

UNIVERSIDAD DON BOSCO  
FACULTAD DE INGENIERÍA  
ESCUELA DE COMPUTACIÓN



Datawarehouse y Minería de Datos

Segundo desafío práctico

Docente: Ing. Karens Medrano

Estudiante:

Kallahan Andrea Salas Bojorquez | SB210537

Martes 06 de diciembre de 2022

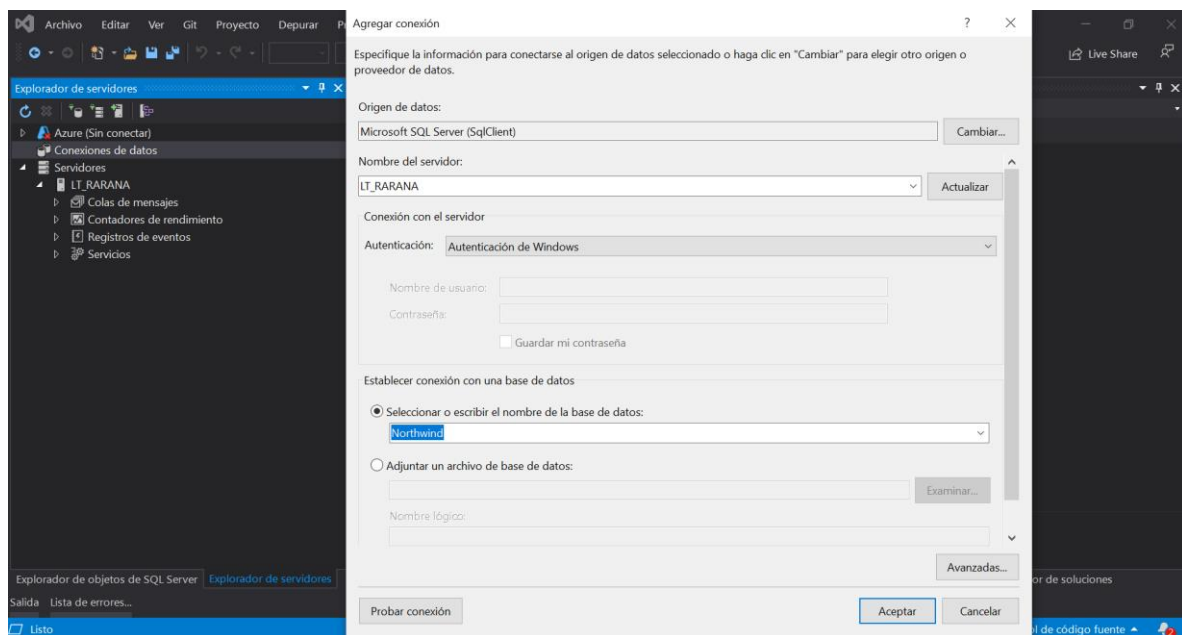
## Ejercicio N1

### 1. Análisis multidimensional OLAP (utilizando la base de datos “Northwind”).

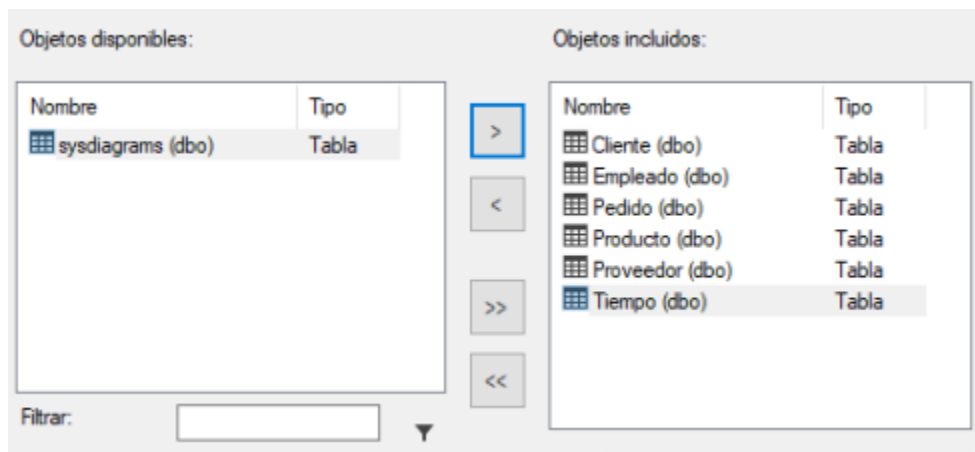
Para iniciar con este ejercicio, debemos definir que vamos a trabajar con la base de datos “Northwind” que se nos ha compartido para poder desarrollar el desafío

Utilizaremos la herramienta de “Proyecto multidimensional y de minería de datos Analysis Services”, para poder dar inicio a la creación de nuestro cubo.

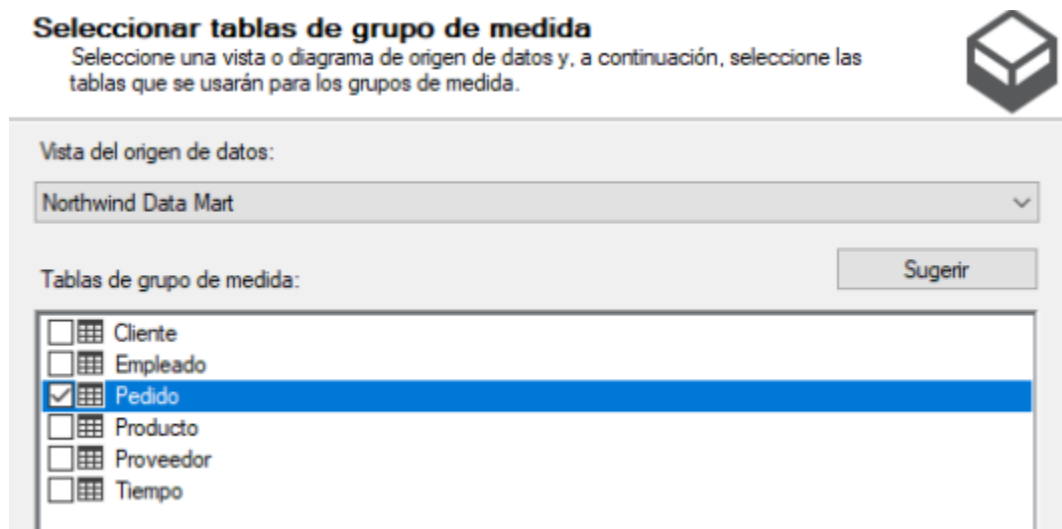
Necesitamos crear la conexión hacia SQL con la base de datos. Esta configuración quedaría de la siguiente manera:



Ahora como segundo punto, seleccionamos las tablas: Cliente, Empleado, Producto, Pedido, Proveedor y Tiempo”, para incluirlos en objetos. Esto es para crear las vistas en la conexión.



Y procedemos a crear el cubo eligiendo la tabla: “pedido”, porque esta es la que conectará a las demás tablas



### Seleccionar medidas


Seleccione las medidas que desea incluir en el cubo.




☐ Medida

☒  Pedido

☒  Precio

☒  Cantidad

☐  Recuento Pedido

Y hay que elegir las dimensiones siguientes:

### Seleccionar nuevas dimensiones

Seleccione las nuevas dimensiones que desea crear, según las tablas disponibles.



☒ Dimensión

☒  Proveedor

☒  Proveedor

☒  Tiempo

☒  Tiempo

☒  Empleado

☒  Empleado

☒  Cliente

☒  Cliente

☒  Producto

☒  Producto

Y solamente queda confirmar la medida y las dimensiones nuevamente

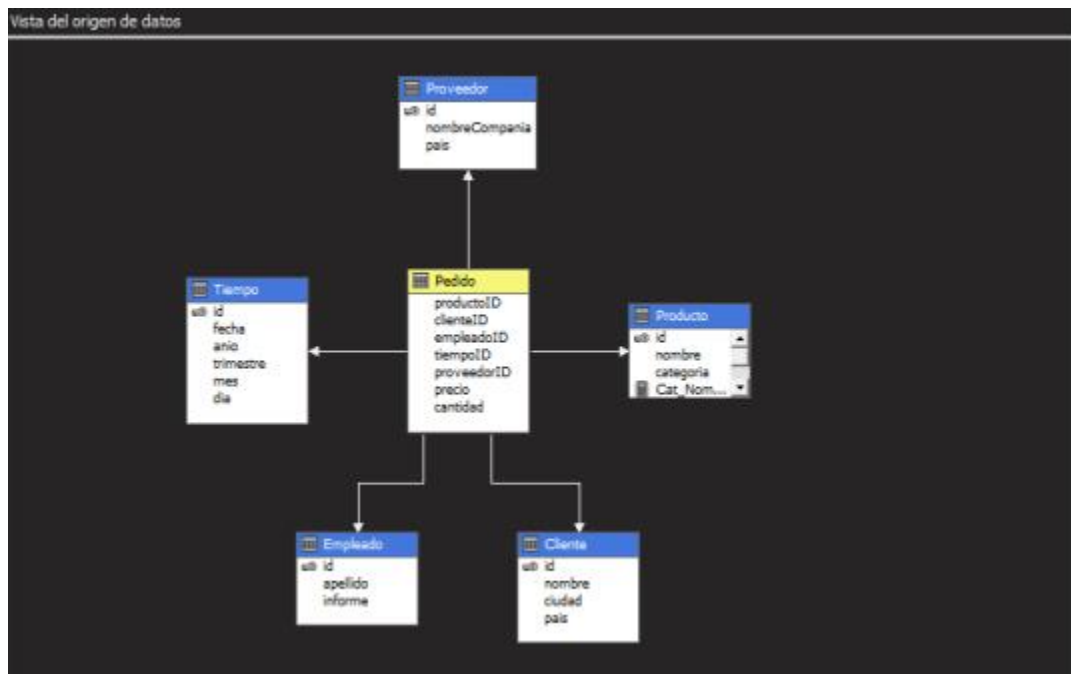
**Finalización del asistente**  
Asigne un nombre al cubo, revise su estructura y, a continuación, haga clic en Finalizar para guardar el cubo.

Nombre del cubo:

Vista previa:

- Grupos de medida
  - Pedido
    - Precio
    - Cantidad
- Dimensiones
  - Proveedor
  - Tiempo
  - Empleado
  - Cliente
  - Producto

El diagrama modelo estrella, queda de la siguiente manera:



Vamos a agregar los atributos a cada dimensión para poder recorrer los datos incluyendo “País Proveedor, Apellido Empleado, Nombre y País del cliente, ID y categoría de producto

Proveedor
id
nombreCompania
pais

Atributos
Proveedor
Id
Pais

Empleado
id
apellido
informe

Atributos
Empleado
Apellido
Id

Cliente
id
nombre
ciudad
pais

Atributos
Cliente
Id
Nombre
Pais

Producto
id
nombre
categoria
Cat_Nombre

Atributos
Producto
Categoria
Id
Nombre

Ahora podemos visualizar los datos deseados (Nombre, país, apellido, categoría, y de donde proviene)

The screenshot shows the 'Northwind Data Mart' application. On the left, a 'Metadatos' pane lists dimensions: Northwind Data Mart, Measures, Pedido, Cantidad, Precio, KPI, Cliente, Empleado, Producto, Proveedor, Id, Pais, and Tiempo. The main area displays a table with columns: Dimensión, Jerarquía, Operador, Expresión de filtro, and Parámetros. Below this, a data table is shown with columns: Nombre, Pais, Apellido, Nombre, Categoría, Pais, and Cantidad. The data table contains 20 rows of sales records.

Nombre	Pais	Apellido	Nombre	Categoría	Pais	Cantidad
Alejandra Camino	Spain	Davolio	Nord-Ost Matjeshering	Seafood	Germany	10
Alejandra Camino	Spain	Davolio	Perth Pasties	Meat/Poultry	Australia	10
Alejandra Camino	Spain	Davolio	Tourtière	Meat/Poultry	Canada	5
Alejandra Camino	Spain	Peacock	Camembert Pierrot	Dairy Products	France	10
Alejandra Camino	Spain	Peacock	Guaraná Fantástica	Beverages	Brazil	6
Alejandra Camino	Spain	Peacock	Nord-Ost Matjeshering	Seafood	Germany	7
Alejandra Camino	Spain	Peacock	Ravioli Angelo	Grains/Cereals	Italy	2
Alejandra Camino	Spain	Peacock	Steeleye Stout	Beverages	USA	4
Alejandra Camino	Spain	Peacock	Teatime Chocolate Bisc...	Confections	UK	1
Alexander Feuer	Germany	Buchanan	Lakkalkööri	Beverages	Finland	10
Alexander Feuer	Germany	Buchanan	Mozzarella di Giovanni	Dairy Products	Italy	30
Alexander Feuer	Germany	Buchanan	Raclette Courdavault	Dairy Products	France	12
Alexander Feuer	Germany	Buchanan	Vegie-spread	Condiments	Australia	6
Alexander Feuer	Germany	Leverling	Pavlova	Confections	Australia	20
Alexander Feuer	Germany	Leverling	Tarte au sucre	Confections	Canada	20
Alexander Feuer	Germany	Leverling	Zaanse koeken	Confections	Netherl...	12
Alexander Feuer	Germany	Peacock	Gorgonzola Telino	Dairy Products	Italy	10
Alexander Feuer	Germany	Peacock	Konbu	Seafood	Japan	20

Agregamos un campo en la tabla "Pedido" para concatenar "categoría" con el nombre

The screenshot shows the 'Crear cálculo con nombre' (Create named calculation) dialog box. It has three main sections: 'Nombre de columna:' with the value 'Cat\_Nombre', 'Descripción:' with an empty text area, and 'Expresión:' with the formula 'categoria+', '+nombre'. At the bottom are buttons for 'Aceptar', 'Cancelar', and 'Ayuda'.

Nombre de columna: Cat\_Nombre

Descripción:

Expresión: categoria+, '+nombre

Aceptar Cancelar Ayuda

Y la vista de la tabla queda así:

Tabla					
id	nombre	categoría	Cat_Nombre		
1	Alice Mutton	Meat/Poultry	Meat/Poultry, Alice Mutton		
2	Aniseed Syrup	Condiments	Condiments, Aniseed Syrup		
3	Boston Crab Meat	Seafood	Seafood, Boston Crab Meat		
4	Camembert Pierrot	Dairy Products	Dairy Products, Camembert Pierrot		
5	Carnarvon Tigers	Seafood	Seafood, Carnarvon Tigers		
6	Chai	Beverages	Beverages, Chai		
7	Chang	Beverages	Beverages, Chang		
8	Chartreuse verte	Beverages	Beverages, Chartreuse verte		
9	Chef Anton's Cajun Seasoning	Condiments	Condiments, Chef Anton's Cajun Seasoning		
10	Chef Anton's Gumbo Mix	Condiments	Condiments, Chef Anton's Gumbo Mix		
11	Chocolade	Confections	Confections, Chocolade		
12	Côte de Blaye	Beverages	Beverages, Côte de Blaye		
13	Escargots de Bourgogne	Seafood	Seafood, Escargots de Bourgogne		
14	Filo Mix	Grains/Cereals	Grains/Cereals, Filo Mix		
15	Flotemysost	Dairy Products	Dairy Products, Flotemysost		
16	Geitost	Dairy Products	Dairy Products, Geitost		
17	Genen Shouyu	Condiments	Condiments, Genen Shouyu		
18	Gnocchi di nonna Alice	Grains/Cereals	Grains/Cereals, Gnocchi di nonna Alice		
19	Gorgonzola Telino	Dairy Products	Dairy Products, Gorgonzola Telino		
20	Grandma's Boysenberry Spread	Condiments	Condiments, Grandma's Boysenberry Spread		
21	Gravad lax	Seafood	Seafood, Gravad lax		
22	Guaraná Fantástica	Beverages	Beverages, Guaraná Fantástica		
23	Gudbrandsdalsost	Dairy Products	Dairy Products, Gudbrandsdalsost		
24	Gula Malacca	Condiments	Condiments, Gula Malacca		

Y con eso podemos concluir el ejercicio N1. Todo se ha realizado de manera exitosa.



## Ejercicio N2

### 2. Análisis del COVID-19 en El Salvador

Para este ejercicio, no tenemos una base de datos inicial. Entonces, necesitamos crearla.

```
C: > Users > Kallahan Andrea > Desktop > D2 > Covid19.sql
1  -- Covid 19 en El Salvador
2  -- SB210537
3  Create database covid19;
4
5  create table territorio(
6      id_zona int identity (1,1) primary key,
7      zona varchar (50)
8  )
9
10
11 create table fecha
12 (
13     id_fecha int identity (1,1) primary key,
14     fecha date,
15     mes varchar(15),
16     año int,
17     día int,
18     trimestre int,
19 )
20
21
22 create table contagios_porcentaje
23 (
24     id_contagios int identity (1,1) primary key,
25     parques_espaciosPublicos int,
26     tiendas_super_farmacia int,
27     comercio_recreacion int,
28     trabajo int,
29     transito int,
30     residencia int,
```

```
create table contagios
(
    id_contagios int identity (1,1) primary key,
    parques_espaciosPublicos int,
    tiendas_super_farmacia int,
    comercio_recreacion int,
    trabajo int,
    transito int,
    residencia int,
    id_fechas int foreign key references fecha(id_fecha),
    id_zonas int foreign key references territorio(id_zona),
)
```

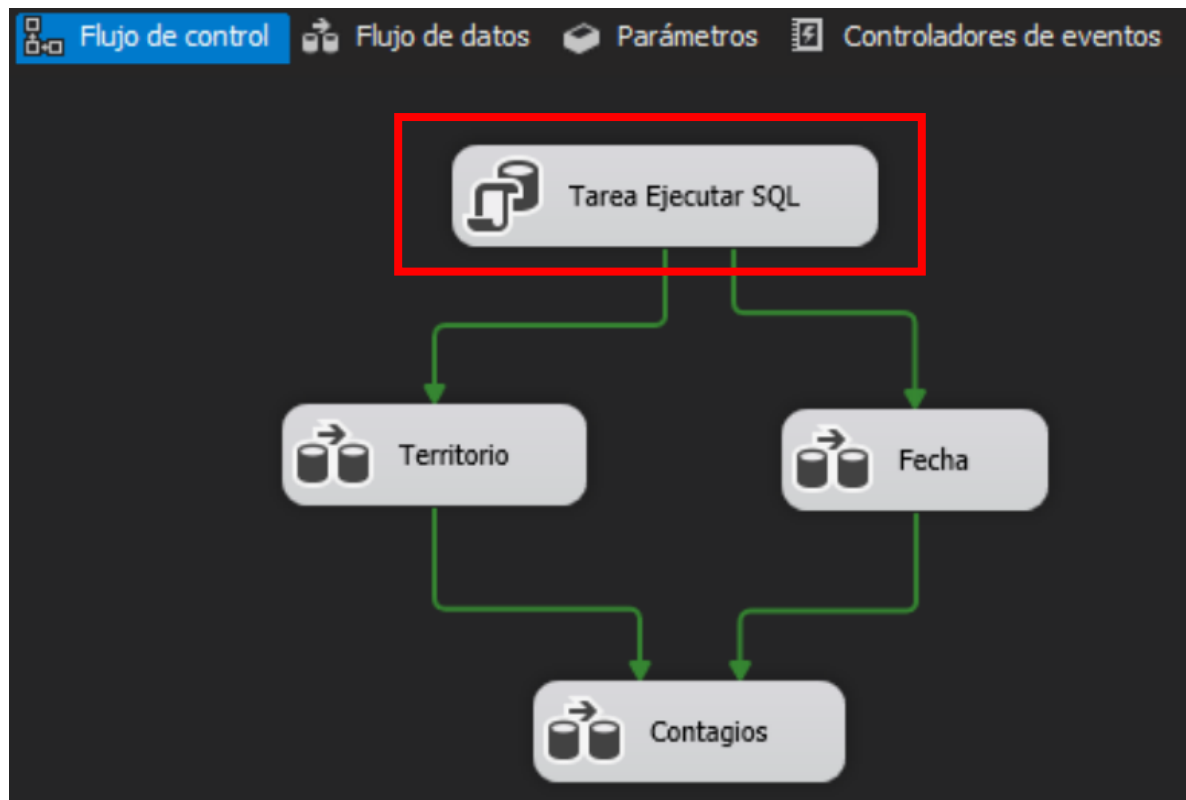
Y una vez creada la base de datos con las tablas, el diagrama queda de la siguiente forma:



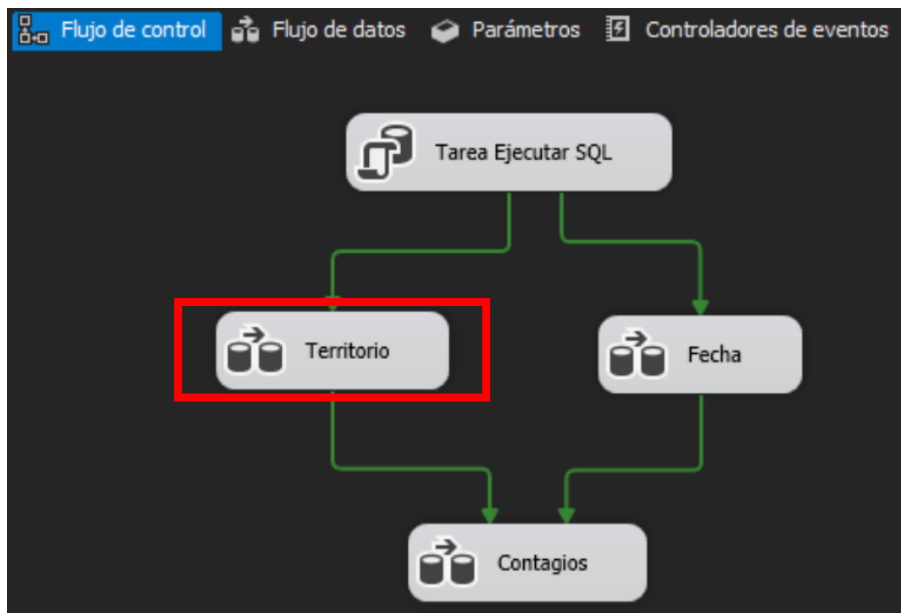
En base a esto, vamos a crear un proyecto en visual studio SSIS y vamos a crear el siguiente diagrama de flujo:



En cada uno de los elementos que se irán describiendo a continuación, se agregarán los siguientes valores en los controles. De esta manera llenaremos las tablas.



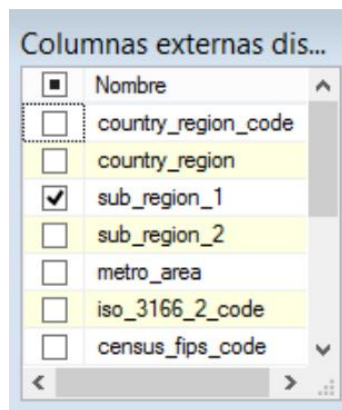
Conjunto de resultados	
ResultSet	Ninguno
General	
Name	Tarea Ejecutar SQL
Description	Tarea Ejecutar SQL
Instrucción SQL	
ConnectionType	OLE DB
Connection	
SQLSourceType	Entrada directa
SQLStatement	delete from contagios_porcentajesdelete from
IsQueryStoredProcedure	False
BypassPrepare	True
Opciones	
TimeOut	0
CodePage	1252
TypeConversionMode	Permitido



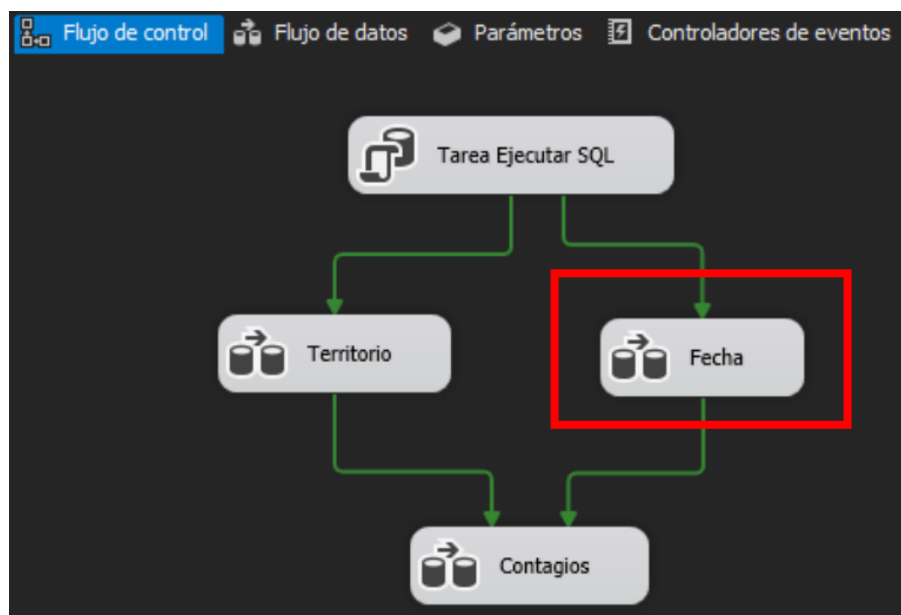
Para columna derivada:

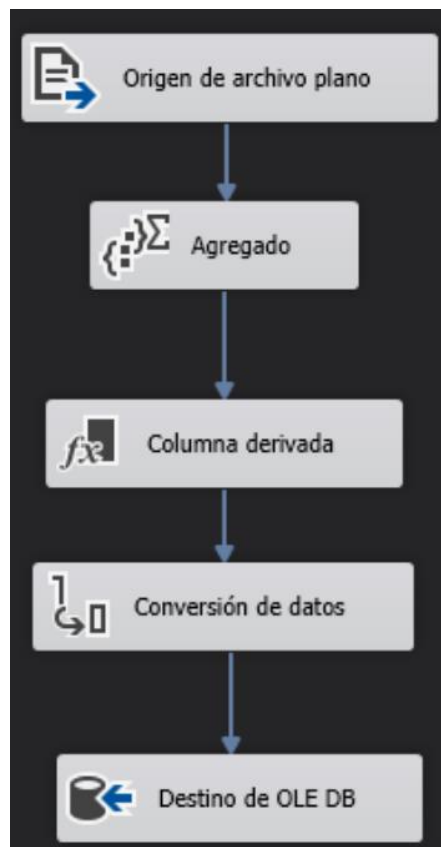
Nombre de columna d...	Columna derivada	Expresión	Tipo de datos
zona	<agregar como colum...	(DT_WSTR,50)sub_region_1 == "" ? "El Salvador" : sub_region_1	cadena Unicode [l

Para origen de archivo:

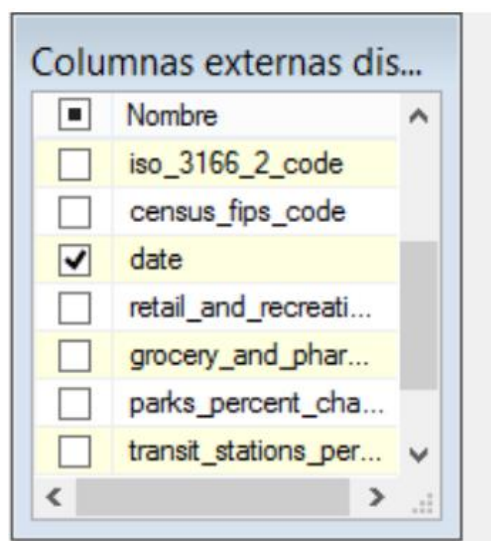


Elemento fecha:





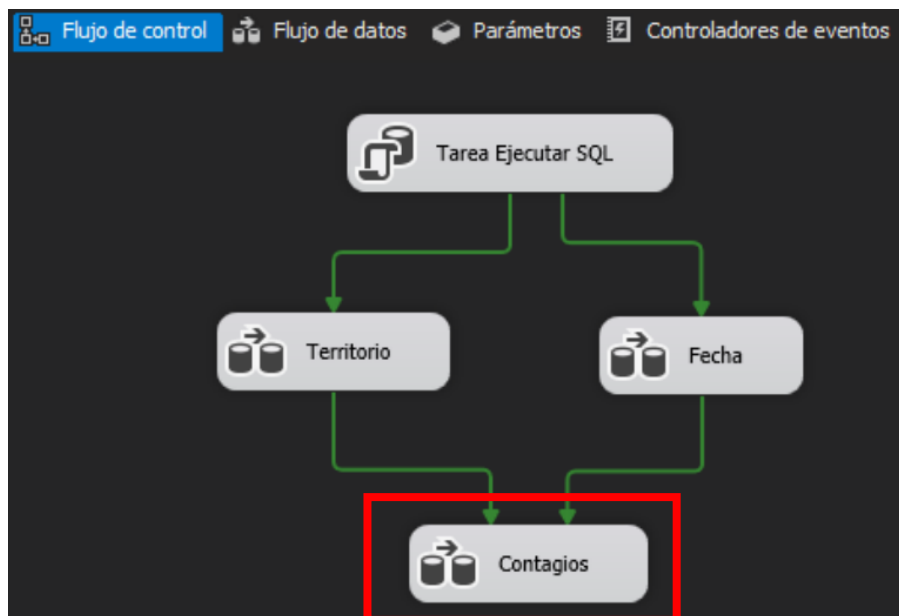
Origen del archivo:



Columna derivada:

Nombre de columna d...	Colu...	Expresión	Tipo de datos	Longitud
año	<agr...	DATEPART("YEAR",(DT_DBDATE)date)	entero de cuatro bytes ...	
trimestre	<agr...	DATEPART("QUARTER",(DT_DBDATE)date)	entero de cuatro bytes ...	
mes	<agr...	(DT_WSTR,15)DATEPART("MONTH",(DT_DBDATE)...	cadena Unicode [DT_...	15
día	<agr...	DATEPART("DAY",(DT_DBDATE)date)	entero de cuatro bytes ...	
fecha_formateada	<agr...	(DT_DATE)date	fecha [DT_DATE]	

Entonces para el último elemento sería lo siguiente:



Procederemos a eliminar los registros de la información del .csv, para poder llenar exitosamente la tabla de relaciones. Agregaremos la sentencia siguiente para tomar solo los campos necesarios:

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

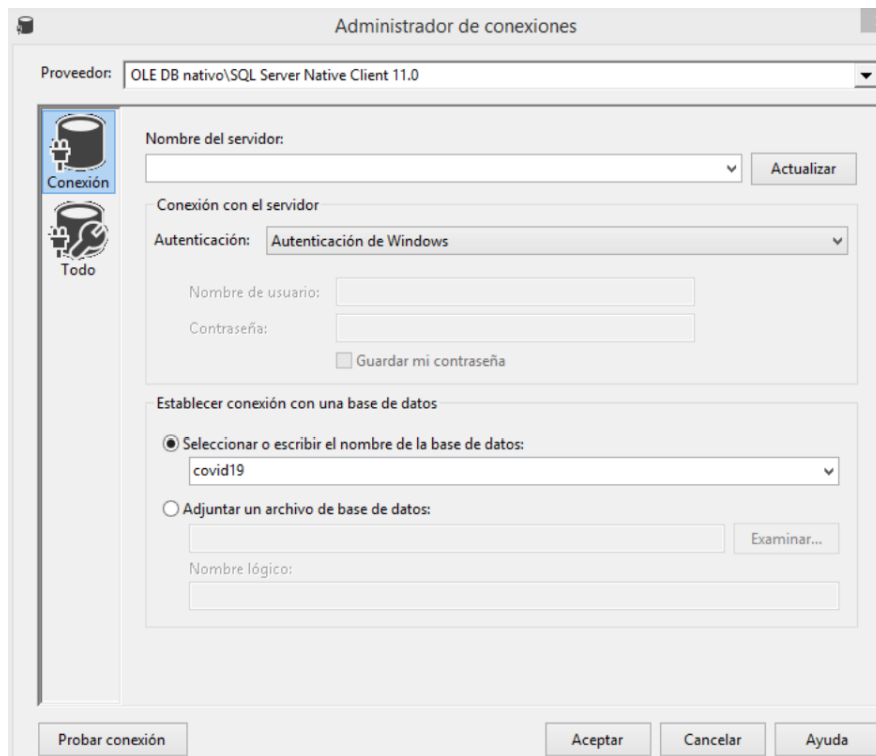
```
SELECT
dt.parks_percent_change_from_baseline,
dt.grocery_and_pharmacy_percent_change_from_baseline,
dt.retail_and_recreation_percent_change_from_baseline,
dt.transit_stations_percent_change_from_baseline,
dt.workplaces_percent_change_from_baseline,
dt.residential_percent_change_from_baseline,
t.id_zona, f.id_fecha
FROM covid.dbo.dato dt INNER JOIN covid19.dbo.fecha f ON
(dt.fecha=f.fecha)

INNER
JOIN covid19.dbo.territorio t ON(dt.sub_region_1=t.zona)
```

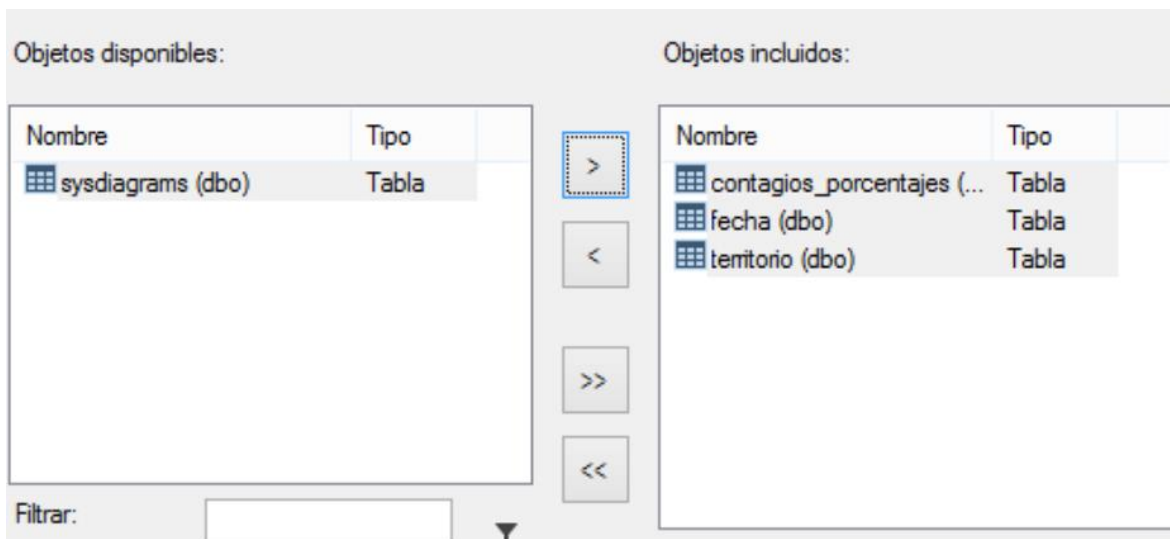
Columna de entrada	Columna de destino
<omitir>	id_contagios
parks_percent_change_from_baseline	parques_espaciospublicos
grocery_and_pharmacy_percent_change_fro...	tiendas_super_farmacia
retail_and_recreation_percent_change_from_...	comercio_recreacion
transit_stations_percent_change_from_baseline	transito
workplaces_percent_change_from_baseline	trabajo
residential_percent_change_from_baseline	residencia
id_zona	id_zonas
id_fecha	id_fechas



Y como siguiente punto, ejecutamos el proyecto y creamos el cubo para este ejercicio.



Agregamos las tablas necesarias



Y seleccionamos la tabla contagios\_porcentajes

Covid19

Tablas de grupo de medida:

☐ territorio

☐ fecha

☒ contagios\_porcentajes

Sugerir

Y sus medidas

### Seleccionar medidas

Seleccione las medidas que desea incluir en el cubo.

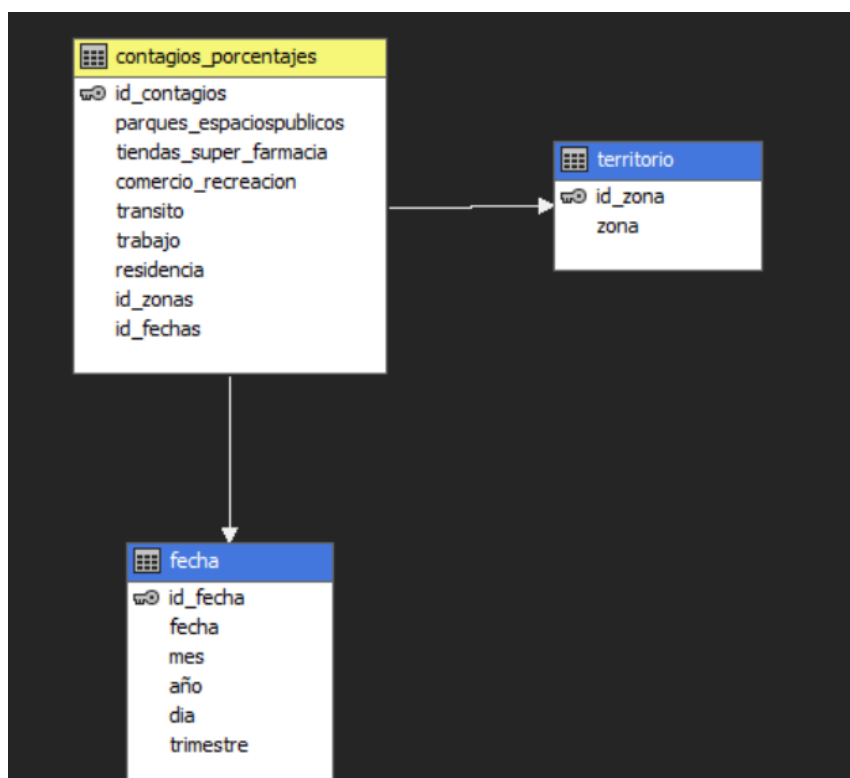


☒ Medida

☒ Contagios Porcentajes

- ☒ Parques Espaciospublicos
- ☒ Tiendas Super Farmacia
- ☒ Comercio Recreacion
- ☒ Transito
- ☒ Trabajo
- ☒ Residencia
- ☒ Recuento Contagios Porcentajes

Y la vista queda de la siguiente forma:



En la opción de “Calculo” en la parte de arriba, se convertirán los datos que ya están en valores en porcentaje. Que se vería de la siguiente manera

Nombre:

[Porcentaje Parques Espacios Publicos]

Propiedades del miembro primario

Jerarquía primaria: Measures

Miembro primario:

Expresión

[Measures].[Parques Espaciospublicos]/100

Propiedades adicionales

Cadena de formato: Percent

Visible: True

Comportamiento si no está vacío:

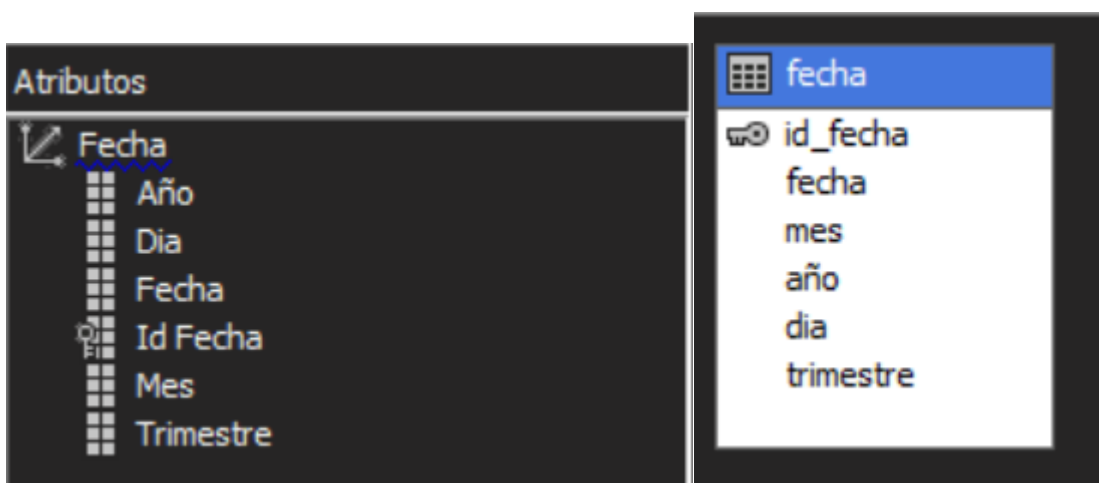
Grupo de medida asociado: (Sin definir)

Carpeta para mostrar:

Expresiones de color

Expresiones de fuente

Modificamos las dimensiones de la siguiente manera:



Y comprobamos que todo esté correcto revisando la tabla:

Editar como texto

Importar...

MDX

Covid 19

Metadatos

Buscar modelo

<Todos>

Porcentaje Transito

KPI

Fecha

Año

Día

Fecha

Id Fecha

Mes

Trimestre

Jerarquía

Territorio

Id Zona

Zona

Dimensión

Jerarquía

Operador

Expresión de filtro

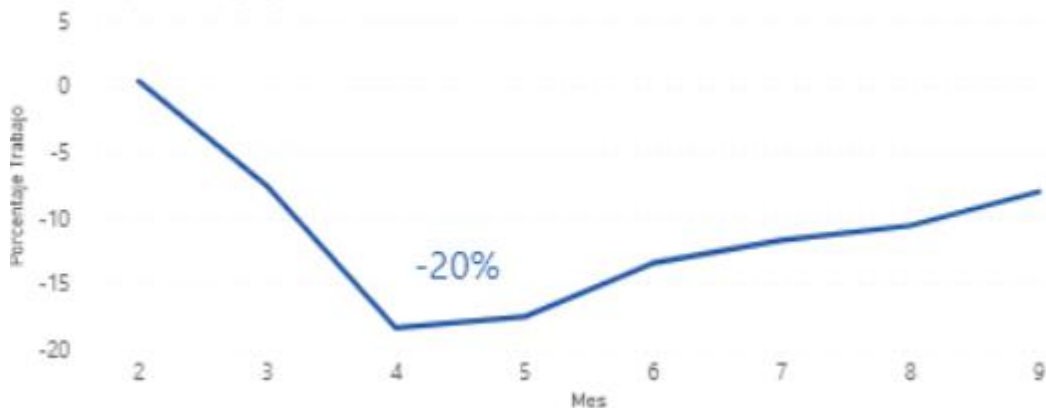
Parám...

<Seleccionar dimensión>

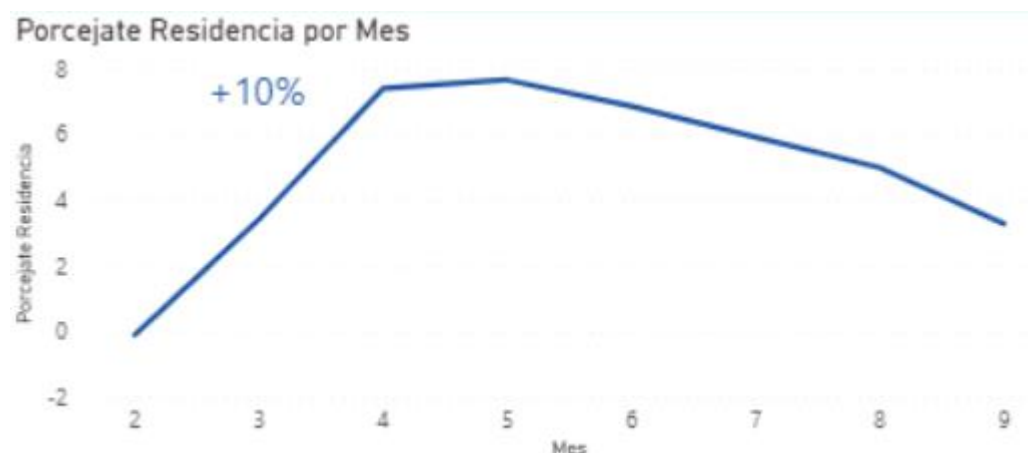
Día	Mes	Año	Trimestre	Fecha	Zona	Porcentaje Transito	Porcentaje Trabajo	Porcentaje Tiendas Far
1	3	2020	1	2020...	Ahuachapán...	0	0,04	0
1	3	2020	1	2020...	Cabañas De...	0	0,08	0
1	3	2020	1	2020...	Chalatenang...	0	0,02	0
1	3	2020	1	2020...	Cuscatlán D...	-0,01	0,03	0
1	3	2020	1	2020...	El Salvador	0,02	0,04	0,08
1	3	2020	1	2020...	La Libertad ...	0,08	0,04	0,09
1	3	2020	1	2020...	La Paz Depa...	-0,11	0,04	-0,02
1	3	2020	1	2020...	La Unión De...	0	0,07	0
1	3	2020	1	2020...	Morazán De...	0	0,1	0
1	3	2020	1	2020...	San Miguel D...	0,09	0,03	0,03
1	3	2020	1	2020...	San Salvado...	0	0,03	0,09
1	3	2020	1	2020...	San Vicente ...	0	0,11	0
1	3	2020	1	2020...	Santa Ana D...	-0,06	0,05	0,06
1	3	2020	1	2020...	Sonsonate D...	0,01	0,04	0,07
1	3	2020	1	2020...	Usulután De...	0	0,06	0,06
1	4	2020	2	2020...	Ahuachapán...	0	-0,53	0
1	4	2020	2	2020...	Cabañas De...	0	-0,53	0

Para terminar el ejercicio, vamos a analizar toda la información en PowerBI, generando unos gráficos para ver de qué manera ha sido afectado cada departamento por las medidas para la pandemia.

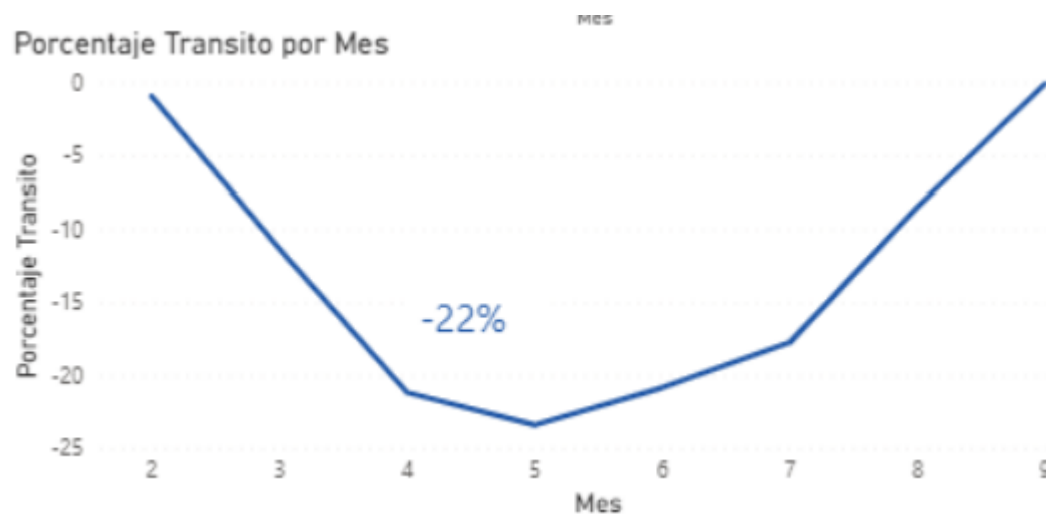
Porcentaje Trabajo por Mes



Porcentaje Transito por Mes



Porcentaje Tiendas Super Farmacia por Mes



En su mayoría, la movilidad se vió reducida. Podemos realizar estos gráficos para cada departamento y crear filtros diferentes dependiendo de lo que se desee analizar y estudiar.

