# Análisis de Datos - Proyecto Amazon

## Columnas y Traducciones

### Customer ID

Traducción: Identificador del Cliente

- Los ID deberían ser únicos.

- Se detectaron 334 valores duplicados de un total de 7,021 registros.

- Es necesario revisar los duplicados para identificar si se trata de errores o de clientes con múltiples entradas válidas.

### Churn Label y Churned

Traducción: Etiqueta de Baja y Baja

- Ambas columnas parecen contener la misma información (una como texto y la otra como valores numéricos: 1 = Yes, 0 = No).

- Validar si es necesario conservar ambas o si una de ellas es redundante.

### Account Length (in months)

Traducción: Duración de la Cuenta (en meses)

- Detectar valores nulos o inconsistentes.

### Customer Service Calls

Traducción: Llamadas al Servicio al Cliente

- Confirmar que los valores sean numéricos y estén dentro de un rango razonable para un plan de datos.

### Avg Monthly GB Download

Traducción: Promedio Mensual de Descarga en GB

- Confirmar que los valores sean numéricos.

### Unlimited Data Plan

Traducción: Plan de Datos Ilimitados

- Es un valor booleano (1 = Sí, 0 = No).

- Revisar si existen valores nulos o inconsistencias (por ejemplo, valores fuera de 0 o 1).

### Extra Data Charges

Traducción: Cargos Adicionales por Datos

- Verificar si los valores se correlacionan con el uso de datos y los planes ilimitados.

- Revisar si hay clientes con cargos excesivos y si afecta al churn.

### State

Traducción: Estado

- Revisar si todos los valores son válidos (códigos de estado correctos).

- Analizar si hay estados con mayor tasa de baja.

### Phone Number

Traducción: Número de Teléfono

- Validar el formato de los números telefónicos (asegurarse de que sean consistentes).

- Comprobar si hay números duplicados, lo que podría indicar errores en los datos.

### Gender

Traducción: Género

- Contiene 3 valores únicos, lo que podría indicar un valor adicional inesperado (por ejemplo, valores vacíos o no clasificados).

### Payment Method

Traducción: Método de Pago

- Tiene 16 valores nulos, lo cual podría requerir imputación o análisis adicional.

### Monthly Charge

Traducción: Cargo Mensual

- Contiene 21 valores nulos, que podrían afectar análisis financieros o de comportamiento del cliente.

### Total Charges

Traducción: Cargos Totales

- También tiene 21 valores nulos. Importante corroborar si coinciden con los de "Monthly Charge".

### Churn Category

Traducción: Categoría de Baja

- Más del 73% de los registros están nulos (5,179 valores nulos).

- Validar su relevancia para el análisis.

### Churn Reason

Traducción: Motivo de Baja

- Similar a "Churn Category", tiene 5,182 valores nulos, lo que podría limitar su utilidad.

### Last Transaction Date

Traducción: Fecha de Última Transacción

- Columna de fecha. Validar si todas las fechas son consistentes y están en un formato válido.

### Applied Discount

Traducción: Descuento Aplicado

- Es un valor booleano (`True/False`).

- Confirmar si la lógica de los descuentos aplicados es consistente.

### Internal Notes

Traducción: Notas Internas

- Contiene solo un valor único. Puede ser irrelevante para el análisis.

## Columnas Numéricas Clave para Revisar Nulos o Valores Extremos:

- Monthly Charge

- Total Charges

- Account Length (in months)

- Customer Service Calls

- Avg Monthly GB Download

- Extra Data Charges

- Customer Tenure (in months)

- Number of Complaints or Support

**Columnas Clave para la Tasa de Abandono**

* **Monthly Charge** (Relevante para churn).
* **Number of Complaints or Support** (Crítica para churn).
* **Customer Segment** (Posible relación con el churn).
* **Extra Data Charges** (Correlación con churn por uso de datos).

**Factores Clave para la Tasa de Abandono**

1. **Duración de la Cuenta (Account Length)**
   * ¿Cuántos meses ha estado el cliente en el servicio antes de abandonar?
   * Analizar si los clientes con menor antigüedad tienen mayor probabilidad de churn.
2. **Antigüedad del Cliente (Customer Tenure)**
   * Comparar con la duración de la cuenta para entender si hay diferencias.
   * Relacionar la antigüedad con la tasa de abandono y los segmentos de cliente.
3. **Género (Gender)**
   * ¿Existe alguna relación entre el género del cliente y la probabilidad de abandono?
4. **Edad (Age)**
   * ¿Clientes de ciertos rangos de edad (jóvenes, mayores) abandonan más que otros?
   * Analizar segmentos específicos como "Menor de 30" o "Senior (mayor)".
5. **Número de Llamadas al Servicio Técnico (Customer Service Calls)**
   * ¿Los clientes que realizan más llamadas al servicio técnico tienen mayor probabilidad de abandonar?
   * Relación entre quejas y satisfacción del cliente.
6. **Número de Quejas (Number of Complaints or Support)**
   * Esta columna es crítica para medir la tasa de abandono.
   * Identificar si las quejas están asociadas con otras variables, como gasto mensual o género.
7. **Método de Pago (Payment Method)**
   * ¿Existen ciertos métodos de pago que correlacionen con mayores tasas de abandono?
8. **Descuentos Aplicados (Applied Discount)**
   * ¿Los clientes con descuentos tienen menor probabilidad de abandonar?
   * Relación entre descuentos, gasto mensual y tiempo de permanencia.
9. **Tiempo de Uso Real (Tiempo en línea o descarga de datos)**
   * Analizar si los clientes que hacen mayor uso del servicio (descarga en GB) tienen menos probabilidades de abandonar.
   * Comparar con el plan de datos contratado.
10. **Segmento del Cliente (Customer Segment)**
    * ¿Los clientes de segmento "bajo" tienen mayor probabilidad de abandonar?
    * Relación con otros factores, como antigüedad o quejas.
11. **Motivo y Categoría de Baja (Churn Reason y Churn Category)**
    * Identificar las razones más comunes de abandono.
    * Analizar si estas razones están asociadas con otros factores como edad, género o tiempo de permanencia.
12. **Cargos Mensuales y Totales (Monthly Charge y Total Charges)**
    * ¿Los clientes con cargos más altos tienen mayor probabilidad de abandonar?
    * Relación entre gasto mensual, descuentos aplicados y quejas.