

# Exploración y segmentación de clientes

## Introducción

En el mundo financiero, la segmentación de clientes es una herramienta clave para personalizar servicios, optimizar estrategias de marketing y gestionar riesgos. Este proyecto se enfoca en el análisis de datos de tarjetas de crédito para identificar patrones en el comportamiento de los clientes mediante técnicas de clustering.

El objetivo es ayudar a la institución financiera Super Bank a entender mejor las diferencias entre segmentos de clientes y extraer información valiosa para la toma de decisiones. Como científicos de datos, deberán aplicar un flujo completo de análisis exploratorio de datos y técnicas de aprendizaje no supervisado, enfocado a un problema real del sector financiero.

## Dataset

Las columnas del dataset 'Super Bank' se detallan a continuación:

**CUST\_ID:** Identificación del titular de la tarjeta de crédito (Categorical).

**BALANCE:** Cantidad de saldo restante en su cuenta para realizar compras.

**BALANCE\_FREQUENCY:** Frecuencia con la que se actualiza el saldo, puntuación entre 0 y 1 (1 = actualizado frecuentemente, 0 = no actualizado frecuentemente).

**PURCHASES:** Cantidad total de compras realizadas desde la cuenta.

**ONEOFF\_PURCHASES:** Monto máximo de una compra realizada en una sola ocasión.

**INSTALLMENTS\_PURCHASES:** Cantidad de compras realizadas a plazos.

**CASH\_ADVANCE:** Adelanto en efectivo otorgado al usuario.

---

**PURCHASES\_FREQUENCY:** Frecuencia con la que se realizan las compras, puntuación entre 0 y 1 (1 = compras frecuentes, 0 = compras poco frecuentes).

**ONEOFFPURCHASESFREQUENCY:** Frecuencia con la que se realizan compras en una sola ocasión (1 = compras frecuentes, 0 = compras poco frecuentes).

**PURCHASESINSTALLMENTSFREQUENCY:** Frecuencia con la que se realizan compras a plazos (1 = realizadas frecuentemente, 0 = no realizadas frecuentemente).

**CASHADVANCEFREQUENCY:** Frecuencia con la que se realiza el adelanto en efectivo.

**CASHADVANCETRX:** Número de transacciones realizadas con adelantos en efectivo.

**PURCHASES\_TRX:** Número de transacciones de compra realizadas.

**CREDIT\_LIMIT:** Límite de crédito de la tarjeta para el usuario.

**PAYMENTS:** Monto de los pagos realizados por el usuario.

**MINIMUM\_PAYMENTS:** Monto mínimo de pagos realizados por el usuario.

**PRCFULLPAYMENT:** Porcentaje de pago completo realizado por el usuario.

**TENURE:** Tiempo de uso del servicio de tarjeta de crédito por parte del usuario.

## Instrucciones

1. Carga el dataset
  - a. Carga el conjunto de datos proporcionado.
  - b. Realiza una evaluación preliminar de los datos. Identifica el número de características, sus tipos de datos y verifica la existencia de valores faltantes.
2. Limpieza de Datos:
  - a. Identifica al menos un mecanismo para imputar datos y justifícalo
  - b. Elimina datos de ruido si existieran
3. Detección y Eliminación de Valores Atípicos:
  - a. Identifica posibles valores atípicos en el conjunto de datos utilizando métodos estadísticos (por ejemplo, Z score, Rango intercuartil, etc.).
  - b. Justifica el método que elegiste para la detección y manejo de valores atípicos.

- 
- c. Elimina los valores atípicos del conjunto de datos. Describe cómo la eliminación de valores atípicos impacta en el conjunto de datos.
  4. Transformación de Datos:
    - a. Normalizar o estandarizar los datos numéricos si es necesario. Explica tu elección de técnica.
    - b. Aplica discretización de ser necesario
  5. Análisis Univariado:
    - a. Selecciona dos variables y realiza un análisis en cada una.
    - b. Incluye y habla sobre las medidas de tendencia central (media, mediana, modo) y medidas de dispersión (rango, varianza, desviación estándar).
    - c. Proporciona al menos una visualización (por ejemplo, histogramas, diagramas de caja) por cada variable seleccionada.
  6. Generación de Insights:
    - a. Basado en el análisis realizado, infiere al menos un insight del conjunto de datos, considerando el contexto del problema.
    - b. Cada insight debe estar respaldado por evidencia estadística y visualizaciones pertinentes.
  7. Clustering:
    - a. Aplicar el método del codo para determinar el número de clusters adecuado.
    - b. Aplicar al menos un método de aprendizaje no supervisado para segmentar clientes.
  8. Caracterización de Clusters
    - a. Interpretar/describir los grupos obtenidos, proporcionando las gráficas adecuadas.

## Entregables

### 1. Notebook

Link de notebook de Google Colab. Deben de estar todos los pasos tomados durante el análisis de datos exploratorio, incluyendo:.

- Visualizaciones generadas. (es decir, ¡no deben borrar los outputs de sus celdas del notebook!)
- Interpretaciones y conclusiones derivadas de cada sección de análisis.
- Insights y cosas extra que hayan realizado
- Bibliografía de recursos consultados

### 2. Presentación

---

Un vídeo de **máximo 10 min** en donde, usando su notebook, presenten su trabajo (metodología, insights, nuevas librerías o métodos usados que no se hayan visto en clases, etc.). Se evaluará:

- Estructura de la presentación: orden, cohesión
- Comunicación: claridad, coherencia, terminología apropiada
- Contenido
- Duración (**máximo 10 minutos**)