Curso   
Power BI Básico

Módulo4:  
Transformación de Datos Básica

Autor: Mauro Chávez Pérez

Fecha: 10.OCT.2024

1. Contexto de la dataset seleccionada

El conjunto de datos seleccionados, dirty\_data, está asociado con los datos relacionados con clientes y su interacción con un almacén, entregando información acerca de la fecha en qué se hacen los pedidos, qué pedidos se realizan, el precio del producto que se pide, cargos de envío, hasta datos de la satisfacción que tuvo el cliente con su último pedido.

1. Variables y descripción



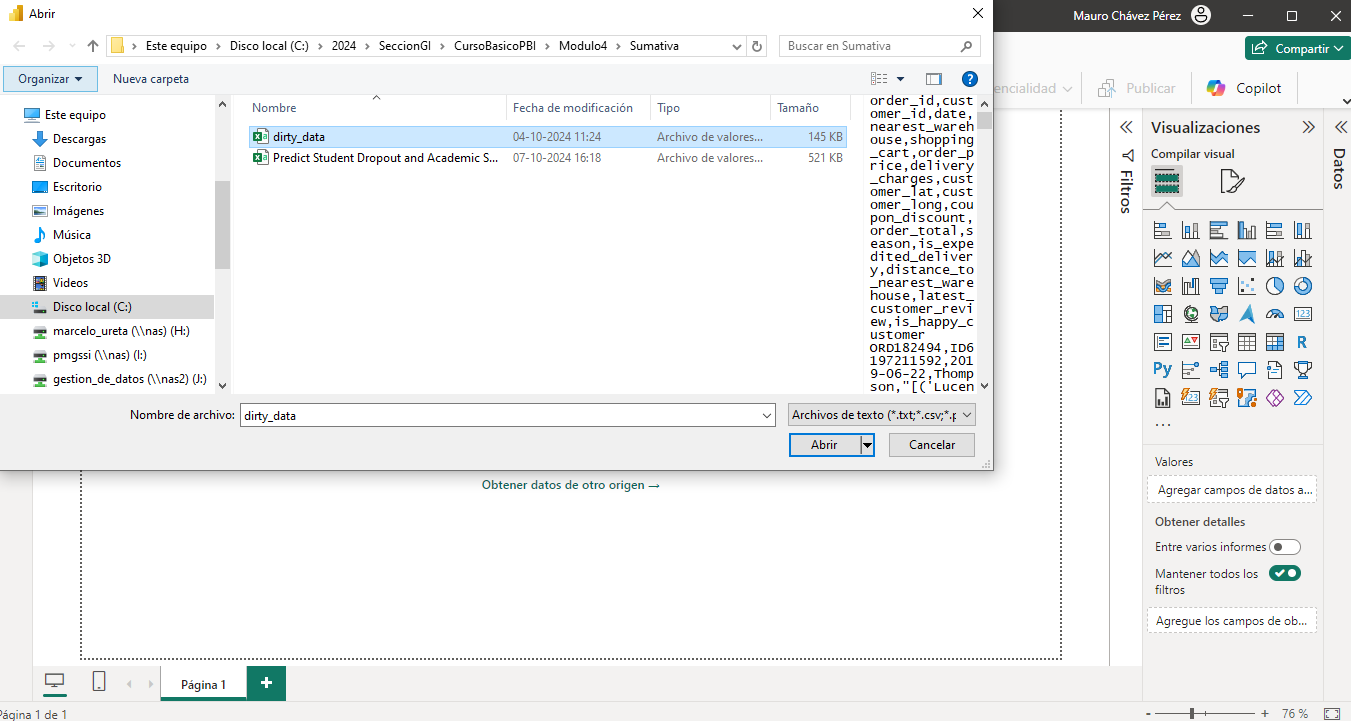
1. Enlace a GitHub

El enlace al repositorio en GitHub, donde está alojado el dataset dirty\_data es el siguiente:

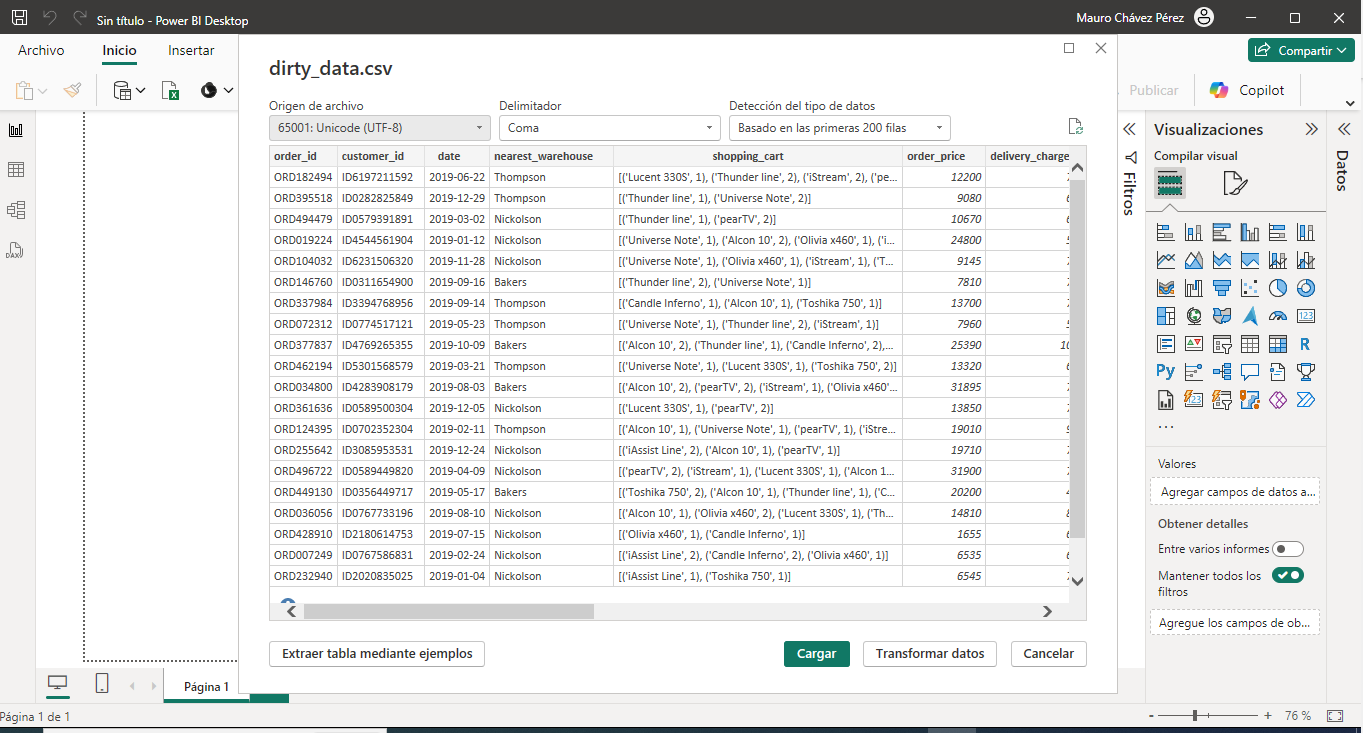
<https://github.com/mgchavezp/Power-BI-Basico.git>

1. Capturas de pantalla

Importación de datos (Accesa localmente el dataset entregado en clases)



Importación datos (Muestra la vista del archivo una vez seleccionado)



Eliminar filas duplicadas

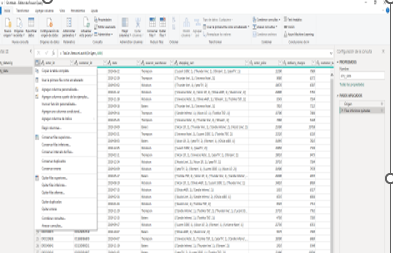
Para tener filas duplicadas, se anexó otra consulta al mismo dataset y se borraron las últimas 400 filas, quedando con 600 filas y 100 filas duplicadas.

Debido a lo anterior quedan dos dataset en el modelo: el original denominado dirty\_dataorig y el trabajado en estas actividades denominado dirty\_data.

Inicialmente se tiene 600 filas. Posicionados en la columna order\_id se selecciona la opción Quitar Duplicados, en el menú destacado

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente



Obteniendo el siguiente resultado:

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Con este resultado de 100 registros menos, entenderemos que existían 100 registros duplicados.

**Cambiar tipos de datos y ajustes varios**.

Columna nearest\_warehouse : se corrigen datos que empiezan con letra minúscula a Mayúsculas, para estandarizar los nombres de los almacenes.

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

Descripción generada automáticamente

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Se modifican formatos de columnas order\_price y delivery\_charges de flotante a decimal fijo, por corresponder cargos monetarios.

(antes)

Tabla

Descripción generada automáticamente

(después)

Tabla

Descripción generada automáticamente

Se eliminan columnas customer\_lat y customer\_long por corresponder a parámetros de latitud y longitud del cliente que no serán utilizadas en este ejercicio.

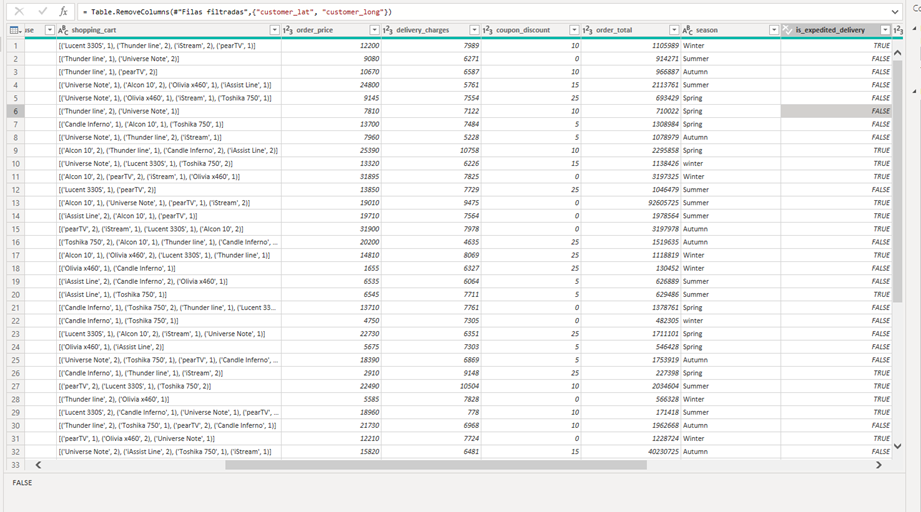
(antes)

Tabla

Descripción generada automáticamente

Posicionados en la columna, presionando botón izquierdo y seleccionado Quitar.

(después)



Rellenar valores faltantes.

No se aplicó ninguna función de relleno.

Creación de columna condicional

Se generó una columna que traduce las estaciones del año, previa estandarización de los nombres de la columna season.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Se recalculó la columna order\_total considerando los datos existentes en las columnas order\_price, delivery\_charges y coupon\_discount como estaban, es decir, no se revisó consistencia de los valores entre ellos. Se realiza sólo con el fin de aplicar lo aprendido. El cálculo realizado obedece a:

(order\_price + delivery\_charges – ((coupon\_discount /100)\* order\_price))

Tabla, Excel

Descripción generada automáticamente