DESAFÍO 1 Importación, Limpieza y Modelado de datos

Grupo 4
Alegre Colombo, María de los Ángeles
González, Valeria Vanesa
Mayol, Hernán
Rodríguez Fittabile, Florencia
Vázquez, Santiago

Contexto: En el siguiente desafío estaremos trabajando con la herramienta POWER BI, donde deberemos importar datos, revisar que los datos sean los correspondientes o si hay que mejorarlos y obtener el modelo de datos relacional correspondiente.

Datos:

El dataset "VENTAS_JUGUETES": contiene información sobre las ventas realizadas, tiendas, productos e inventario.

Pasos a seguir:

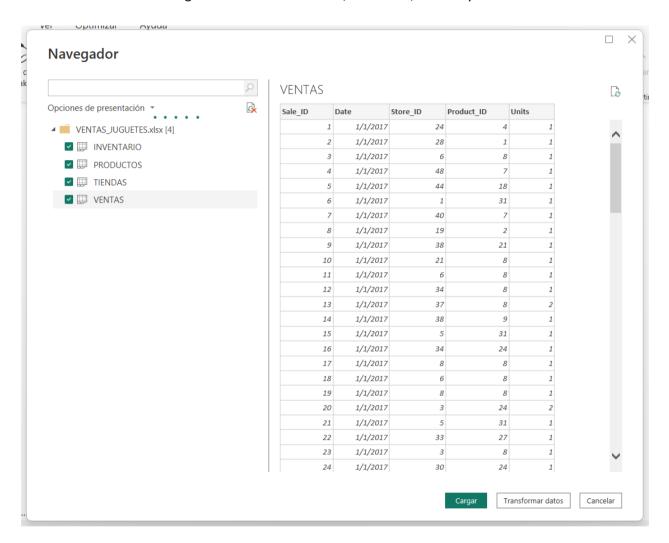
- 1. Importa los datos a Power BI: Carga los datos del archivo de ventas en Power BI. link
- 2. Limpieza de datos: Verifica que los datos estén limpios y coherentes. Realiza las transformaciones necesarias para asegurarte de que los formatos de fecha y valores numéricos sean correctos.
- 3. Crea una tabla calendario.
- 4. Modelado de datos: revisa que se hayan generado las relaciones de forma automática, caso contrario, generarlas de manera manual. Revisa las cardinalidades.

Presentación: deberán cargar a la plataforma un archivo que describa los pasos seguidos para realizar el desafío con las capturas correspondientes de las tablas luego de la limpieza y del DER.

Descripción y capturas correspondientes:

1- IMPORTACIÓN DE DATOS

Para el siguiente punto procedimos a importar los datos del archivo de Excel de Ventas y cargarlos en PowerBI. Se visualizan las siguientes tablas: Inventario, Productos, Tiendas y Ventas.



2- PREVIO A LA LIMPIEZA

Para esta instancia se visualiza cada tabla en su estado original previo a la limpieza de datos.

Imagen 1_Productos

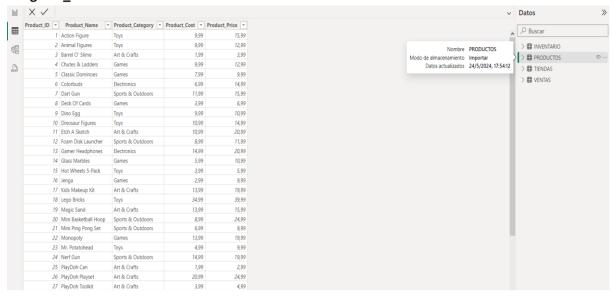


Imagen 2_Tiendas

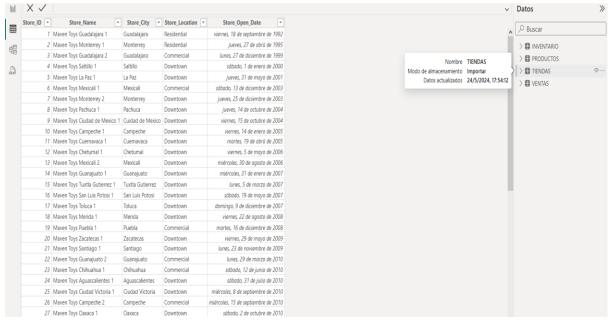


Imagen 3_Ventas

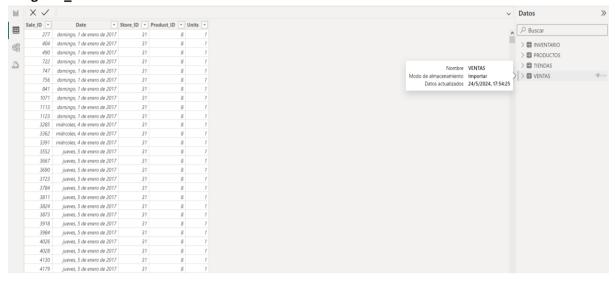
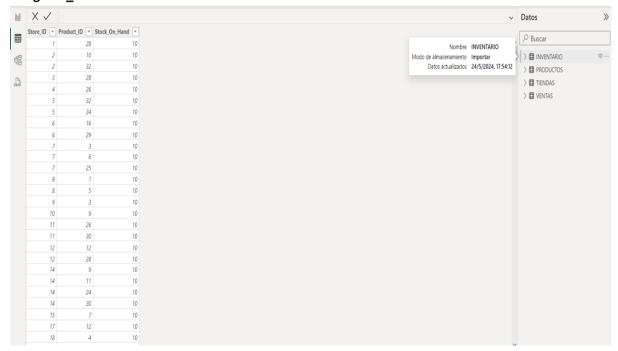
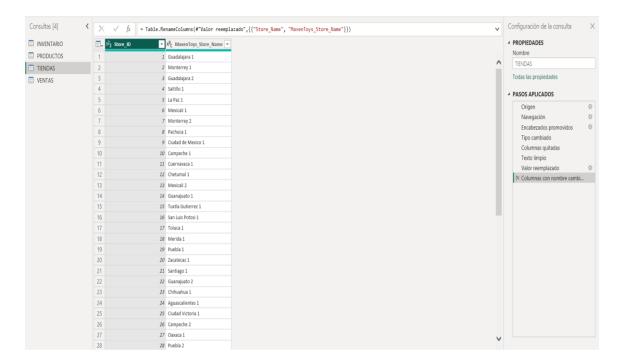


Imagen 4_Inventario



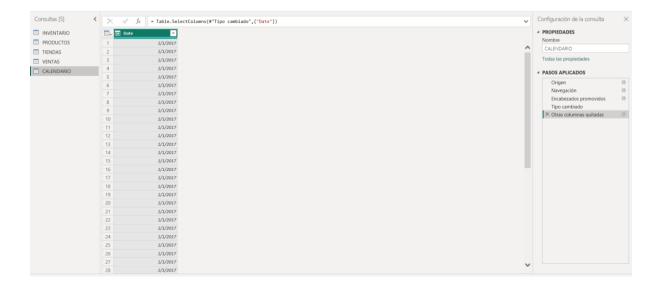
3- DATOS LIMPIOS

De las 4 tablas (Tiendas, Inventarios, Productos y Ventas), en **Tiendas** consideramos necesario suprimir las columnas **store_city**, **store_location y store_open_date**. Además, en la columna **store_date** eliminamos la denominación "**Maven Toys**", ya que, la misma era repetitiva y no aportaba datos relevantes. Asimismo, renombramos la columna por **MavenToys_Store_Name**.



Seguidamente, procedimos a la creación de una **tabla Calendario** basándonos en la columna date de la tabla **Ventas**, ya que, nos permite tener todos los periodos de tiempo necesarios en el análisis y así, tener un detalle más preciso.

Esta entidad **Calendario**, no poseía ningún atributo identificador ni modo de relacionarlo con los otros, por lo cual, no aparece en el diagrama entidad-relación.





4- MODELO DE DATOS RELACIONAL:

