**Proyecto de Data Science**

**“Predicción de abandono**

**para clientes de una**

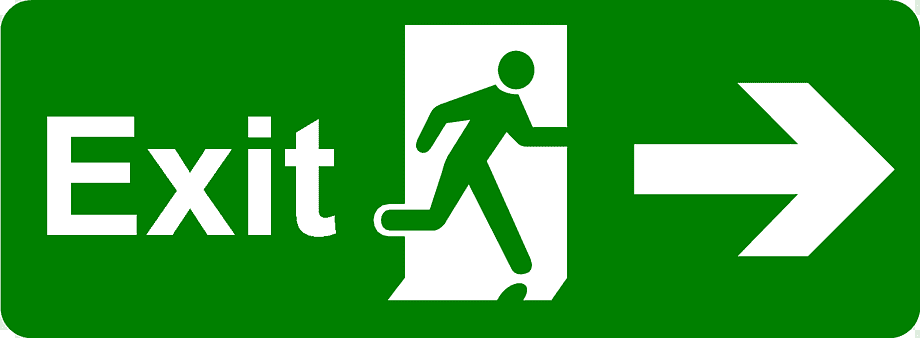
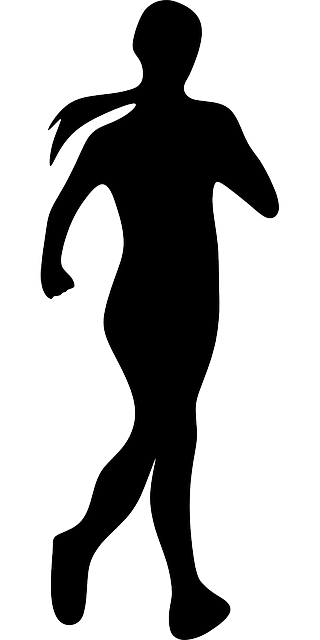
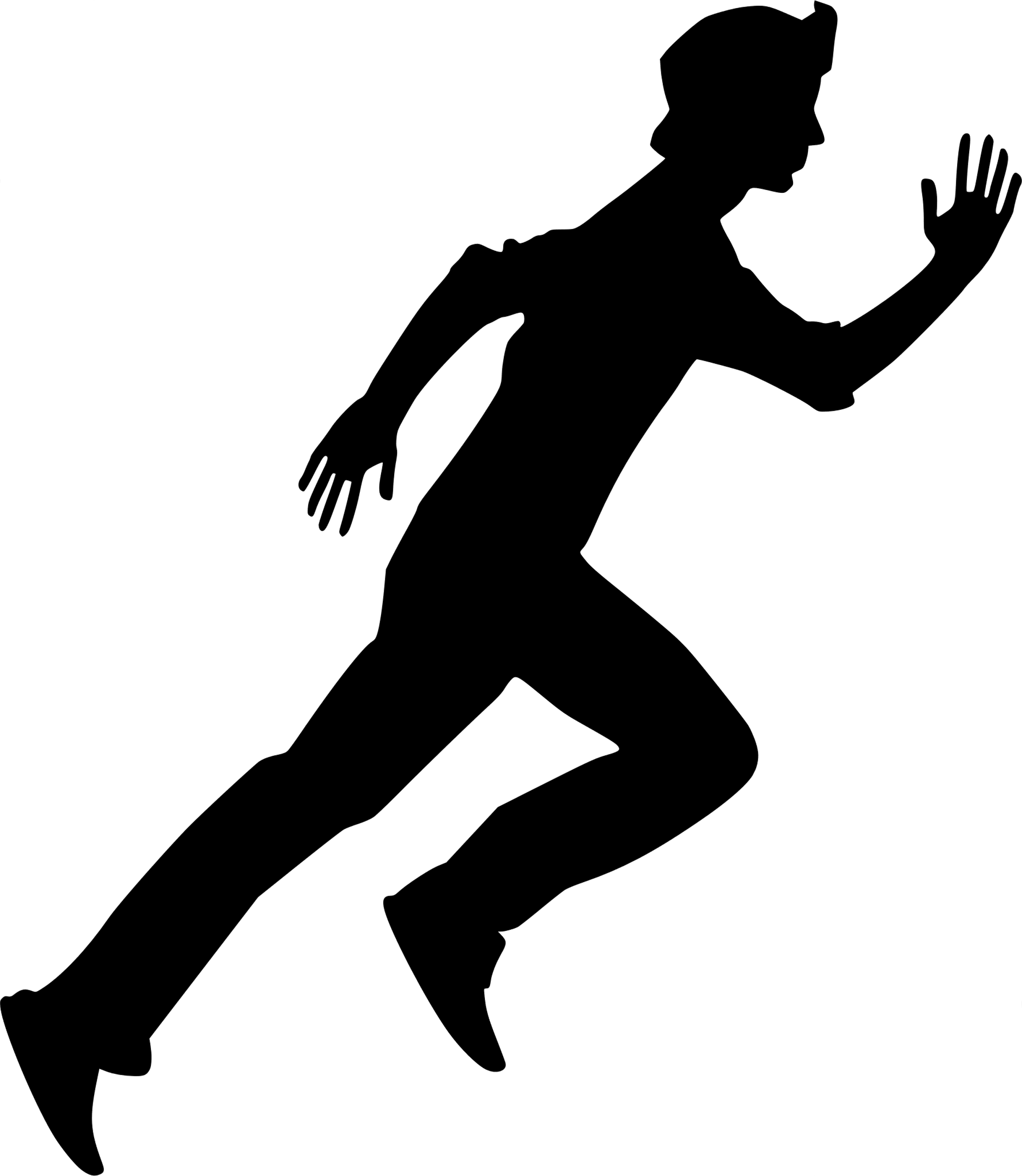
**tarjeta de crédito”**

**Rizzo Agustín**

27/3/2024

# A white text on a black background Description automatically generatedINTRODUCCIÓN:

El concepto de abandono o “churn” de clientes, definido como la cantidad de consumidores que deja de contratar un servicio en un periodo determinado de tiempo ha cobrado mucha importancia en el último tiempo. Entender cuáles son los principales motivos por los que un cliente decide dar de baja el servicio se considera una habilidad fundamental dentro de la inteligencia de mercado puesto que permite a las empresas tomar acciones de negocio oportunas para retener al cliente y anticiparse de esa forma a posibles pérdidas. Si bien un análisis retrospectivo de tipo diagnóstico puede ser de utilidad en una primera instancia para aproximarse a las causas del “churn”, para poder explotar el concepto de forma completa se requieren herramientas analíticas adicionales, particularmente aquellas que permitan hacer un análisis de tipo predictivo. Es aquí donde la Ciencia de Datos, mediante el uso de algoritmos de clasificación predictivos como el Machine Learning (ML) se vuelve una disciplina sumamente atractiva para ser incorporada dentro del stack tecnológico del buiseness intelligence.



# OBJETIVO:

Diseñar, entrenar y testear un modelo de ML para predecir que clientes son candidatos más probables para dar de baja el servicio de una tarjeta de crédito.

# FUENTE DEL DATASET:

<https://www.kaggle.com/datasets/anwarsan/credit-card-bank-churn>

# DESCRIPCIÓN DEL DATASET

El set de datos se compone de 10127 filas o registros y 21 columnas o variables. A continuación se ofrece una breve descripción de las distintas variables junto con el tipo de dato que la componen:

* **CLIENTNUM** - Numero de cliente. Identificador único del cliente titular de la cuenta.
* **Customer\_Age** - Variable demográfica: edad del cliente en años
* **Gender** - Variable demográfica - M=Hombre, F=Mujer
* **Dependent\_count** - Variable demográfica - Número de dependientes
* **Education\_Level** - Variable demográfica - Calificación educativa del titular de la cuenta (ejemplo: escuela secundaria, graduado universitario, etc.)
* **Marital\_Status** - Variable demográfica: Casado, Soltero, Divorciado, Desconocido
* **Income\_Category** - Variable demográfica: categoría de ingresos anuales del titular de la cuenta (< $40.000, $40.000 - 60.000, $60.000 - $80.000, $80.000-$120.000, > $120.000, Desconocido)
* **Card\_Category** - Variable de producto: tipo de tarjeta (azul, plateada, dorada, platino)
* **Months\_on\_book** - Periodo de relación con el banco.
* **Total\_Relationship\_Count** - Número total de productos en poder del cliente
* **Months\_Inactive\_12\_mon** - Nº de meses inactivos en los últimos 12 meses
* **Contacts\_Count\_12\_mon** - Nº de Contactos en los últimos 12 meses
* **Credit\_Limit** - Límite de crédito en la tarjeta de crédito
* **Total\_Revolving\_Bal** - Saldo rotatorio total de la tarjeta de crédito
* **Avg\_Open\_To\_Buy** - Línea de Crédito Abierta para Comprar (Promedio de los últimos 12 meses)
* **Total\_Amt\_Chng\_Q4\_Q1** - Cambio en el monto de la transacción (Q4 sobre Q1)
* **Total\_Trans\_Amt** - Monto total de la transacción (últimos 12 meses)
* **Total\_Trans\_Ct** - Recuento total de transacciones (últimos 12 meses)
* **Total\_Ct\_Chng\_Q4\_Q1** - Cambio en el recuento de transacciones (cuarto trimestre respecto al primer trimestre)
* **Avg\_Utilization\_Ratio** - Índice de utilización promedio de la tarjeta
* **Attrition\_Flag** - Variable de evento interno (actividad del cliente): si la cuenta está cerrada, entonces 1, de lo contrario 0