



Actualmente tenemos un problema con un proyecto en Wise Athena. Un cliente nos ha proporcionado un set de datos pero no tenemos recursos suficientes para limpiar y analizar estos datos sin comprometer las fechas de entrega ya pactadas con otros clientes.

El conjunto de datos se compone de:

- Un maestro o catálogo de productos (maestro_productos_prueba.csv): contiene toda la información disponible relativa a los productos que vende el cliente y competidores
- Un maestro o catálogo de clientes (maestro_clientes_prueba): contiene toda la información disponible sobre las tiendas y distribuidores que venden los productos del cliente
- Archivos de venta de sellout para dos proveedores diferentes (sellout-proveedor1.csv y sellout-proveedor2.csv): contiene histórico de ventas y precio de sellout (del retailer a consumidor final) para cada producto y cada tienda.
- Archivo de venta de sellin (sellin.csv): contiene el histórico de ventas y precio de sellin (del fabricante al retailer) para cada producto y tienda

SellOut es la información de los procesos de venta que ocurren desde los distribuidores de nuestro cliente hacia los compradores finales (es decir, cuando alguien compra un producto de nuestro cliente en un supermercado)

SellIn es la información de los procesos de venta de nuestro cliente hacia sus distribuidores (es decir, cuando un supermercado compra un lote de productos de nuestro cliente con el fin de vendérselos a los compradores finales)

El objetivo es realizar una limpieza y ajuste de estos datos, y generar una estructura de datos que sea fácilmente consultable por el equipo de data scientists para generar sus modelos. Esto requiere que los datos no sólo estén limpios, sino que sean coherentes entre si y estén perfectamente ordenados. Un ejemplo de consulta que el equipo de data scientists puede requerir es "necesito todos los registros de sellout y sellin entre las fechas X e Y para los productos A,B,C y D en esta lista concreta de tiendas: T1, T2, T3, T4, T5 y T6" siendo X,Y, A, B, C, D, T1, T2, T3, T4, T5, y T6 variables.

Se valorará positivamente si una vez generada la estructura de datos para consulta, se incluyen elementos iniciales de exploración de los datos para proporcionar una primera intuición del dataset.