**Resumen de la segunda sesión: Limpieza de datos**

En esta sesión, llevamos a cabo varios procedimientos esenciales para la limpieza de datos, incluyendo la eliminación de columnas irrelevantes, la definición adecuada de los tipos de datos y la normalización de los mismos. A continuación, se detallan las acciones emprendidas:

* **Eliminación de registros duplicados:** Identificamos y eliminamos 14 registros duplicados.
* **Normalización de datos:** Añadimos una hoja adicional para realizar la normalización utilizando la función BuscarV en lugar de encadenar funciones SI.
* **Descripción de las columnas:**
* **Customer ID:** Identificador único de cada cliente. Implementamos un formato condicional para resaltar los identificadores duplicados, encontrando coincidencias duplicadas.
* **Churn Label:** Etiqueta que indica si el cliente ha abandonado el servicio (Yes) o no (No).
* **Churned:** Variable binaria (0 o 1) que refleja si el cliente ha dejado de utilizar el servicio (1) o continúa utilizando el servicio (0). Ambas columnas tenían datos idénticos, por lo que se decidió eliminar la columna "Churned".
* **Account Length (in months):** Tiempo de suscripción del cliente al servicio.
* **Under 30:** Indica si el cliente tiene menos de 30 años, validando que los valores fueran coherentes con la columna 'Age'.
* **Senior:** Indica si el cliente tiene más de 65 años. Se detectaron inconsistencias en clientes de 65 años, por lo que se acordó contactar al cliente para resolverlas.
* **Group:** Indica si el cliente tiene un plan familiar contratado.
* **Number of Customers in Group:** Número de personas en el mismo plan familiar.
* **Device Protection & Online Backup:** Indica si el cliente dispone de protección de dispositivos o copias de seguridad en línea.
* **Contract Type:** Tipo de contrato (mensual, anual, etc.).
* **Payment Method:** Método de pago utilizado. Se estandarizaron los registros en formato Nombre Propio.
* **Monthly Charge:** Cargo mensual del cliente. Se unificaron los datos a valores numéricos eliminando unidades monetarias adicionales.
* **Total Charges:** Cargos totales acumulados durante la suscripción.
* **Contact Date:** Fecha del último contacto con el cliente. Se formateó para que Excel interpretara correctamente como fecha.
* **Hobby:** Pasatiempos del cliente. Columna eliminada por falta de relevancia.
* **Favorite TV Show:** Programa de televisión favorito del cliente. Columna eliminada por falta de relevancia.
* **Internal Notes:** Notas internas sobre el cliente. Columna eliminada por falta de relevancia.
* **Columnas relevantes para el 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