



UNIVERSIDAD DON BOSCO

Materia: Datawarehouse y Minería de Datos

Docente: Ing. Karens Medrano

Estudiante: Eduardo Ezequiel López Rivera

Carné: LR230061

Ciclo: 01-2025

Problemática

La Floristería **"Fiorella"** quiere saber cómo se compran sus productos y tiene la data de tres departamentos del país. Por ello, solicita su opinión sobre qué productos sobresalen, qué combinaciones son mejores y desea este estudio tanto por departamento como a nivel nacional.

Finalmente, los datos se cargarán en 4 archivos de Excel: **"Preferencias_Departamentales.xlsx"**, **"Preferencias_Nacionales.xlsx"**, **"Ventas_Departamentales.xlsx"** y **"Ventas_Nacional.xlsx"** incorporando el análisis de tendencias y patrones de compra.

Contenido

Problemática	2
Configuración de proyecto en visual studio 2022	4
<i>TOMAR EN CUENTA</i>	5
Solución	8
Desarrollo 10%	8
Desarrollo 25%	9
Desarrollo 40%	10
Desarrollo 50%	11
Desarrollo 60%	12
Desarrollo 75%	13
Desarrollo 85%	14
Desarrollo 100%	15
Ejecución exitosa.....	16

Configuración de proyecto en visual studio 2022

En caso de ejecutar el proyecto, clonar el siguiente repositorio:

https://github.com/eduardoezequieel/desafio1_DMD_EduardoLopez.git

Existen 4 carpetas

- Fiorella_Data_CSV (Están contenidos los archivos .csv correspondientes a la data que se va a utilizar como origen).
- Fiorella_Data_Excel (Están contenidos los archivos .xlsx correspondientes a la data resultante de las operaciones, es decir, el destino).
- Integration Services Project1 (Contiene el proyecto de visual studio 2022).
- Moldes (Son las plantillas de los archivos en Excel, contiene únicamente las columnas, solo están de respaldo).

TOMAR EN CUENTA

Crear base de datos, puede tener cualquier nombre, pero se deben crear las siguientes tablas:

```
CREATE TABLE [VentasFiorella] (  
    [id] nvarchar(50),  
    [RosasInt] int,  
    [ClavelesInt] int,  
    [MacetasInt] int,  
    [TierraInt] int,  
    [GirasolesInt] int,  
    [HortensialInt] int,  
    [GlobosInt] int,  
    [TarjetasInt] int,  
    [LiriosInt] int,  
    [AuroraInt] int,  
    [TulipanesInt] int,  
    [OrquideasInt] int,  
    [CarmesInt] int,  
    [ListonInt] int,  
    [Departamento] nvarchar(12)  
);
```

```
CREATE TABLE Combinaciones_Productos_Departamento (  
    Departamento NVARCHAR(255),  
    Producto1 NVARCHAR(255),  
    Producto2 NVARCHAR(255),  
    CantidadCompras INT  
);
```

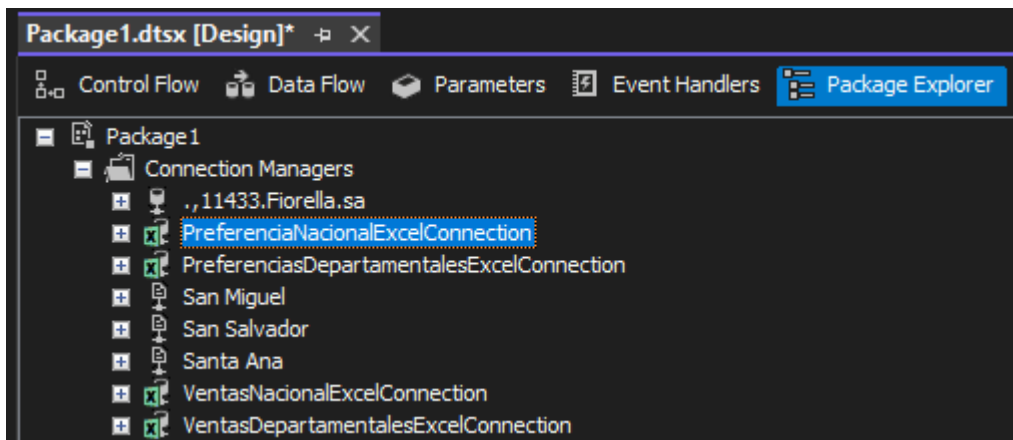
```
CREATE TABLE Combinaciones_Productos_Pais (

    Producto1 NVARCHAR(255),

    Producto2 NVARCHAR(255),

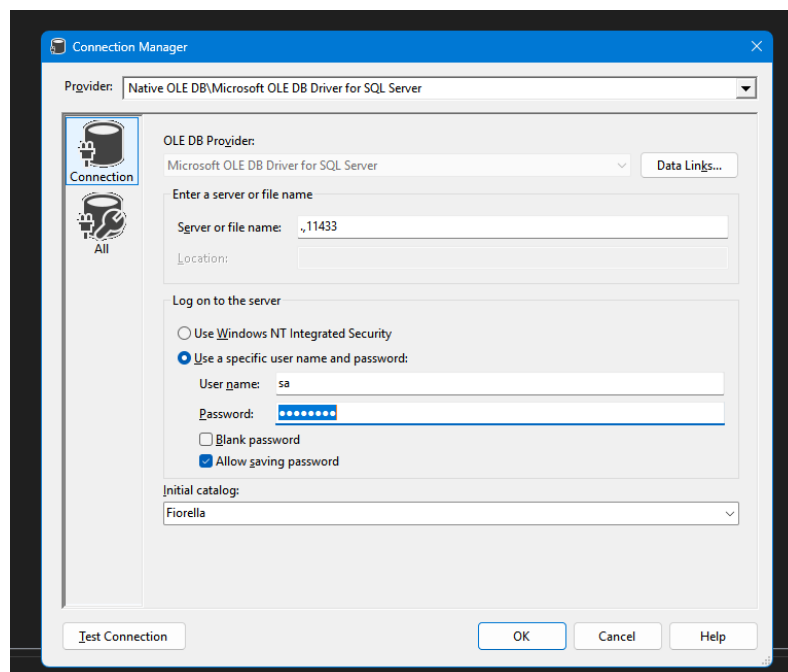
    CantidadCompras INT

);
```



Es necesario revisar cada una de las conexiones del proyecto y asignarlas a sus archivos correspondientes, ya que se usan rutas absolutas, estas no funcionarán justo después de clonar el proyecto. Se detalla a continuación cada una de las conexiones:

- **.,11433.Fiorella.sa:** Contiene la conexión hacia la base de datos en SQL Server, rellenar los campos a conveniencia y luego de verificar que funciona la conexión, por favor verificar que exista la base de datos con la tabla principal

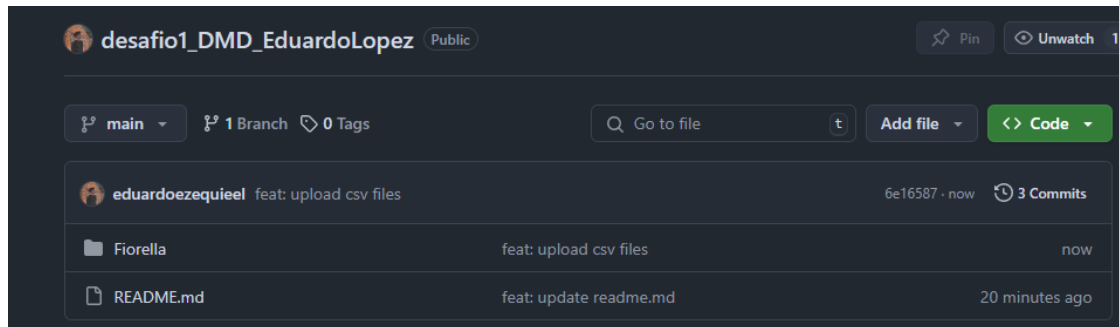


- **PreferenciaNacionalExcelConnection:** Ruta al archivo Preferencia_Nacional.xlsx
- **PreferenciasDepartamentalesExcelConnection:** Ruta al archivo Preferencias_Departamentales.xlsx
- **San Miguel:** Ruta al archivo .csv de las ventas en San Miguel.
- **Santa Ana:** Ruta al archivo .csv de las ventas en Santa Ana.
- **San Salvador:** Ruta al archivo .csv de las ventas en San Salvador.
- **VentasNacionalExcelConnection:** Ruta al archivo Ventas_Nacional.xlsx
- **VentasDepartamentalesExcelConnection:** Ruta al archivo Ventas_Departamentales.xlsx

Solución

Desarrollo 10%

Se crea un repositorio Git, alojado en GitHub para un control de versiones adecuado.



Desarrollo 25%

Se agregan Flat File Sources, correspondiendo a cada una de las sucursales de la floristería, y se realizan conversiones de tipos de datos, las columnas de booleanos a integers, finalmente agregamos una nueva columna para identificar a que sucursal pertenece la venta.

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

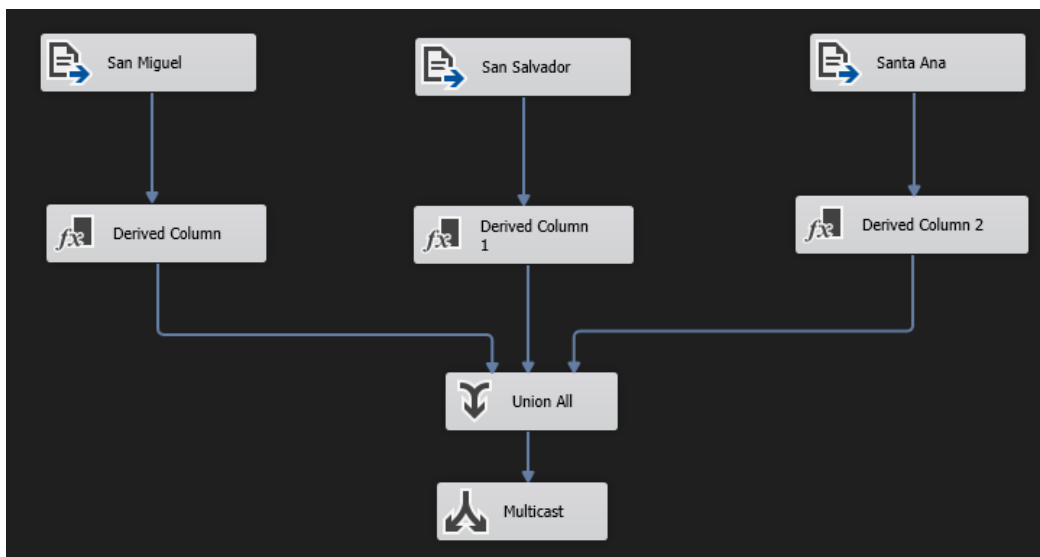
Columns	Funciones matemáticas
Rosas	Funciones de cadena
Claveles	Funciones de fecha y hora
Macetas	Funciones NULL
Tierra	Conversiones de tipo
Girasoles	Operadores
Hortensia	
Globos	
Tarjetas	
Lirios	
Aurora	
Tulpanes	

Derived Column Name	Derived Column	Expression	Data Type	Length	Precision	Scale	Code Page
RosasInt	<add as new column>	(Rosas ? 1 : 0)	four-byte signed integ...				
ClavelesInt	<add as new column>	(Claveles ? 1 : 0)	four-byte signed integ...				
MacetasInt	<add as new column>	(Macetas ? 1 : 0)	four-byte signed integ...				
Tierralnt	<add as new column>	(Tierra ? 1 : 0)	four-byte signed integ...				
GirasolesInt	<add as new column>	(Girasoles ? 1 : 0)	four-byte signed integ...				
Hortensialnt	<add as new column>	(Hortensia ? 1 : 0)	four-byte signed integ...				
GlobosInt	<add as new column>	(Globos ? 1 : 0)	four-byte signed integ...				
TarjetasInt	<add as new column>	(Tarjetas ? 1 : 0)	four-byte signed integ...				
LiriosInt	<add as new column>	(Lirios ? 1 : 0)	four-byte signed integ...				
Auroralnt	<add as new column>	(Aurora ? 1 : 0)	four-byte signed integ...				
TulpanesInt	<add as new column>	(Tulpanes ? 1 : 0)	four-byte signed integ...				
OrquideasInt	<add as new column>	(Orquideas ? 1 : 0)	four-byte signed integ...				
CarmesInt	<add as new column>	(Carmesi ? 1 : 0)	four-byte signed integ...				
ListonInt	<add as new column>	(Liston ? 1 : 0)	four-byte signed integ...				
Departamento	<add as new column>	"San Miguel"	Unicode string [DT_WS...	10			



Desarrollo 40%

Unimos las 3 fuentes de datos con un **union all** y agregamos un **multicast** para poder tener varias salidas de datos



Output Column Name	Union All Input 4	Union All Input 1	Union All Input 2
Rosas	Rosas	Rosas	Rosas
Claveles	Claveles	Claveles	Claveles
Macetas	Macetas	Macetas	Macetas
Tierra	Tierra	Tierra	Tierra
Girasoles	Girasoles	Girasoles	Girasoles
Hortensia	Hortensia	Hortensia	Hortensia
Globos	Globos	Globos	Globos
Tarjetas	Tarjetas	Tarjetas	Tarjetas
Lirios	Lirios	Lirios	Lirios
Aurora	Aurora	Aurora	Aurora
Tulipanes	Tulipanes	Tulipanes	Tulipanes
id	id	id	id
Orquideas	Orquideas	Orquideas	Orquideas
Carmesi	Carmesi	Carmesi	Carmesi
Liston	Liston	Liston	Liston
RosasInt	RosasInt	RosasInt	RosasInt
ClavelesInt	ClavelesInt	ClavelesInt	ClavelesInt
MacetasInt	MacetasInt	MacetasInt	MacetasInt
Tierralnt	Tierralnt	Tierralnt	Tierralnt
GirasolesInt	GirasolesInt	GirasolesInt	GirasolesInt
Hortensialnt	Hortensialnt	Hortensialnt	Hortensialnt
GlobosInt	GlobosInt	GlobosInt	GlobosInt
TarjetasInt	TarjetasInt	TarjetasInt	TarjetasInt
LiriosInt	LiriosInt	LiriosInt	LiriosInt
Auroralnt	Auroralnt	Auroralnt	Auroralnt
TulipanesInt	TulipanesInt	TulipanesInt	TulipanesInt
OrquideasInt	OrquideasInt	OrquideasInt	OrquideasInt
CarmesiInt	CarmesiInt	CarmesiInt	CarmesiInt
ListonInt	ListonInt	ListonInt	ListonInt
Departamento	Departamento	Departamento	Departamento

Desarrollo 50%

Se agregan dos **aggregate**, uno para calcular las ventas por producto por departamento y otro a nivel nacional. Luego aplicamos un **unpivot** para pasar las columnas de cada producto, a una sola columna llamada **Producto** y la suma de estas a una llamada **Cantidad**.

Input Column	Output Alias	Operation	Cc
RosasInt	Rosas	Sum	
ClavelesInt	Claveles	Sum	
MacetasInt	Macetas	Sum	
Tierralnt	Tierra	Sum	
GirasolesInt	Girasoles	Sum	
Hortensialnt	Hortensia	Sum	
GlobosInt	Globos	Sum	
TarjetasInt	Tarjetas	Sum	
LiriosInt	Lirios	Sum	
Auroralnt	Aurora	Sum	
TulipanesInt	Tulipanes	Sum	
OrquideasInt	Orquideas	Sum	
CarmesInt	Carmesi	Sum	
ListonInt	Liston	Sum	
Departamento	Departamento	Group by	

Input Column	Output Alias	Operation	Cc
RosasInt	Rosas	Sum	
ClavelesInt	Claveles	Sum	
MacetasInt	Macetas	Sum	
Tierralnt	Tierra	Sum	
GirasolesInt	Girasoles	Sum	
Hortensialnt	Hortensia	Sum	
GlobosInt	Globos	Sum	
TarjetasInt	Tarjetas	Sum	
LiriosInt	Lirios	Sum	
Auroralnt	Aurora	Sum	
TulipanesInt	Tulipanes	Sum	
OrquideasInt	Orquideas	Sum	
CarmesInt	Carmesi	Sum	
ListonInt	Liston	Sum	

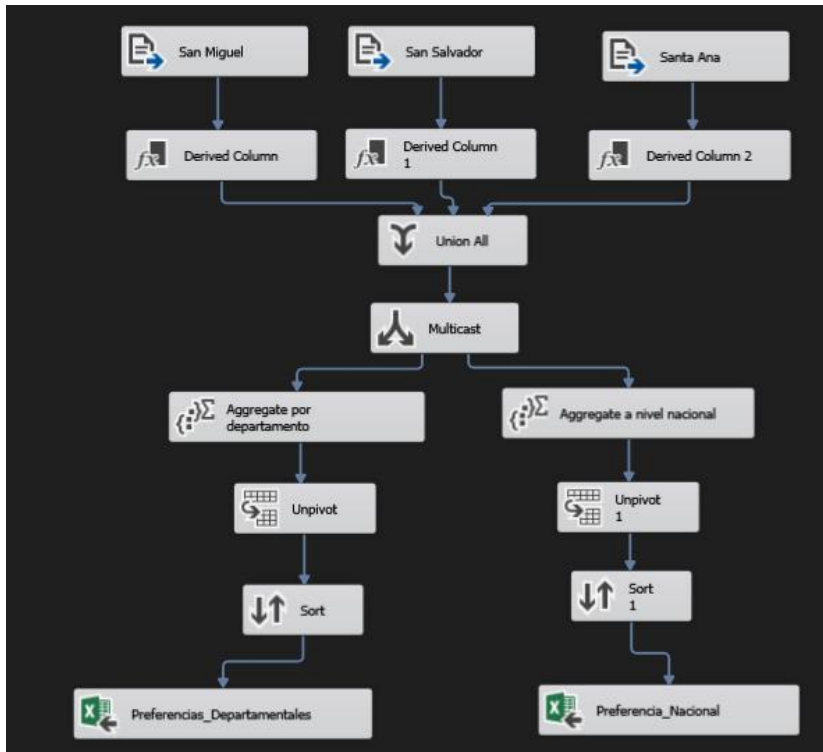
Input Column	Destination Column	Pivot Key Value
Hortensia	Cantidad	Hortensia
Globos	Cantidad	Globos
Tarjetas	Cantidad	Tarjetas
Lirios	Cantidad	Lirios
Aurora	Cantidad	Aurora
Tulipanes	Cantidad	Tulipanes
Orquideas	Cantidad	Orquideas
Carmesi	Cantidad	Carmesi

Pivot key value column name:

Producto

Desarrollo 60%

Se agrega una salida para un archivo Excel que contendrá la información a nivel nacional y otro a nivel departamental, se crean las conexiones con los archivos que previamente poseen las columnas para que el modelado corresponda.



Excel Destination Editor

Configure the properties that enable the insertion of data via an Excel provider.

Connection Manager

Mappings

Error Output

Specify a connection manager, data source, or data source view for the Excel destination. Then, select the mode used to access data within the destination. After selecting the data access mode, select from among the additional data access options that appear.

Excel connection manager:

PreferenciasDepartamentalesExcelConnection

New...

Data access mode:

Table or view

Name of the Excel sheet:

Hoja1\$

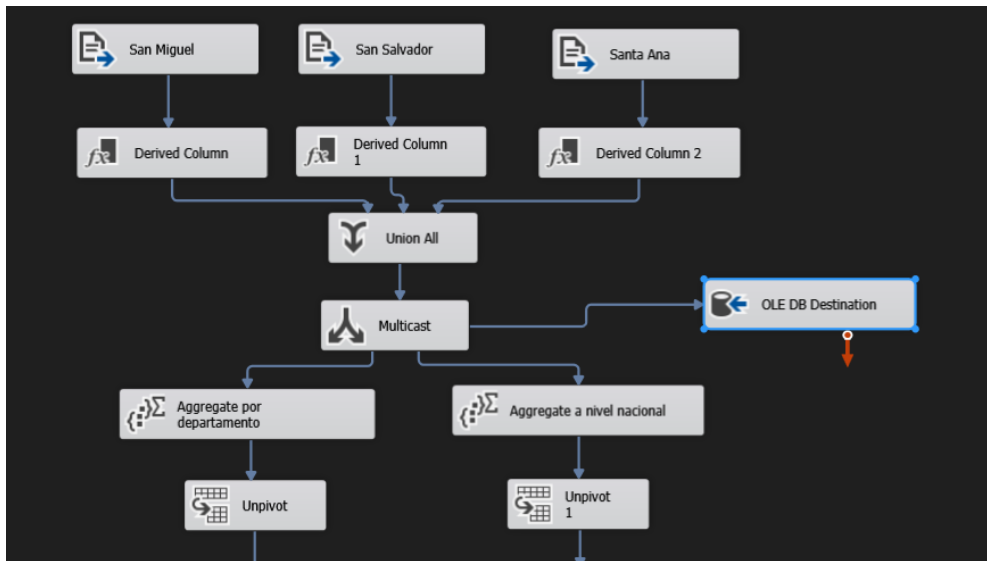
New...

View Existing

OK Cancel Help

Desarrollo 75%

Se agrega una tercera salida, en este caso un OLE DB Destination para guardar la data y poder hacer consultas mas avanzadas mas adelante. Es necesario asegurarnos de que exista la tabla de ventas (VentasFiorella)



OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager
Mappings
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:
.,11433.Fiorella.sa New...

Data access mode:
Table or view - fast load

Name of the table or the view:
[VentasFiorella] New...

☐ Keep identity ☒ Table lock
☐ Keep nulls ☐ Check constraints

Rows per batch:

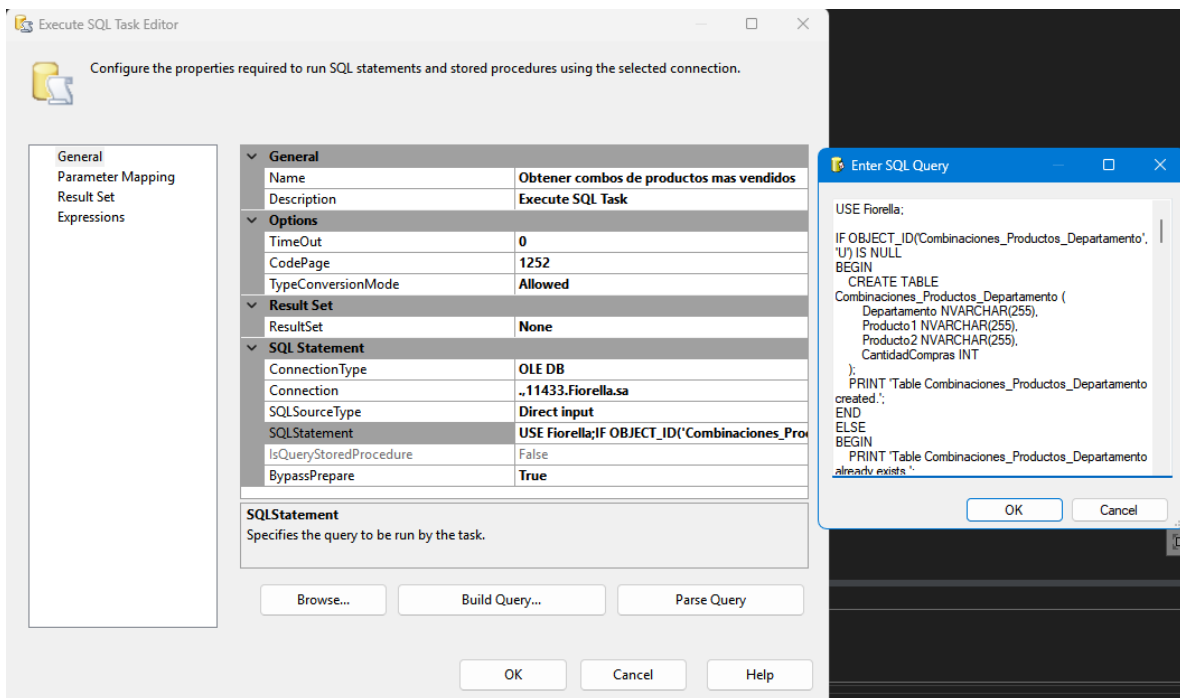
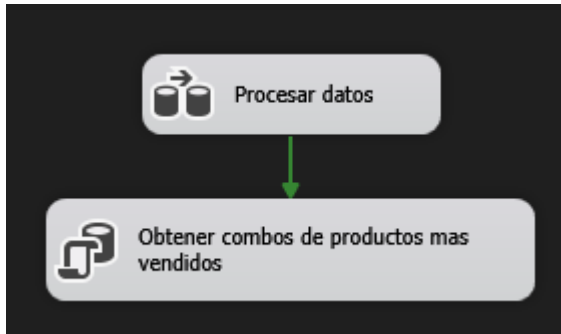
Maximum insert commit size:

View Existing

OK Cancel Help

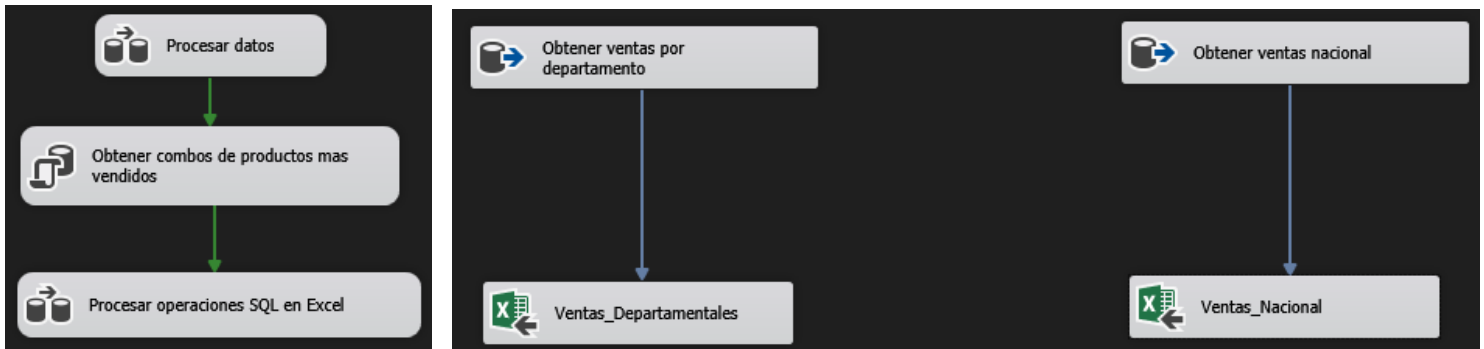
Desarrollo 85%

Creamos un Execute SQL Task en nuestro Control Flow para hacer una consulta SQL y obtener los combos de productos mas vendidos a nivel departamental y nacional.



Desarrollo 100%

Agregamos un data flow para leer los datos que se guardaron en el paso anterior y pasarlos los archivos correspondientes en formato .xlsx



OLE DB connection manager:

.,11433.Fiorella.sa

New...

Data access mode:

Table or view

Name of the table or the view:

[dbo].[Combinaciones_Productos_Departamento]

Excel connection manager:

VentasDepartamentalesExcelConnection

New...

Data access mode:

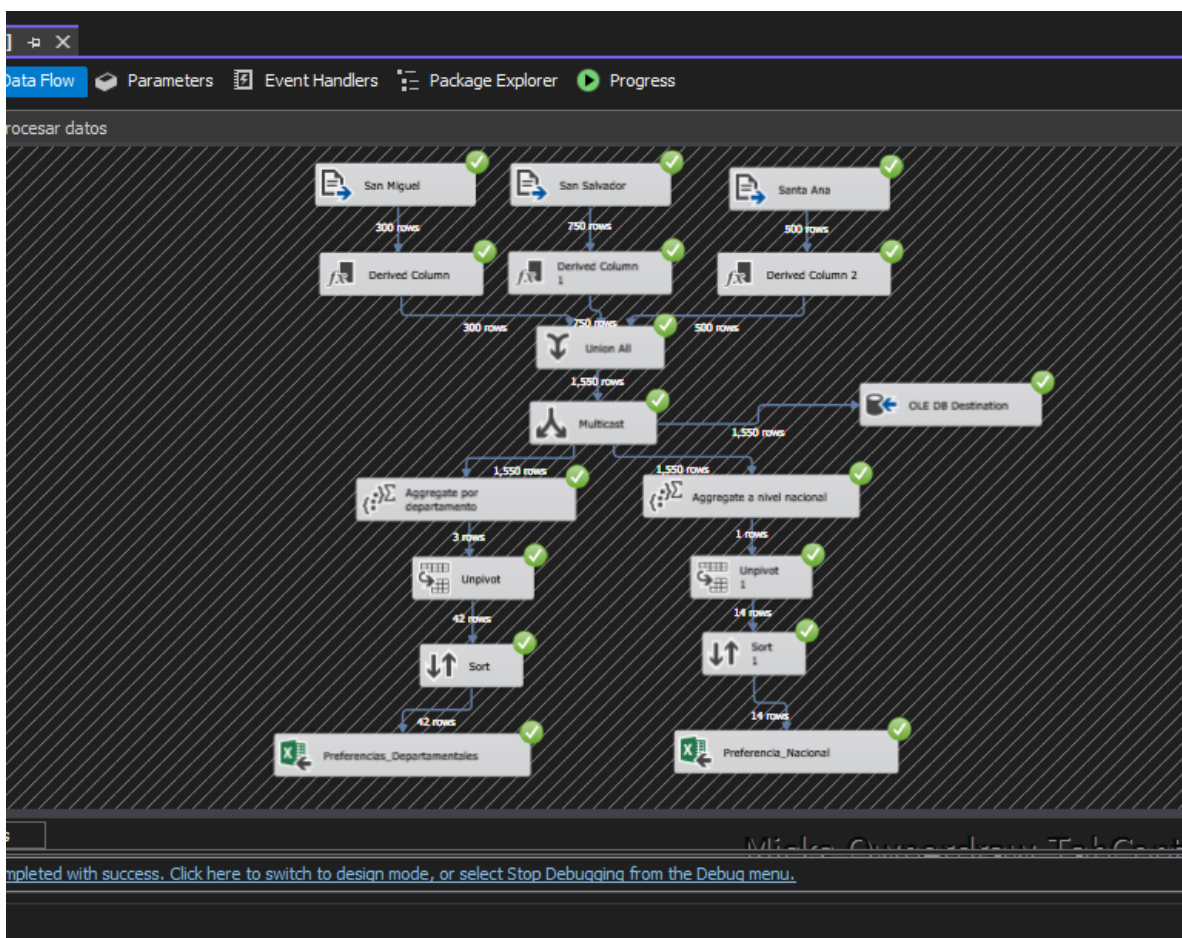
Table or view

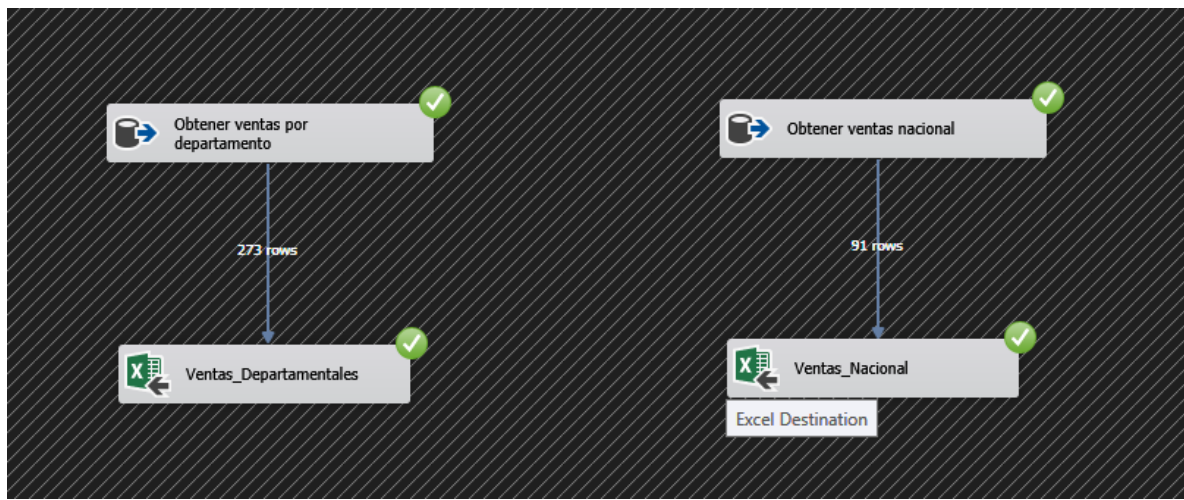
Name of the Excel sheet:

Hoja1\$

New...

Ejecución exitosa





	A	B	C
1	Producto	Cantidad	
2	Liston	975	
3	Rosas	945	
4	Globos	892	
5	Aurora	804	
6	Orquideas	797	
7	Lirios	795	
8	Girasoles	787	
9	Tarjetas	779	
10	Macetas	778	
11	Hortensia	774	
12	Tulipanes	753	
13	Carmesi	747	
14	Tierra	745	
15	Claveles	733	
16			

	A	B	C	D
1	Producto	Cantidad	Departamento	
2	Aurora	160	San Miguel	
3	Lirios	160	San Miguel	
4	Carmesi	158	San Miguel	
5	Orquideas	158	San Miguel	
6	Rosas	157	San Miguel	
7	Hortensia	157	San Miguel	
8	Globos	151	San Miguel	
9	Girasoles	150	San Miguel	
10	Tulipanes	149	San Miguel	
11	Liston	149	San Miguel	
12	Tarjetas	143	San Miguel	
13	Macetas	141	San Miguel	
14	Tierra	141	San Miguel	
15	Claveles	137	San Miguel	
16	Liston	690	San Salvador	
17	Rosas	612	San Salvador	
18	Globos	587	San Salvador	
19	Macetas	392	San Salvador	
20	Aurora	384	San Salvador	
21	Tarjetas	384	San Salvador	
22	Orquideas	380	San Salvador	
23	Hortensia	374	San Salvador	
24	Girasoles	371	San Salvador	
25	Tierra	368	San Salvador	
26	Lirios	365	San Salvador	
27	Tulipanes	357	San Salvador	
28	Carmesi	353	San Salvador	
29	Claveles	350	San Salvador	
30	Lirios	270	Santa Ana	
31	Girasoles	266	Santa Ana	
32	Aurora	260	Santa Ana	
33	Orquideas	259	Santa Ana	
34	Tarjetas	252	Santa Ana	
35	Tulipanes	247	Santa Ana	
36	Claveles	246	Santa Ana	

	A	B	C	D	E
1	Departamento	Producto1	Producto2	CantidadCompras	
2	Santa Ana	Orquideas	Rosas	90	
3	Santa Ana	Carmesi	Claveles	108	
4	San Miguel	Macetas	Rosas	69	
5	San Miguel	Carmesi	Girasoles	86	
6	San Salvador	Hortensia	Rosas	309	
7	San Miguel	Orquideas	Tierra	74	
8	Santa Ana	Carmesi	Tulipanes	117	
9	San Miguel	Hortensia	Orquideas	83	
10	Santa Ana	Carmesi	Lirios	128	
11	San Miguel	Aurora	Tulipanes	80	
12	Santa Ana	Claveles	Hortensia	129	
13	Santa Ana	Lirios	Tierra	123	
14	Santa Ana	Macetas	Tarjetas	131	
15	San Salvador	Lirios	Macetas	205	
16	San Miguel	Lirios	Tulipanes	79	
17	San Salvador	Macetas	Tierra	194	
18	San Salvador	Aurora	Tierra	188	
19	San Salvador	Liston	Tulipanes	320	
20	San Salvador	Globos	Macetas	310	
21	San Miguel	Hortensia	Rosas	83	
22	San Miguel	Aurora	Rosas	90	
23	San Miguel	Rosas	Tulipanes	74	
24	San Salvador	Carmesi	Macetas	182	
25	Santa Ana	Aurora	Tarjetas	145	
26	San Miguel	Aurora	Claveles	78	
27	San Miguel	Claveles	Tierra	61	
28	Santa Ana	Claveles	Tarjetas	120	
29	San Miguel	Aurora	Globos	71	
30	San Miguel	Aurora	Lirios	77	
31	San Salvador	Globos	Hortensia	301	
32	San Salvador	Orquideas	Tarjetas	198	
33	San Miguel	Claveles	Lirios	79	
34	San Salvador	Claveles	Orquideas	169	
35	Santa Ana	Carmesi	Rosas	82	
36	Santa Ana	Liston	Rosas	53	

Producto 1	Producto2	CantidadCompras
Claveles	Liston	463
Claveles	Globos	410
Claveles	Rosas	443
Macetas	Tarjetas	394
Claveles	Lirios	386
Aurora	Claveles	368
Liston	Tierra	468
Lirios	Liston	477
Orquideas	Tierra	377
Globos	Tulipanes	427
Girasoles	Tierra	365
Claveles	Tierra	352
Girasoles	Orquideas	411
Rosas	Tarjetas	483
Carmesi	Hortensia	362
Aurora	Tierra	384