



## **DATAWAREHOUSE Y MINERÍA DE DATOS**

**CATEDRATICO KAREN MEDRANO**

### **DESAFIO 1**

<b>N°</b>	<b>ESTUDIANTES</b>	<b>CARNET</b>
1	MENA CORTEZ, BRYAN ANTONIO	MC211787
2	LOPEZ GUTIERREZ, CALEB VERENICE	LG211551

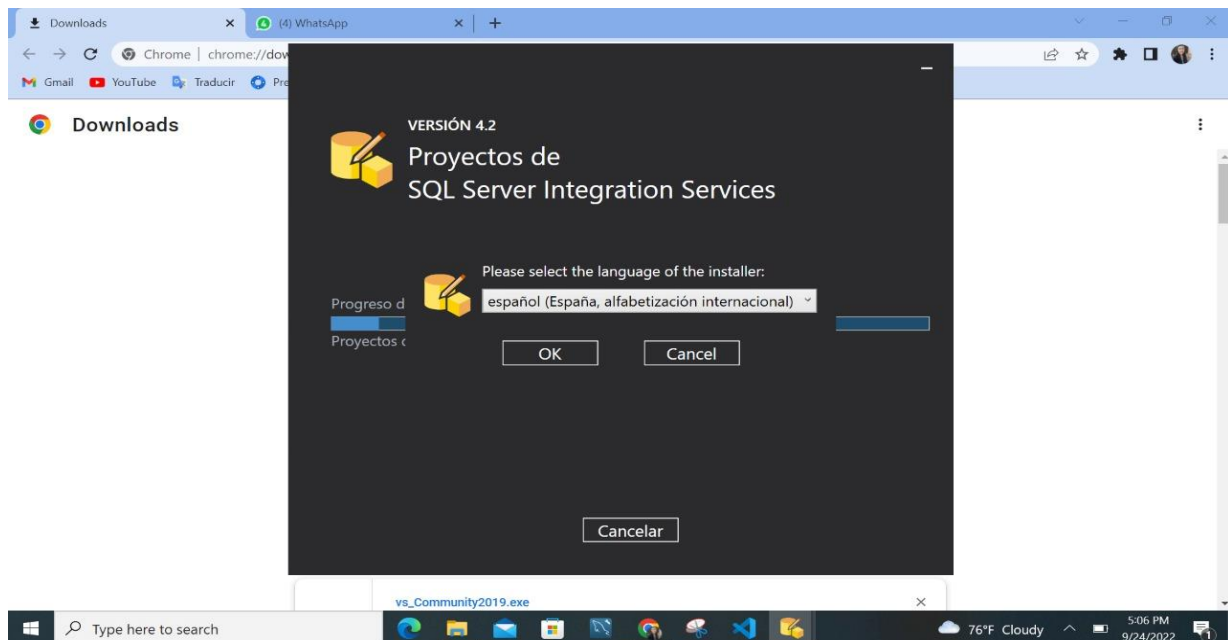
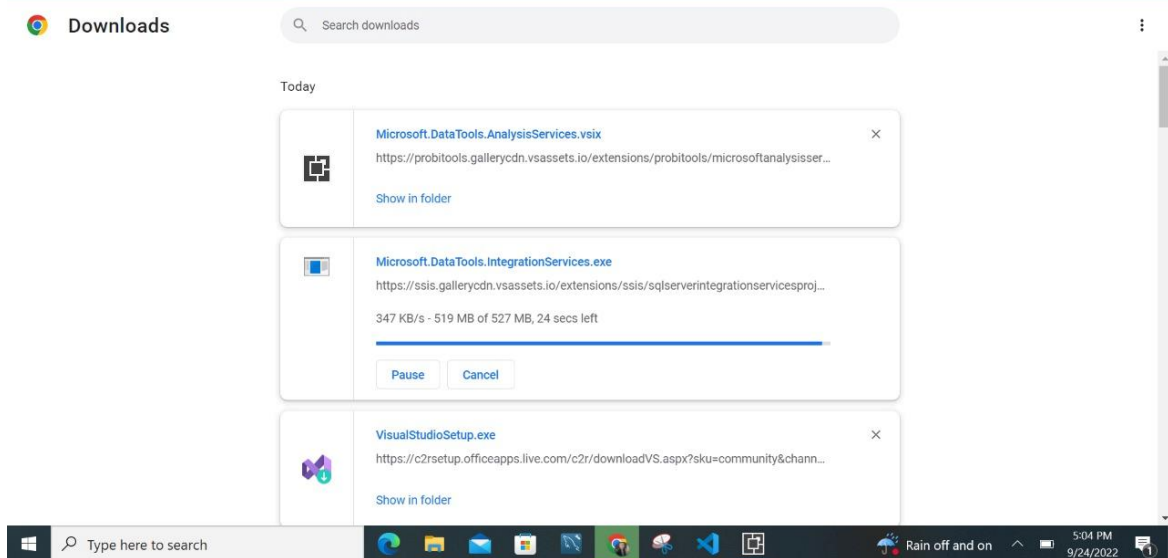
**24 DE SEPTIEMBRE DEL AÑO 2022**

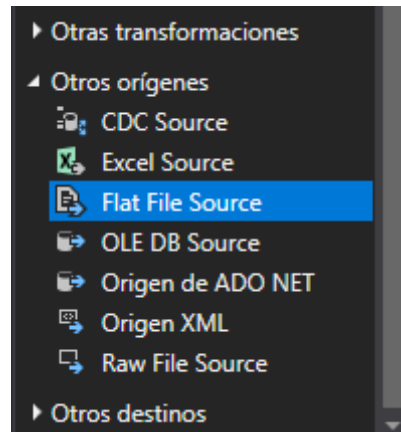
## Ejercicios:

Para los dos primeros ejercicios se comparten archivos csv, para el tercero son archivos sql, para ser cargados por medio de un ETL en la base de datos y ser analizados, se deben de compartir las consultas sql hechas para análisis de cada problema expuesto.

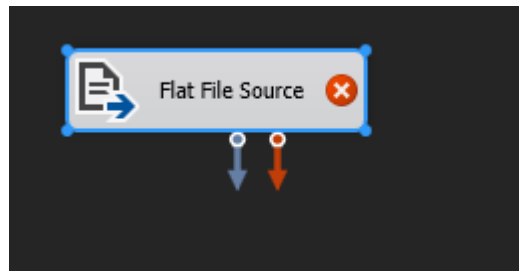
1. **(40%)** El Spa, "**Diego**", necesita segmentar sus clientes, para realizar una campaña de fidelización, y le pide a usted que efectué un análisis de sus tres sucursales, que defina cuantos grupos y que características tienen.

- Descargamos las extensiones:

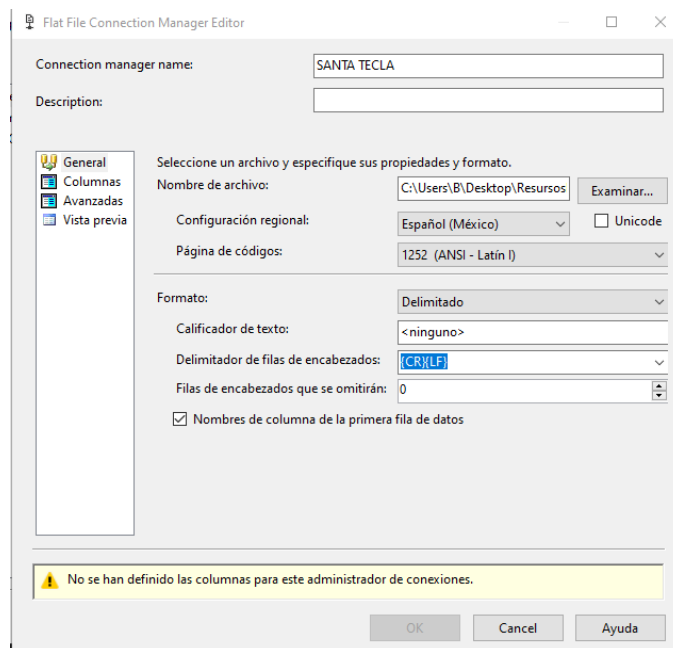




Seleccionamos uno de los tipos de herramientas de Toolbox, flat file Source y luego lo colocamos en el área de flat flow.



Se selecciona el Flat File Source y se ingresa a cargar los datos del archivo csv



Se llena los datos necesarios para realizar la conexión.

Flat File Connection Manager Editor

Connection manager name: SANTA TECLA

Description:

General  
Columnas  
Avanzadas  
Vista previa

Especifique los caracteres que delimitan el archivo de origen:

Delimitador de filas: {LF}

Delimitador de columnas: Coma {,}

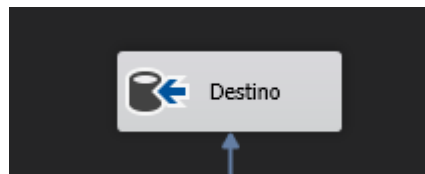
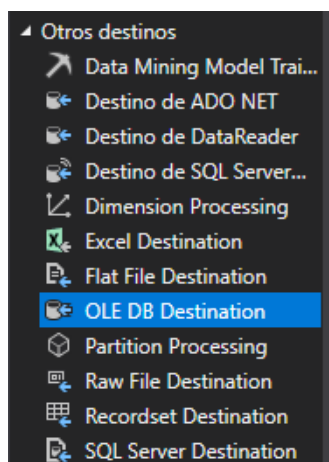
Vista previa de las filas 2-101:

id	Sexo	ingresos	PromVisit	Ec
Cynthy Crich	0	1175.45	6.88	5
Felide Biglin	0	2956.11	6.08	3
Carey Billington	0	2454.94	3.27	2
Correy Dello	0	1253.40	1.93	4
Simmonds Pennazzi	0	1153.21	2.94	4
Karylin Benit	0	2560.18	6.59	2
Bibbie Betjeman	0	2945.27	6.36	3

Actualizar Restablecer columnas

OK Cancel Ayuda

Se muestran los datos cargados del archivo csv haciendo clic en columnas



Agregamos OLE DB Destination

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:  
DESKTOP-3U20PBM\DESAFIO1.EJERCICIO1 New...

Data access mode:  
Table or view - fast load

Name of the table or the view:  
New...

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

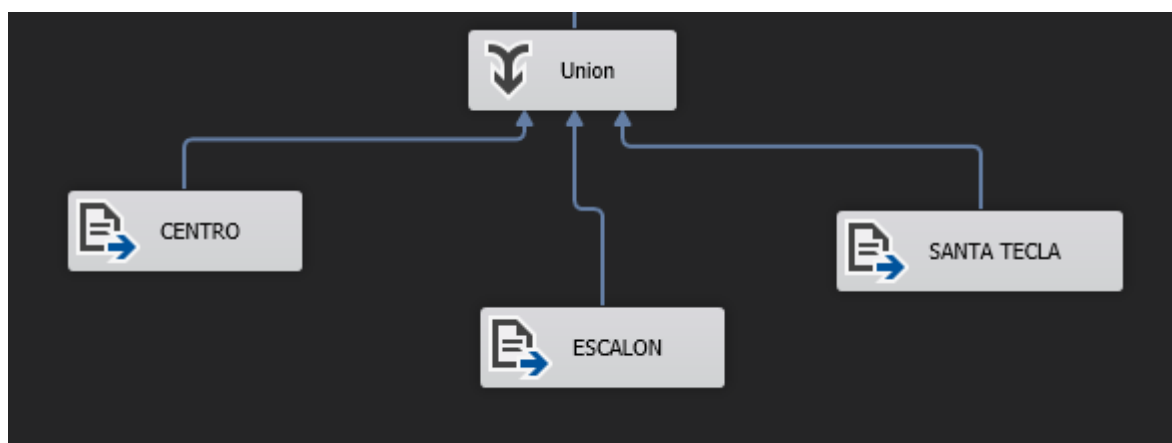
Rows per batch:   
Maximum insert commit size:

View Existing

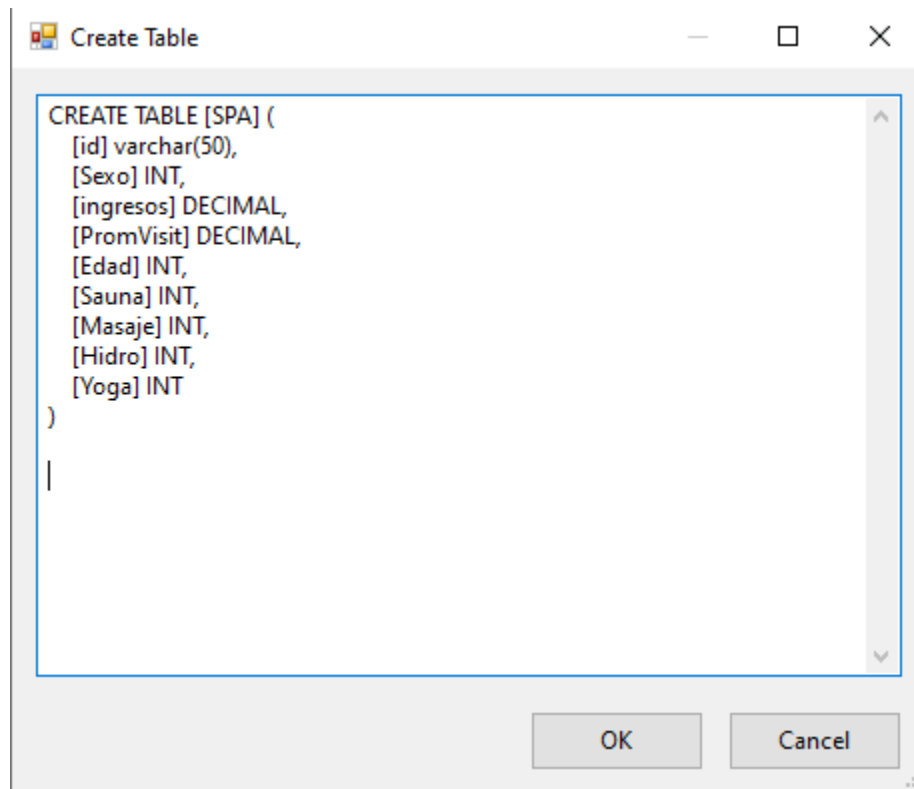
Select a table or view from the list.

OK Cancel Ayuda

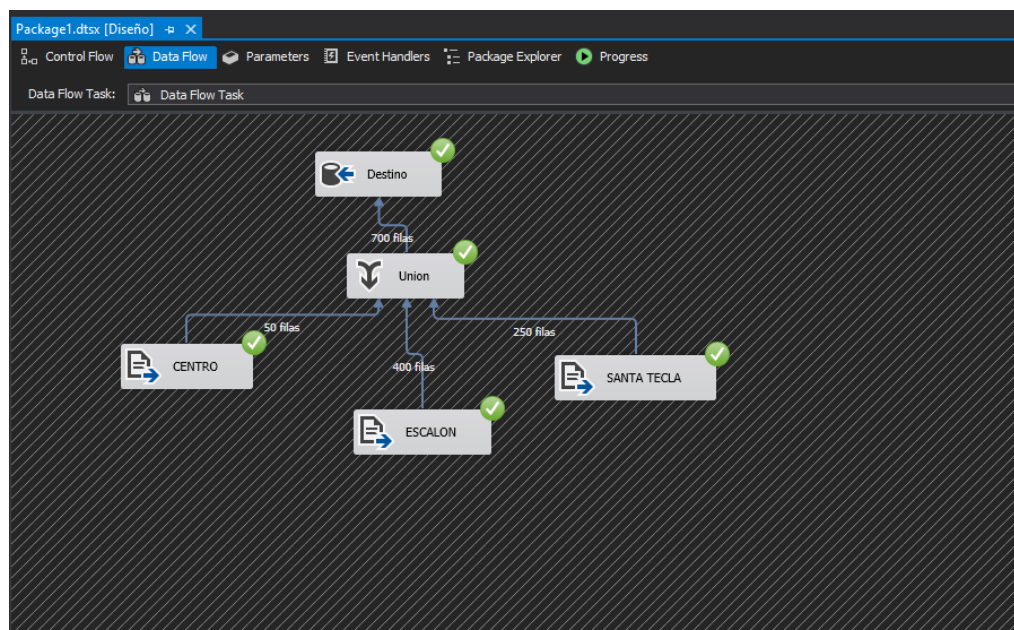
Configuramos el OLE DB Destination



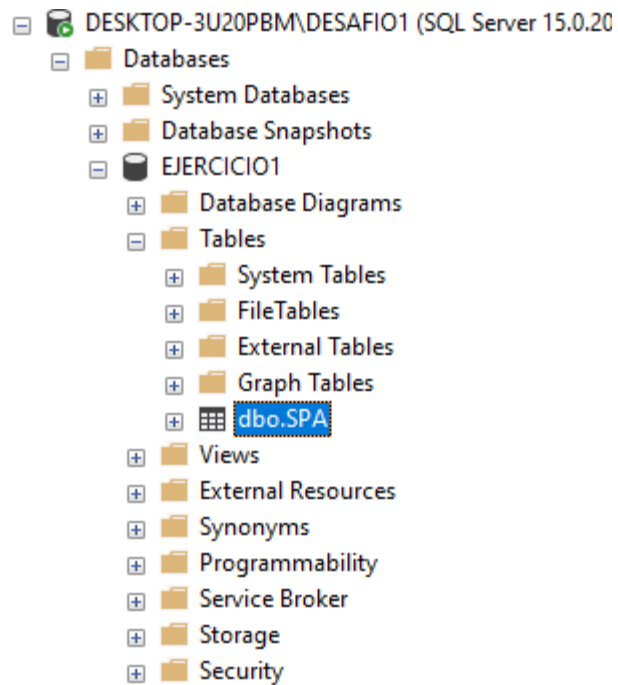
Visualizamos el seguimiento del trabajo en ETL



SQL para la creación de tabla



Visualizamos los resultados



Orden de los archivos en SQL Server y la tabla creada.

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```
SELECT avg(ingresos) as Promedio
FROM [EJERCICIO1].[dbo].[SPAFL0AT]
where Sexo=0
```

90 %

Results Messages

	Promedio
1	1654.69482323232

Consulta 1 del promedio

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```
SELECT avg(ingresos) as Promedio
FROM [EJERCICIO1].[dbo].[SPAFL0AT]
where Sexo=1
```

90 %

Results Messages

	Promedio
1	1712.43842105263

El promedio de los ingresos del sexo 1 son mayores.

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```

SELECT avg(PromVisit) as Promedio_De_Visitas
FROM [EJERCICIO1].[dbo].[SPAFLOAT]
where Sexo=0

```

90 %

Results Messages

	Promedio_De_Visitas
1	3.92989898989899

---

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```

SELECT avg(PromVisit) as Promedio_De_Visitas
FROM [EJERCICIO1].[dbo].[SPAFLOAT]
where Sexo=1

```

90 %

Results Messages

	Promedio_De_Visitas
1	3.88924342105263

Los clientes de sexo 0 visitan el spa de manera más frecuente.

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```

SELECT sum(Sauna) as Sauna,
       sum(Masaje) as Masaje,
       sum(Hidro) as Hidro,
       sum(Yoga) as Yoga
FROM [EJERCICIO1].[dbo].[SPAFLOAT]

```

90 %

Results Messages

	Sauna	Masaje	Hidro	Yoga
1	284	366	271	361

El servicio más pedido son los masajes, seguidos de yoga, después sauna y el menos usado es el Hidro.

SQLQuery24.sql -...OP-3U20PBM\B (61))\*

```

SELECT avg(Edad) as Edad
FROM [EJERCICIO1].[dbo].[SPAFLOAT]
where Sexo=0

```

90 %

Results Messages

	Edad
1	41



```

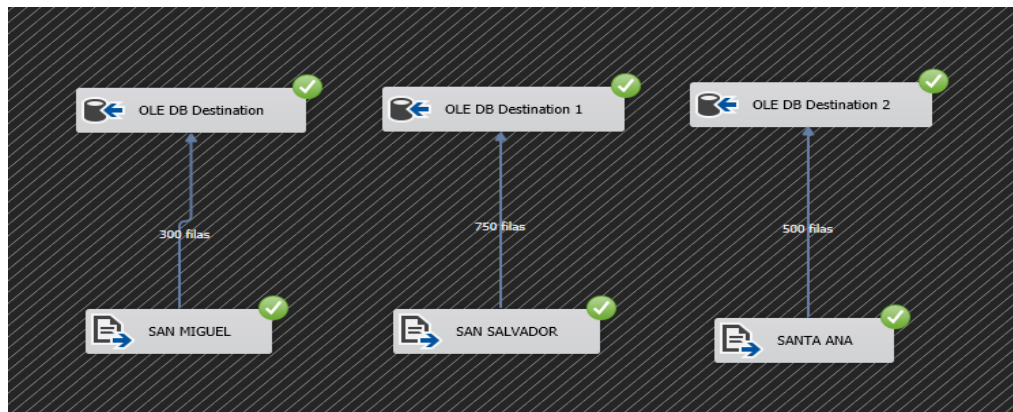
SQLQuery24.sql -...OP-3U20PBM\B (61)) *
SELECT avg(Edad) as Edad
FROM [EJERCICIO1].[dbo].[SPAFLOAT]
where Sexo=1

```

Edad
43

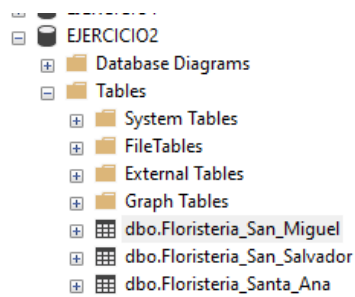
El promedio de edad de los clientes con sexo 0 es de 41 y los clientes de sexo 1 es de 43.

2. **(40%)** La Floristería "**Fiorella**" quiere saber cómo se compran sus productos, y tiene la data de tres departamentos del país, por lo cual les pide su opinión sobre qué productos sobresalen, que combinaciones son mejores y quieren este estudio por departamento y también por país.



Comenzaremos haciendo el análisis por departamento, por lo que aún no uniremos todos los datos, eso lo dejaremos para más adelante, cuando nos toque analizar los datos como país.

En el ejercicio numero 2 comenzamos exportando cada csv como una tabla en nuestra nueva base de datos llamada "Ejercicio2"



Al ejecutar el paquete en visual studio, ya podremos manipular las tablas en SQL Server.

SQLQuery30.sql -...OP-3U20PBM\B (54))\*

```

SELECT
    sum(Rosas) as Rosas
    ,sum(Claveles) as Claveles
    ,sum(Macetas) as Macetas
    ,sum(Tierra) as Tierra
    ,sum(Girasoles) as Girasoles
    ,sum(Hortensia) as Hortensia
    ,sum(Globos) as Globos
    ,sum(Tarjetas) as Tarjetas
    ,sum(Orquideas) as Orquideas
    ,sum(Carmesi) as Carmesi
    ,sum(Lirios) as Lirios
    ,sum(Aurora) as Aurora
    ,sum(Tulipanes) as Tulipanes
    ,sum(Liston) as Liston
FROM [EJERCICIO2].[dbo].[Floristeria_San_Miguel]

```

90 %

Results Messages

	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquideas	Carmesi	Lirios	Aurora	Tulipanes	Liston
1	157	137	141	141	150	157	151	143	158	158	160	160	149	149

Con esta consulta podemos observar que en San miguel los productos más vendidos son los Lirios y Aurora, los menos vendidos son las Macetas y la tierra, por lo que se podrían vender combos de donde se combinen los Lirios con Macetas y las Aurora con Tierra y así mover los productos menos vendidos.

SQLQuery30.sql -...OP-3U20PBM\B (54))\*

```

SELECT
    sum(Rosas) as Rosas
    ,sum(Claveles) as Claveles
    ,sum(Macetas) as Macetas
    ,sum(Tierra) as Tierra
    ,sum(Girasoles) as Girasoles
    ,sum(Hortensia) as Hortensia
    ,sum(Globos) as Globos
    ,sum(Tarjetas) as Tarjetas
    ,sum(Orquideas) as Orquideas
    ,sum(Carmesi) as Carmesi
    ,sum(Lirios) as Lirios
    ,sum(Aurora) as Aurora
    ,sum(Tulipanes) as Tulipanes
    ,sum(Liston) as Liston
FROM [EJERCICIO2].[dbo].[Floristeria_San_Salvador]

```

90 %

Results Messages

	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquideas	Carmesi	Lirios	Aurora	Tulipanes	Liston
1	612	350	392	368	371	374	587	384	380	353	365	384	357	690

Esta consulta es de San Salvador y aquí los productos más vendidos son el Listón, las Rosas y los globos, los menos vendidos son los Claveles y el Carmesí, por lo que los combos pueden ser arreglos que mezclen Rosas y claveles, listones y carmesí.

SQLQuery30.sql -...OP-3U20PBM\B (54))\*

```

SELECT
    sum(Rosas) as Rosas
    ,sum(Claveles) as Claveles
    ,sum(Macetas) as Macetas
    ,sum(Tierra) as Tierra
    ,sum(Girasoles) as Girasoles
    ,sum(Hortensia) as Hortensia
    ,sum(Globos) as Globos
    ,sum(Tarjetas) as Tarjetas
    ,sum(Orquideas) as Orquideas
    ,sum(Carmesi) as Carmesi
    ,sum(Lirios) as Lirios
    ,sum(Aurora) as Aurora
    ,sum(Tulipanes) as Tulipanes
    ,sum(Liston) as Liston
FROM [EJERCICIO2].[dbo].[Floristeria_Santa_Ana]

```

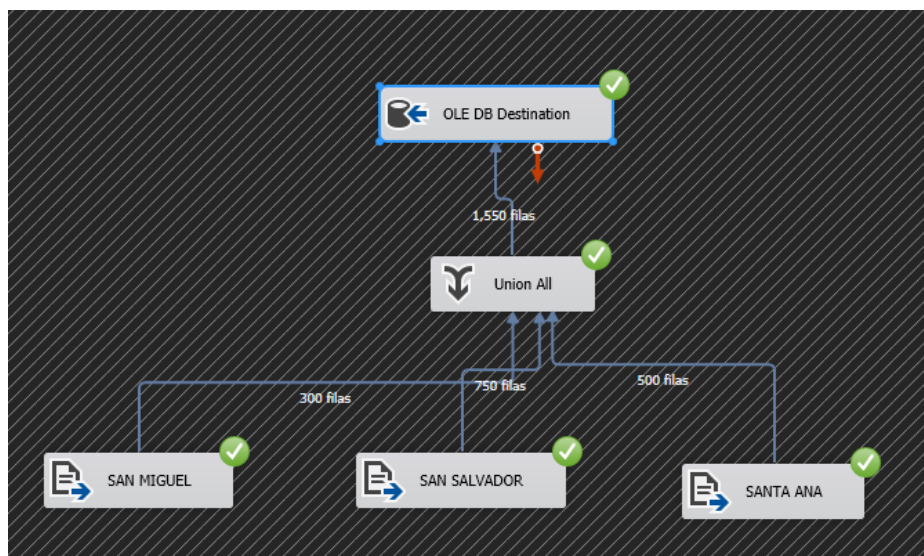
90 %

Results Messages

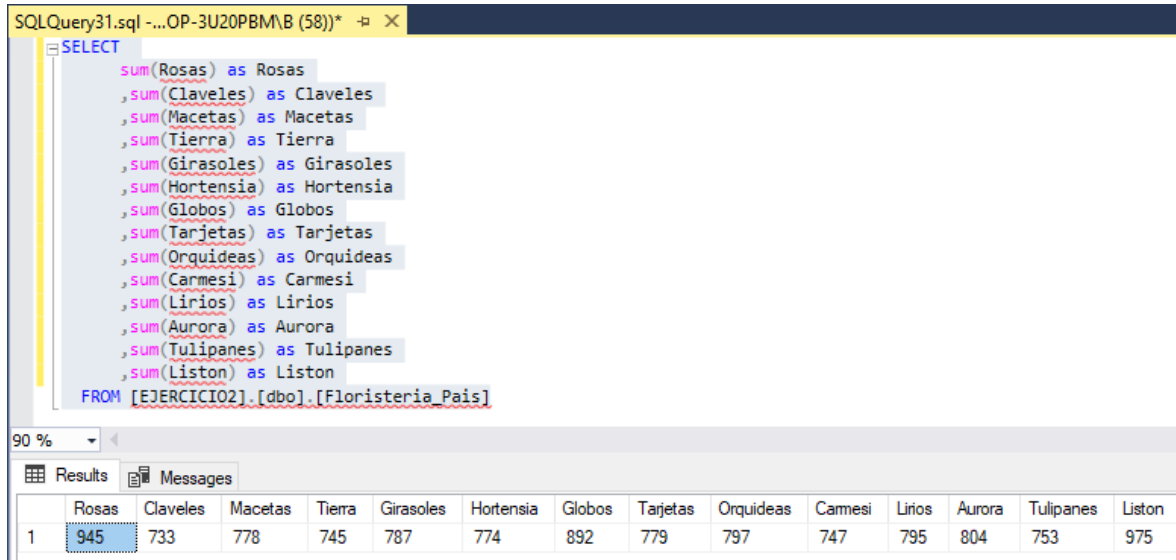
	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquideas	Carmesi	Lirios	Aurora	Tulipanes	Liston
1	176	246	245	236	266	243	154	252	259	236	270	260	247	136

En la consulta de Santa Ana observamos que los productos más vendidos son los Lirios y las Auroras, los menos vendidos son Listón y los Globos, los combos podrían ser “por la compra de Lirios o Auroras los globos y los listones con 50% de descuento”.

Para poder análisis los datos de todo el país como un conjunto, lo que haremos será desde visual studio unirlos y generar una nueva tabla donde ya podremos hacer consultas a nivel de país, el esquema nos queda de la siguiente manera:



Teniendo la tabla de país, hacemos la suma de cada producto vendido y obtenemos los siguientes datos:



The screenshot shows a SQL query window with the following text:

```
SELECT
    sum(Rosas) as Rosas
    ,sum(Claveles) as Claveles
    ,sum(Macetas) as Macetas
    ,sum(Tierra) as Tierra
    ,sum(Girasoles) as Girasoles
    ,sum(Hortensia) as Hortensia
    ,sum(Globos) as Globos
    ,sum(Tarjetas) as Tarjetas
    ,sum(Orquideas) as Orquideas
    ,sum(Carmesi) as Carmesi
    ,sum(Lirios) as Lirios
    ,sum(Aurora) as Aurora
    ,sum(Tulipanes) as Tulipanes
    ,sum(Liston) as Liston
FROM [EJERCICIO2].[dbo].[Floristeria_Pais]
```

Below the query window, the 'Results' tab is active, displaying a table with 15 columns and 1 row. The first column is empty, and the subsequent columns contain the sum of each product.

	Rosas	Claveles	Macetas	Tierra	Girasoles	Hortensia	Globos	Tarjetas	Orquideas	Carmesi	Lirios	Aurora	Tulipanes	Liston
1	945	733	778	745	787	774	892	779	797	747	795	804	753	975

A nivel de País los productos más vendidos son el Listón y las rosas, los menos vendidos son los Clavel y el Carmesí, por lo tanto, se podrían combinar dichos productos y mejorar las ventas además de rotar el inventario que menos se vende.