

# Credit Card Bank Churn

Análisis del comportamiento de los clientes para reducir la tasa de abandono y mejorar las estrategias de fidelización en los servicios bancarios

ALUMNO: JOSÉ CARRIZO

COMISIÓN: 67395

PROFESOR: GUSTAVO BENÍTEZ

TUTOR: LUCIANO LUSACHI

TRABAJO FINAL 2025



# ÍNDICE

• • • • •	1	Preguntas	3
• • • • •	2	Hipótesis	4
• • • • •	3	Descripción del dataset	5
• • • • •	4	EDA	6
• • • • •	5	Insights - Takeways	7
• • • • •	6	Feature Engineering & Selection	9
• • • • •	7	Modelos	10
• • • • •	8	Modelos y optimizaciones	11
• • • • •	9	Predicciones claves	12
• • • • •	10	Conclusiones Finales	13

# Preguntas

Comprender los desafíos de retención de clientes en bancos

1

¿Qué características del cliente predicen mejor la baja del servicio de tarjeta de crédito?

2

¿Influye la edad o la antigüedad del cliente en el banco en la decisión de darse de baja?

3

¿Los clientes con mayor uso de tarjetas y saldos rotativos tienden a quedarse?

4

¿El nivel de ingresos afecta la probabilidad de abandono?

# Hipótesis

Conjeturas de investigación para explorar los comportamientos de abandono en los clientes

1

Los clientes con límites de crédito más bajos y con menores transacciones son más propensos a darse de baja debido a una menor interacción y satisfacción.

2

Los clientes más jóvenes y con menor antigüedad son más propensos a darse de baja, posiblemente debido a la falta de compromiso a largo plazo o menor lealtad.

3

Los clientes con mayor uso de tarjetas tienen menores probabilidades de abandonar la cuenta de la tarjeta de crédito, ya que dependen mucho de sus gastos.

# Descripción del Dataset

Comprender los componentes y la estructura de nuestro conjunto de datos para el análisis

El dataset utilizado en este proyecto se titula "BankChurners.csv".

Contiene información anonimizada de clientes de tarjetas de crédito de una institución financiera.

Incluye, además, datos demográficos (como edad y sexo), comportamiento crediticio (como límite de crédito y saldo rotativo), métricas de actividad del cliente (incluyendo transacciones totales y tasa de utilización) y estado de abandono (activo o deserción).



# Análisis Exploratorio de Datos (EDA)

Comprender patrones y tendencias en los datos de abandono de servicios de tarjetas de crédito.

El objetivo de esta fase es comprender la estructura, la distribución y las relaciones dentro del dataset relacionados con los clientes de tarjetas de crédito y su comportamiento de abandono. Proporciona información crucial sobre cómo las variables demográficas y de comportamiento que interactúan con la pérdida de clientes, sentando las bases para el modelado predictivo y la generación de recomendaciones comerciales.



# Insights & Takeways

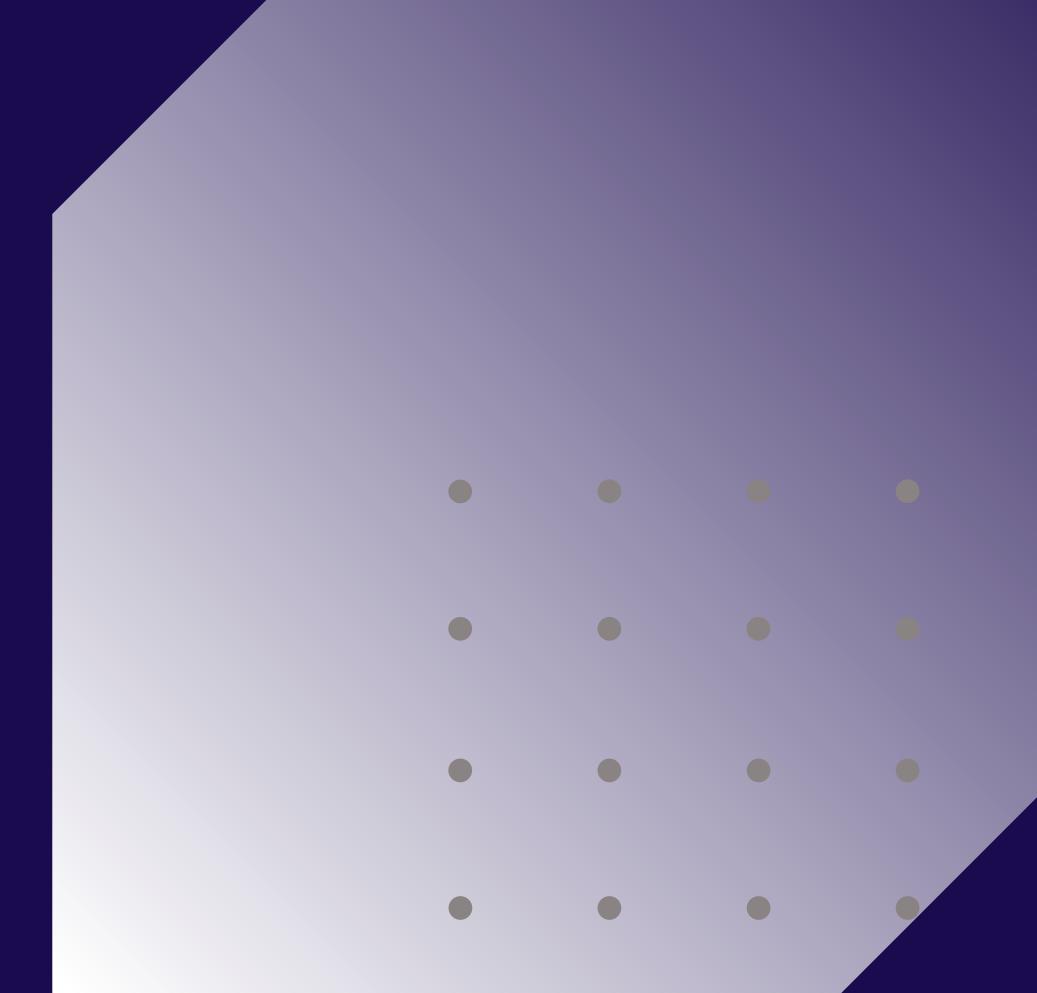


## 1. Tasa de Abandono y Desequilibrio de Clase

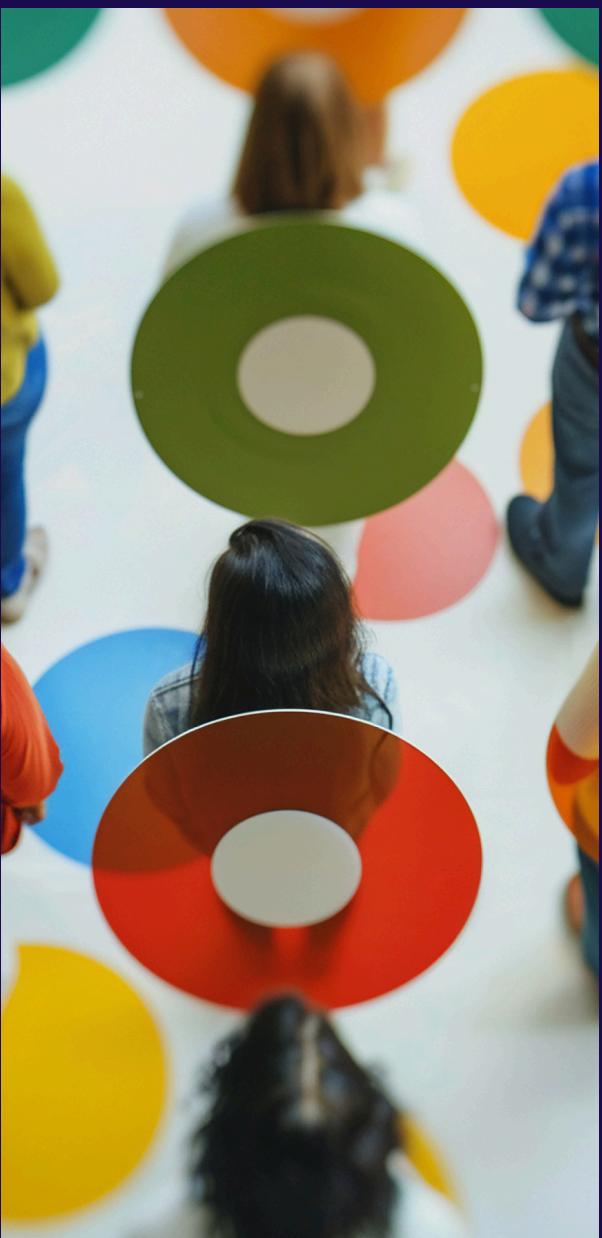
Aproximadamente el 16% de los clientes abandonaron el servicio, lo que revela un desequilibrio de clase significativo. Esto debe abordarse durante el modelado para garantizar predicciones precisas.

## 2. Observaciones del Perfil del Cliente

- La mayoría de los clientes tienen entre 40 y 55 años y estudios de posgrado.
- La mayoría posee tarjetas Blue estándar y gana menos de \$40,000 al año.
- Los clientes que abandonan tienden a tener límites de crédito bajo y menos transacciones, lo que sugiere poco “feeling” con el banco.



# Insights & Takeways



## 3. Patrones de Comportamiento y Abandono

- Los clientes que abandonan tienen menos transacciones mensuales y cantidades totales de transacción inferiores.
- Los índices de utilización tienden a ser menores para los clientes que abandonan la plataforma; especialmente, en las categorías de bajos ingresos.
- Una alta interacción del cliente se correlaciona con un menor riesgo de abandono.

## 4. Calidad de los datos

- El dataset no presenta valores faltantes y se encontraron valores cero principalmente en las características basadas en el uso (por ejemplo, saldo rotativo), lo que podría reflejar una inactividad real.



# Feature Engineering & Selection

Creación de features de datos valiosas para el análisis

- 1 Aplicación de One-Hot Encoding
- 2 Se agrega la feature “EngagementScore” que combina dos métricas vinculadas a la actividad transaccional.
- 3 Definición de la variable objetivo y generación de la lista “features” con los nombres de columna que NO sean igual a la variable objetivo.
- 4 Puesta a punto para la gestión del desbalance de clases y entrenamiento de modelos de clasificación.

# Modelos

Descripción general de los modelos de predicción de abandono utilizados en el análisis

1

**Logistic Regression:** gestiona la tarea de la clasificación de abandono de forma correcta así como también si sobreestima o subestima la tasa de abandono. Es adecuado como referencia en la comparación con otros modelos más complejos.

2

**Random Forest:** buena gestión tanto de los datos categóricos como de los numéricos. Además, resiste al “overfitting” debido a la naturaleza del dataset.

3

**XGBoost:** muestra excelentes rendimientos en dataset tabulares y gestión de los valores faltantes, la no linealidad y las interacciones.

# Optimización de los modelos

Técnicas para mejorar el rendimiento y la precisión del modelo

1

Se seleccionó **XGBoost** para el ajuste de hiperparámetros puesto que presenta un desempeño superior en todas las métricas de evaluación (sobre todo, con **F1 Score** y **ROC AUC**).

2

Se aplicó GridSearchCV con Cross Validation de 3 pasos para explotar combinaciones de hiperparámetros claves.

3

Se utilizó ROC AUC como métrica de puntuación para garantizar un rendimiento equilibrado, tanto para los clientes que abandonan como para los que permanecen.

# Predictores claves

Comprender los factores más influyentes en el abandono del servicio por parte de clientes

1

La identificación de las features más impactantes mejora el rendimiento del modelo.

2

**Las características redundantes pueden generar overfitting y complejidad.**

3

La selección optimizada de funciones mejora el modo en que interpretamos y comprendemos los datos.

# Conclusiones finales

- El dataset presentó una tasa de pérdida del 16%, lo que confirme un evidente desequilibrio de clases que se gestionó con éxito mediante SMOTE y “class weighting”.
- Características como el total de transacciones, la puntuación de interacción, el límite de crédito y la tasa de utilización mostraron una fuerte relación con la pérdida de clientes.
- XGBoost presentó el mejor rendimiento general, con el AUC ROC y F1 Score más altos, lo que lo convierte en el modelo más eficaz para identificar la pérdida de clientes.
- El ajuste de hiperparámetros mediante GridSearchCV mejoró aún más la precisión y la generalización del modelo.



# Muchas gracias por la atención

José Carrizo

Ingeniero en Sistemas

carrizoja@gmail.com

