**Recap segunda sesión: Limpieza de datos**

**Eliminación de columnas, establecimiento del tipo de los datos, normalización de los datos.**

**En esta sesión de trabajo hemos seguido explorando el conjunto de datos eliminando los datos irrelevantes. Además, comprobamos que los datos tenían el tipo que les correspondía y normalizamos los datos para que siguieran el mismo formato.**

**Lo primero que hicimos fue comprobar si teníamos algún registro duplicado. Eliminamos 14 registros.**

**A lo largo de la sesión añadimos una hoja para poder hacer la normalización de los datos con la función BuscarV en lugar de encadenar funciones SI.**

**Descripción de las columnas:**

1. **Customer ID: Identificador único de cada cliente.**
   * ***Utilizamos un formato condicional para resaltar si había algún identificador duplicado. Encontramos que sí había duplicados.***
2. **Churn Label: Etiqueta que indica si el cliente se fue (*Yes*) o no (*No*).**
3. **Churned: Variable binaria (0 o 1) que indica si el cliente abandonó el servicio (1) o se mantuvo (0).**
   * ***Comprobamos que ambas columnas tenían exactamente los mismos datos. Todos los 'yes' eran 1 y los 'no' eran 0. Eliminamos la columna Churned ya que en este caso nos interesan más los valores de 'yes' y 'no'.***
4. **Account Length (in months): Tiempo que el cliente ha estado suscrito al servicio, en meses.**
5. **Customer Service Calls: Número de llamadas realizadas al servicio de atención al cliente.**
6. **Avg Monthly GB Download: Promedio mensual de datos descargados por el cliente en GB.**
7. **Unlimited Data Plan: Indica si el cliente tiene un plan de datos ilimitado (1: Sí, 0: No).**
   * ***Cambiamos el formato de los datos de 0 y 1 a 'no' y 'yes'.***
8. **Extra Data Charges: Cargos adicionales por uso de datos más allá del límite del plan.**
9. **State: Estado de residencia del cliente.**
10. **Phone Number: Número de teléfono del cliente.**
    * ***Eliminamos esta columna ya que no nos aporta valor en el análisis.***
11. **Gender: Género del cliente.**
    * ***Cambiamos los datos para que todas las categorías estuvieran en inglés.***
12. **Age: Edad del cliente.**
13. **Under 30: Indica si el cliente es menor de 30 años.**
    * ***Comprobamos que los valores afirmativos de esta columna fueran coherentes con los valores de la columna Age.***
14. **Senior: Indica si el cliente es mayor de 65 años.**
    * ***Realizamos la misma comprobación que en la columna anterior. Encontramos inconsistencias y contactaremos con el cliente para determinar los valores correctos.***
15. **Comprobación Senior: Columna creada para verificar si los valores de la columna Senior coincidían con Age.**
16. **Group: Indica si el cliente tiene contratado un plan familiar.**
17. **Number of Customers in Group: Número de clientes que pertenecen al mismo plan familiar.**
18. **Device Protection & Online Backup: Indica si el cliente tiene protección de dispositivo y respaldo en línea.**
19. **Contract Type: Tipo de contrato (mensual, anual, etc.).**
20. **Payment Method: Método de pago utilizado por el cliente.**
    * ***Estandarizamos los registros en formato de Nombre Propio.***
21. **Monthly Charge: Cargo mensual que el cliente paga por el servicio.**
    * ***Convertimos los valores monetarios a números eliminando las unidades.***
22. **Total Charges: Cargos totales acumulados durante la suscripción.**
23. **Churn Category: Categoría que indica el motivo de la deserción (*churn*).**
24. **Churn Reason: Razón específica por la que el cliente abandonó el servicio.**
25. **Contact Date: Fecha del último contacto con el cliente.**
    * ***Se formateó la fecha para que Excel interpretara los datos como datos de tipo fecha.***
26. **Last Transaction Date: Fecha de la última transacción realizada por el cliente.**
27. **Tenure (in months): Tiempo total de suscripción del cliente.**
    * ***Eliminamos esta columna, ya que aportaba información similar a otra.***
28. **Customer Segment: Clasificación del cliente según su nivel de uso o importancia (Alto, Medio, Bajo).**
    * ***Se tradujeron los datos para que todos estuvieran en inglés.***
29. **Average Monthly Expenses: Gasto promedio mensual del cliente.**
    * ***Redujimos el número de decimales en esta columna.***
30. **Number of Complaints or Support: Número de quejas o interacciones con el servicio de soporte.**
31. **Preferred Contact Method: Método preferido para ser contactado (email, teléfono, etc.).**
    * ***Se formatearon los datos para estandarizar el idioma.***
32. **Applied Discount: Indica si se ha aplicado algún descuento al cliente.**
    * ***Se formatearon los datos para que todos estuvieran en inglés.***
33. **Hobby: Pasatiempos del cliente.**
    * ***Eliminamos esta columna ya que no nos aporta valor en el análisis.***
34. **Favorite TV Show: Programa de televisión favorito del cliente.**
    * ***Eliminamos esta columna ya que no nos aporta valor en el análisis.***
35. **Internal Notes: Notas internas relacionadas con el cliente.**
    * ***Eliminamos esta columna ya que no nos aporta valor en el análisis.***

**Columnas relevantes para nuestro análisis:**

* **Churn label.**
* **Duración de la cuenta.**
* **Frecuencia de llamadas a atención al cliente.**
* **Número de quejas o soporte.**
* **Método de pago utilizado.**
* **Tipo de contrato.**
* **Segmento de cliente y su relación con el abandono.**