# 📊 Proyecto: Análisis de Renuncia de Clientes - Telecom X

## 📌 Descripción del Proyecto

Este proyecto tiene como objetivo analizar y predecir la **evasión de clientes (churn)** en la empresa de telecomunicaciones **Telecom X**, utilizando técnicas de análisis de datos, estadística y modelos de machine learning.

El análisis se divide en tres fases principales:

1. **Preparación de los Datos**
2. **Modelado Predictivo**
3. **Interpretación y Conclusiones**

📂 **Fuente de datos:**  
Los datos provienen de un proceso de limpieza previo realizado en un proyecto anterior y están disponibles en formato CSV:

👉 [Dataset - Telecom X](https://raw.githubusercontent.com/JuanWilson70/Telecom-X-parte2/refs/heads/main/datos_normalizados_TelecomX.csv)

## ⚙️ Tecnologías Utilizadas

* **Lenguaje:** Python 3
* **Entorno:** Google Colab (Jupyter Notebook)
* **Librerías principales:**
  + numpy
  + pandas
  + matplotlib
  + seaborn
  + plotly
  + scikit-learn
  + imblearn
  + statsmodels

## 🗂️ Plan de Trabajo

1. **Carga de datos** desde CSV y creación del DataFrame.
2. **Análisis exploratorio** de los datos (EDA).
3. **Preprocesamiento**: limpieza, reemplazo de valores y eliminación de columnas irrelevantes.
4. **Codificación de variables categóricas** (One-Hot Encoding / Label Encoding).
5. **Análisis del churn rate** (proporción de evasión).
6. **Balanceo de clases** con oversampling/undersampling (imblearn).
7. **Normalización y estandarización** de variables numéricas.
8. **Análisis de correlación** entre variables.
9. **Evaluación de multicolinealidad** mediante VIF (Variance Inflation Factor).
10. **Separación de datasets** (originales, filtrados por correlación y VIF).
11. **Construcción de modelos predictivos:**

* Árbol de Decisión
* Random Forest
* K-Nearest Neighbors (KNN)

1. **Evaluación de modelos** con métricas: Recall, Precision, F1, Accuracy, AUC.
2. **Validación cruzada** con pipeline y balanceo de datos.
3. **Comparación y selección del modelo Champion**.
4. **Interpretación de resultados**, conclusiones y recomendaciones.
5. **Guardado del modelo final** para uso futuro.

## 📖 Diccionario de Datos

* **Contrato 1 año / 2 años** → Tiempo de contrato.
* **Pago tarjeta crédito / chequera electrónica / cheque** → Métodos de pago.
* **Evasión** → Variable objetivo (1 = cliente dejó la empresa, 0 = se mantiene).
* **Adulto Mayor** → Cliente con edad ≥ 65 años.
* **Cónyuge** → Si el cliente está casado o no.
* **Cargas** → Si el cliente tiene dependientes.
* **Contrato** → Tipo de contrato (mensual, anual, etc.).
* **Servicio telefónico** → Suscripción a telefonía.
* **Múltiples líneas** → Cliente con más de una línea.
* **Internet** → Suscripción a internet.
* **Seguridad online, Respaldo online, Protección equipo, Soporte técnico** → Servicios adicionales.
* **TV cable, Películas online** → Servicios de entretenimiento.
* **Factura online** → Preferencia de recibir factura digital.
* **Factura\_mes** → Total mensual facturado.
* **Total** → Total gastado en la empresa.
* **Cuentas diarias** → Gasto mensual dividido entre 30 días.

## ✅ Resultados Esperados

* Identificar las **principales variables** que explican la evasión de clientes.
* Construir un modelo predictivo robusto para anticipar la renuncia de clientes.
* Generar **recomendaciones estratégicas** que permitan a la empresa reducir el churn.

✍️ Autor:Eduardo Urrutia **Proyecto Telecom X - Análisis de Datos**