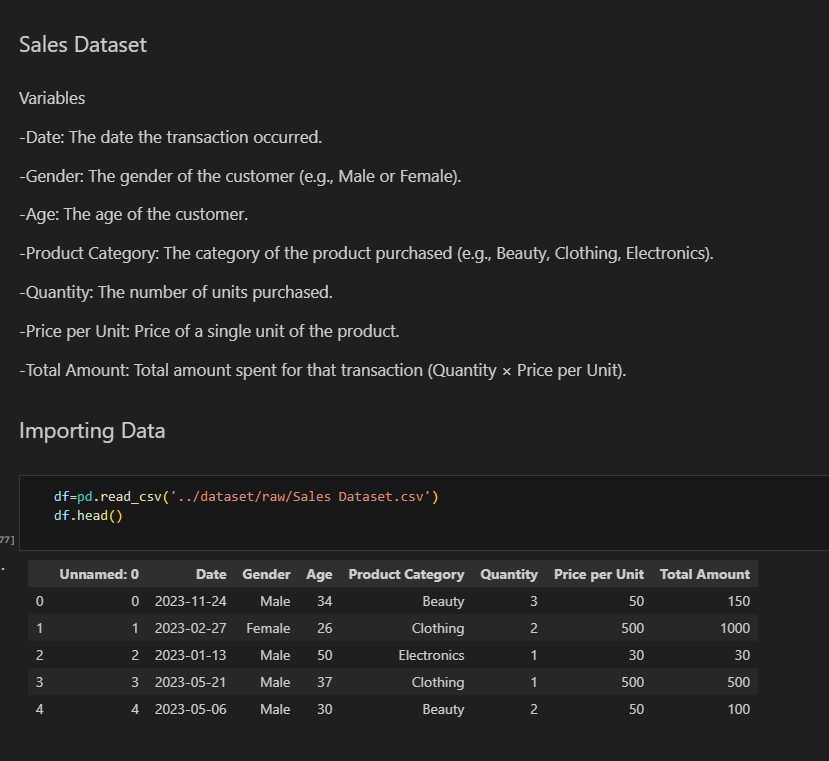
Para mi 1er proyecto para mi portafolio de data science escogi este dataset:



Primero en mi notebook visualize las primeras filas de dataframe, el nombre de las columnas, tipos de datos.

Para limpiar los datos

Elimine la columna “ Unnamed: 0” ya que solo ofrece uso de memoria innecesaria  
Con la funcion info(memory\_usage=”deep”), pude ver que no hay valores null y la cantidad de memoria que usa el dataframe(163.6 kb)

Optimizando datos

Converti algunas variables a valores que sean mas pequeños, como int8 y float32.

Separe la columna Gender a dos columnas Male y Female con valores True y False (Dummy variables)

Coloque la columna Date como el index ya que muchas de las graficas que hare sera usando ese columna.

Al final de la optimizacion volvi a chequear la memoria usada por el dataframe ahora con los cambios realizados y bajo hasta 83kb, siendo esto un 49% menos que antes.

### ****1. Preguntas generales (overview del dataset):****

¿Cuántas transacciones hay en total?

¿En qué rango de fechas se encuentran los datos?

¿Hay datos faltantes o valores atípicos (outliers)?

¿Cuál es la distribución de edad de los clientes?

### ****2. Por género:****

¿Cuál es la distribución de transacciones entre hombres y mujeres?

¿Quién gasta más en promedio por transacción: hombres o mujeres?

¿Qué categorías de productos son más populares según el género?

¿Hay diferencias significativas en las cantidades compradas por género?

### ****3. Por tiempo (Date):****

¿Cómo varían las ventas a lo largo del tiempo? ¿Hay tendencias o estacionalidades?

¿En qué días o meses se vende más?

¿Existen picos de ventas que puedan estar relacionados con eventos o campañas?

### ****4. Por producto:****

¿Qué categorías de productos generan más ingresos?

¿Qué productos se venden más en cantidad?

¿Qué producto tiene el mayor precio por unidad promedio?

¿Hay productos que se venden mucho pero generan poca ganancia?

### ****5. Por edad:****

### ¿Qué grupo de edad compra más?

¿Hay diferencias en las categorías preferidas según la edad?

¿Cómo se relaciona la edad con el total gastado?

### ****6. Métricas y relaciones clave:****

¿Cómo se relacionan cantidad y total gastado? ¿Mayor cantidad siempre implica mayor gasto?

¿Existe correlación entre edad y gasto?

¿Cuál es el ticket promedio por transacción?

¿Qué combinación de género + categoría + edad genera mayor ingreso?

### ****7. Para objetivos de negocio / marketing:****

¿Quiénes son mis mejores clientes (por ingresos totales)?

¿Qué segmentos están poco explotados?

¿Dónde enfocar una campaña de marketing: en género, edad o producto?

¿Qué categorías podrían ampliarse o promocionarse más?