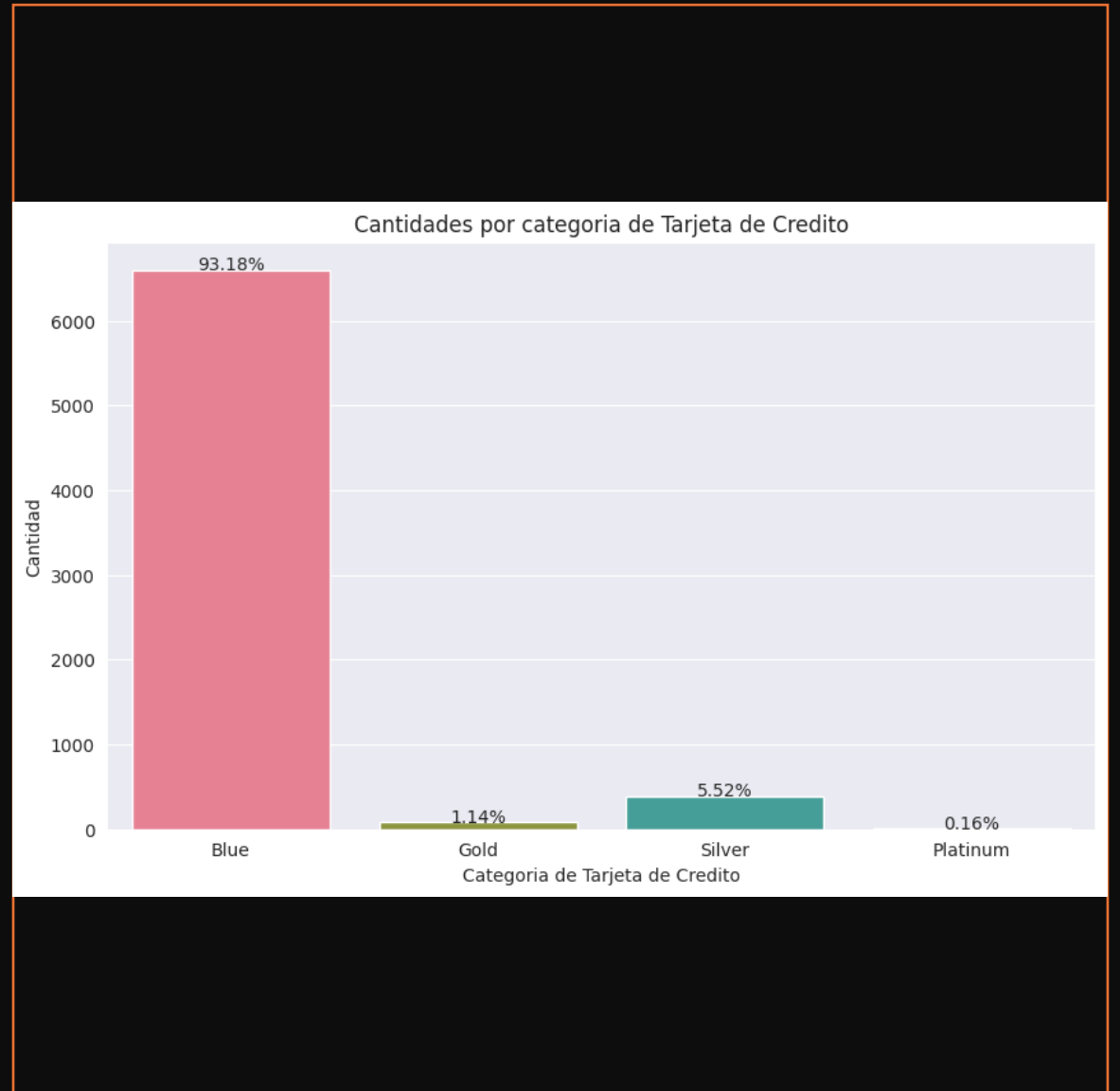
---

Comenzaremos viendo la cantidad de clientes separados en diferentes categorías:

- Por categoría de tarjeta de crédito.
- Por nivel educativo.
- Por estado civil.
- Por nivel de ingresos.
- Por franja de edad

# Porcentajes de clientes por categoria de tarjeta

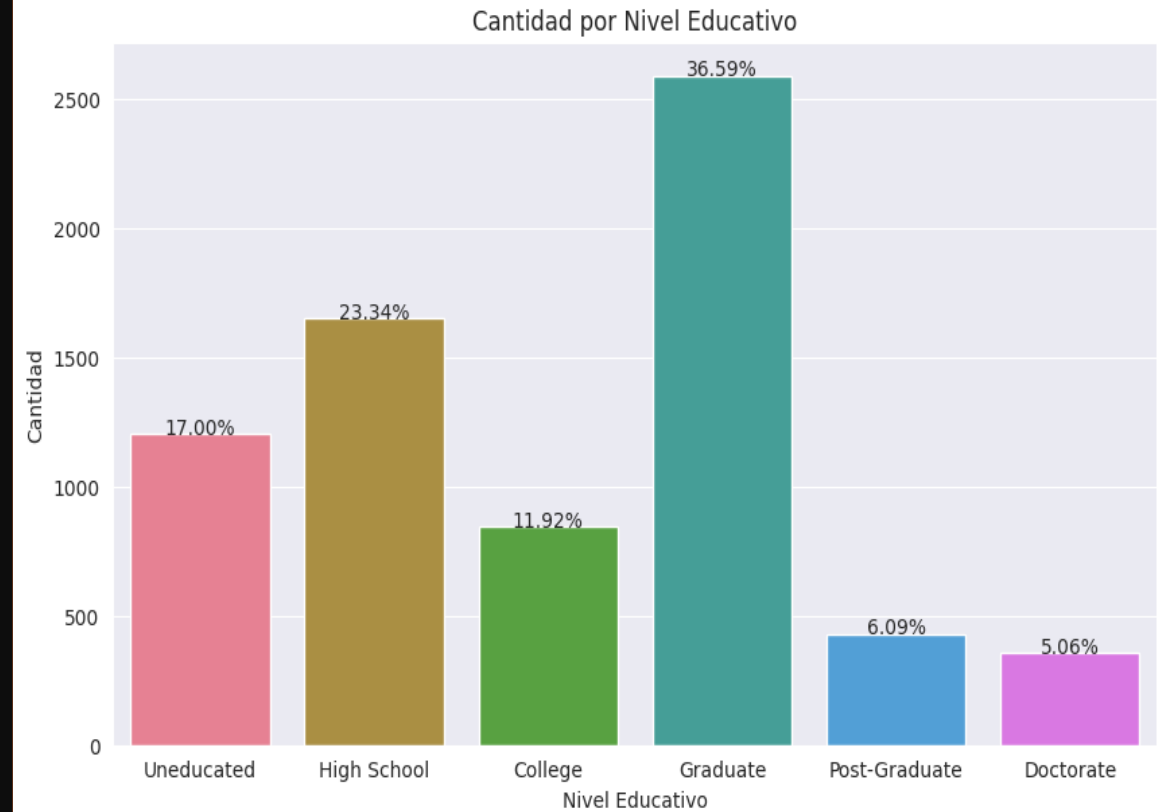
Se observa que la mayoría de los clientes se encuentran en la categoría mas baja (blue)





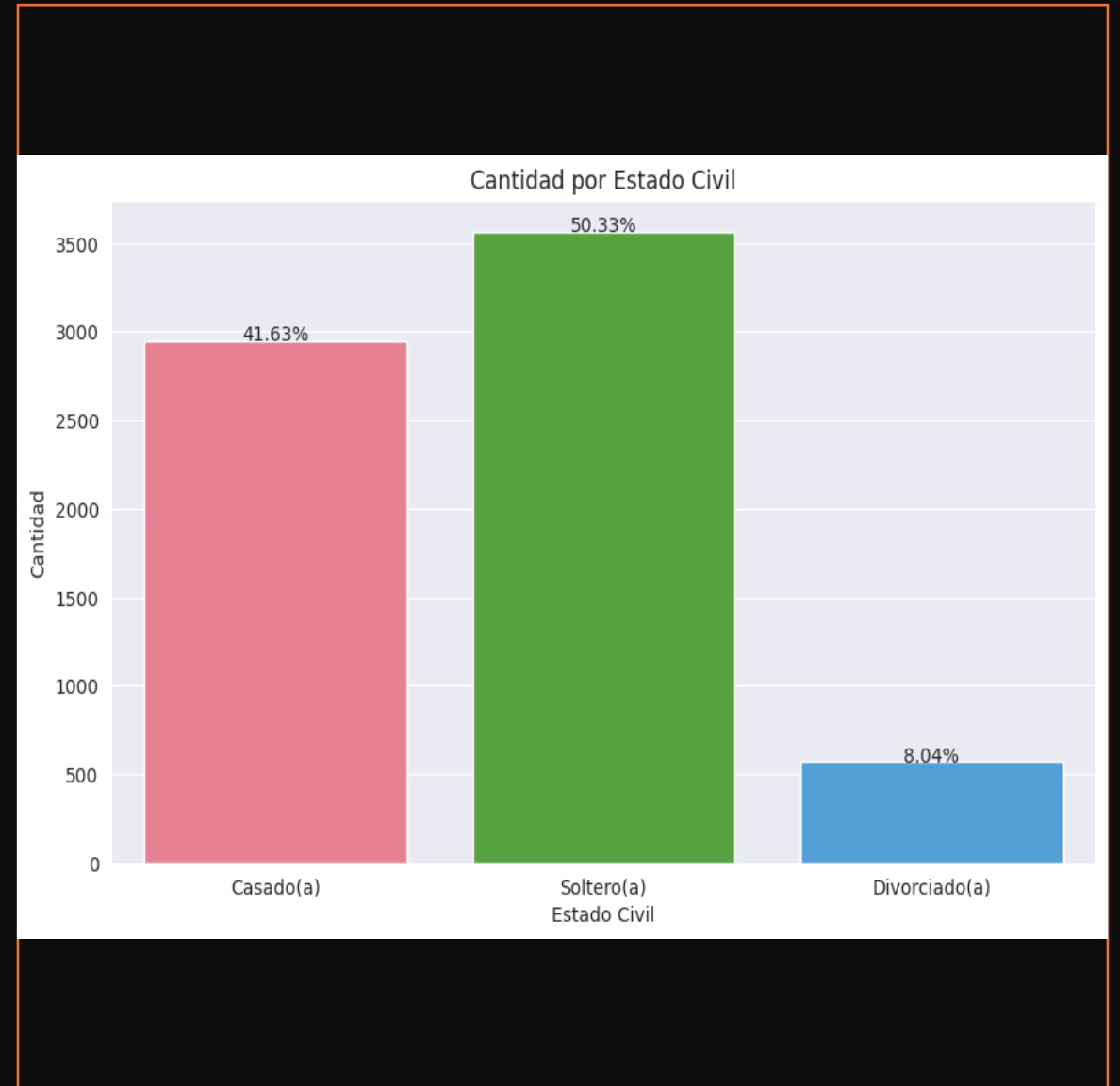
# Porcentajes de clientes por nivel educativo

En este caso aquellos que se encuentran en la categoría “Graduate” son quienes mas contratan el servicio, siendo aproximadamente 2 de cada 5 clientes.



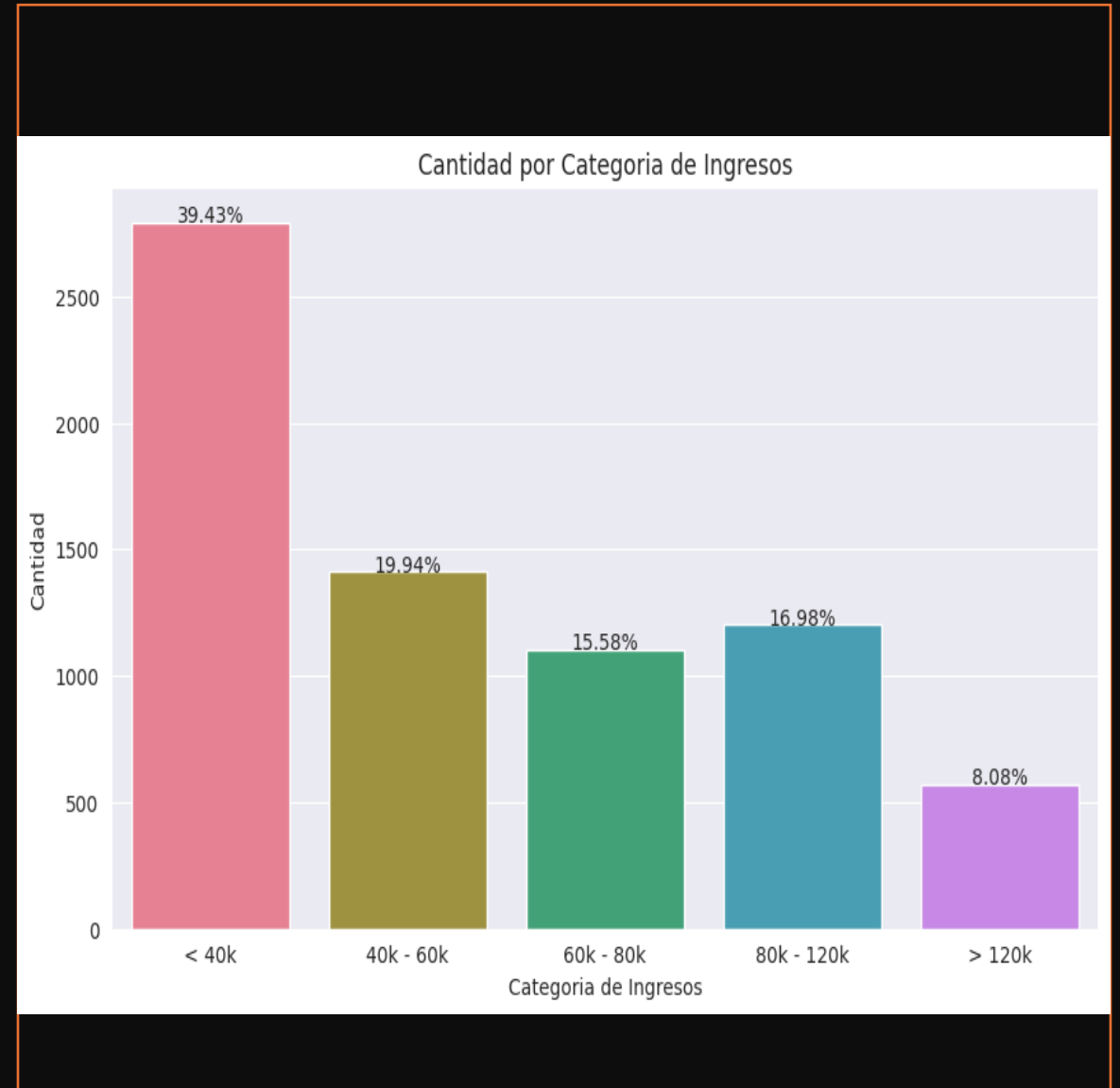
# Porcentajes de clientes por estado civil

Se observa que la mitad de los clientes corresponden a la categoría “Soltero”



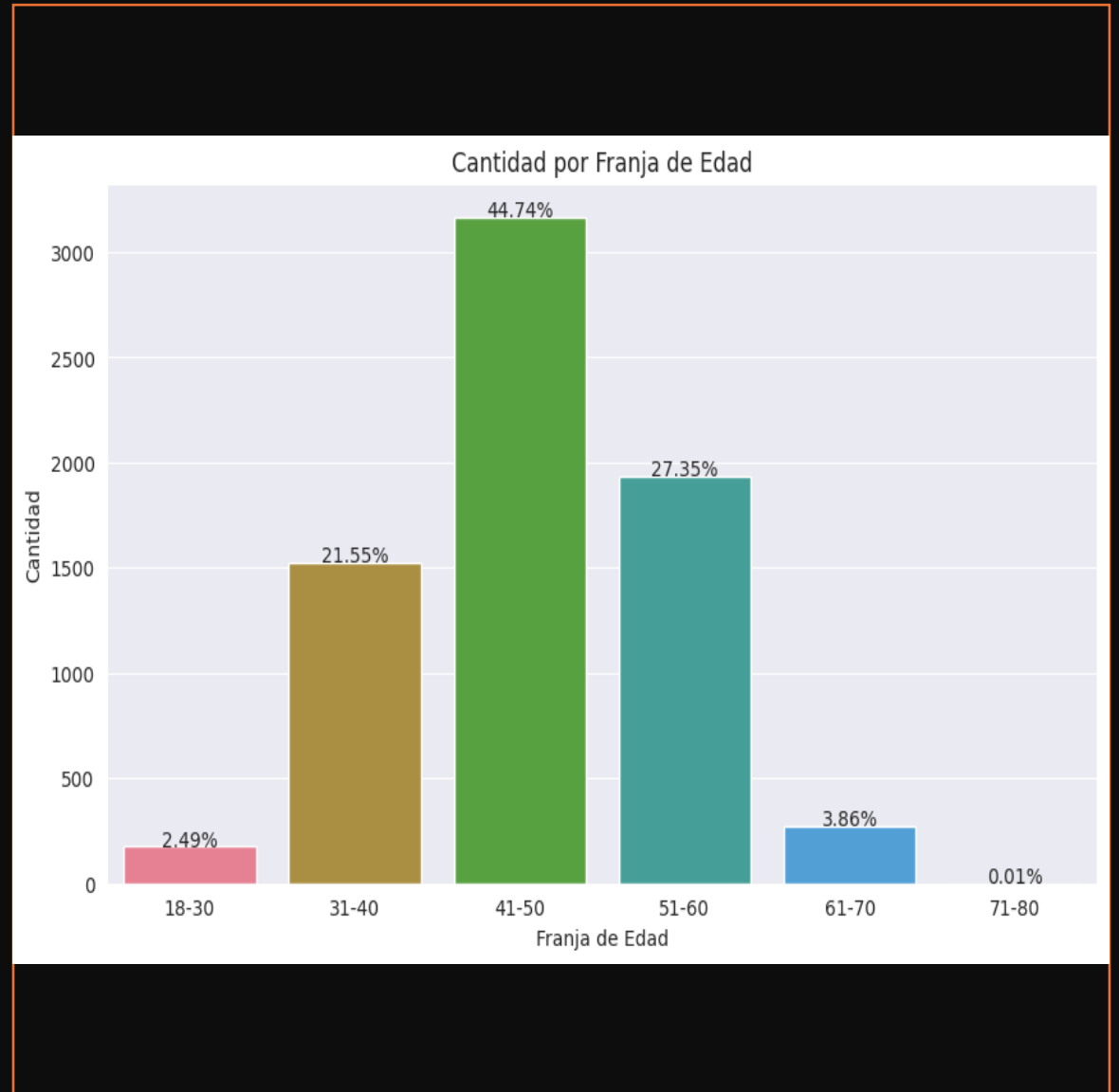
# Porcentajes de clientes por nivel de ingresos

En este gráfico podemos ver que la mayor parte de los clientes pertenecen al decil más bajo de ingresos encontrándose por debajo de los \$40.000

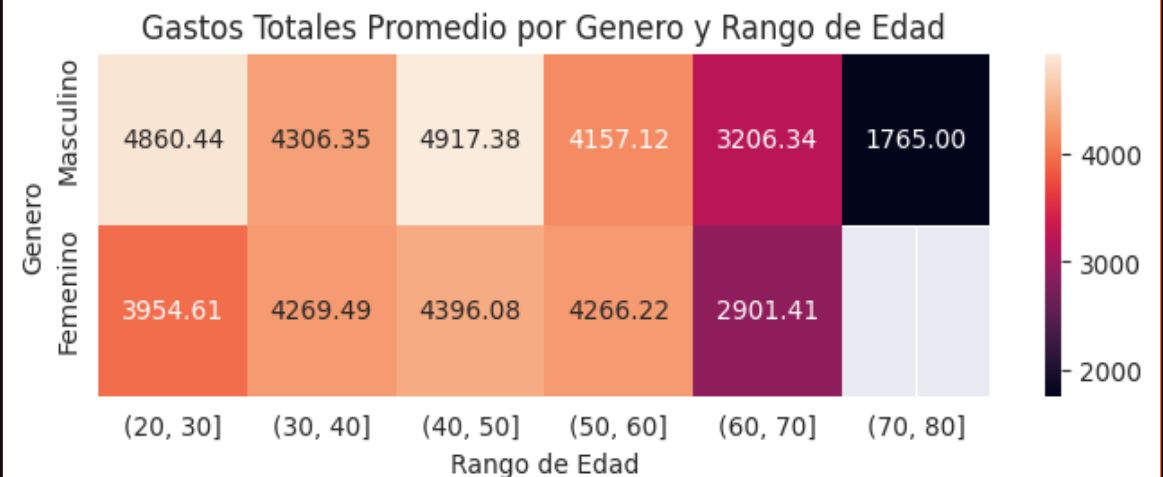


# Porcentajes de clientes por franja de edad

En cuanto a las edades de nuestros clientes la mayoría se encuentra entre los 41 y 50 años.



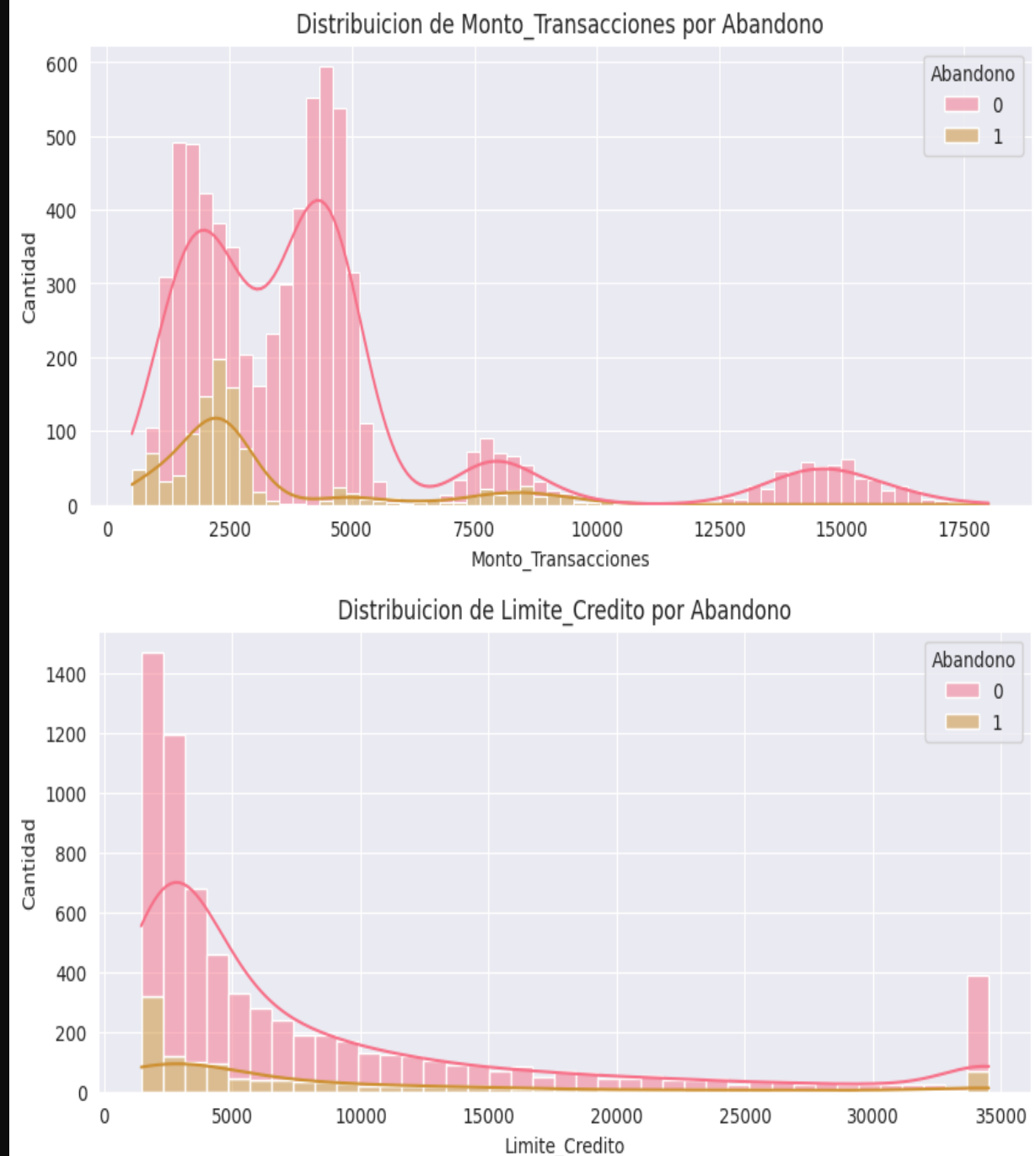
En el siguiente gráfico se muestra la correlación entre los gastos totales promedio y el rango de edad, discriminado por el género de los clientes.



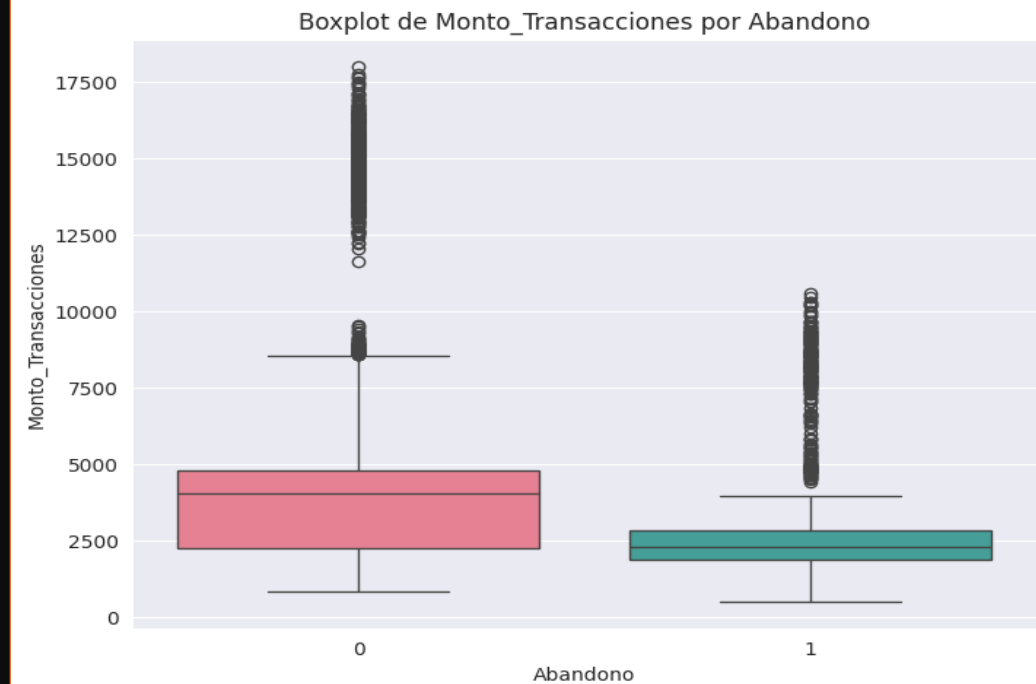
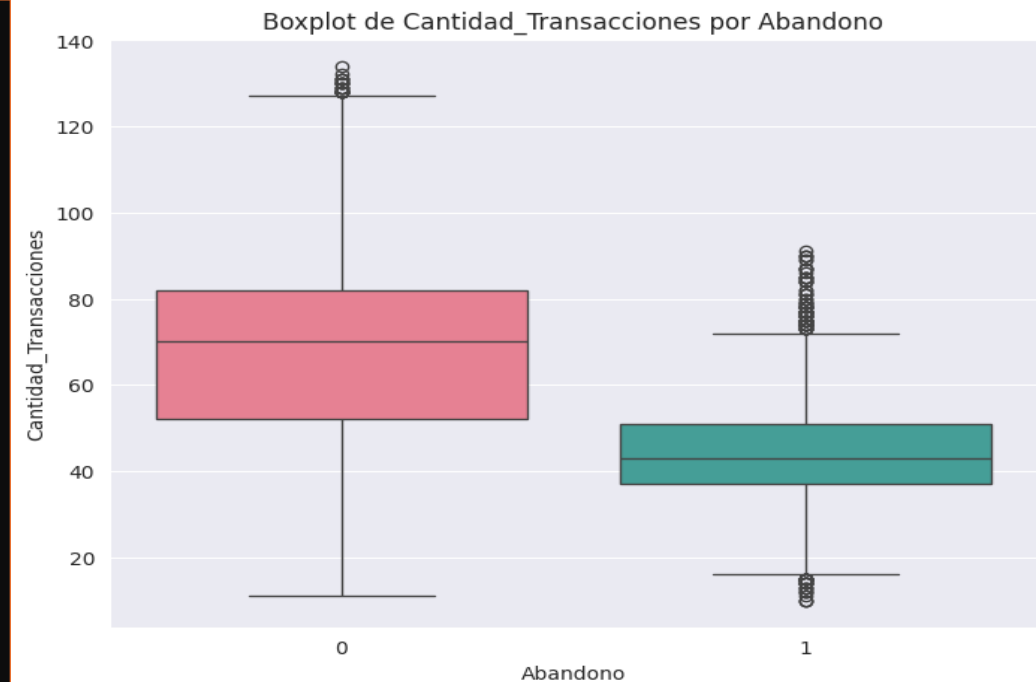


Para tener nociones más concretas sobre el objetivo del estudio y a fin de responder los interrogantes planteados en la hipótesis, a continuación, se mostrarán histogramas y boxplot de las distintas variables en relación con el abandono del servicio.

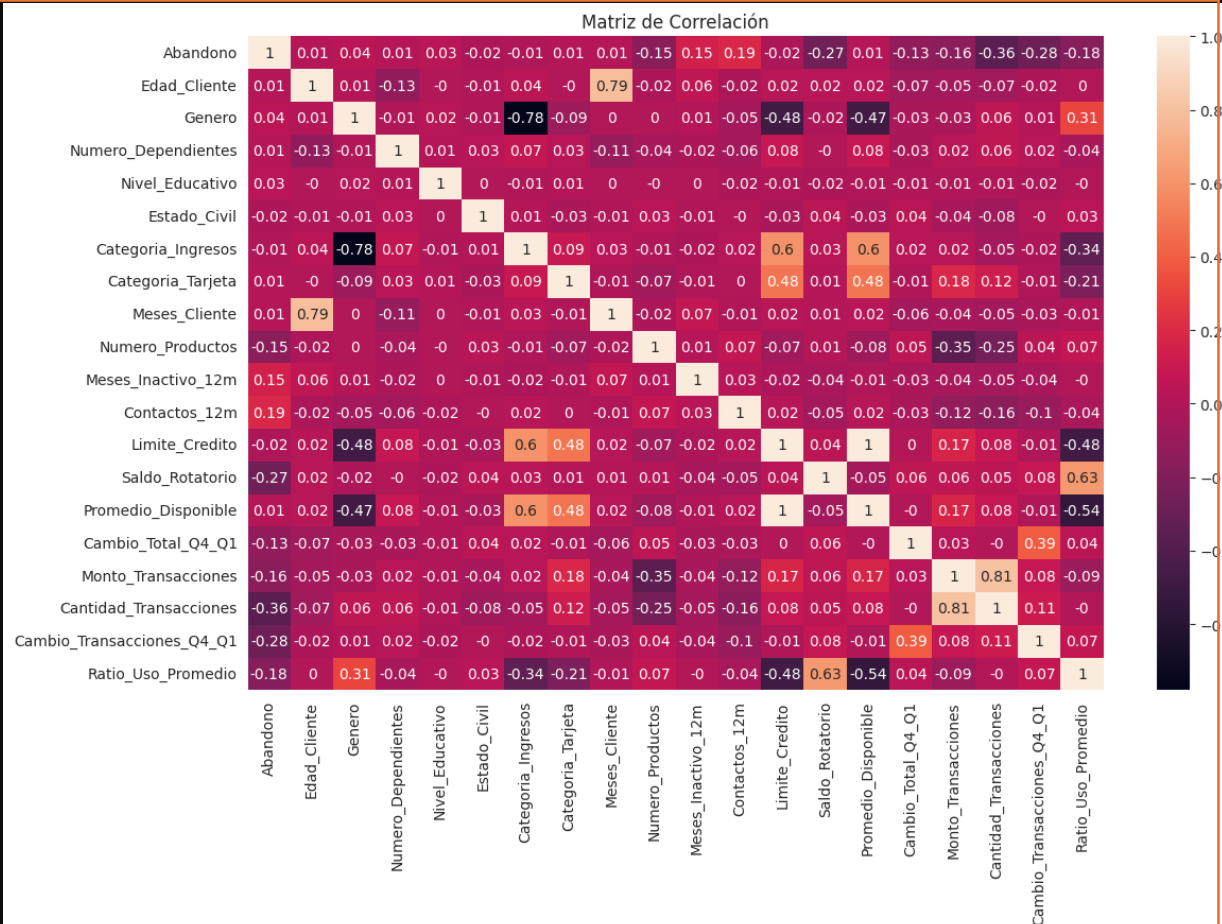
En los siguientes gráficos se muestran las distribuciones de abandono del servicio por límite de crédito y monto de transacciones. Con estos datos podemos ver que los clientes con mayor probabilidad de abandonar el servicio se da en aquellos con transacciones en el orden de los \$2500 y con un límite de crédito bajo.



En los gráficos de caja (boxplot) se observa que la mayor cantidad de datos outliers (valores atípicos) para la categoría de abandono del servicio se encuentran en la “Cantidad de Transacciones” y “Monto de Transacciones” siendo muchos más en este último.



La siguiente matriz de correlación nos muestra que entre las categorías que tienen una mayor correlación con el abandono se encuentran como ya hemos visto la “Cantidad de Transacciones” y “Monto de Transacciones” a las que se puede agregar el “Saldo Rotatorio” y el “Ratio de Uso Promedio”



A continuación, se detallan las particularidades de abandono del servicio que se desprenden de los análisis posteriores

- Categoría de ingresos predominante: menos de \$40k.
- Género: Los hombres del rango de edad entre 40 a 50 años son quienes poseen más gastos promedio.
- Categoría de tarjeta dominante: Blue.
- Límite de tarjetas: cuanto menor sea el límite de tarjetas, mayor será la deserción.
- Saldo rotatorio: cuando el saldo = 0 es donde vemos mayor concentración de abandono.
- Valor de transacción: mayor concentración de abandono en clientes que realizaron transacciones por un valor de \$2500.
- Número de transacciones: mayor concentración de abandono en clientes que realizaron entre 40 y 60 veces en los últimos 12 meses.

A fin de dar una respuesta a la hipótesis planteada en el principio, se puede inferir que a priori tanto el límite de crédito como el valor de las transacciones realizadas son un factor para tener en cuenta a la hora de que los clientes abandonen el servicio.

Con el fin de mejorar estos análisis en un futuro se desarrollarán modelos de predicción basados en “Machine Learning” para poder predecir de una manera más eficiente a aquellos potenciales clientes que abandonen el servicio.