

UNIVERSIDAD DON BOSCO  
FACULTAD DE INGENIERÍA  
ESCUELA DE COMPUTACIÓN



Datawarehouse y Minería de Datos  
Entrega final del proyecto de cátedra  
Docente: Ing. Karens Medrano

Estudiantes:

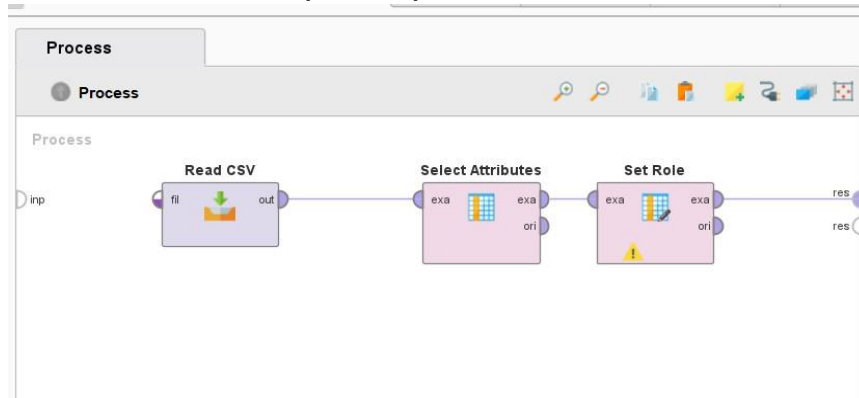
Josué Alfonso Ayala Martínez | AM210971  
Kallahan Andrea Salas Bojórquez | SB210537

Sábado 03 de diciembre de 2022

## Esquelas de tránsito.

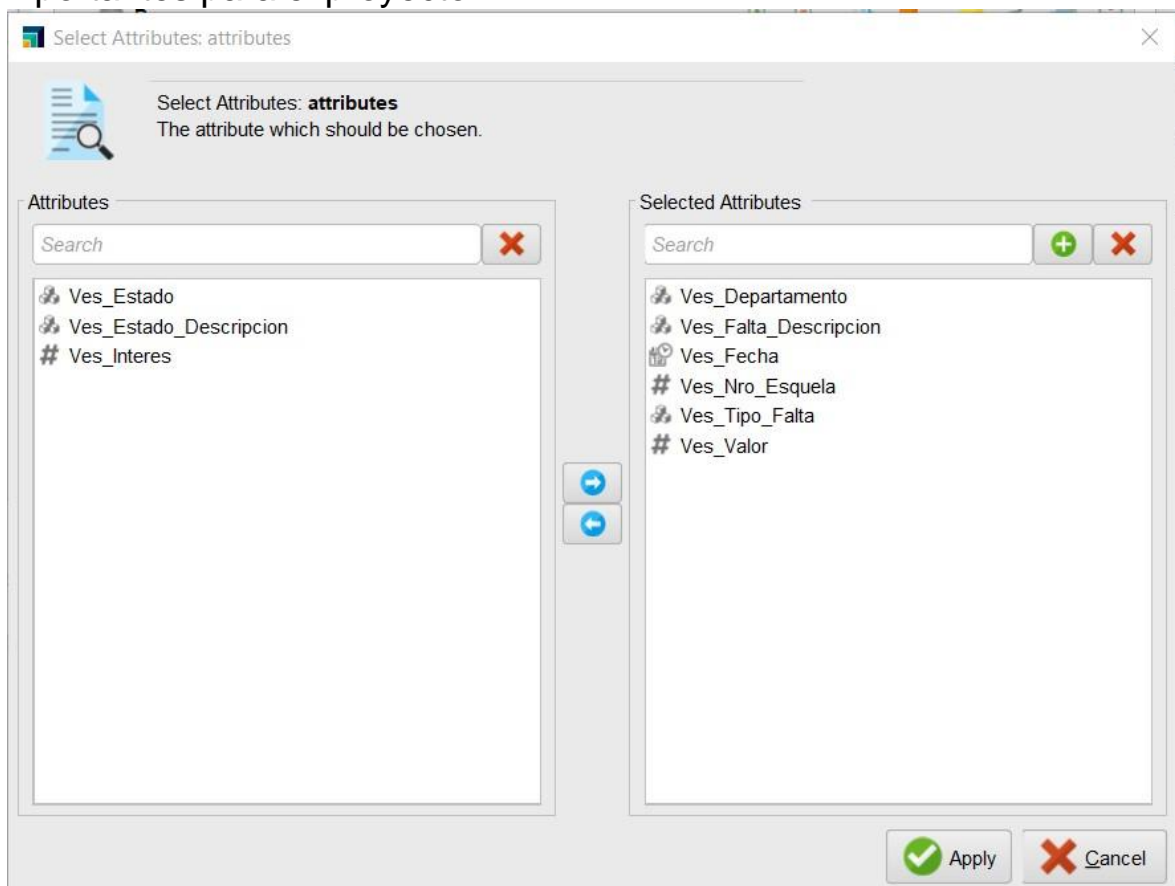
### 1. Reglas de asociación

Para analizar la data de Esquelas para el cierre del año 2018,



dispondremos a utilizar la herramienta RapidMiner. Para comenzar debemos crear el siguiente diagrama:

En el apartado “Select Attribute”, seleccionaremos las columnas importantes para el proyecto



Para set Role nos dirigimos a declarar Ves\_Nro\_Esqueda de tipo "id"

Edit Parameter List: set additional roles

Edit Parameter List: **set additional roles**  
This parameter defines additional attribute role combinations.

attribute name	target role
Ves_Nro_Esqueda	id

Luego, procedemos a ejecutarlo y a comprobar que se haya realizado correctamente la separación de la información.

Result History

ExampleSet (Set Role)

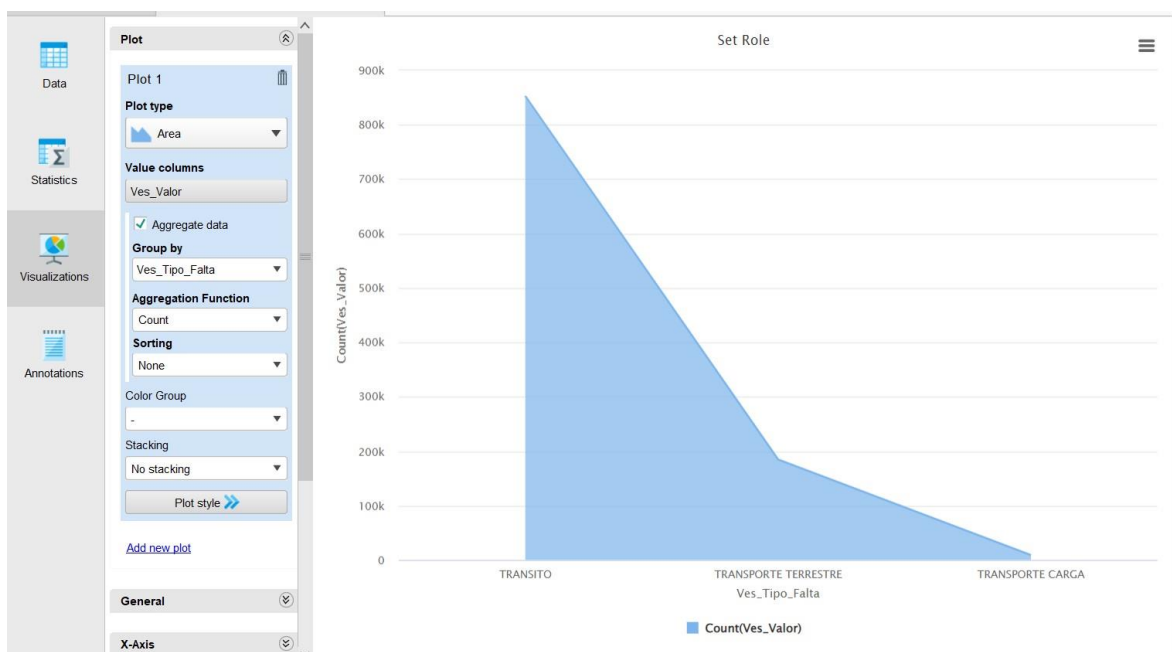
Open in Turbo Prep Auto Model

Filter (1,046,926 / 1,046,926 examples): all

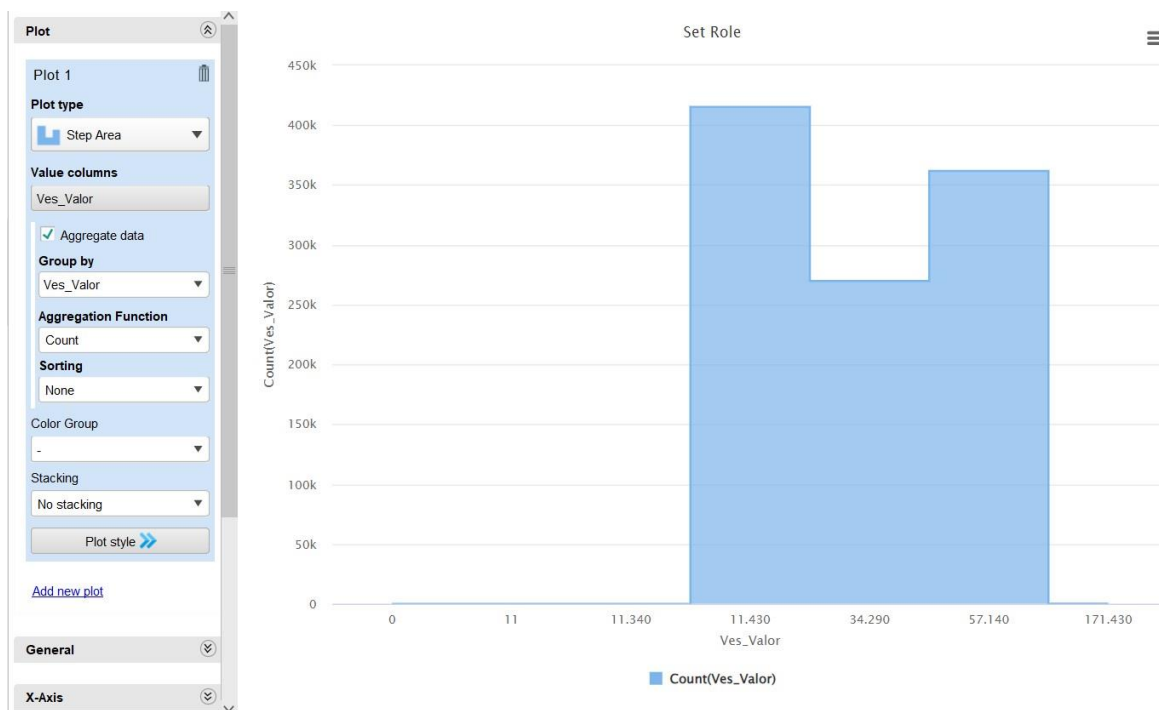
Row No.	Ves_Nro_Es...	Ves_Fecha	Ves_Tipo_Fa...	Ves_Falta_D...	Ves_Departa...	Ves_Valor
1	110637	Jun 4, 2003, 1...	TRANSITO	NO UTILIZAR ...	CHALATENA...	34.290
2	107742	Jun 4, 2003, 1...	TRANSITO	NO UTILIZAR ...	CHALATENA...	34.290
3	106645	Jun 4, 2003, 1...	TRANSITO	ESTACIONA...	SAN SALVAD...	34.290
4	154061	Jun 4, 2003, 1...	TRANSITO	BAJAR O SU...	SAN SALVAD...	34.290
5	154102	Jun 3, 2003, 1...	TRANSITO	DISPUTARSE...	SAN SALVAD...	57.140
6	119779	Jun 6, 2003, 1...	TRANSITO	CIRCULAR V...	SAN SALVAD...	57.140
7	106938	Jun 4, 2003, 1...	TRANSITO	TRANSPORT...	SAN SALVAD...	57.140
8	154457	Jun 4, 2003, 1...	TRANSITO	TRANSPORT...	SAN SALVAD...	57.140
9	154951	Jun 6, 2003, 1...	TRANSPORT...	NO RESPET...	SAN SALVAD...	57.140
10	154976	Jun 6, 2003, 1...	TRANSPORT...	NO RESPET...	SAN SALVAD...	57.140
11	109210	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290
12	109211	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290
13	109207	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290
14	105578	Jun 9, 2003, 1...	TRANSITO	CIRCULAR C...	SAN SALVAD...	11.430
15	109469	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290
16	154826	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290
17	154952	Jun 6, 2003, 1...	TRANSPORT...	CONDU CIR ...	SAN SALVAD...	34.290
18	110968	Jun 6, 2003, 1...	TRANSPORT...	TRANSPORT...	SAN SALVAD...	34.290

ExampleSet (1,046,926 examples, 1 special attribute, 5 regular attributes)

Podemos realizar un gráfico en el apartado de estadísticas para poder ver (por ejemplo) los tipos de faltas

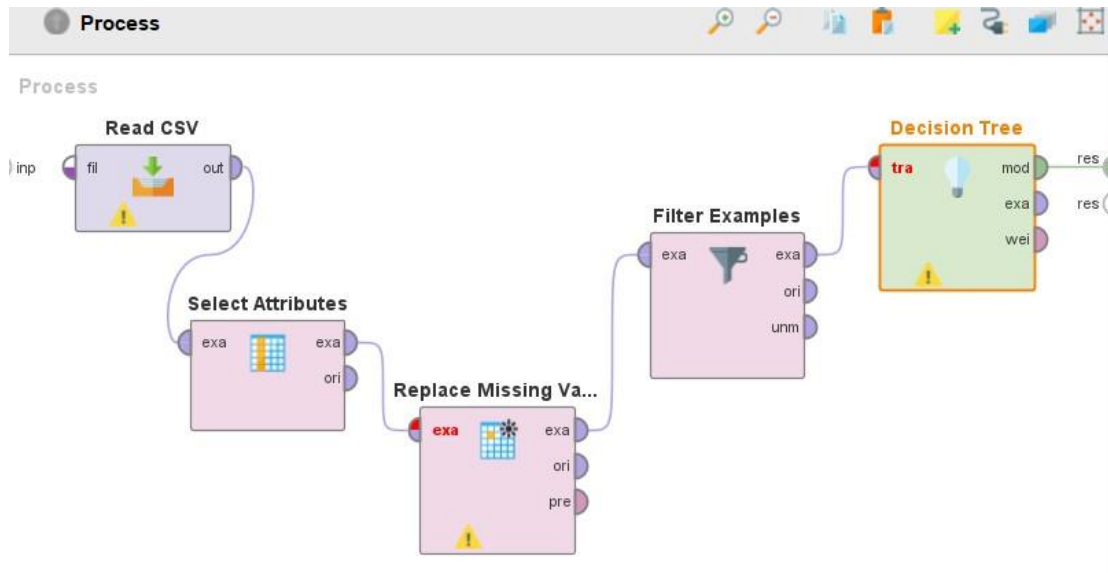


O según el monto de las esquelas

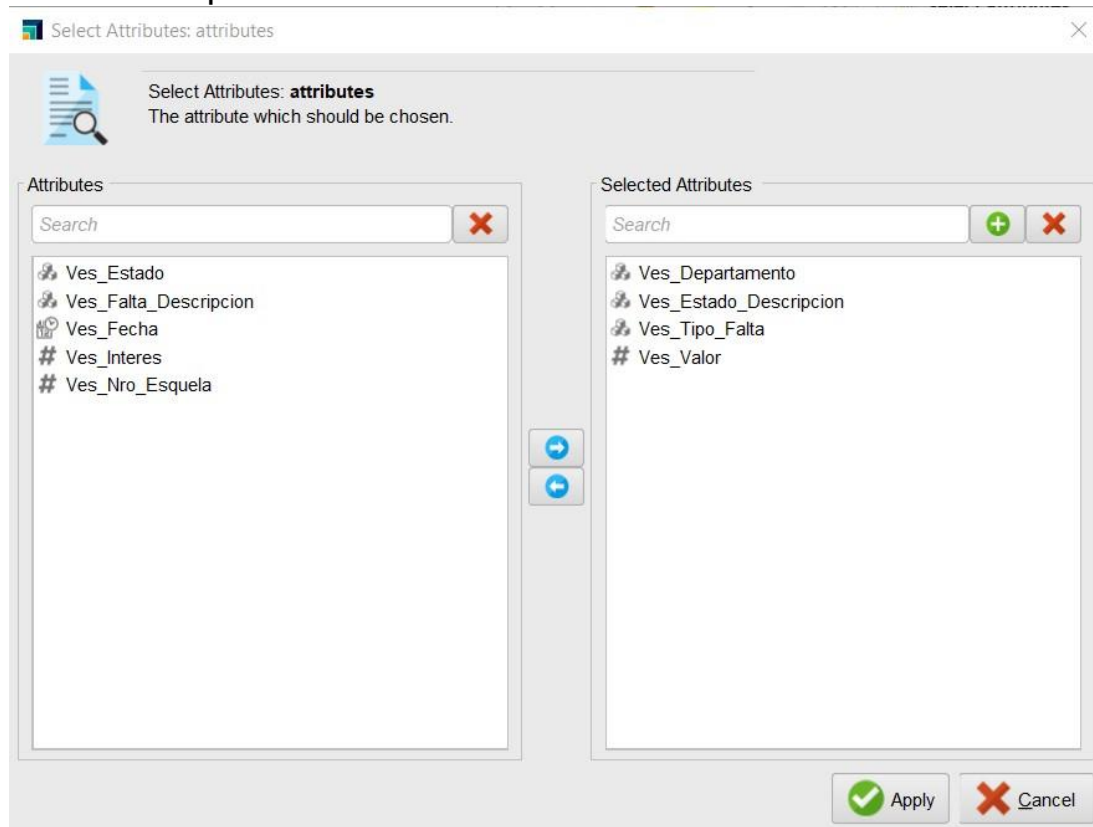


## 2. Árbol de decisiones

Creamos un nuevo espacio de trabajo en la herramienta RapidMiner y procedemos a crear el siguiente diagrama:



En “Select Attributes”, procedemos a seleccionar las columnas que utilizaremos para el análisis.



Dentro de Replace Missing Values colocaremos CANCELADA, en todos esos espacios que se encuentren vacíos.

**Parameters** X

**Replace Missing Values**

☐ create view ⓘ

attribute filter type: single ⓘ

attribute: Ves\_Estado\_Descripcion ⓘ

☐ invert selection ⓘ

☐ include special attributes ⓘ

default: value ⓘ

columns: Edit List (0)... ⓘ

replenishment val...: CANCELADA ⓘ

Ahora dentro de “Filter Examples” filtramos la información a procesar colocando lo siguiente, tratando de mostrar en su mayoría las faltas de tránsito:


Create Filters: filters X





Create Filters: **filters**  
Defines the list of filters to apply.


Ves_Tipo_Falta	equals	TRANSITO	✎	✖
Ves_Valor	=	11	✎	✖


De esta misma manera, colocamos en tipo de falta “TRANSPORTE CARGA” y “TRANSPORTE TERRESTRE” para obtener los tres diagramas distintos.





Create Filters: filters

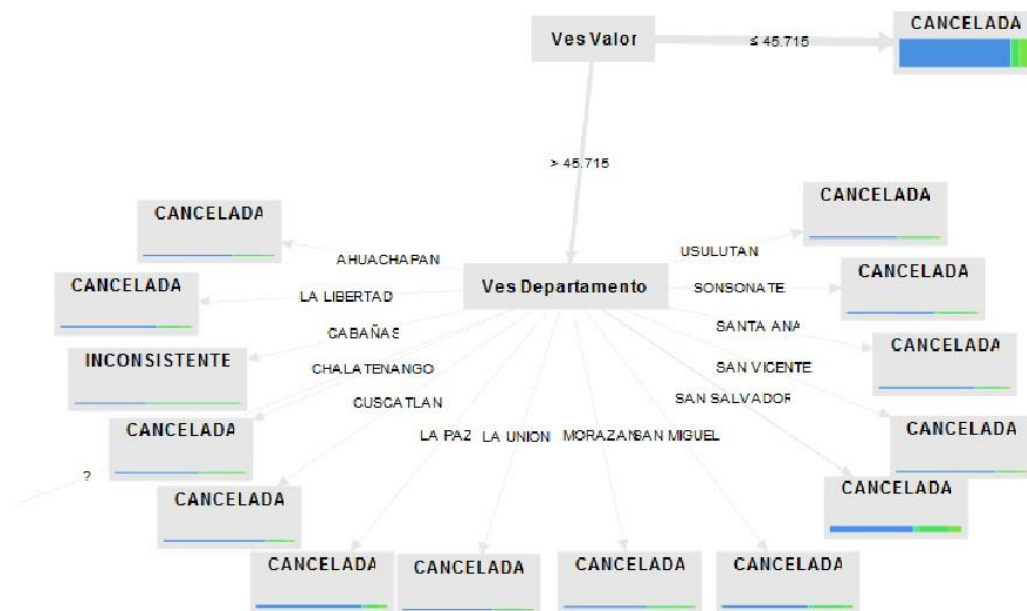
 Create Filters: **filters**  
Defines the list of filters to apply.

Ves_Tipo_Falta	equals	TRANSPORTE CARGA		
Ves_Valor	=	11		

 Create Filters: filters ✕

 Create Filters: **filters**  
Defines the list of filters to apply.

Ves_Tipo_Falta	equals	TRANSPORTE TERRESTRE		
Ves_Valor	=	11		



## Parque vehicular

### 1. Implementación de ETL

Para iniciar con esta fase del proyecto, se debe crear inicialmente la base de datos y para eso utilizaremos SQL Server. Luego de eso, crearemos las tablas que utilizaremos para la BDD ( tabla auxiliar, de dimensiones y de hechos) y crearemos el diagrama del análisis.

El Query debería de quedar de la siguiente manera:

```
-- DMD Segunda Entrega
--ETL

create database Pvehicular
use Pvehicular
go

create table [DataPvehicular](
    [TIPO_PLACA] NVARCHAR(50),
    [CANTIDAD_DE_CILINDROS] NUMERIC(20,0),
    [CANTIDAD_DE_PUERTAS] NUMERIC(20,0)
    [VALOR_DEL_VEHICULO] DECIMAL(28,0)
    [FECHA_DE_INGRESO] DATE,
    [ANIO_INGRESO] BIGINT,
    [MES_INGRESO] BIGINT,
    [CLASE] NVARCHAR(50),
    [PERTENENCIA] NVARCHAR(50),
    [MARCA] NVARCHAR(50),
    [MODELO] NVARCHAR(50),
    [CAPACIDAD] DECIMAL(28,0)
    [DES_CAPACIDAD] NVARCHAR(50),
    [COMBUSTIBLE] NVARCHAR(50),
    [CONDICION_INGRESO] NVARCHAR(50),
    [PROPIETARIO_DEPARTAMENTO] NVARCHAR(50),
    [PROPIETARIO_MUNICIPIO] NVARCHAR(50),
    [ESTADO] NVARCHAR(50),
)
```



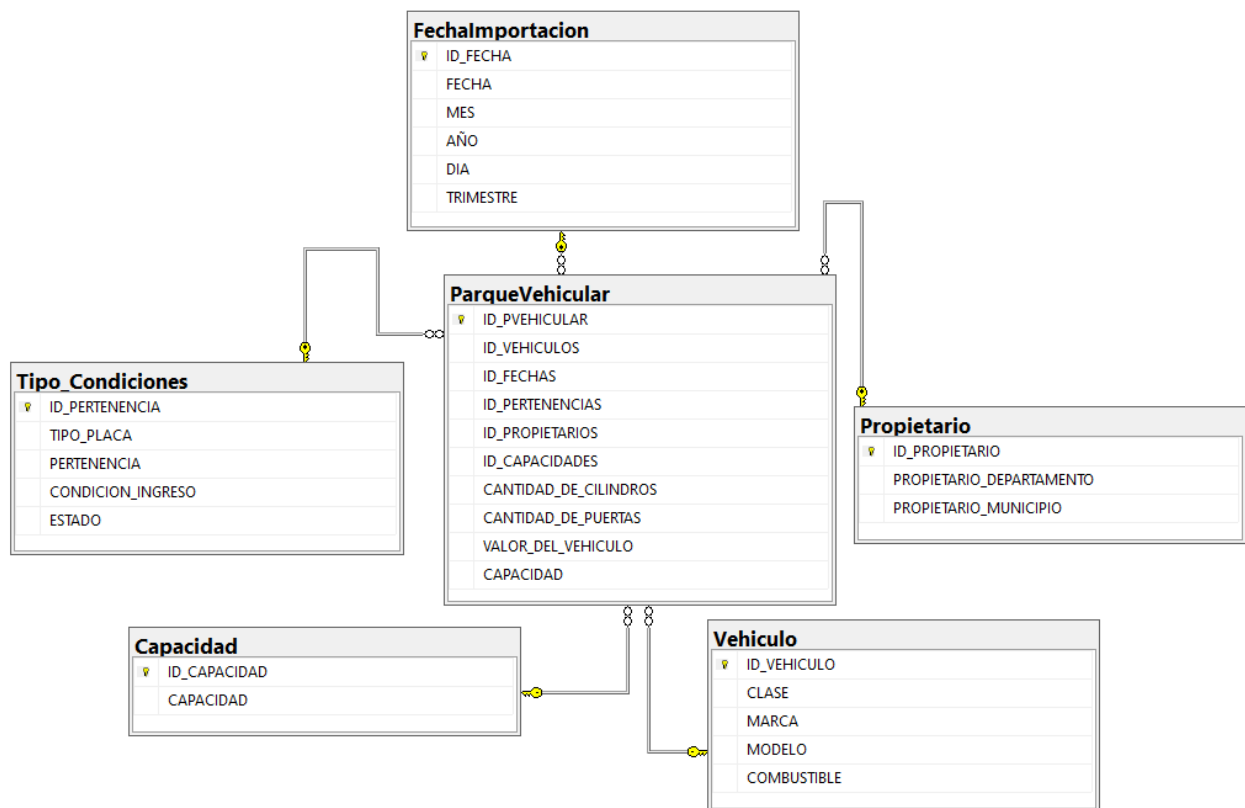
```
create table Vehiculo(  
    ID_VEHICULO INT IDENTITY(1,1) PRIMARY KEY,  
    CLASE NVARCHAR(50),  
    MARCA NVARCHAR(50),  
    MODELO NVARCHAR(50),  
    COMBUSTIBLE NVARCHAR(50),  
)  
  
create table FechaImportacion(  
    ID_FECHA INT IDENTITY(1,1) PRIMARY KEY,  
    FECHA DATE  
    MES VARCHAR(15)  
    AÑO INT  
    DIA INT  
    TRIMESTRE INT  
)  
  
create table Tipo_Condiciones(  
    ID_PERTENENCIA INT IDENTITY(1,1) PRIMARY KEY  
    TIPO_PLACA NVARCHAR(50),  
    PERTENENCIA NVARCHAR(50),  
    CONDICION_INGRESO NVARCHAR(50),  
    ESTADO NVARCHAR(50),  
)  
  
create table Propietario(  
    ID_PROPIETARIO INT IDENTITY (1,1) PRIMARY KEY,  
    PROPIETARIO_DEPARTAMENTO NVARCHAR(50),  
    PROPIETARIO_MUNICIPIO NVARCHAR(50),  
)  
  
create table Capacidad(  
    ID_CAPACIDAD INT IDENTITY(1,1) PRIMARY KEY,  
    CAPACIDAD NVARCHAR(50),  
)
```

```

create table ParqueVehicular(
  ID_PVEHICULAR INT IDENTITY(1,1) PRIMARY KEY,
  ID_VEHICULOS INT FOREIGN KEY REFERENCES Vehiculo(ID_VEHICULO),
  ID_FECHAS INT FOREIGN KEY REFERENCES FechaImportacion(ID_FECHA),
  ID_PERTENENCIAS INT FOREIGN KEY REFERENCES Tipo_Condiciones(ID_PERTENENCIA),
  ID_PROPIETARIOS INT FOREIGN KEY REFERENCES Propietario(ID_PROPIETARIO)
  ID_CAPACIDADES INT FOREIGN KEY REFERENCES Capacidad(ID_CAPACIDAD),
  CANTIDAD_DE_CILINDROS numeric (20,0)
  CANTIDAD_DE_PUERTAS numeric (20,0)
  VALOR_DEL_VEHICULO DECIMAL (28,0)
  CAPACIDAD decimal (28,0)
)

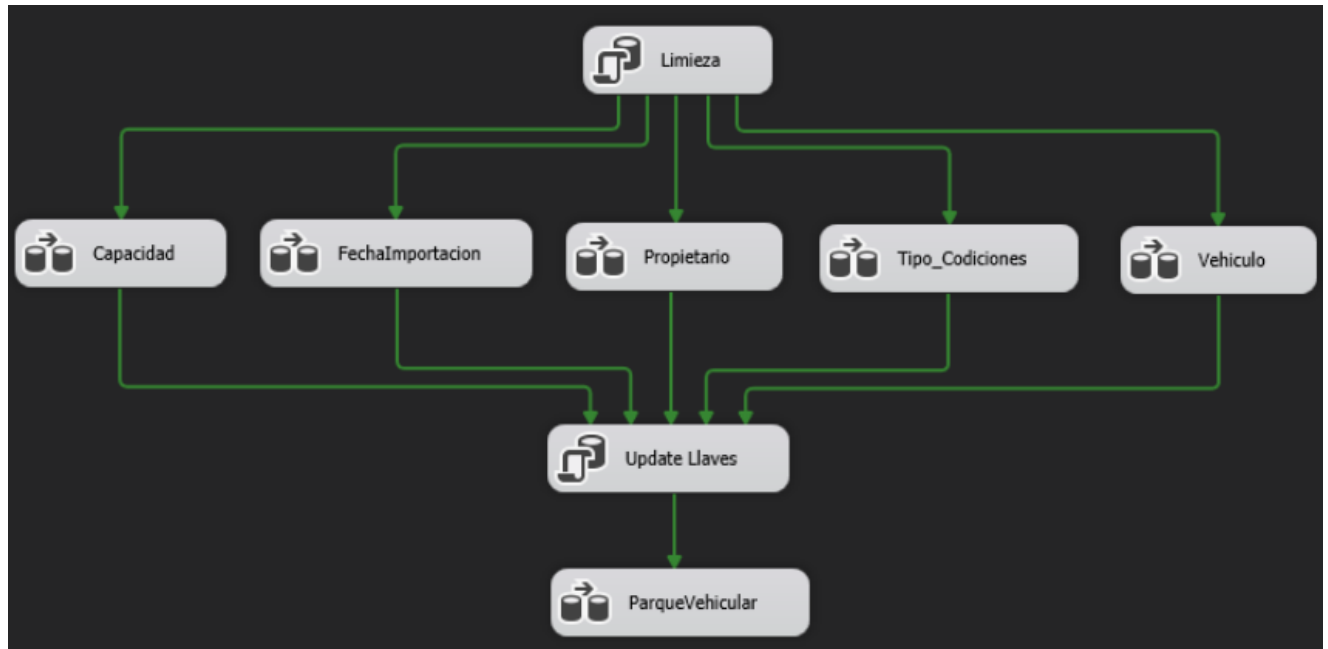
```

Y el diagrama nos queda de la siguiente forma:



Se logra ver que la tabla de hechos (ParqueVehicular) es la principal, y está unida a las tablas de dimensiones. De esta manera, forman el diagrama en estrella donde las dimensiones son los atributos de los datos en la tabla de hechos.

Acto seguido, crearemos un nuevo proyecto con SSIS en Visual Studio 2019 y agregaremos los componentes del diagrama de flujo.

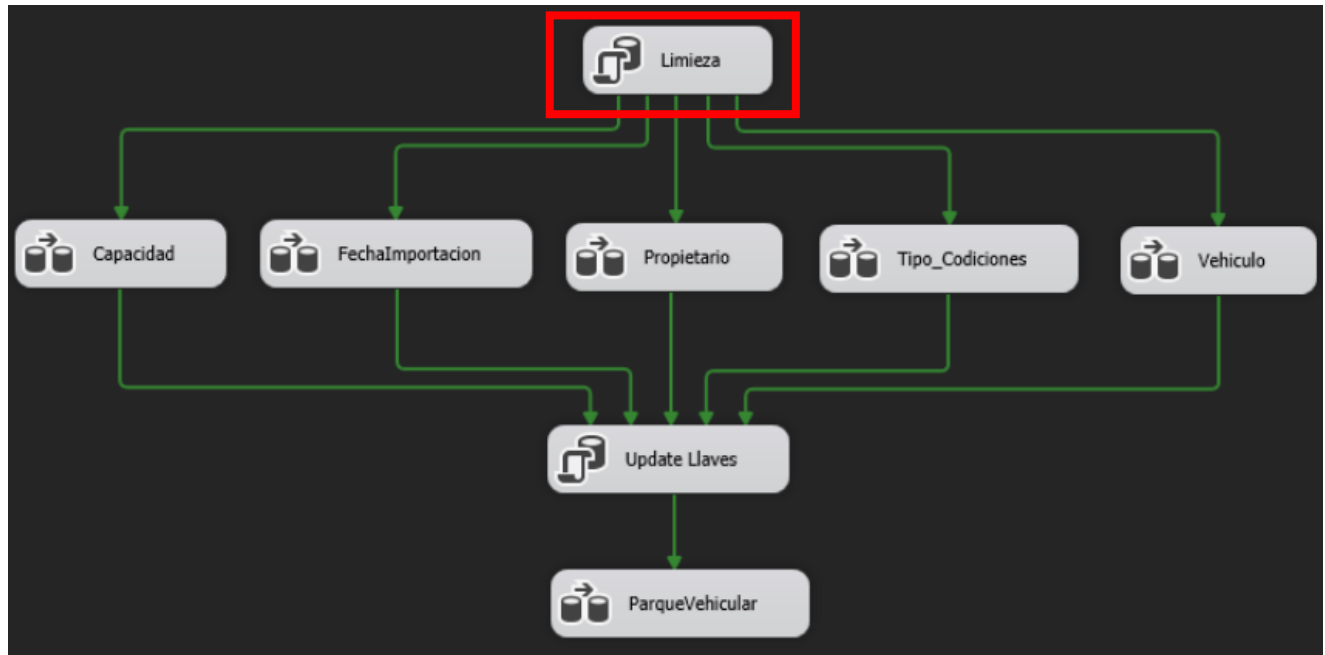


Es importante que se vea así y en ese orden, por las relaciones de las llaves foráneas que definimos al crear las tablas. Inicialmente se llenan las tablas de FechaImportacion, Capacidad, Propietario, Vehículo y Tipo\_Condicion.

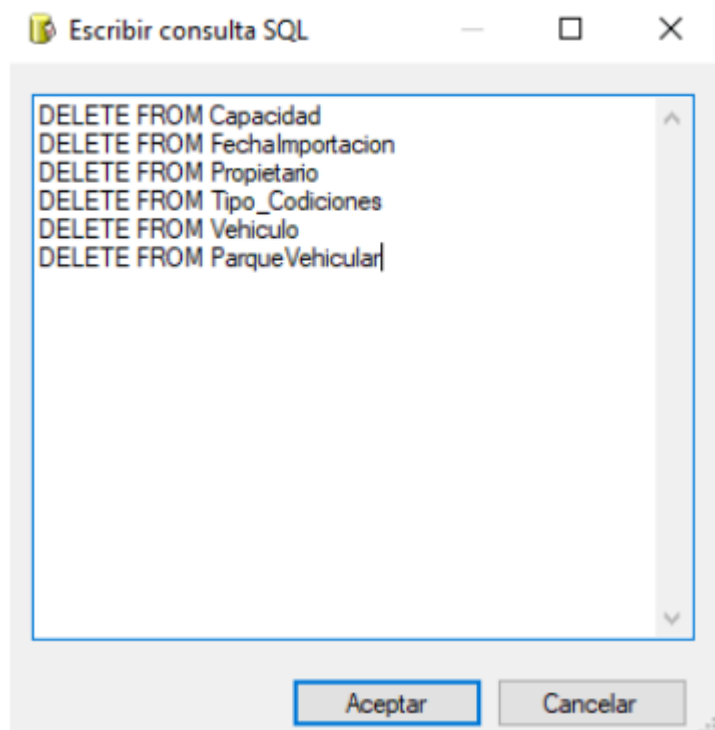
Luego de procesar las tablas de dimensiones, seguimos el flujo actualizando las llaves foráneas para pasar la información al flujo final que alimentara la tabla de Hechos.

Definiremos entonces para qué funciona cada uno de los elementos del diagrama de flujo.

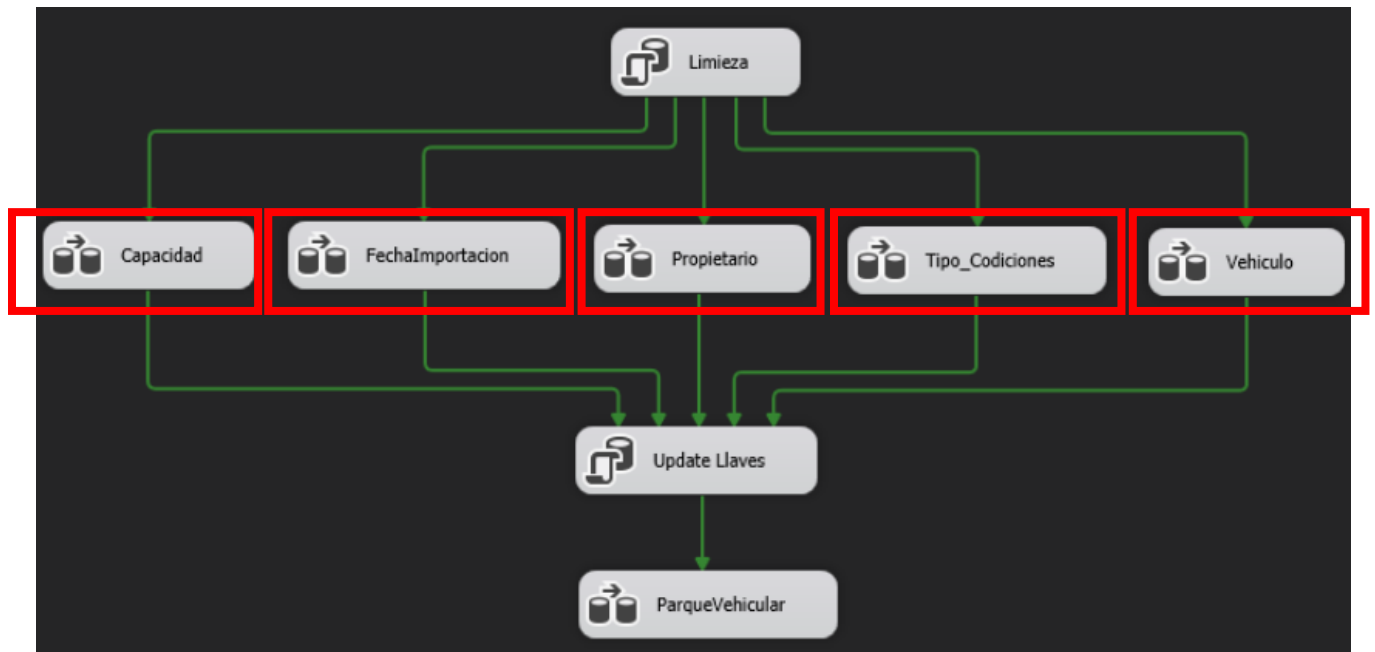
## 1. Limpieza



Con este elemento, podemos borrar información previa que pueda existir en el sistema, para poder pasarle la nueva información de nuestra nueva base de datos. (Aunque en este momento no hay nada, pero es una buena práctica)

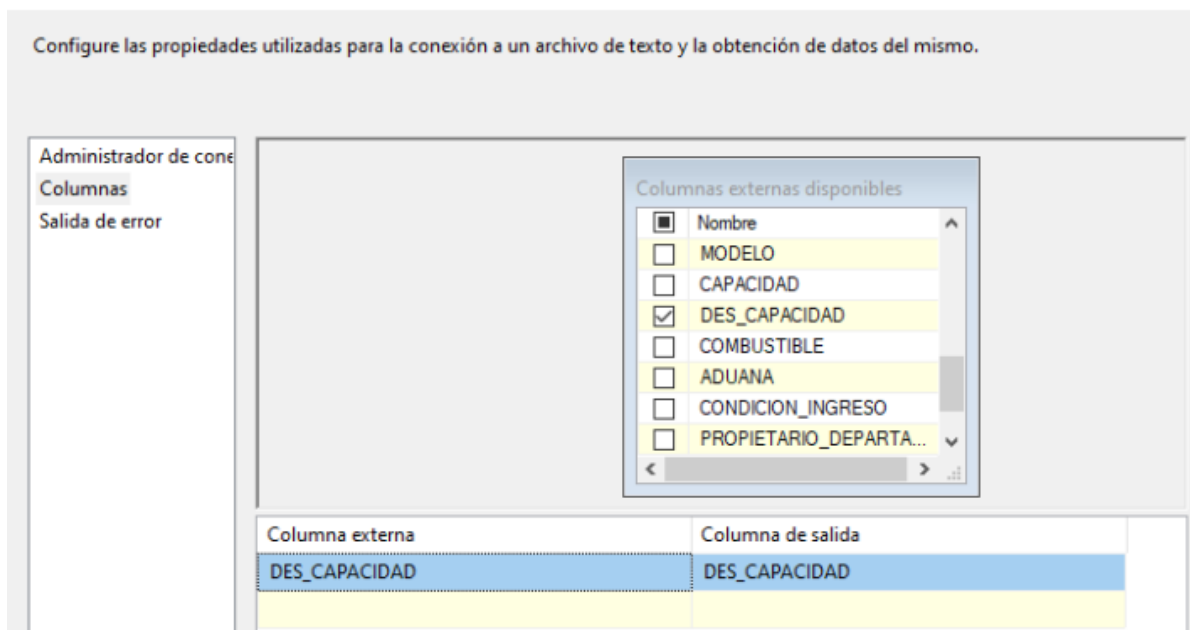


A continuación de eso, al mismo tiempo se llenarán las tablas de Propietario, FechaImportacion, Tipo\_Condicion y Vehículo. Obtendríamos algo como la siguiente imagen:



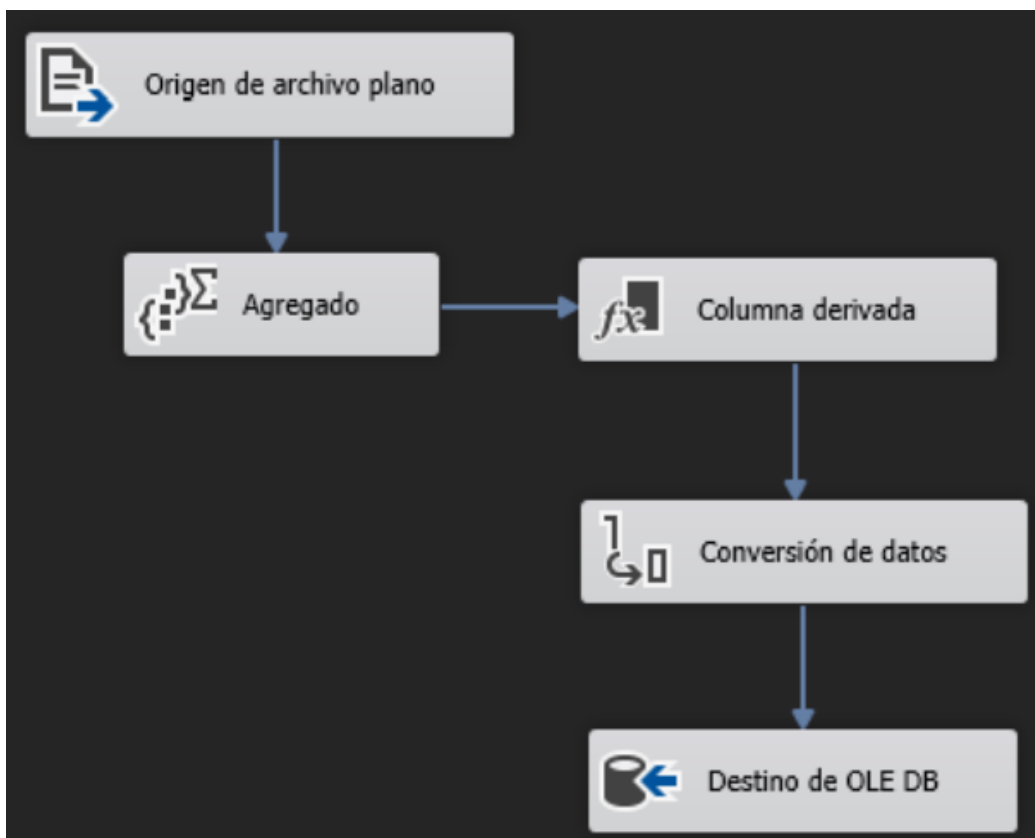
## 2. Capacidad

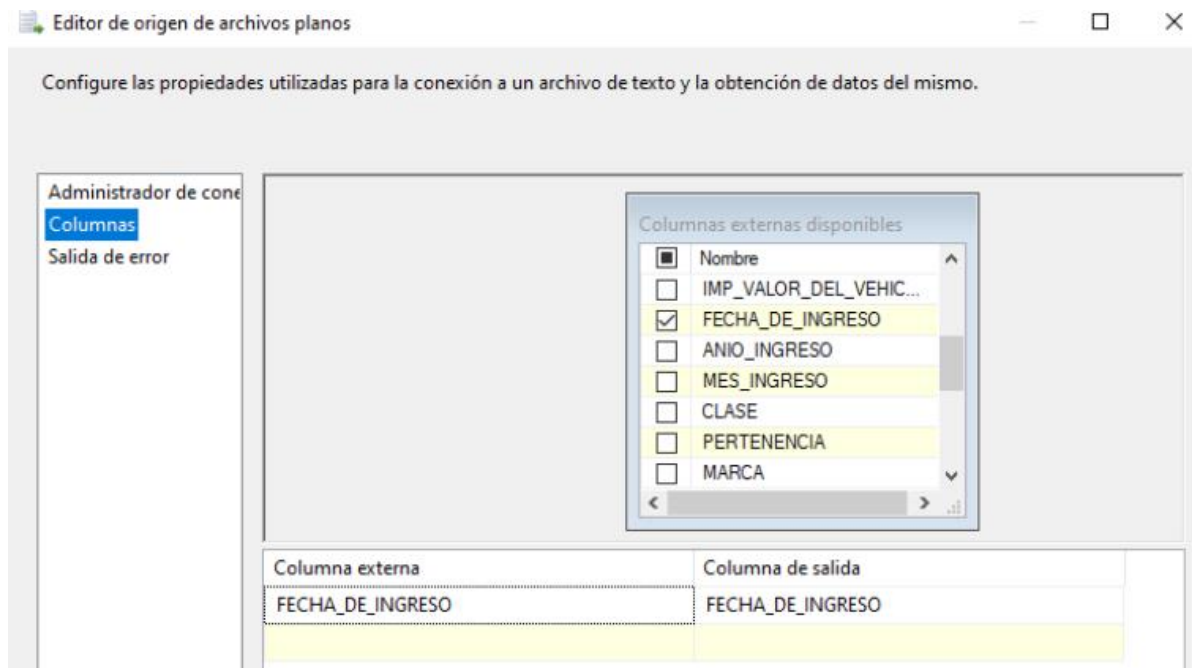




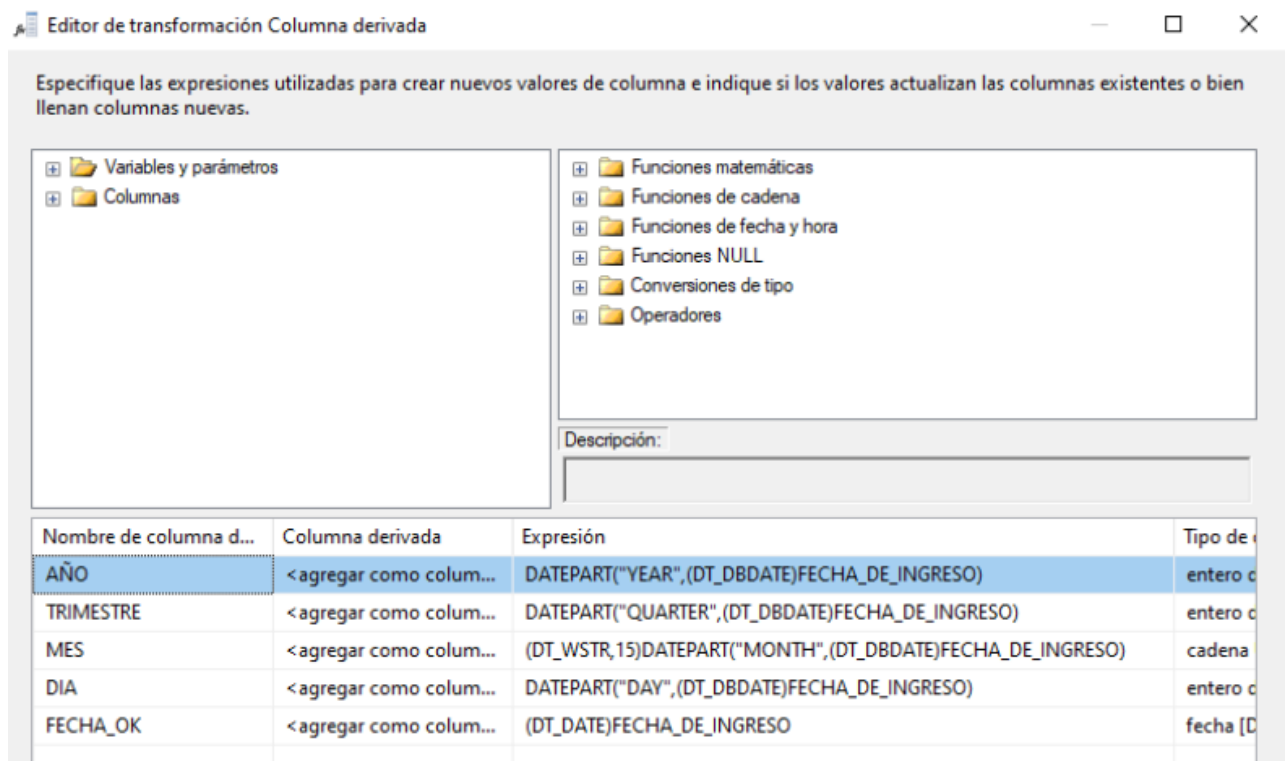
Seleccionamos solamente el campo “DES\_CAPACIDAD” del archivo .csv, y procederemos a agrupar los datos en la base de datos P Vehicular.

### 3. FechaImportacion





Seguimos el mismo proceso, y seleccionamos solamente “FECHA\_DE\_INGRESO” del mismo archivo csv, pero en el apartado de “columna derivada” procedemos a establecer lo siguiente:



## 4. Propietario

Editor de origen de archivos planos

Configure las propiedades utilizadas para la conexión a un archivo de texto y la obtención de datos del mismo.

Diagrama de flujo de configuración:

- Origen de archivo plano
- Agregado
- Columna derivada
- Destino de OLE DB

Columnas externas disponibles:

- ☒ Nombre
- ☐ DES\_CAPACIDAD
- ☐ COMBUSTIBLE
- ☐ ADUANA
- ☐ CONDICION\_INGRESO
- ☒ PROPIETARIO\_DEPARTA...
- ☒ PROPIETARIO\_MUNICIPIO
- ☐ ESTADO

Columna externa	Columna de salida
PROPIETARIO_DEPARTAMENTO	PROPIETARIO_DEPARTAMENTO
PROPIETARIO_MUNICIPIO	PROPIETARIO_MUNICIPIO

Variables y parámetros:

- Columnas

Funciones disponibles:

- Funciones matemáticas
- Funciones de cadena
- Funciones de fecha y hora
- Funciones NULL
- Conversiones de tipo
- Operadores

Descripción:

Nombre de columna d...	Columna derivada	Expresión
DEPARTAMENTO_OK	<agregar como colum...	PROPIETARIO_DEPARTAMENTO == "CABAÑAS" ? "CABAÑAS" : PROPIETARIO_DEPARTAMENTO



## 5. Tipo\_Codiciones

Columnas externas disponibles

<input type="checkbox"/>	Nombre
<input checked="" type="checkbox"/>	TIPO_PLACA
<input type="checkbox"/>	ANIO_DE_FABRICACION
<input type="checkbox"/>	CILINDRADA
<input type="checkbox"/>	CANTIDAD_DE_CILINDROS
<input type="checkbox"/>	CANTIDAD_DE_PUERTAS
<input type="checkbox"/>	VALOR_DEL_VEHICULO
<input type="checkbox"/>	COLORES

Columna externa	Columna de salida
TIPO_PLACA	TIPO_PLACA
PERTENENCIA	PERTENENCIA
CONDICION_INGRESO	CONDICION_INGRESO
ESTADO	ESTADO



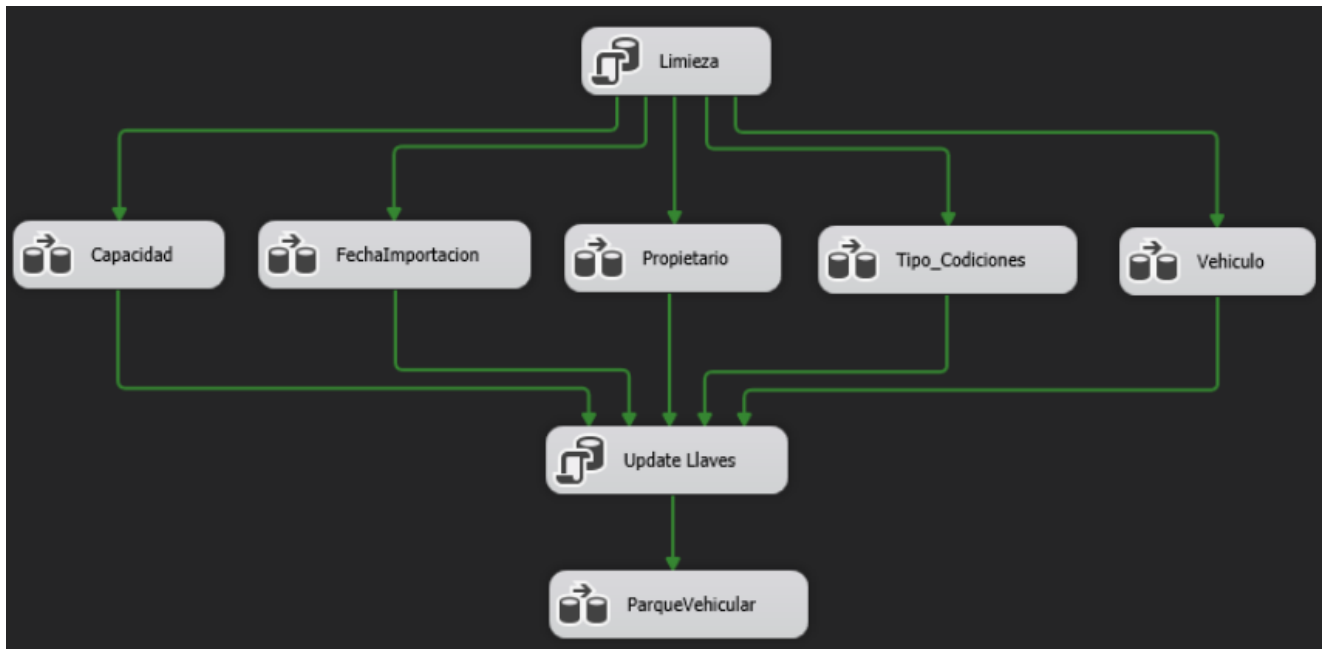
## 6. Vehículo

Columnas externas disponibles	
<input checked="" type="checkbox"/>	Nombre
<input type="checkbox"/>	MES_INGRESO
<input checked="" type="checkbox"/>	CLASE
<input type="checkbox"/>	PERTENENCIA
<input checked="" type="checkbox"/>	MARCA
<input checked="" type="checkbox"/>	MODELO
<input type="checkbox"/>	CAPACIDAD
<input type="checkbox"/>	DES_CAPACIDAD

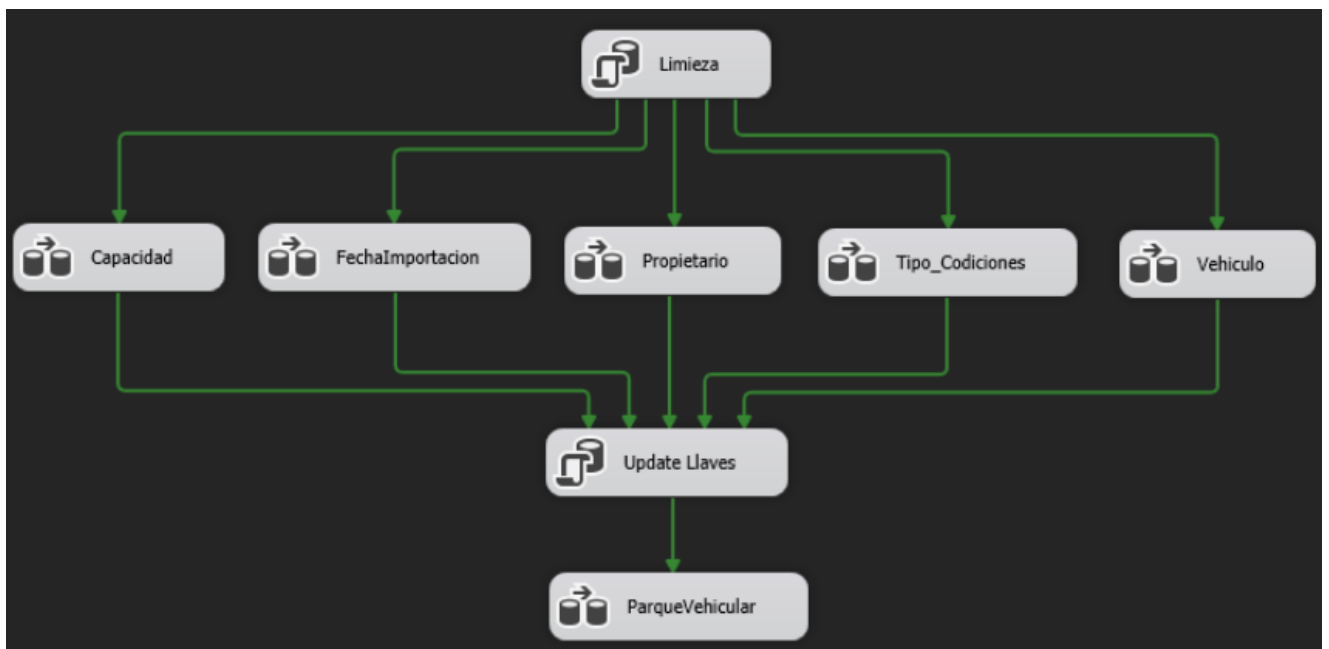
Columna externa	Columna de salida
CLASE	CLASE
MARCA	MARCA
MODELO	MODELO
COMBUSTIBLE	COMBUSTIBLE

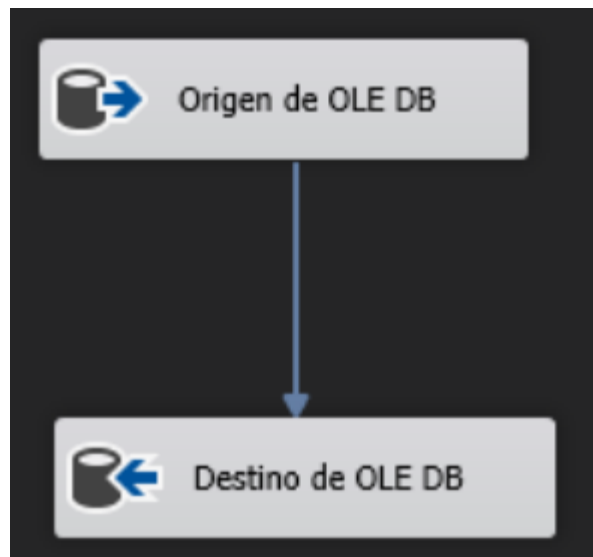


Media vez configuradas las tablas, actualizamos la tabla auxiliar para establecer la conexión entre las tablas de dimensiones y la tabla de hechos.



Entonces, llegamos a ParqueVehicular. Y acá nos encontramos con el segundo diagrama. Solo pasaremos los atributos que se definieron en la tabla de hechos.





En “origen de OLE DB”, utilizaremos el modo de acceso de datos comando SQL, para extraer los campos necesarios y las FK.

Texto de comando SQL:

```
SELECT V.ID_VEHICULO,
       F.ID_FECHA, TC.ID_PERTENENCIA,
       P.ID_PROPIETARIO, C.ID_CAPACIDAD,
       DV.CANTIDAD_DE_CILINDROS,
       DV.CANTIDAD_DE_PUERTAS, DV.VALOR_DEL_VEHICULO,
       DV.CAPACIDAD FROM DataPVehicular DV
INNER JOIN Vehiculo V ON V.ID_VEHICULO=DV.ID_VEHICULO_temp
INNER JOIN FechaImportacion F ON F.ID_FECHA=DV.ID_FECHA_temp
INNER JOIN Tipo_Codiciones TC ON
TC.ID_PERTENENCIA=DV.ID_PERTENENCIA_temp
INNER JOIN Propietario P ON
P.ID_PROPIETARIO=DV.ID_PROPIETARIO_temp
INNER JOIN Capacidad C ON
C.ID_CAPACIDAD=DV.ID_CAPACIDADE_temp|
```

Parámetros...

Generar consulta...

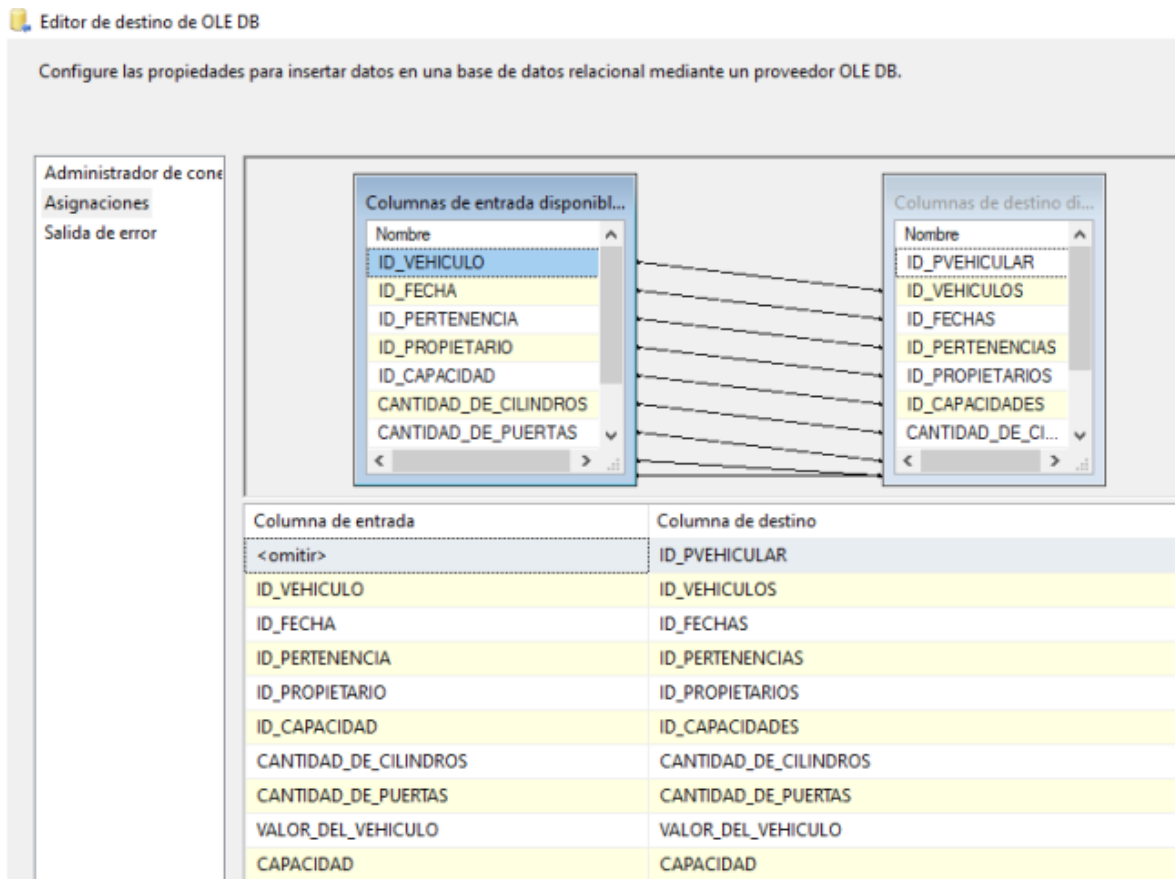
Examinar...

Analizar consulta

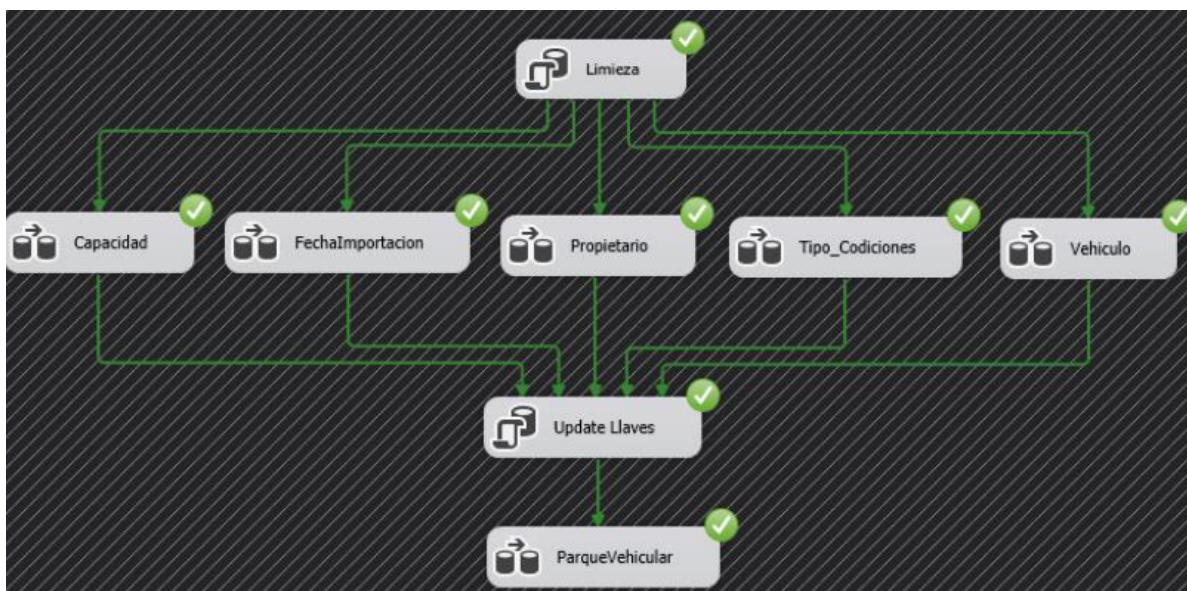
Vista previa...

Asignamos las columnas según los campos de la tabla.

Dejaremos “ID\_PVEHICULAR” sin relacionarlo, porque ahí está la llave primaria de la tabla hechos



Y ejecutamos



Realizamos un SELECT \* FROM solo para verificar que la información se haya transferido de manera exitosa

```
SELECT * FROM Vehiculo
```

	ID_VEHICULO	CLASE	MARCA	MODELO	COMBUSTIBLE
1	23350	CABEZAL	INTERNATIONAL	SERIES 118	DIESEL
2	23351	MOTOCICLETA	YAMAHA	DELTA YZT	GASOLINA
3	23352	CABEZAL	WHITE GMC	VOLVO	DIESEL
4	23353	MOTOCICLETA	SUZUKI	DS 100 X	GASOLINA
5	23354	MOTOCICLETA	HONDA	CBR 300 R	GASOLINA
6	23355	AUTOMOVIL	HYUNDAI	SONATA	GASOLINA
7	23356	CABEZAL	INTERNATIONAL	S 2300	DIESEL
8	23357	REMOLQUE <15	HITCH	N/D	SIN COMBUSTIBLE
9	23358	ALQUILER	DAEWOO	FINO DX	GASOLINA
10	23359	CAMION PESADO	FREIGHTLINER	MAXIMIZER II	DIESEL
11	23360	PICK UP	JEEP	PATRIOT	GASOLINA

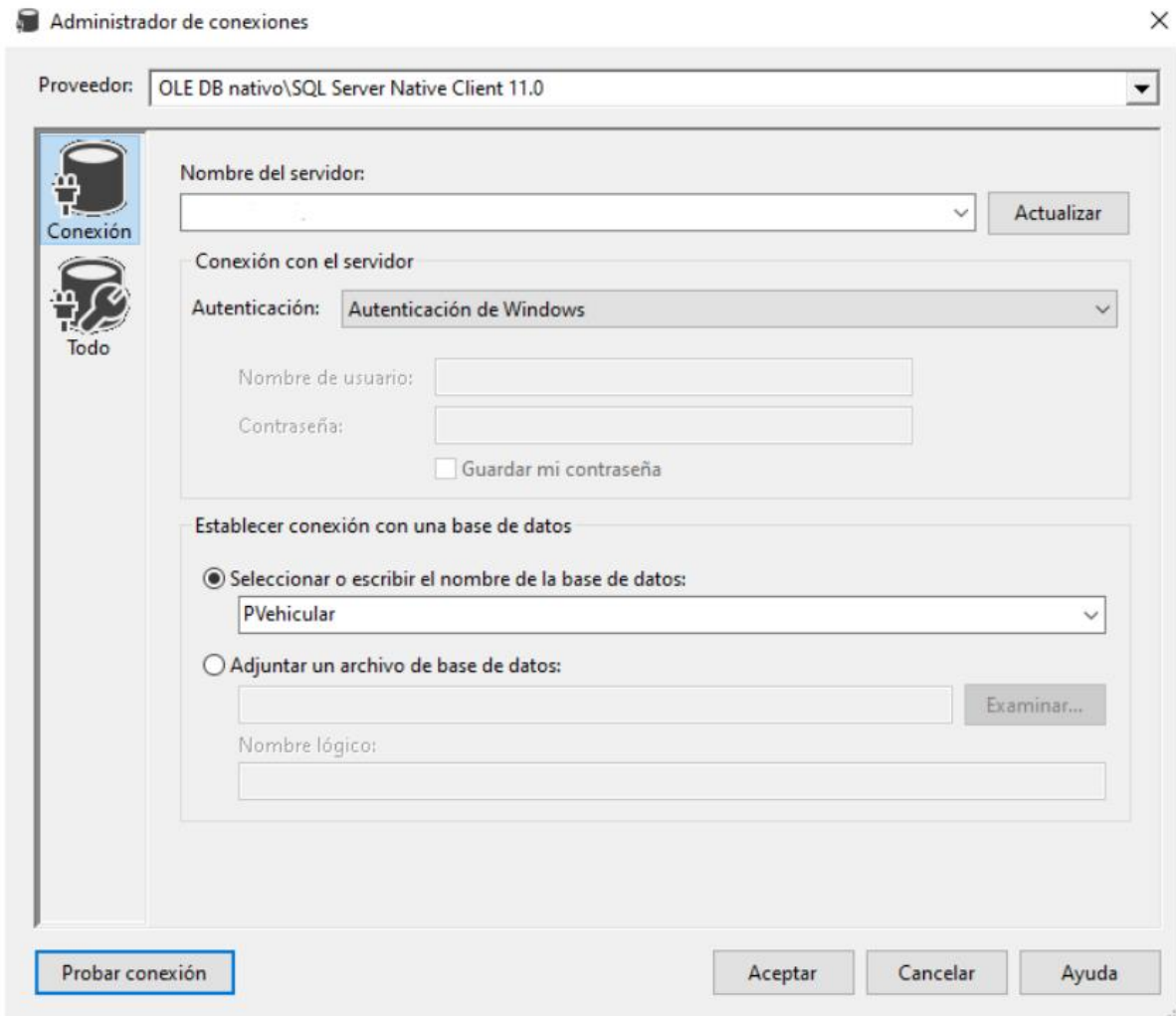
```
SELECT * FROM FechaImportacion
```

	ID_FECHA	FECHA	MES	AÑO	DIA	TRIMESTRE
1	967596	2018-04-09	4	2018	9	2
2	967597	2013-04-24	4	2013	24	2
3	967598	2017-12-06	12	2017	6	4
4	967599	2012-07-19	7	2012	19	3
5	967600	2015-10-07	10	2015	7	4
6	967601	2012-09-07	9	2012	7	3
7	967602	2006-04-08	4	2006	8	2
8	967603	2016-01-25	1	2016	25	1
9	967604	2016-09-06	9	2016	6	3
10	967605	2014-10-31	10	2014	31	4
11	967606	2008-07-22	7	2008	22	3

Solo en algunas tablas para confirmar que esté bien. Y efectivamente, todo se ha transferido correctamente.

## 2. Cubo OLAP

Para esto, iniciaremos un nuevo proyecto con una conexión a la base de datos que ya creamos para el ejemplo anterior “PVeicular”



Administrador de conexiones

Proveedor: OLE DB nativo\SQL Server Native Client 11.0

Nombre del servidor:

Actualizar

Conexión con el servidor

Autenticación: Autenticación de Windows

Nombre de usuario:

Contraseña:

☐ Guardar mi contraseña

Establecer conexión con una base de datos

☒ Seleccionar o escribir el nombre de la base de datos:

PVeicular

☐ Adjuntar un archivo de base de datos:

Examinar...

Nombre lógico:

Probar conexión

Aceptar Cancelar Ayuda



Y procedemos a crear las vistas en la conexión con el proyecto.  
Seleccionamos las siguientes tablas y las incluimos en “objetos”

Asistente para vistas del origen de datos

### Seleccionar tablas y vistas

Seleccione los objetos de la base de datos relacional que deben incluirse en la vista del origen de datos.

Objetos disponibles:

Nombre	Tipo
DataPVehicular (dbo)	Tabla
DataPVehicular3 (dbo)	Tabla
sysdiagrams (dbo)	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Objetos incluidos:

Nombre	Tipo
Capacidad (dbo)	Tabla
FechaImportacion (dbo)	Tabla
ParqueVehicular (dbo)	Tabla
Propietario (dbo)	Tabla
Tipo_Codiciones (dbo)	Tabla
Vehiculo (dbo)	Tabla

Agregar tablas relacionadas

>

<

>>

<<

< Back   Next >   Finish >>|   Cancel



Asistente para vistas del origen de datos

### Finalización del asistente

Proporcione un nombre y haga clic a continuación en Finalizar para crear la nueva vista del origen de datos.

Nombre:

P Vehicular\_

Vista previa:

P Vehicular\_

- Capacidad (dbo)
- FechaImportacion (dbo)
- ParqueVehicular (dbo)
- Propietario (dbo)
- Tipo\_Codiciones (dbo)
- Vehiculo (dbo)

< Back   Next >   **Finalizar**   Cancel

Y creamos el cubo como tal, conectando todas las tablas

Asistente para cubos

### Seleccionar tablas de grupo de medida

Seleccione una vista o diagrama de origen de datos y, a continuación, seleccione las tablas que se usarán para los grupos de medida.

Vista del origen de datos:

P Vehicular

Tablas de grupo de medida:


Sugerir

- ☐ Capacidad
- ☐ FechaImportacion
- ☐ Propietario
- ☐ Tipo\_Codiciones
- ☐ Vehiculo
- ☒ ParqueVehicular


< Back   **Next >**   Finish >>   Cancel






**Seleccionar medidas**

Seleccione las medidas que desea incluir en el cubo.



☒ Medida

☒  Parque Vehicular


- ☒  CANTIDAD DE CILINDROS
- ☒  CANTIDAD DE PUERTAS
- ☒  VALOR DEL VEHICULO
- ☒  CAPACIDAD
- ☒  Recuento Parque Vehicular

Seleccionamos las dimensiones que están conformadas por las tablas Vehiculo, Tipo\_Codiciones, FechaImportacion, Capacidad, y Propietario






Asistente para cubos

**Seleccionar dimensiones existentes**

Seleccione las dimensiones existentes que se incluirán en el cubo.

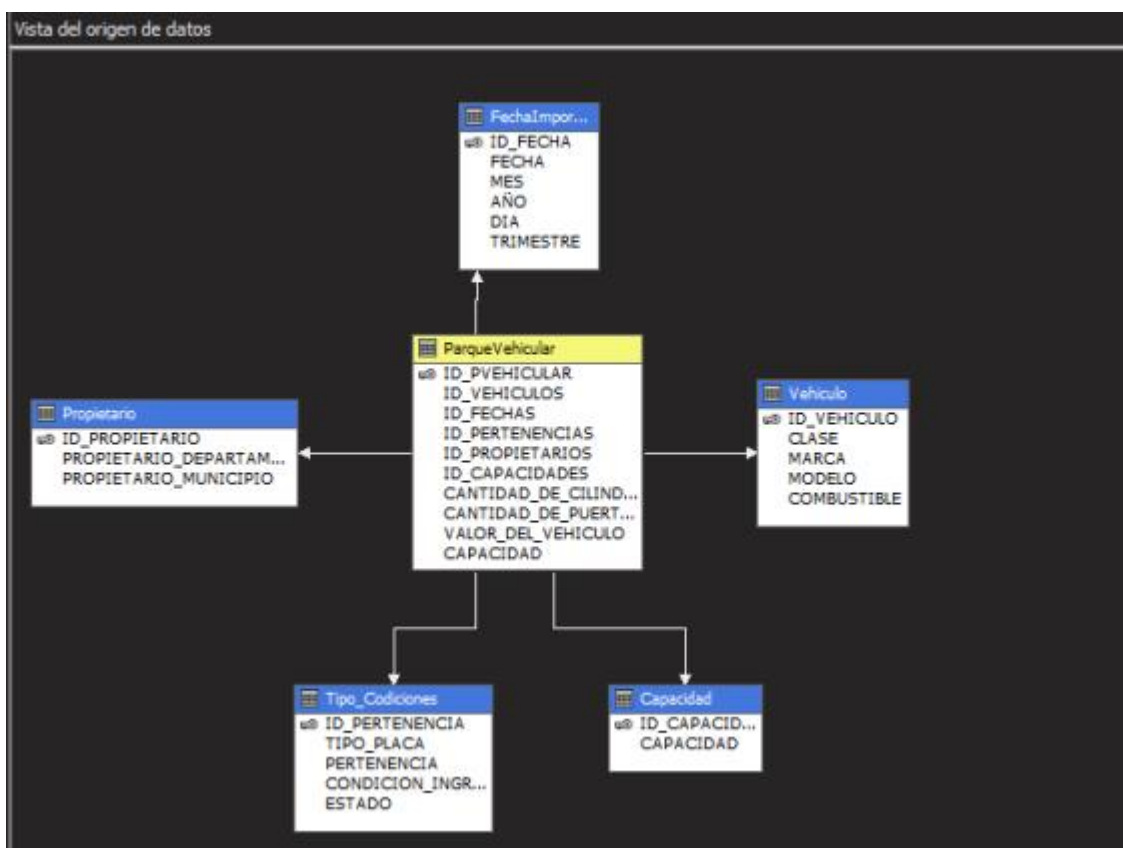


☒ Dimensión

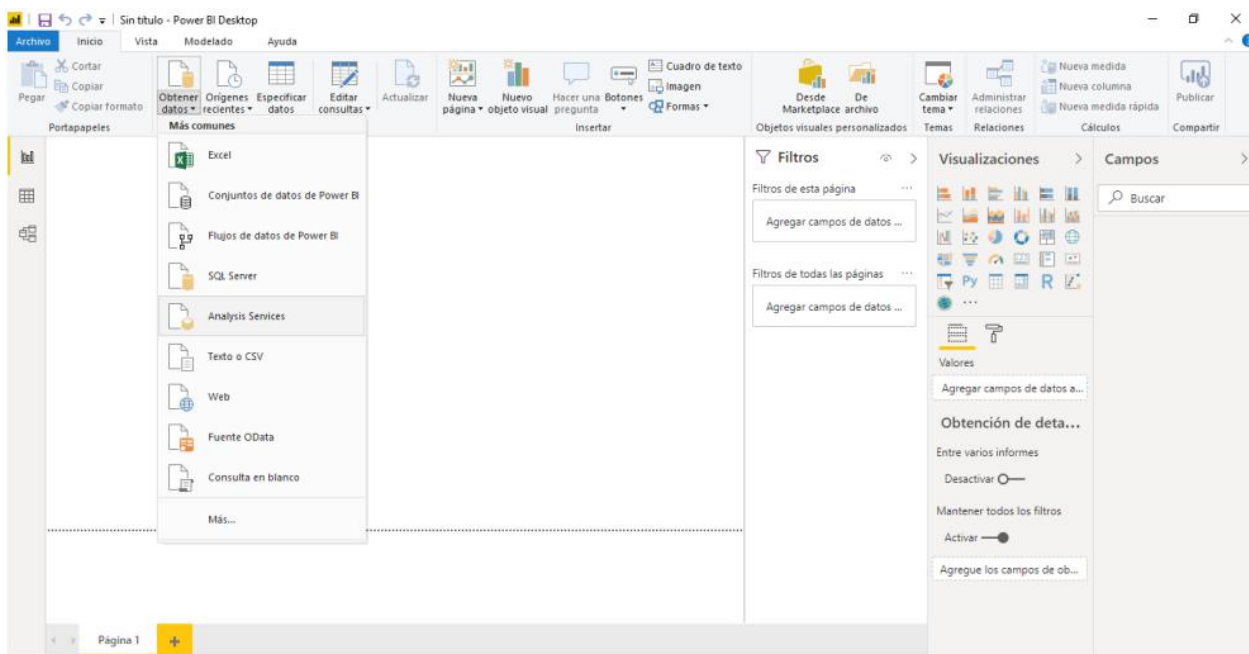
- ☒  Vehiculo
- ☒  Tipo Codiciones
- ☒  Fecha Importacion
- ☒  Capacidad
- ☒  Propietario

< Back   Next >   Finish >>   Cancel

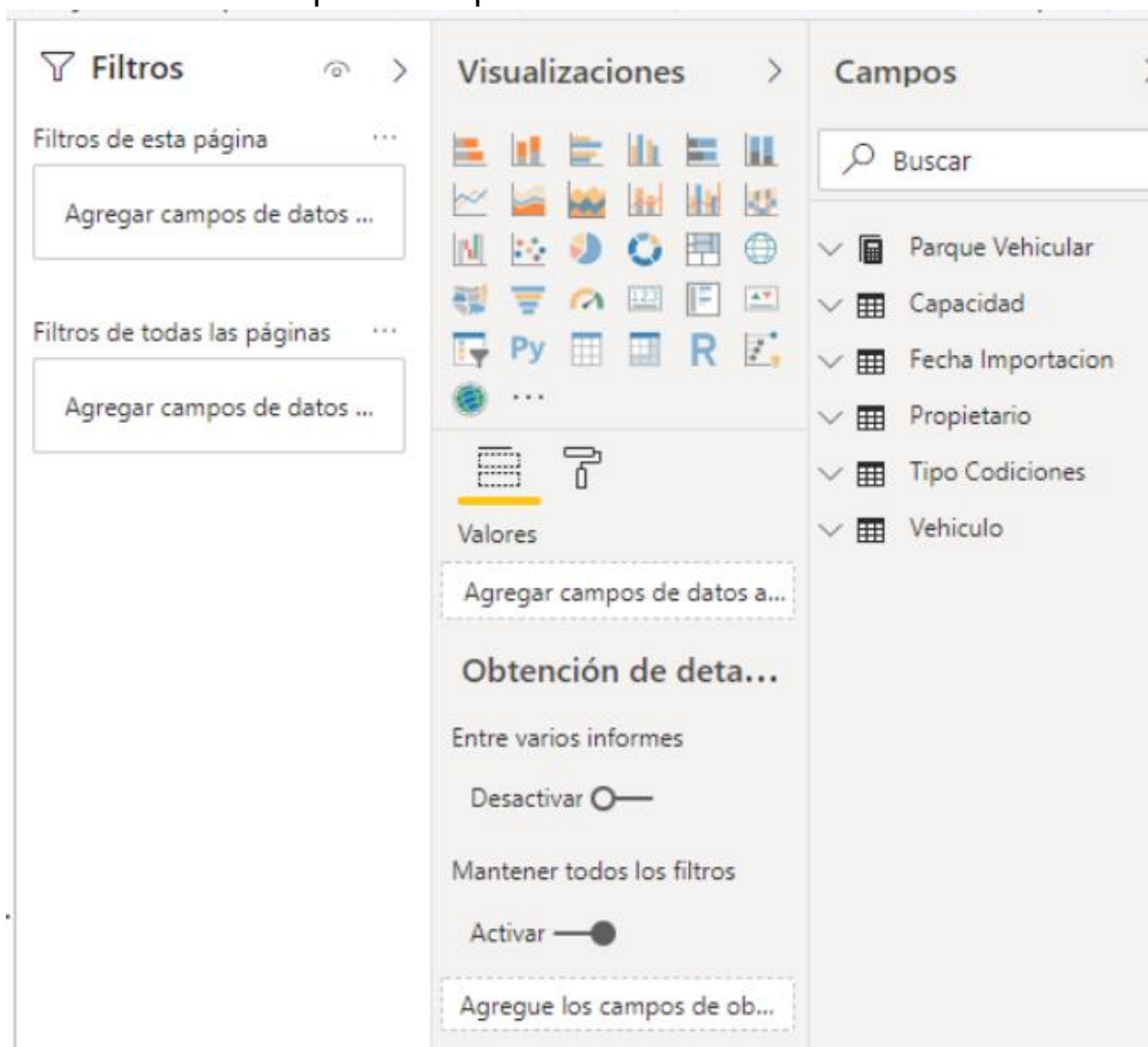
En el cubo creado, podemos encontrar el diagrama modelo estrella, similar al que se hizo en SQL SERVER



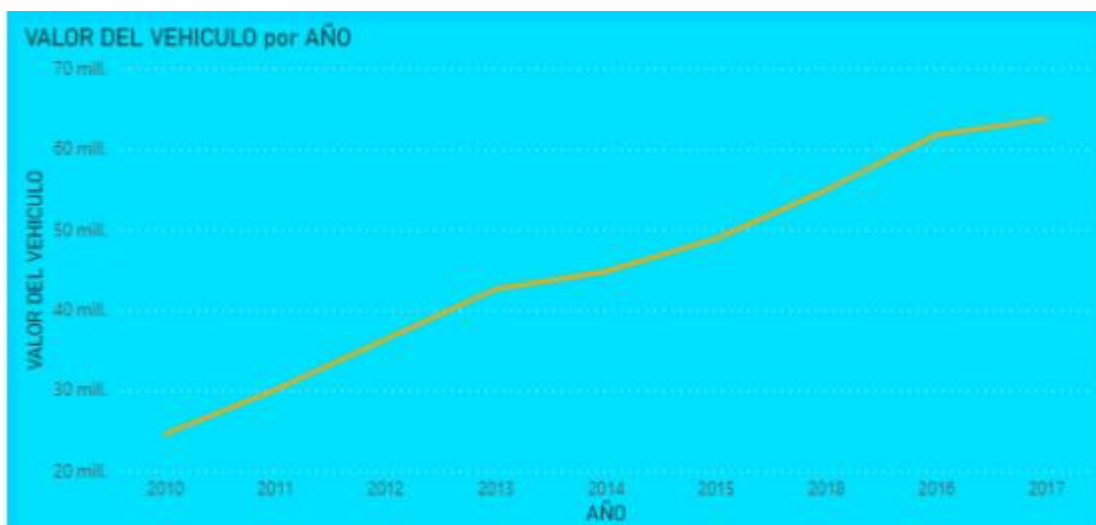
Y procederemos a analizar todos estos datos en la herramienta en “PowerBI”

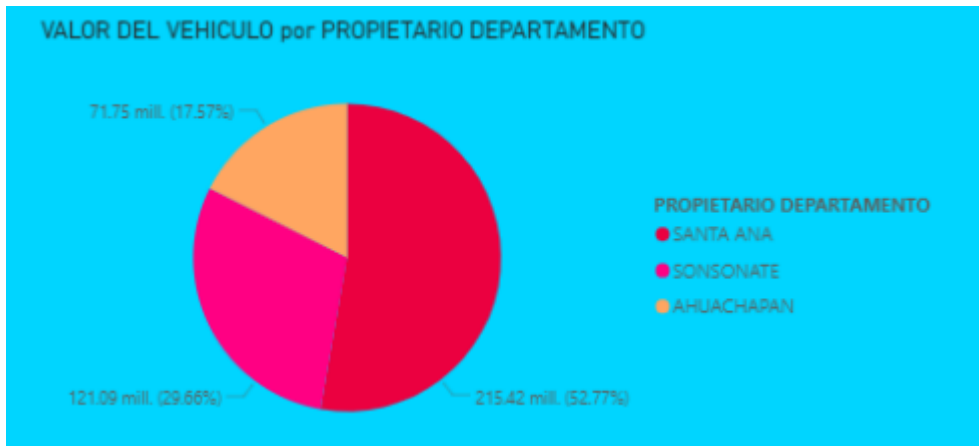


Realizamos una conexión con la base de datos “PVeicular”.  
Al realizar esto, podemos ver las dimensiones y medidas del cubo y otros elementos importantes para analizar la información

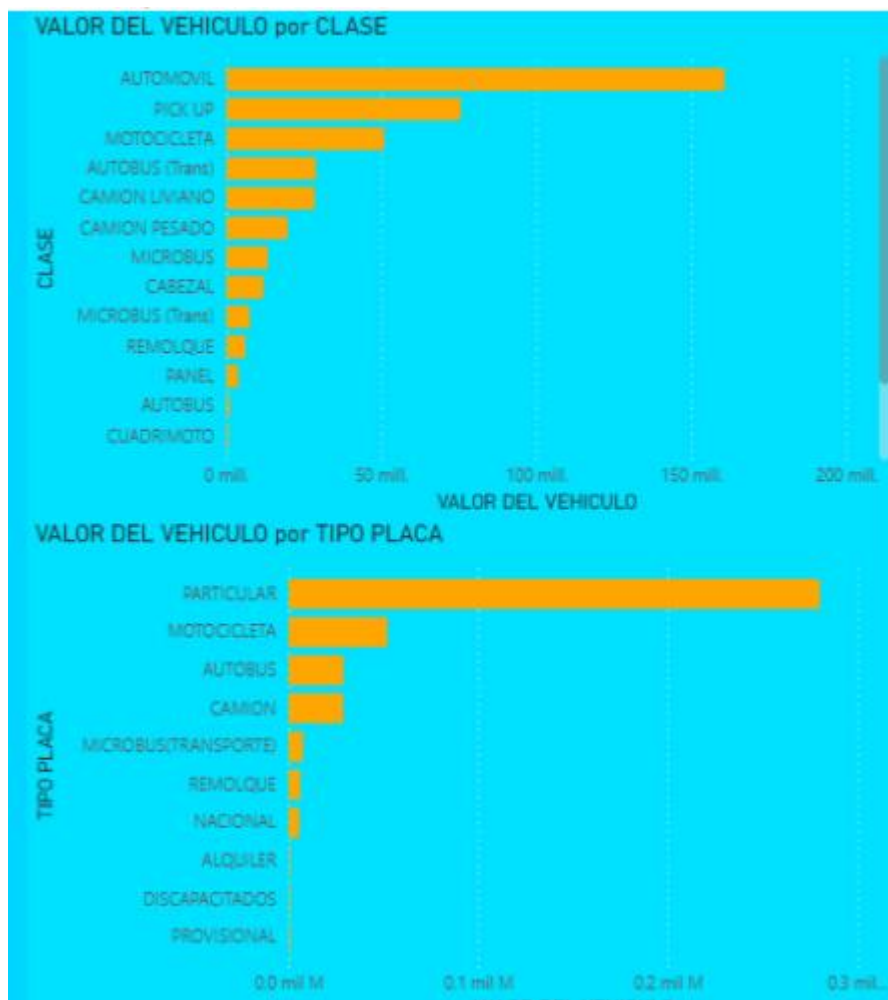


Elegiremos mostrar un gráfico de pastel, para poder ver la participación y el costo del vehículo según el departamento, una gráfica lineal también para ver el valor por año y el año más rentable. Seleccionamos el filtro departamento y el año para el filtrado de la información.





Y así podemos crear también por ejemplo gráficos de barras, comparando valor del vehículo con el tipo de vehículo, placa y etc.



De esta manera, podemos ver la cantidad de vehículos según el tipo por año (por ejemplo)