



**UNIVERSIDAD
DON BOSCO**

UNIVERSIDAD DON BOSCO

Materia: Datawarehouse y Minería de Datos

Docente: Ing. Karens Medrano

Estudiante: Diego René López Martínez

Carné: LM231893

Ciclo: 01-2025

ÍNDICE

Problemática-----	3
Creación del repositorio GIT-----	4
Creación de Base de Datos-----	5
Agregar los “Flat File Sources”-----	7
Solución-----	10

PROBLEMÁTICA

La empresa "TechStore" solicita un proceso ETL utilizando tres archivos CSV: productos.csv, ventas.csv y clientes.csv. Se debe extraer la información, unir las ventas con los clientes y productos, y transformar los datos calculando el monto total de cada venta, clasificando a los clientes en "Frecuentes" y "Ocasionales", y generando un Código de País extrayendo los dos primeros caracteres del ID Cliente. Finalmente, los datos se cargarán en dos archivos de Excel: "Clientes_Frecuentes.xlsx" y "Clientes_Ocasionales.xlsx", incorporando la segmentación y el código de país generado.

CREACIÓN DEL REPOSITORIO EN GIT

Creamos un repositorio desde la app de escritorio y escogemos el nombre y un breve descripción del proyecto, en este caso, vamos a escoger una carpeta de nuestro escritorio, en el que deberíamos de recopilar toda información de nuestro proyecto, tablas, excels, projects, y así poderlos subir a GIT-HUB y los demás puedan clonarlo.

Create a new repository [X]

Name
Desafio_

Description
Desafio_1 de DATA WAREHOUSE

Local path
C:\Users\diego\OneDrive\Escritorio\Desafio_1_LM2 [Choose...]

☐ Initialize this repository with a README

Git ignore
None [v]

License
None [v]

The repository will be created at C:\Users\diego\OneDrive\Escritorio\Desafio_1_LM231893\Desafio_

[Create repository] [Cancel]

CREACIÓN DE BASE DE DATOS – ANTICIPADAMENTE

Crear base de datos, no importa que nombre tenga, pero se tienen que generar estas tablas para que pueda funcionar!

```
CREATE TABLE [Ventas] (  
    [id_venta] bigint,  
    [id_cliente] bigint,  
    [nombre_cliente] nvarchar(100),  
    [id_producto] bigint,  
    [precio] numeric(18,0),  
    [cantidad] numeric(18,0),  
    [monto_venta] numeric(37,0),  
    [codigo_pais] nvarchar(2),  
    [nombre_producto] nvarchar(50),  
    [ciudad] nvarchar(50)  
)
```

```
CREATE TABLE ClientesFrecuentes (
```

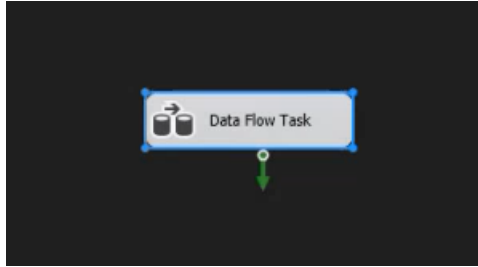
```
id_cliente INT,  
nombre_cliente NVARCHAR(255),  
frecuencia_compra INT,  
monto_total NUMERIC(18, 2)  
);
```

-- Crear la tabla para clientes ocasionales

```
CREATE TABLE ClientesOcasionales (  
id_cliente INT,  
nombre_cliente NVARCHAR(255),  
frecuencia_compra INT,  
monto_total NUMERIC(18, 2)  
);
```

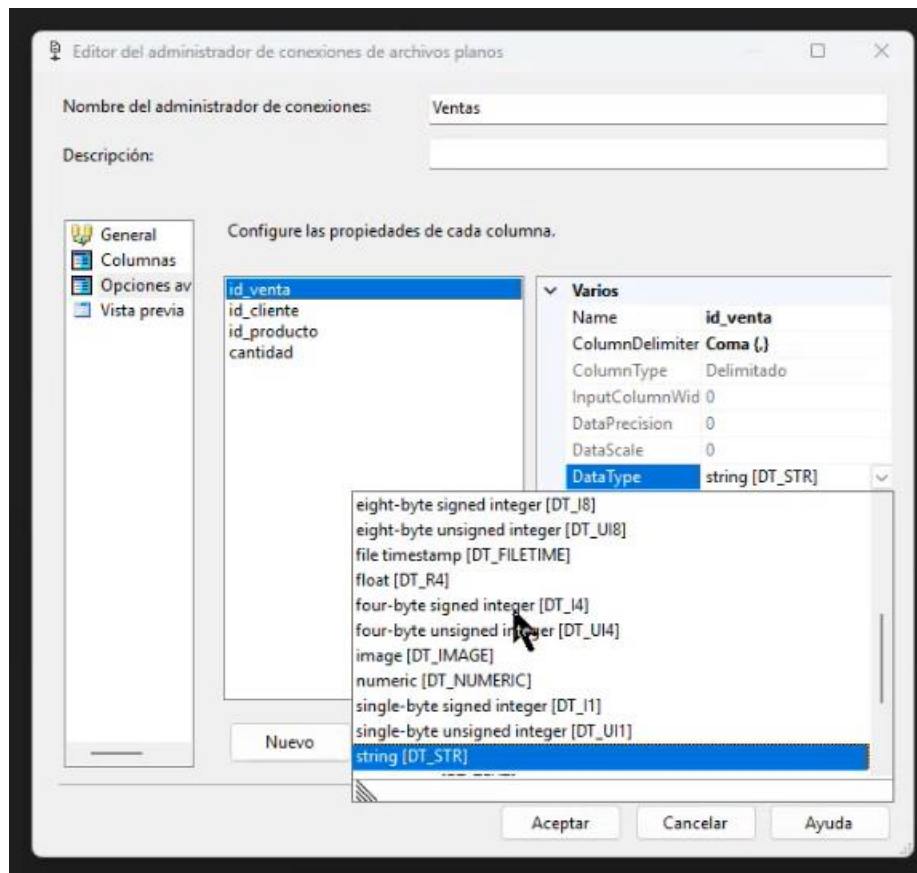
“AGREGAR FLAT FILE SOURCES”

Agregamos un Data Flow Task

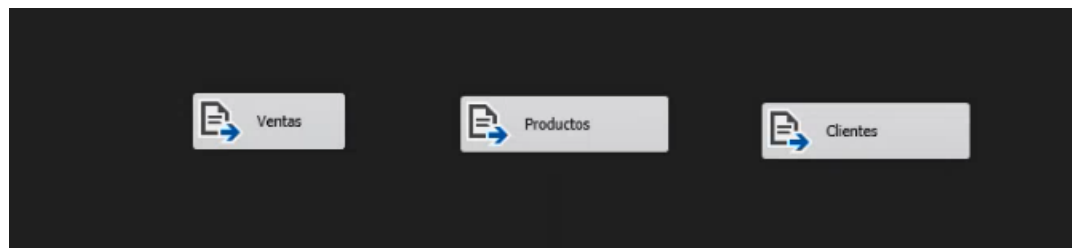
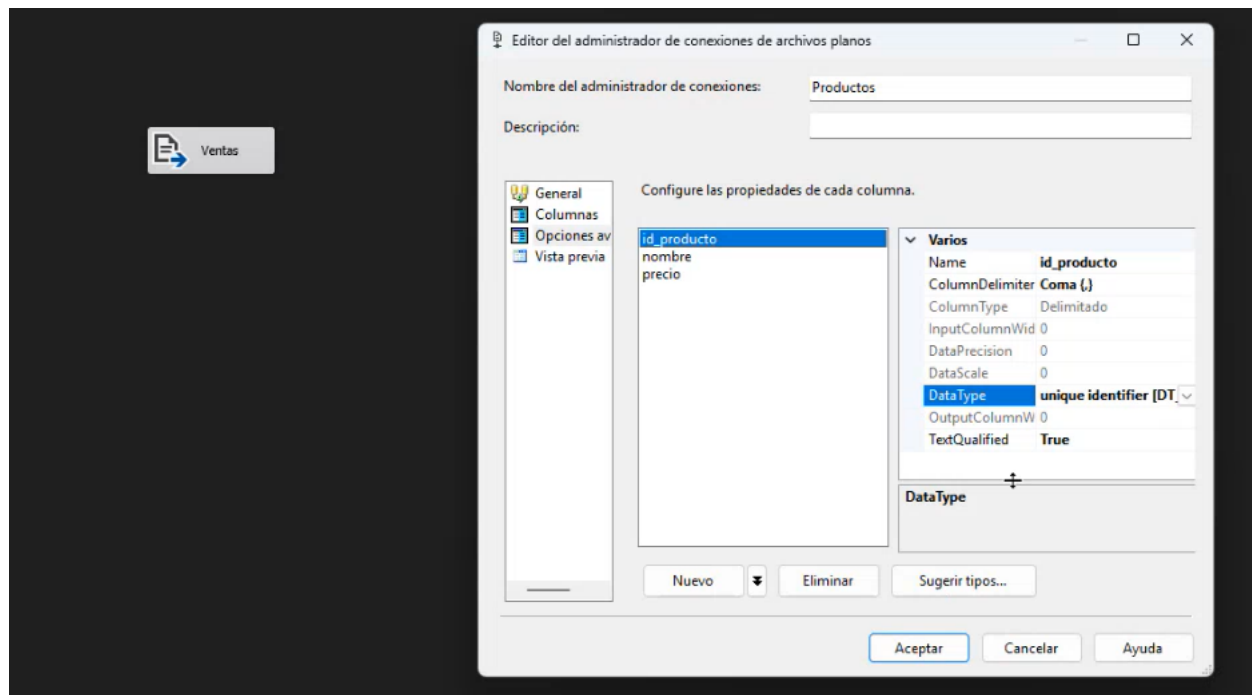


(En este apartado vamos a procesar datos.)

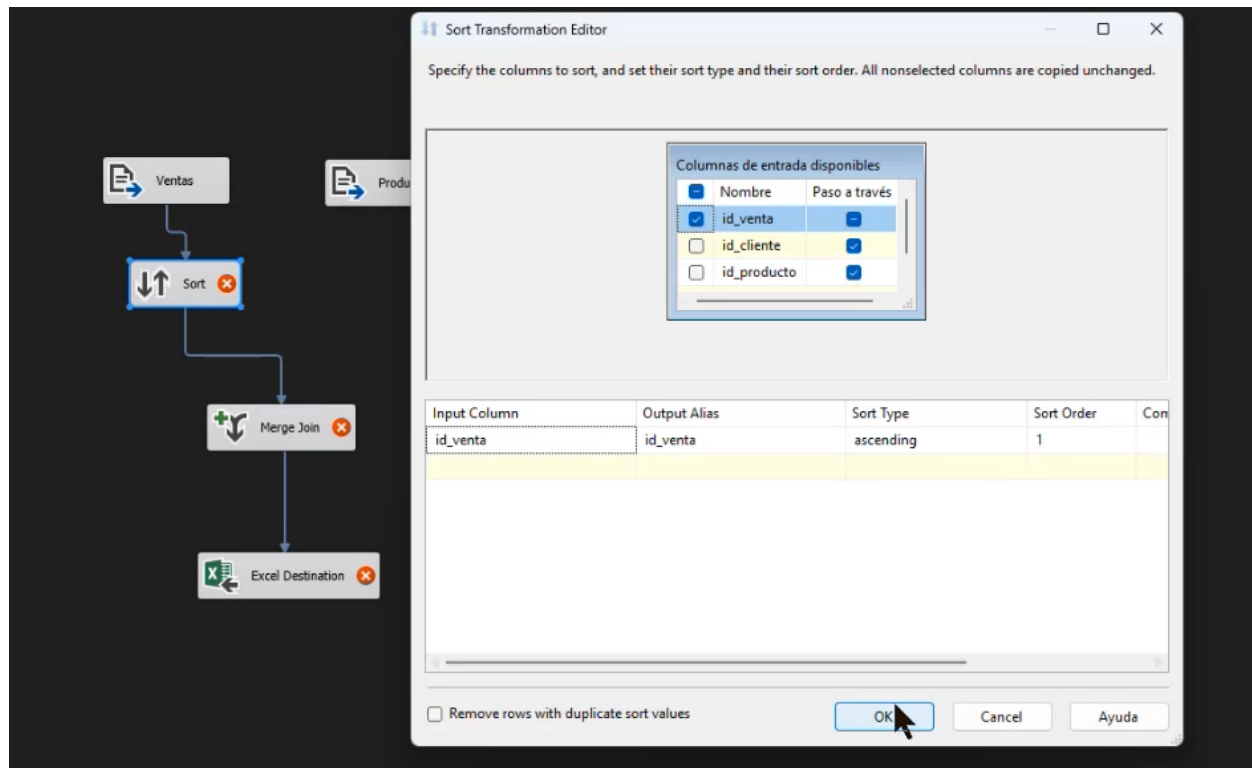
Agregamos los “Flat File Sources” y modificamos los valores de cada columna según corresponda, en el caso de “id_venta” serían integer, ya que son números.



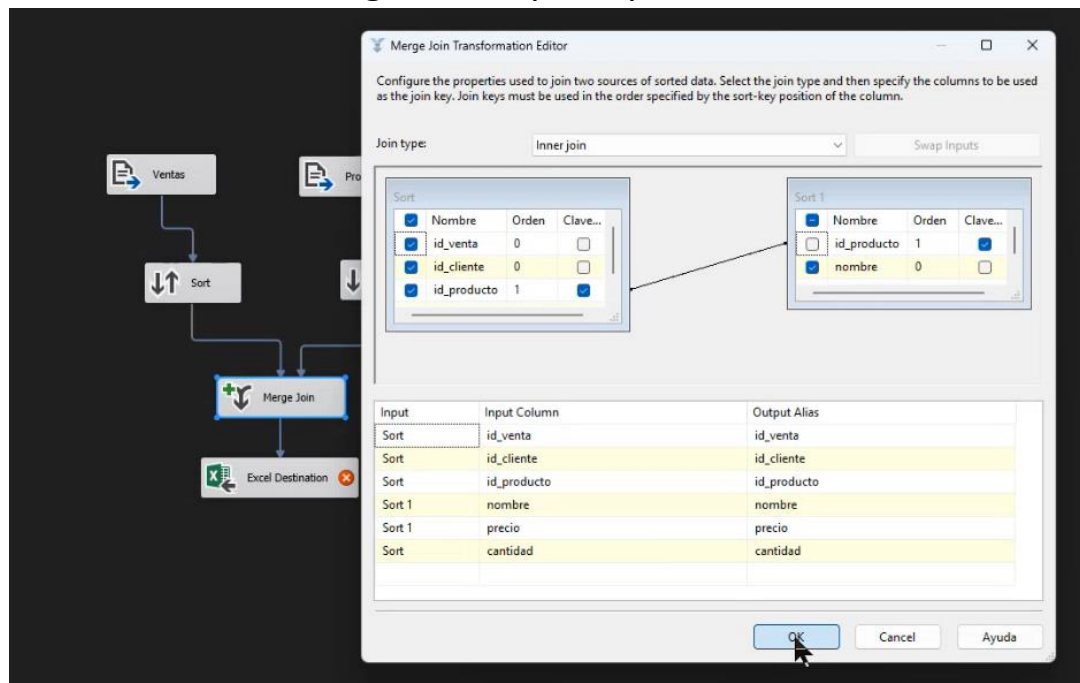
Y así sucesivamente con los siguientes .csv



Nos aseguramos de ordenar los datos con un “Sort” y ordenar los datos de las csv para poder unirlos con un “Merge Join” más adelante



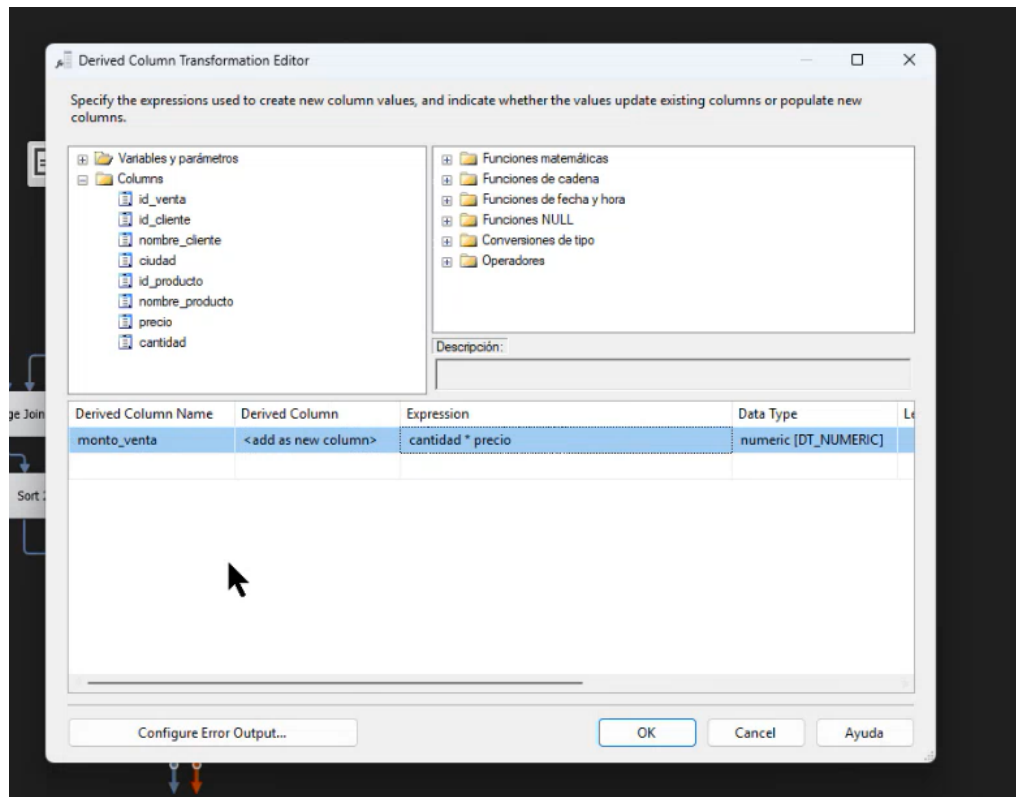
Utilizamos un “Merge Join” para poder unir los datos de las tablas



Luego de ordenarlas y utilizar el “Merge Join”, verificamos que el sistema funcione perfectamente

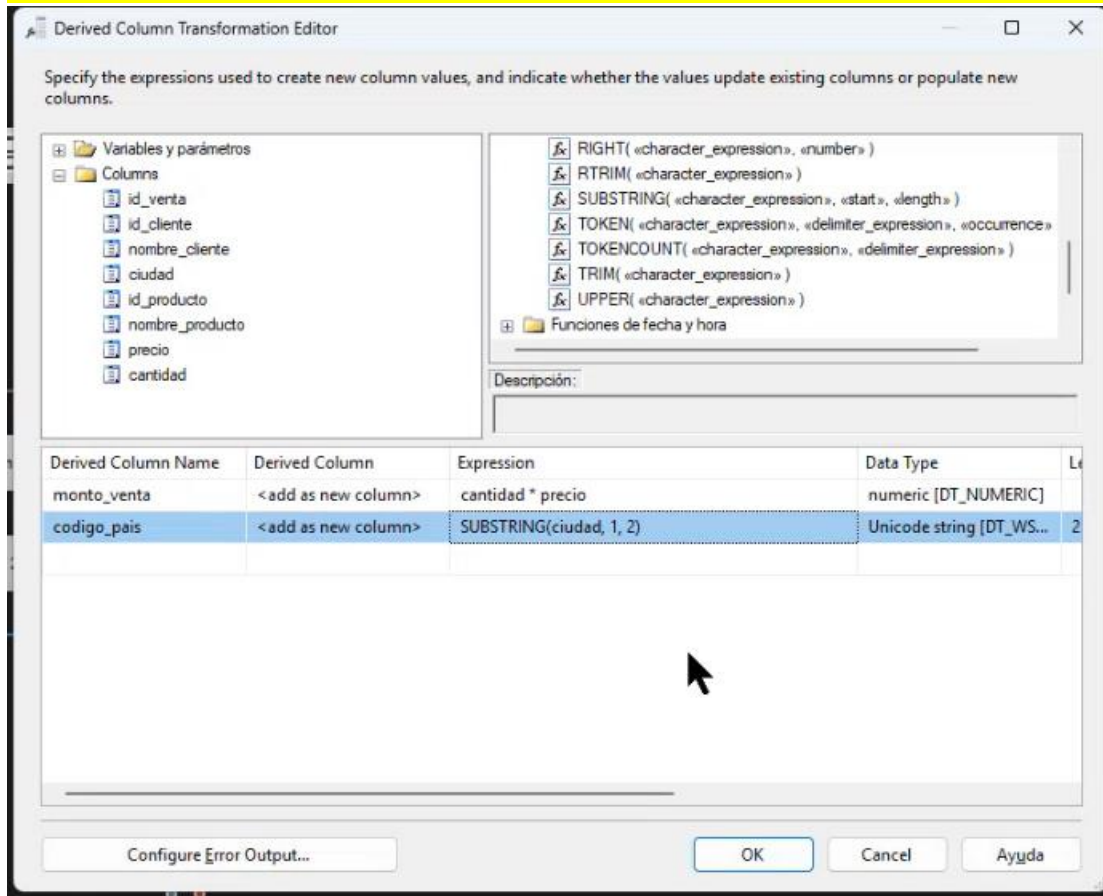


Sacamos el monto total de las ventas por cliente con lo siguiente, que es parte que lo solicita el ejercicio

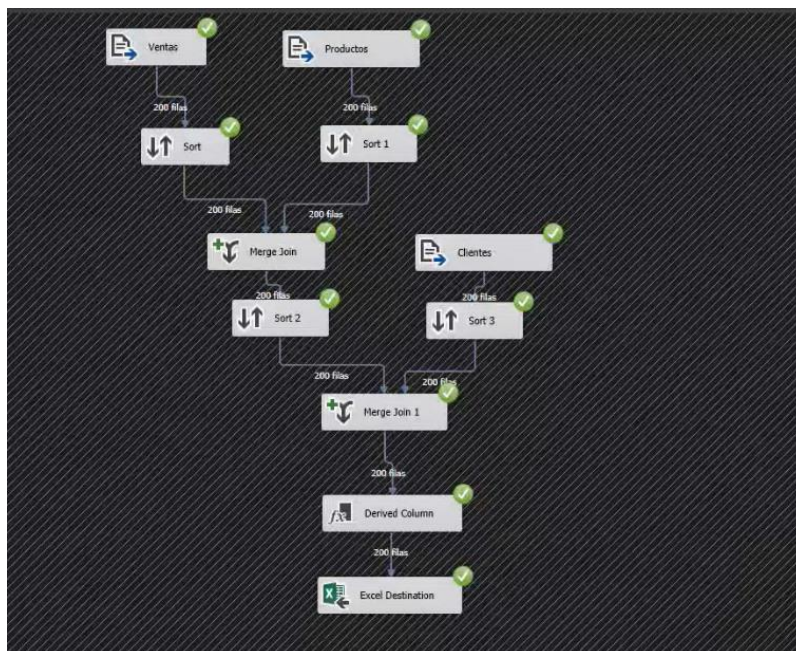


Utilizamos el “SUBSTRING (ciudad 1,2)” para substraer el “Código país” de la “Ciudad”, en este caso he elegido ciudad ya que en el ejercicio decía del “ID_CLIENTE” pero verificando los datos, todos los ID_CLIENTES comienzan con “cliente” entonces no le encontré sentido que se repitieran los dos caracteres de “CL” entonces intuí que sería los

de ciudad, ya que es similar a país y varían sus primeros caracteres.



Verificamos que nos corre el programa :D



Separar		Copiar datos							
id_venta	i...	nombre_cliente	ciudad	i...	nombre_produ...	precio	cantidad	monto_venta	codigo_pais
1100	1	Cliente_1	Valencia	1...	Producto_111	867	6	5202	Va
1199	3	Cliente_3	Valencia	2...	Producto_294	362	10	3620	Va
1122	6	Cliente_6	Barcel...	1...	Producto_114	32	6	192	Ba
1073	9	Cliente_9	Madrid	1...	Producto_123	935	5	4675	Ma
1115	9	Cliente_9	Madrid	2...	Producto_275	508	8	4064	Ma
1039	10	Cliente_10	Barcel...	1...	Producto_151	317	4	1268	Ba
1185	11	Cliente_11	Madrid	1...	Producto_143	740	8	5920	Ma
1036	11	Cliente_11	Madrid	2...	Producto_283	116	9	1044	Ma
1015	12	Cliente_12	Sevilla	1...	Producto_118	696	1	696	Se
1162	12	Cliente_12	Sevilla	2...	Producto_247	163	4	652	Se
1014	14	Cliente_14	Barcel...	1...	Producto_153	191	1	191	Ba
1054	16	Cliente_16	Barcel...	1...	Producto_155	11	6	66	Ba
1108	18	Cliente_18	Bilbao	2...	Producto_283	116	6	696	Bi
1118	19	Cliente_19	Bilbao	3...	Producto_300	311	6	1866	Bi
1040	20	Cliente_20	Sevilla	2...	Producto_249	598	10	5980	Se
1152	21	Cliente_21	Bilbao	2...	Producto_263	356	2	712	Bi
1077	21	Cliente_21	Bilbao	1...	Producto_141	394	10	3940	Bi
1187	21	Cliente_21	Bilbao	1...	Producto_147	516	10	5160	Bi
1009	22	Cliente_22	Madrid	2...	Producto_249	598	7	4186	Ma
1070	23	Cliente_23	Madrid	1...	Producto_145	333	7	2331	Ma
1007	23	Cliente_23	Madrid	2...	Producto_266	176	8	1408	Ma
1092	24	Cliente_24	Valencia	2...	Producto_206	989	7	6923	Va
1200	26	Cliente_26	Barcel...	1...	Producto_180	91	5	455	Ba
1194	27	Cliente_27	Barcel...	2...	Producto_232	938	7	6566	Ba
1005	28	Cliente_28	Madrid	2...	Producto_218	492	2	984	Ma
1074	29	Cliente_29	Valencia	2...	Producto_274	729	3	2187	Va
1064	30	Cliente_30	Madrid	1...	Producto_190	493	6	2958	Ma
1170	30	Cliente_30	Madrid	2...	Producto_216	495	10	4950	Ma
1191	32	Cliente_32	Barcel...	1...	Producto_150	523	2	1046	Ba
1097	34	Cliente_34	Sevilla	1...	Producto_111	867	4	3468	Se
1047	38	Cliente_38	Sevilla	2...	Producto_296	440	7	3080	Se
1190	38	Cliente_38	Sevilla	1...	Producto_144	307	9	2763	Se
1141	38	Cliente_38	Sevilla	1...	Producto_188	152	9	1368	Se
1105	41	Cliente_41	Valencia	1...	Producto_198	553	4	2212	Va
1182	44	Cliente_44	Bilbao	2...	Producto_291	588	4	2352	Bi
1163	44	Cliente_44	Bilbao	1...	Producto_148	541	3	1623	Bi
1002	48	Cliente_48	Valencia	1...	Producto_161	674	9	6066	Va
1144	49	Cliente_49	Barcel...	2...	Producto_246	756	2	1512	Ba
1132	49	Cliente_49	Barcel...	2...	Producto_244	377	5	1885	Ba
1085	49	Cliente_49	Barcel...	1...	Producto_170	44	3	132	Ba
1111	49	Cliente_49	Barcel...	2...	Producto_252	339	5	1695	Ba
1003	51	Cliente_51	Bilbao	2...	Producto_258	622	5	3110	Bi
1065	51	Cliente_51	Bilbao	2...	Producto_287	247	10	2470	Bi
1116	51	Cliente_51	Bilbao	2...	Producto_201	943	8	7544	Bi
1145	52	Cliente_52	Valencia	2...	Producto_290	11	8	88	Va
1186	52	Cliente_52	Valencia	2...	Producto_223	343	10	3430	Va

Adjunto Número total de filas: 0. Búferes: 0

Filas mostradas = 200

Adjunto Número total de filas: 0. Búferes: 0

Filas mostradas = 200

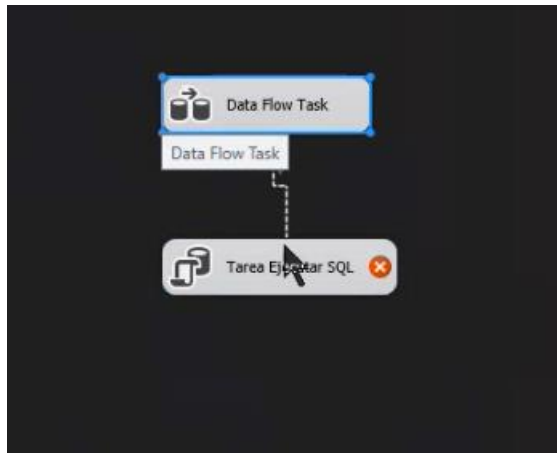
Portapapeles		Fuente	Alineación	Número	Estilos de celda	Formato	Edición	Complementos												
C9		fx		9																
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	ciudad	precio	cantidad	nombre_c	id_venta	id_cliente	nombre_c	id_produc	monto_ve	codigo_pais										
2	Valencia	867	6	Producto_1100	1	Cliente_1_111			5202	Va										
3	Valencia	362	10	Producto_1199	3	Cliente_3_294			3620	Va										
4	Barcelona	32	6	Producto_1122	6	Cliente_6_114			192	Ba										
5	Madrid	935	5	Producto_1073	9	Cliente_9_123			4675	Ma										
6	Madrid	508	8	Producto_1115	9	Cliente_9_275			4064	Ma										
7	Barcelona	317	4	Producto_1039	10	Cliente_1_151			1268	Ba										
8	Madrid	740	8	Producto_1185	11	Cliente_1_143			5920	Ma										
9	Madrid	116	9	Producto_1036	11	Cliente_1_283			1044	Ma										
10	Sevilla	696	1	Producto_1015	12	Cliente_1_118			696	Se										
11	Sevilla	163	4	Producto_1162	12	Cliente_1_247			652	Se										
12	Barcelona	191	1	Producto_1014	14	Cliente_1_153			191	Ba										
13	Barcelona	11	6	Producto_1054	16	Cliente_1_155			66	Ba										
14	Bilbao	116	6	Producto_1108	18	Cliente_1_283			696	Bi										
15	Bilbao	311	6	Producto_1118	19	Cliente_1_300			1866	Bi										
16	Sevilla	598	10	Producto_1040	20	Cliente_2_249			5980	Se										
17	Bilbao	356	2	Producto_1152	21	Cliente_2_263			712	Bi										
18	Bilbao	394	10	Producto_1077	21	Cliente_2_141			3940	Bi										
19	Bilbao	516	10	Producto_1187	21	Cliente_2_147			5160	Bi										
20	Madrid	598	7	Producto_1009	22	Cliente_2_249			4186	Ma										
21	Madrid	333	7	Producto_1070	23	Cliente_2_145			2331	Ma										
22	Madrid	176	8	Producto_1007	23	Cliente_2_266			1408	Ma										
23	Valencia	688	7	Producto_1092	24	Cliente_2_206			6923	Va										

Hoja1

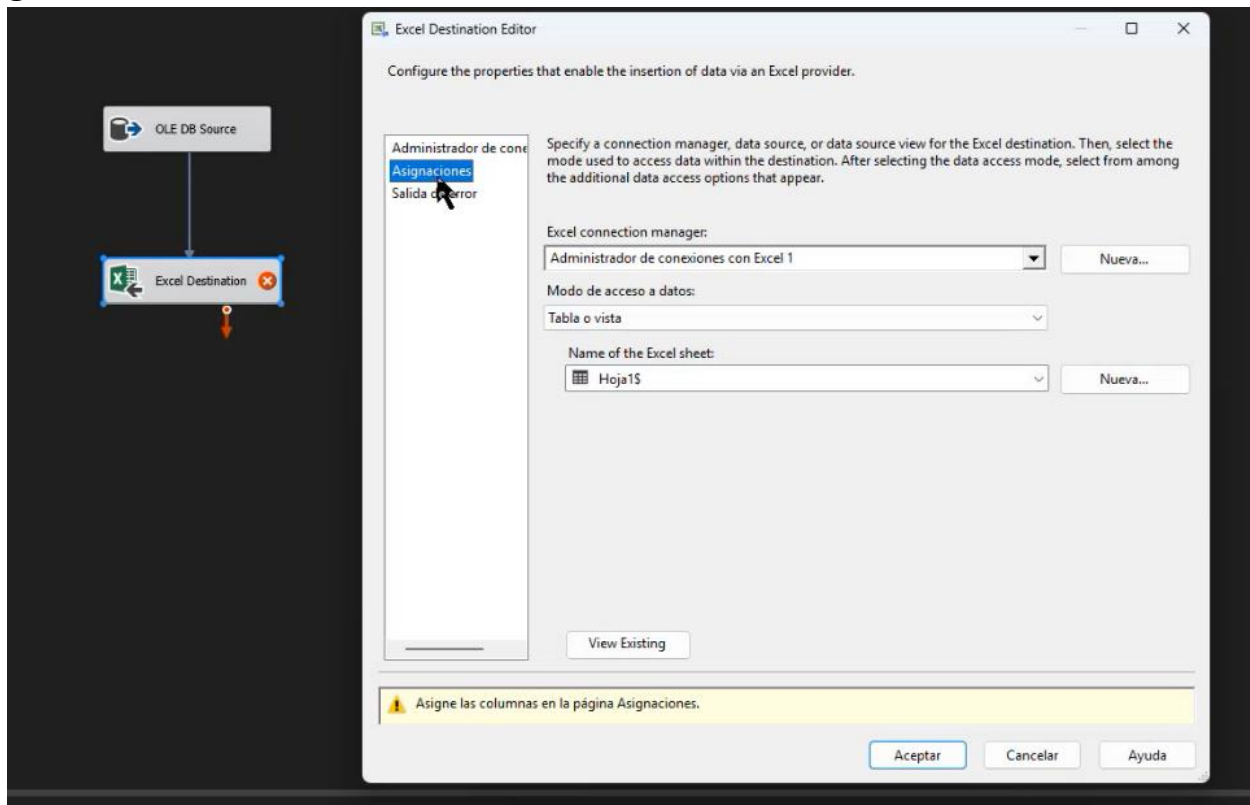
Accesibilidad: todo correcto

100%

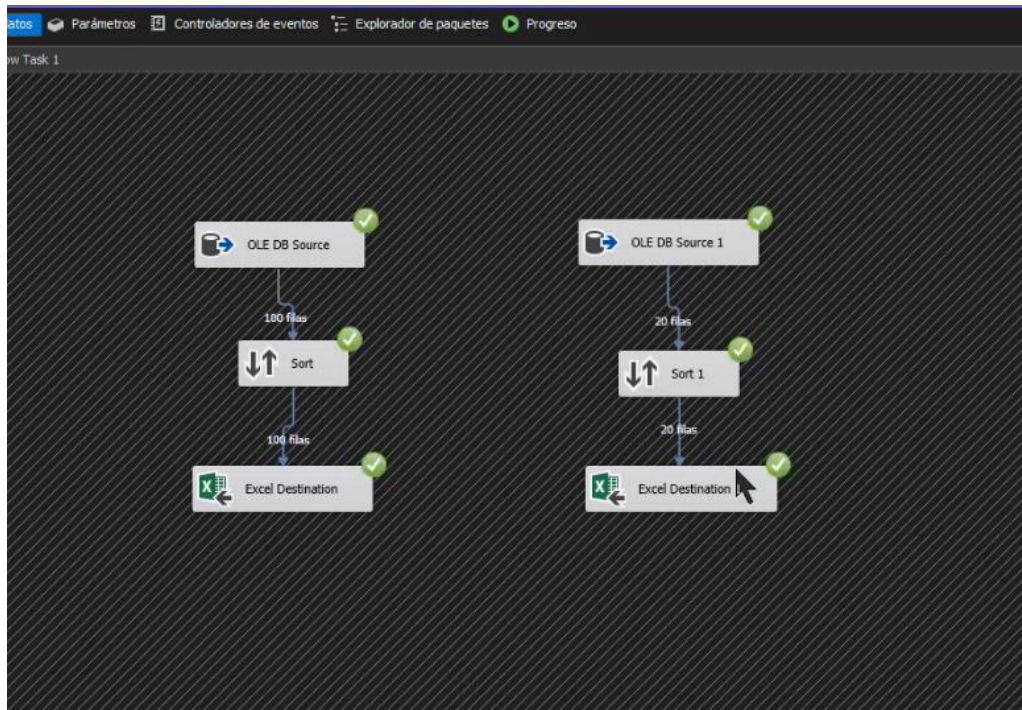
Ahora separamos los clientes frecuentes y ocasionales (en este caso los vamos a diferenciar con lo que tanto compran y el monto que han gastado en total), utilizaremos una “Tarea Ejecutar SQL” para poder crear las diferentes tablas.



Para cada tabla conectamos su Excel en el que lo va a exportar, luego de generar las consultas/tareas en SQL



y así con el otro Excel, también agregamos un SORT, para ordenarlos de manera ascendente y descendente como se solicita en el ejercicio



Verificamos que funciona el programa :D

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id_cliente	nombre_cliente	frecuencia_compra	monto_total									
2													
3													
4													
5													
6													
7	200	Cliente_200	5	14522.00									
8	49	Cliente_49	4	5224.00									
9	158	Cliente_158	4	12265.00									
10	196	Cliente_196	3	5065.00									
11	174	Cliente_174	3	9432.00									
12	167	Cliente_167	3	6980.00									
13	160	Cliente_160	3	7873.00									
14	138	Cliente_138	3	7673.00									
15	130	Cliente_130	3	6812.00									
16	122	Cliente_122	3	10366.00									
17	109	Cliente_109	3	13063.00									
18	105	Cliente_105	3	5502.00									
19	81	Cliente_81	3	7001.00									
20	51	Cliente_51	3	13124.00									
21	38	Cliente_38	3	7211.00									
22	198	Cliente_198	3	8276.00									
23	51	Cliente_51	3	8612.00									