## **Análisis Predictivo en Clientes de un Servicio de Streaming**

### **Introducción**

Comenzamos con la exploración del conjunto de datos con el objetivo de determinar si es posible predecir la baja de clientes (churn) en un servicio de streaming. Además, buscamos identificar qué variables tienen mayor relevancia en esta problemática y cuáles no aportan valor significativo.

**Hipótesis**

**¿Cuáles son las variables que influyen en la cancelación de clientes?**

**Hipótesis:**

* Los clientes con menor tiempo de suscripción son más propensos a cancelar.
* Una baja satisfacción del cliente está correlacionada con una mayor tasa de churn.
* Factores como el tipo de dispositivo utilizado, la región y el historial de pagos pueden influir en la cancelación.

### **Exploración de los Datos (EDA)**

Para abordar este problema, contamos con un conjunto de datos de **3,500 clientes**, donde analizamos diversas variables relacionadas con su comportamiento y suscripción.

#### **1. Tipos de Variables en el DataFrame**

Nuestro conjunto de datos contiene variables de distintos tipos:

**Variables categóricas:**

* Device Used Most Often, Genre Preference, Region, Payment History, Subscription Plan, Churn Status, Promotional Offers Used

**Variables numéricas:**

* Subscription Length (Months), Customer Satisfaction Score (1-10), Daily Watch Time (Hours), Engagement Rate (1-10), Monthly Income ($), Number of Profiles Created

#### **2. Análisis Exploratorio**

Para comprender mejor el comportamiento de los clientes, realizamos un análisis exploratorio mediante gráficos:

#### **3.Exploracion con Gráficos**

### **3.1. Distribución de la Variable Objetivo**

Se observó que la variable Churn Status está desbalanceada, con una mayor proporción de clientes que permanecen activos en comparación con los que se dan de baja.

### **3.2. Distribución de Variables Numéricas**

Se analizaron las variables numéricas principales, observando tendencias en satisfacción, ingresos mensuales, edad, tiempo de uso y tiempo de suscripción. Algunas variables presentan sesgos, mientras que otras tienen distribuciones más uniformes.

### **3.3. Tiempo de Suscripción y Churn**

Se evidenció una relación entre el tiempo de suscripción y el churn: los clientes con más meses de permanencia tienden a continuar en el servicio. Es un indicador clave para la retención.

### **3.4. Satisfacción del Cliente**

Aunque no de forma concluyente en todos los casos, los datos muestran una tendencia en la que clientes con menor satisfacción tienen mayor probabilidad de cancelación. Este indicador requiere análisis complementario.

### **3.5. Dispositivo Utilizado**

A través de un heatmap, se determinó que los usuarios que más utilizan smartphones y laptops presentan mayores tasas de churn. Aquellos que acceden desde Smart TVs y tablets, en cambio, tienen una tasa menor de baja.

### **3.6. Distribución de la Satisfacción**

La mayoría de los usuarios manifiesta un nivel de satisfacción entre 6 y 9. Los niveles más bajos (<5) aparecen vinculados con mayor churn, lo que valida parcialmente la hipótesis inicial.

### **3.7. Tipo de Plan y Churn**

El plan básico muestra una mayor proporción de cancelaciones, lo que puede estar asociado a una percepción menor de valor o una menor fidelización por parte de estos clientes.

## **Conclusión Preliminar**

El análisis exploratorio sugiere que ciertas variables tienen un papel importante en la cancelación del servicio. Entre ellas, se destacan:

* Tiempo de suscripción
* Satisfacción del cliente
* Tipo de dispositivo
* Plan de suscripción
* Historial de pagos

Estas variables serán En las que nos enfocaremos en la construcción de un modelo de Machine Learning orientado a predecir la probabilidad de churn y anticipar acciones para prevenirlo.