# 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

--------------------

**Actualización: Anotaciones Clave**

**Segmento del Cliente**

* **Traducción**: Customer Segment
  + Los segmentos (alto, medio, bajo) podrían referirse al nivel de uso del cliente.
  + Comparar con el "Plan Contratado" para analizar posibles relaciones.
  + Por ejemplo, clientes de segmento bajo podrían tener planes más económicos o limitados.

**Número de Quejas o Soporte**

* **Traducción**: Number of Complaints or Support
  + Esta columna es crítica para medir la tasa de abandono.
  + Clientes con más quejas podrían estar más propensos al churn.
  + Confirmar valores numéricos y revisar posibles valores atípicos.

**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)