**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 estos 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 que 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 no interesaban más los valores de 'Yes' y 'No' que los numéricos.**
4. **Account Length (in months): Tiempo que el cliente ha estado suscrito al servicio.**
5. **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'.**
6. **Senior: Indica si el cliente es mayor de 65 años.**
   * **Realizamos la misma comprobación que en la columna anterior. Aquí vimos que había clientes con 65 años que tenían asignado 'Yes' y otros que tenían asignado 'No', por lo tanto, se quedó en contactar con el cliente para ver qué deberíamos poner a estos clientes.**
7. **Group: Indica si el cliente tenía contratado un plan familiar.**
8. **Number of Customers in Group: Número de clientes que pertenecen al mismo plan familiar.**
9. **Device Protection & Online Backup: Indica si el cliente tiene protección de dispositivos o respaldo en línea.**
10. **Contract Type: Tipo de contrato (mensual, anual, etc.).**
11. **Payment Method: Método de pago utilizado por el cliente.**
    * **Esta columna tenía registros en minúscula y mayúscula. Se estandarizaron poniéndolo en formato Nombre Propio.**
12. **Monthly Charge: Cargo mensual que el cliente paga por el servicio.**
    * **Esta columna tenía valores numéricos y otros con una unidad monetaria. Se convirtió en dato numérico y se añadió al nombre de la columna. Se unificaron los datos a numéricos.**
13. **Total Charges: Cargos totales acumulados durante la suscripción.**
14. **Contact Date: Fecha del último contacto con el cliente.**
    * **Se formateó la fecha para que Excel interpretara los datos como datos de tipo fecha.**
15. **Hobby: Pasatiempos del cliente.**
    * **Eliminamos esta columna ya que no nos aporta valor en el análisis.**
16. **Favorite TV Show: Programa de televisión favorito del cliente.**
    * **Eliminamos esta columna ya que no nos aporta valor en el análisis.**
17. **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**