

**UNIVERSIDAD DON BOSCO**  
**INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**  
**DATAWAREHOUSE Y MINERÍA DE DATOS**  
**CICLO 02-2020**



**Entrega del tercer desafío practico**

**PRESENTADO POR:**

**GIRÓN CORNEJO, EMMANUEL ANTONIO GC180313**  
**ZEPEDA LÓPEZ, EDGARDO ANIBAL L180073**

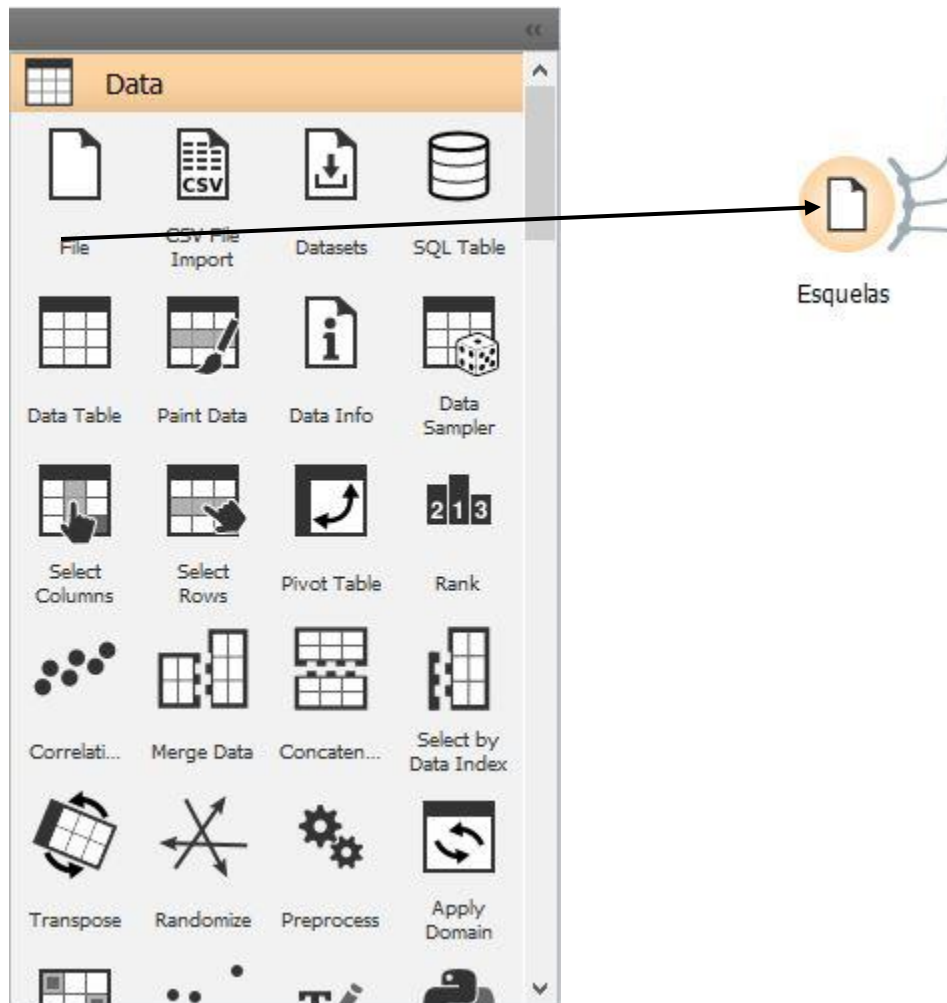
**DOCENTE: Alexander Sigüenza**

**Grupo: 01L**

**Soyapango, 15 de noviembre de 2020**

## ARBOL DE DECISION (HECHO EN ORANGE DATAMINING).

. Para el análisis de los algoritmos utilizados, empezaremos con el árbol de decisión hecho en Orange, primero en el programa se nos muestra un panel que dice data donde se puede observar varios tipos de orígenes de datos. En este caso tomaremos el primero el que dice file para poder cargar cualquier tipo de archivo dataset y tipo table luego lo arrastramos al lienzo en blanco.



. Luego damos click al origen de datos, y se nos muestra un formulario, debemos ir al apartado de archivo con puntos suspensivos para cargar nuestro dataset. En este caso se usará el de esquelas. Y luego le damos aceptar. Ya luego de haber cargado el archivo se nos mostrara los nombres de las columnas, el tipo de dato que en este caso lo toma por categoría de datos, los roles si son de función o objetivo, los valores que incluye el dataset.

The image shows a two-step process for loading a dataset. First, a file explorer window displays the desktop contents, with 'Esquelas.csv' highlighted. An arrow indicates the selection of this file. Second, the 'Esquelas' panel in the data tool is shown, where 'Esquelas.csv' is loaded. The panel includes a file list, a 'Reload' button, and a table of columns.

**File Explorer Details:**

- Location: Este equipo > Escritorio
- Files: ProyectoETL-master, Recursos de guias DMD, SQL Query, Ultimo intento de reporte, **Esquelas.csv**, Park\_Vehicular\_datos\_abiertos, Recurso actividad documental Control\_Ventas, Supermarket, Supermarket
- File Name: Esquelas.csv
- File Type: All readable files (\*.basket \*.bsk)

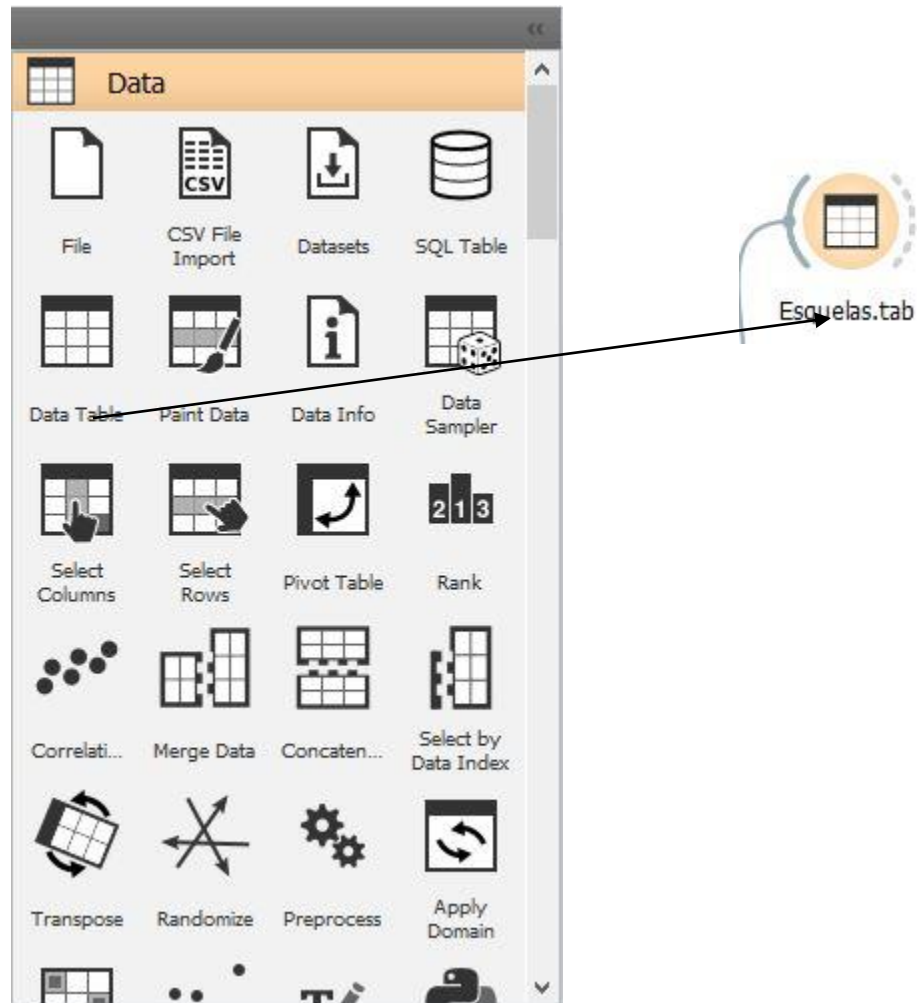
**Esquelas Panel Details:**

- File: Esquelas.csv
- Info: 44472 instance(s), 8 feature(s) (no missing values), Data has no target variable, 1 meta attribute(s)
- Columns (Double click to edit):

	Name	Type	Role	Values
1	NumEsquela	numeric	feature	
2	Fecha	datetime	feature	
3	TipoFalta	categorical	feature	TRANSITO, TRANSPORTE CARGA, TRANSPORTE TERRESTRE
4	Departamento	categorical	feature	AHUACHAPAN, CABAÑAS, CHALATENANGO, CUSCATLAN, LA ...
5	Estado	categorical	feature	ANC, CARG, CBR, INC, INI, IPR, IPT, IRI, PPZ, REV
6	EstadoDes	categorical	feature	ANALISIS COBRADA, CANCELADA, CARGADA, EN PAGOS A PLAZO, ...
7	Valor	numeric	feature	

Buttons: Browse documentation datasets, Reset, Apply

- Volvemos al panel y en este caso tomamos un modulo que dice data table y los arrastramos al lienzo y luego lo unimos al origen de datos.



. Luego le damos click al modulo y se nos mostraran todos los datos que incluye el dataset, esto nos permite poder analizar mejor el contenido del dataset. También del lado izquierdo se nos muestra sección de variables que nos permite mostrar las variables posibles a utilizar en el dataset y la opción de instancia de clases por color, quiere decir que esto nos ayudara a demostrar que tipo de reglas se pueden hacer en el dataset que se va utilizar y por ultimo la sección de selección que nos permite ver todas las filas de la tabla. Siempre se dejará enmarcada como se muestra en el data table.

Esquejas.tab

Info

44472 instances (no missing data)  
8 features  
No target variable.  
1 meta attribute

Variables

☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection

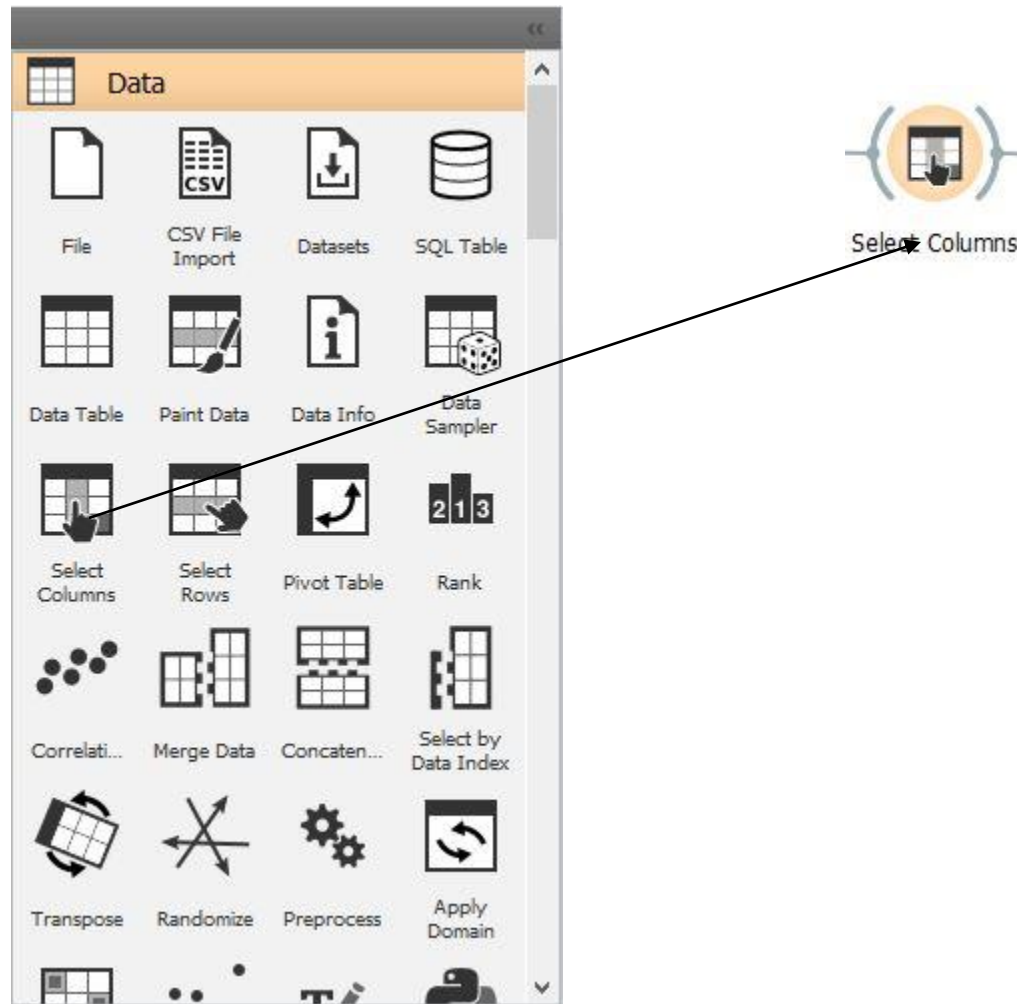
☒ Select full rows

Restore Original Order

☒ Send Automatically

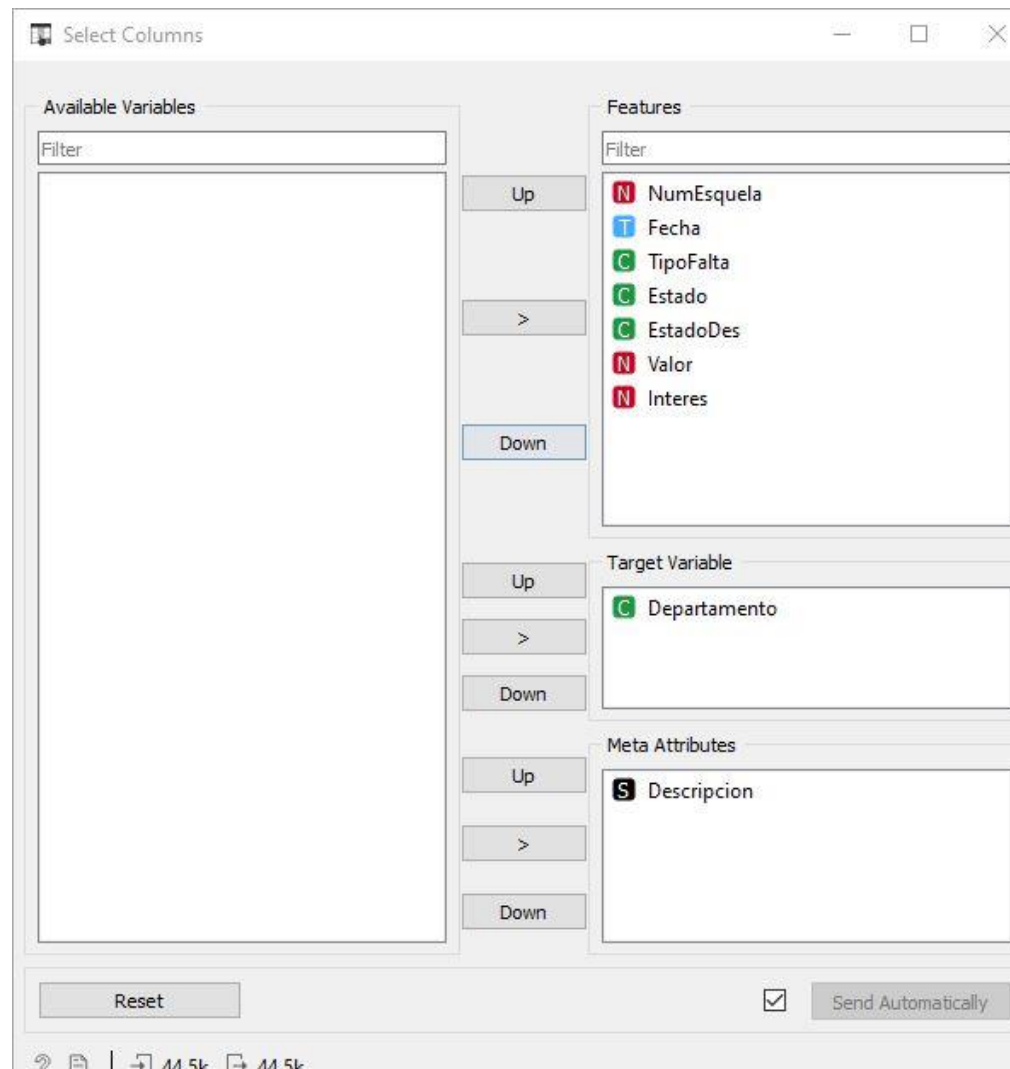
	Descripcion	NumEsquela	Fecha	TipoFalta	Departamento	Estado	EstadoDes	Valor	Interes
1	CONDUCIR EN ...	100403	2003-06-11	TRANSITO	LA LIBERTAD	CBR	CANCELADA	57.14	0.00
2	CIRCULAR A ...	100880	2003-08-06	TRANSITO	SAN SALVADOR	IPR	IMPROCEDENTE	57.14	0.00
3	CARECER ...	100881	2003-08-08	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
4	REBASAR LA ...	100883	2003-08-20	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
5	REBASAR LA ...	100884	2003-08-20	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
6	NO PORTAR LA ...	100885	2003-08-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
7	ESTACIONARSE ...	100886	2003-08-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	57.14	0.00
8	ESTACIONARSE ...	100887	2003-08-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	57.14	0.00
9	INCUMPLIR LA...	100889	2003-08-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
10	INCUMPLIR LA...	100890	2003-08-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
11	REBASAR LA ...	100892	2003-08-26	TRANSITO	SAN SALVADOR	CARG	CARGADA	34.29	21.94
12	REBASAR LA ...	100893	2003-08-26	TRANSITO	SAN SALVADOR	CARG	CARGADA	34.29	21.94
13	VIRAR EN "U" ...	100895	2003-08-28	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
14	EFFECTUAR ...	100899	2003-09-15	TRANSPORTE ...	SAN SALVADOR	INI	PENDIENTE DE ...	34.29	21.94
15	CIRCULAR CO...	101186	2003-07-18	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
16	NO UTILIZAR E...	101187	2003-07-18	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
17	NO PORTAR LA ...	101190	2003-07-18	TRANSITO	SAN SALVADOR	CARG	CARGADA	11.43	7.31
18	NO PORTAR ...	101192	2003-07-19	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
19	CIRCULAR A ...	101196	2003-07-20	TRANSITO	SAN SALVADOR	CBR	CANCELADA	57.14	0.00
20	CIRCULAR A ...	101199	2003-07-22	TRANSITO	SAN SALVADOR	CBR	CANCELADA	57.14	0.00
21	CIRCULAR CO...	101380	2003-06-11	TRANSITO	SAN SALVADOR	CBR	CANCELADA	57.14	0.00
22	NO CONCEDER...	101388	2003-06-19	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
23	NO CONCEDER...	101389	2003-06-19	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
24	EFFECTUAR ...	101391	2003-06-19	TRANSPORTE ...	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
25	EFFECTUAR ...	101392	2003-06-19	TRANSPORTE ...	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
26	NO RESPETAR ...	101393	2003-06-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	34.29	0.00
27	CARECER ...	101394	2003-06-26	TRANSITO	SAN SALVADOR	CBR	CANCELADA	11.43	0.00
28	EFFECTUAR ...	101397	2003-06-28	TRANSPORTE ...	SAN SALVADOR	CBR	CANCELADA	34.29	0.00

. Volvemos al panel nuevamente y esta vez arrastramos al lienzo el modulo para seleccionar columnas y lo unimos directamente al origen de datos file.

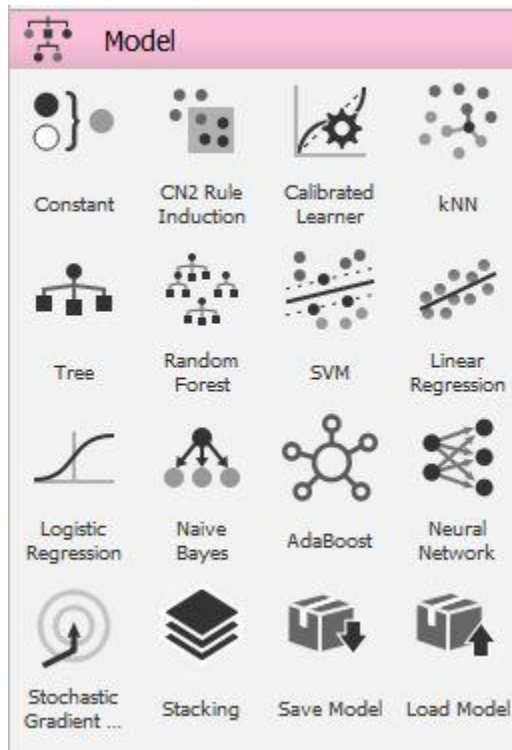




. Luego damos click al módulo y nos aparecerá un formulario que nos da varias secciones para la división de datos por columnas. En el lado izquierdo es el apartado de variables donde se van a utilizar quiere decir que se puede no usar y usar las variables de las columnas, del lado derecho esta el apartado de funciones donde aparecerá la data, la parte de abajo esta Target Variable que quiere decir la variable principal o la que predomina en el origen de datos. Y de ultimo el atributo meta quiere decir que es un atributo que normalmente en Orange lo pone como tipo text. Y para el análisis dejamos de la siguiente manera: Como variable objetivo el departamento.



. Volvemos al panel y nos vamos al apartado de modelos, donde seleccionamos el módulo tree, luego lo arrastramos al lienzo y desde el select column lo unimos al módulo tree, luego se nos muestra un formulario para el árbol donde nos sale la configuración y características del árbol de decisión, la dejamos los valores por defecto.



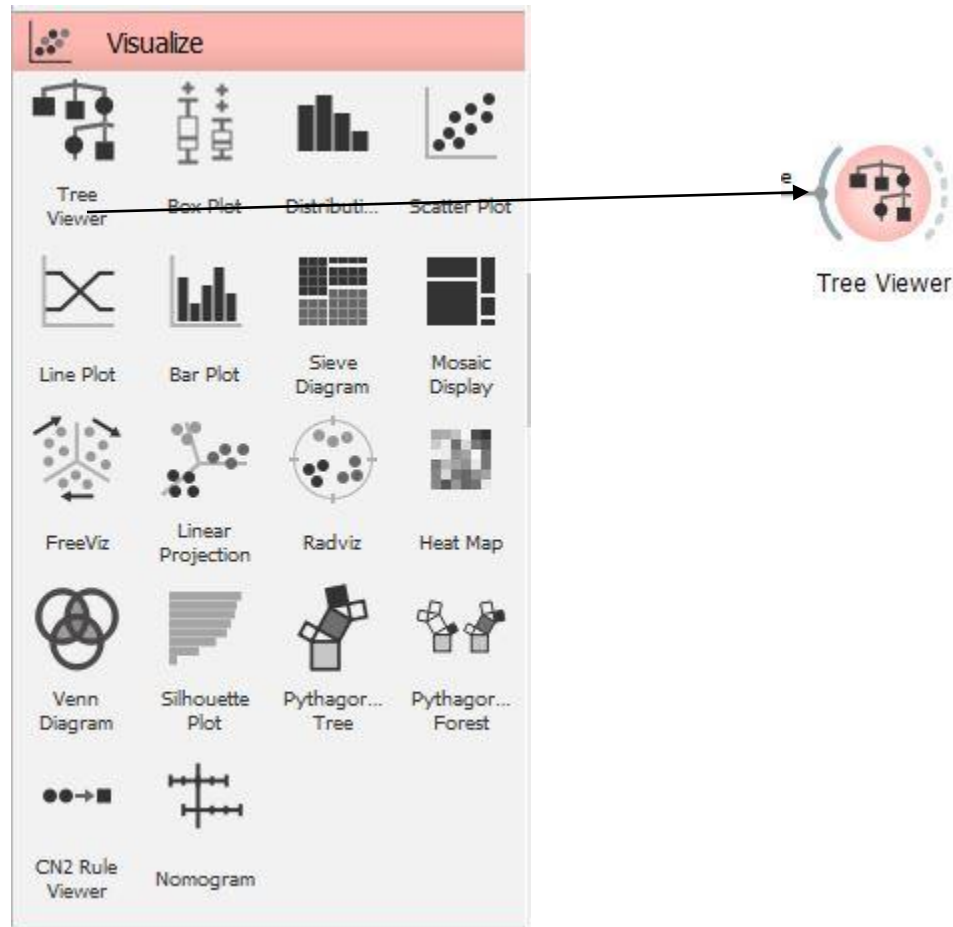
The 'Tree' configuration window shows the following settings:

- Name: Tree
- Parameters:
  - ☐ Induce binary tree
  - ☒ Min. number of instances in leaves: 4
  - ☒ Do not split subsets smaller than: 8
  - ☒ Limit the maximal tree depth to: 120
- Classification:
  - ☒ Stop when majority reaches [%]: 95
- ☒ Apply Automatically

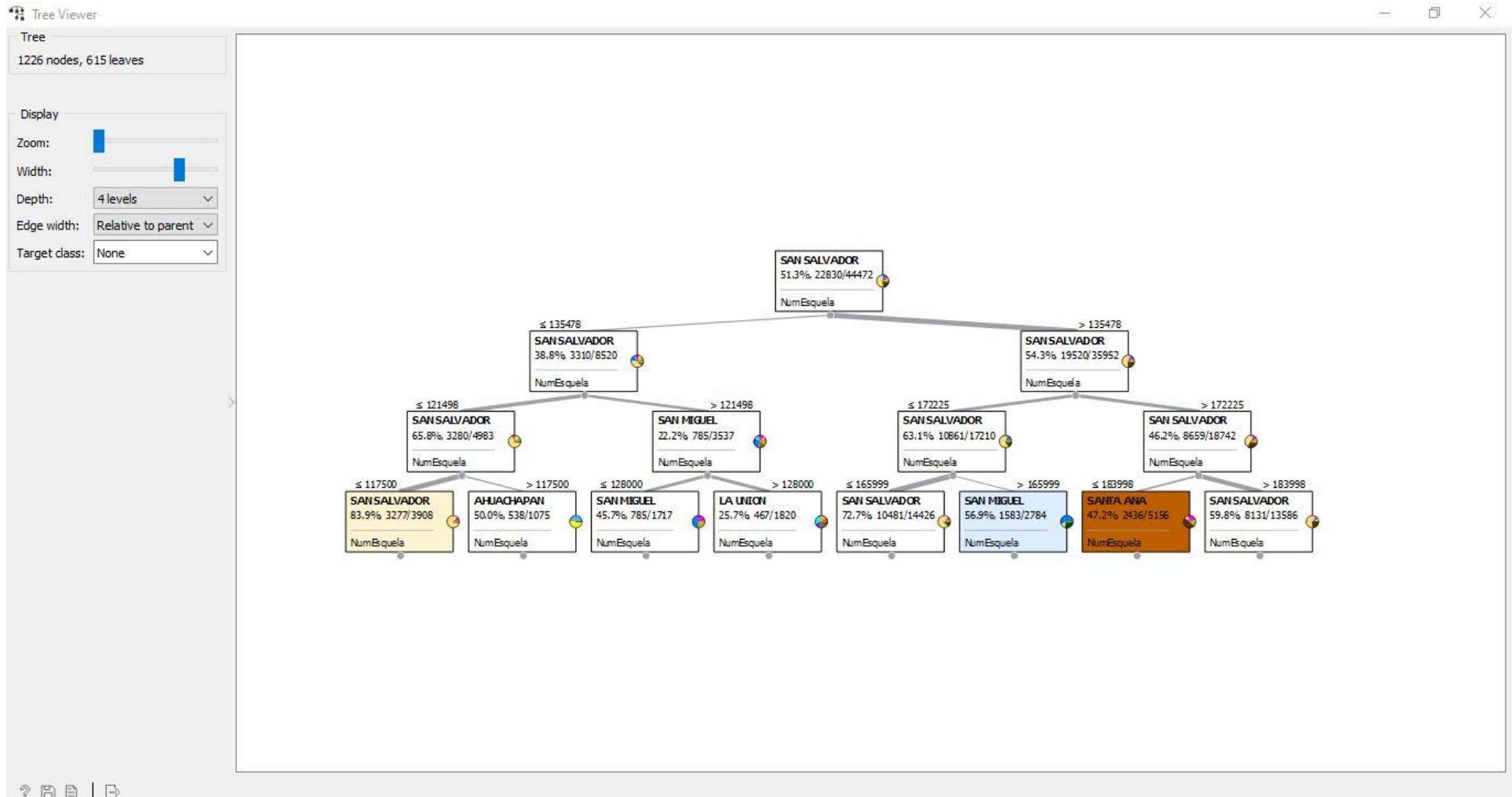
At the bottom, there is a status bar showing a question mark, a document icon, and a file size of 44.5k.



. Luego en el panel vamos al apartado de visualizaciones y arrastramos el módulo de tree viewer, en este modulo es donde podemos visualizar el árbol de decisión cumpliendo el análisis del algoritmo, para unir la visualización se debe unir desde el modulo tree al módulo de tree viewer.

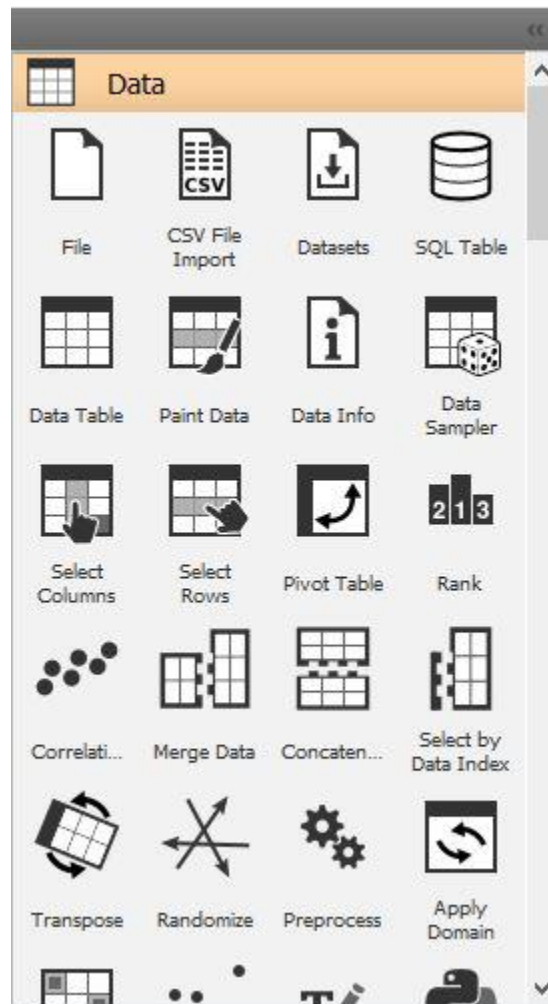


. Luego le damos click al módulo y podemos observar el árbol de decisiones donde en el panel izquierdo podemos elegir los niveles que conforman el árbol. Donde podemos notar la variable objetivo que cumple con la decisión que toma el árbol por departamento y la cantidad de escuelas por departamento.

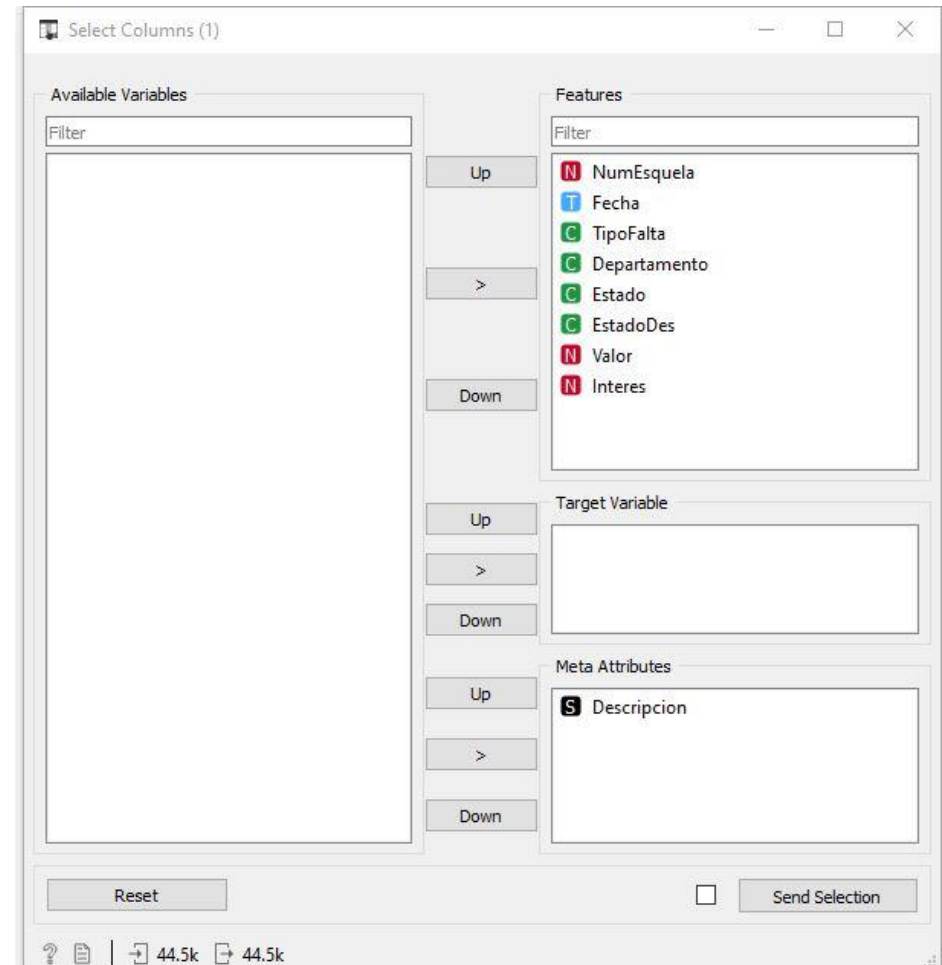
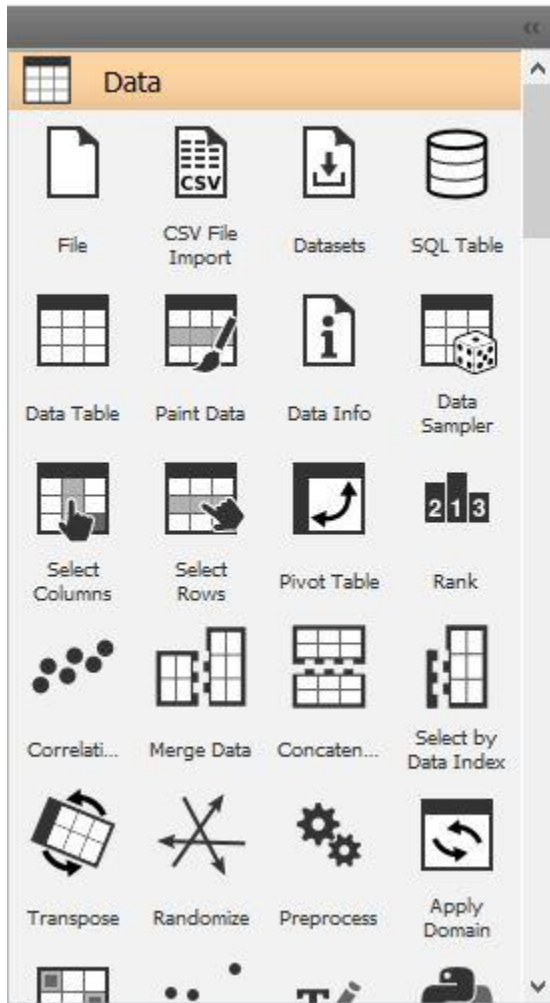


## REGLAS DE ASOCIACIÓN (HECHO EN ORANGE DATAMINING).

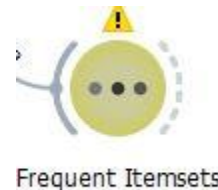
. Arrastramos otro archivo file al lienzo y con los mismos pasos de poder cargar un dataset. Y también un data table para poder unir del origen de datos al data table.



. De la misma forma, tomamos un select column para poder organizar los datos y el análisis en este caso será diferente ya que para reglas de asociación en Orange datamining, no es necesario tener una variable objetivo para que el algoritmo cumpla su función. Simplemente va generando las reglas necesarias del origen de datos y como se puede observar se toma todos los datos como función y el atributo por defecto.



. Ahora en el panel, vamos al apartado de asociación, y arrastramos al lienzo el módulo de frequent itemsets. Este modulo nos muestra la cantidad de ítems que se encuentran en el origen de datos, en este caso se conecta desde el select column. Luego de eso le damos click y se nos mostrara un panel con características de itemsets este nos permite definir la cantidad de support y máximo número de ítems que se quiere mostrar. Le damos click en find itemsets y nos mostrara el support y el porcentaje de cada ítem del dataset en este caso tiene un 10% de support mínimo y un máximo de ítems de 10,000.



\*\*\* Frequent Itemsets

Info

Number of itemsets: 31

Selected itemsets: 0

Selected examples: 0

Expand all Collapse all

Find itemsets

Minimal support: 10%

Max. number of itemsets: 10000

Find Itemsets

Filter itemsets

Contains:

Min. items: 1 Max. items: 999

Apply these filters in search

Send Selection Automatically

Itemsets	Support	%
▼ TipoFalta=TRANSITO	34776	78.2
▼ Departamento=SAN SALVADOR	16142	36.3
▼ Estado=CBR	13951	31.37
EstadoDes=CANCELADA	13951	31.37
EstadoDes=CANCELADA	13951	31.37
▼ Departamento=SANTA ANA	5225	11.75
▼ Estado=CBR	4560	10.25
EstadoDes=CANCELADA	4560	10.25
EstadoDes=CANCELADA	4560	10.25
▼ Estado=CBR	29921	67.28
EstadoDes=CANCELADA	29921	67.28
EstadoDes=CANCELADA	29921	67.28
▼ TipoFalta=TRANSPORTE TERRESTRE	9379	21.09
▼ Departamento=SAN SALVADOR	6588	14.81
▼ Estado=CBR	4745	10.67
EstadoDes=CANCELADA	4745	10.67
EstadoDes=CANCELADA	4745	10.67
▼ Estado=CBR	7009	15.76
EstadoDes=CANCELADA	7009	15.76
EstadoDes=CANCELADA	7009	15.76
▼ Departamento=SAN SALVADOR	22830	51.34
▼ Estado=CBR	18786	42.24
EstadoDes=CANCELADA	18786	42.24
EstadoDes=CANCELADA	18786	42.24
▼ Departamento=SANTA ANA	5640	12.68
▼ Estado=CBR	4879	10.97
EstadoDes=CANCELADA	4879	10.97
EstadoDes=CANCELADA	4879	10.97
▼ Estado=CBR	37220	83.69
EstadoDes=CANCELADA	37220	83.69
EstadoDes=CANCELADA	37220	83.69

Data has continuous attributes which will be skipped.



. Ahora volvemos al panel y esta vez seleccionamos el modulo de Association Rules, las reglas de asociación lo arrastramos al lienzo y lo conectamos directamen al origen de datos que es el file. Y ahora le damos click al modulo ya conectado. En el panel izquierdo es donde podemos configurar el minimo de support y minimo de confidence y también la cantidad maxima de reglas la cual se ha tomado en support minimo del 16% y de confidence 23%. El numero maximo de reglas se deja por defecto de 10,000.



Association Rules

Info

Number of rules: 50  
Filtered rules: 50  
Selected rules: 0  
Selected examples: 0

Find association rules

Minimal support: 16%  
Minimal confidence: 23%  
Max. number of rules: 10000

☐ Induce classification (itemset → class) rules

☐ Find Rules

Filter rules

Antecedent

Contains:

Min. items: 1 Max. items: 999

Consequent

Contains:

Min. items: 1 Max. items: 999

☒ Apply these filters in search

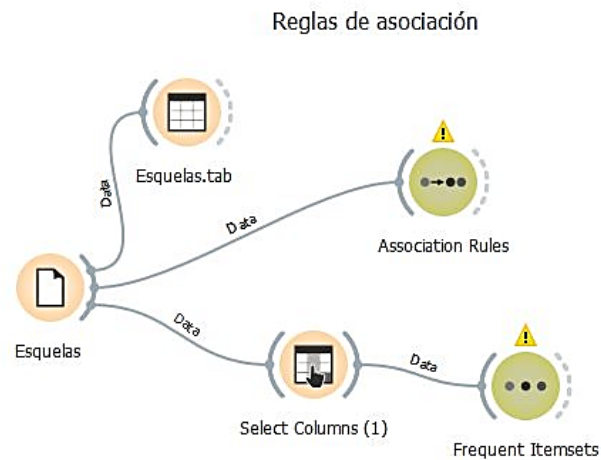
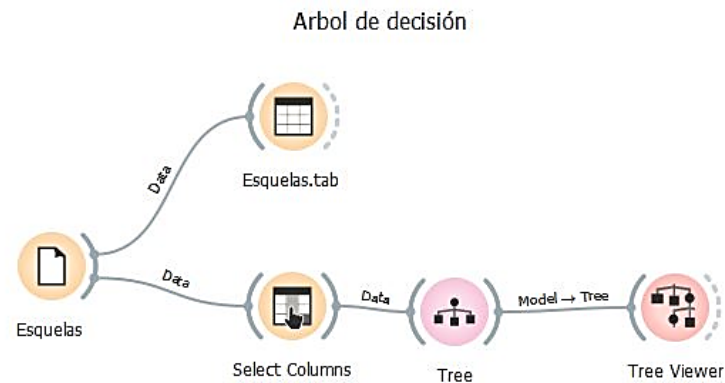
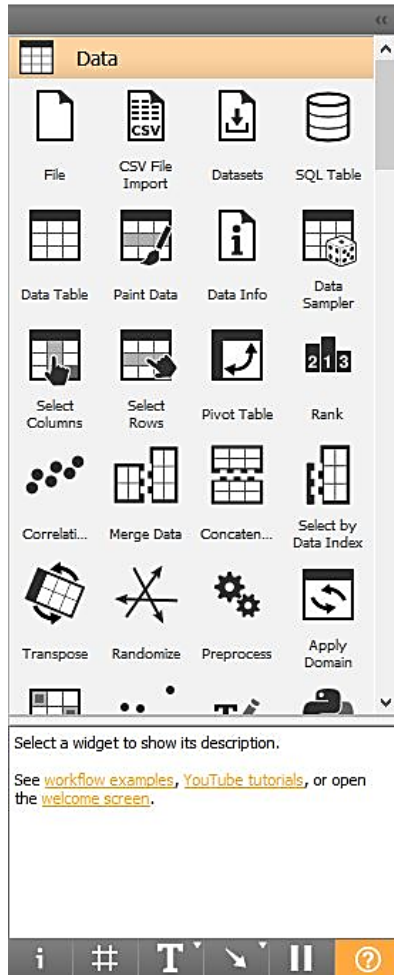
☐ Send selection

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.837	1.000	0.837	1.000	1.195	0.136	EstadoDes= CANCELADA	Estado= CBR
0.837	1.000	0.837	1.000	1.195	0.136	Estado= CBR	EstadoDes= CANCELADA
0.673	0.804	0.837	0.934	1.028	0.018	Estado= CBR	TipoFalta= TRANSITO
0.673	0.860	0.782	1.070	1.028	0.018	TipoFalta= TRANSITO	Estado= CBR
0.673	0.804	0.837	0.934	1.028	0.018	EstadoDes= CANCELADA	TipoFalta= TRANSITO
0.673	0.860	0.782	1.070	1.028	0.018	TipoFalta= TRANSITO	EstadoDes= CANCELADA
0.673	0.804	0.837	0.934	1.028	0.018	Estado= CBR, EstadoDes= CANCELADA	TipoFalta= TRANSITO
0.673	1.000	0.673	1.244	1.195	0.110	TipoFalta= TRANSITO, EstadoDes= CANCELADA	Estado= CBR
0.673	0.804	0.837	0.804	1.195	0.110	EstadoDes= CANCELADA	TipoFalta= TRANSITO, Estado= CBR
0.673	1.000	0.673	1.244	1.195	0.110	TipoFalta= TRANSITO, Estado= CBR	EstadoDes= CANCELADA
0.673	0.804	0.837	0.804	1.195	0.110	Estado= CBR	TipoFalta= TRANSITO, EstadoDes= CANCELADA
0.673	0.860	0.782	1.070	1.028	0.018	TipoFalta= TRANSITO	Estado= CBR, EstadoDes= CANCELADA
0.422	0.823	0.513	1.630	0.983	-0.007	Departamento= SAN SALVADOR	Estado= CBR
0.422	0.505	0.837	0.613	0.983	-0.007	Estado= CBR	Departamento= SAN SALVADOR
0.422	0.505	0.837	0.613	0.983	-0.007	EstadoDes= CANCELADA	Departamento= SAN SALVADOR
0.422	0.823	0.513	1.630	0.983	-0.007	Departamento= SAN SALVADOR	EstadoDes= CANCELADA
0.422	1.000	0.422	1.981	1.195	0.069	Departamento= SAN SALVADOR, EstadoDes= CANCELADA	Estado= CBR
0.422	0.505	0.837	0.505	1.195	0.069	EstadoDes= CANCELADA	Departamento= SAN SALVADOR, Estado= CBR
0.422	0.505	0.837	0.613	0.983	-0.007	Estado= CBR, EstadoDes= CANCELADA	Departamento= SAN SALVADOR
0.422	1.000	0.422	1.981	1.195	0.069	Departamento= SAN SALVADOR, Estado= CBR	EstadoDes= CANCELADA
0.422	0.823	0.513	1.630	0.983	-0.007	Departamento= SAN SALVADOR	Estado= CBR, EstadoDes= CANCELADA
0.422	0.505	0.837	0.505	1.195	0.069	Estado= CBR	Departamento= SAN SALVADOR, EstadoDes= CANCELADA
0.363	0.707	0.513	1.523	0.904	-0.038	Departamento= SAN SALVADOR	TipoFalta= TRANSITO
0.363	0.464	0.782	0.656	0.904	-0.038	TipoFalta= TRANSITO	Departamento= SAN SALVADOR
0.314	0.743	0.422	1.851	0.950	-0.017	Departamento= SAN SALVADOR, Estado= CBR	TipoFalta= TRANSITO
0.314	0.864	0.363	2.306	1.033	0.010	TipoFalta= TRANSITO, Departamento= SAN SALVADOR	Estado= CBR
0.314	0.611	0.513	1.311	0.908	-0.032	Departamento= SAN SALVADOR	TipoFalta= TRANSITO, Estado= CBR
0.314	0.401	0.782	0.540	0.950	-0.017	TipoFalta= TRANSITO	Departamento= SAN SALVADOR, Estado= CBR
0.314	0.466	0.673	0.763	0.908	-0.032	TipoFalta= TRANSITO, Estado= CBR	Departamento= SAN SALVADOR
0.314	0.375	0.837	0.434	1.033	0.010	Estado= CBR	TipoFalta= TRANSITO, Departamento= SAN SALVADOR

Data has continuous attributes which will be skipped.

## DIAGRAMAS.

Se puede notar que en reglas de asociación en los módulos aparece un símbolo de advertencia, no es nada grave en este caso. Lo que pasa es que cuando variables son continuas, son reorganizados por la cantidad de datos que contiene el dataset.





## DMD104G01L Ejercicio Recuperación de datos en un parque vehicular

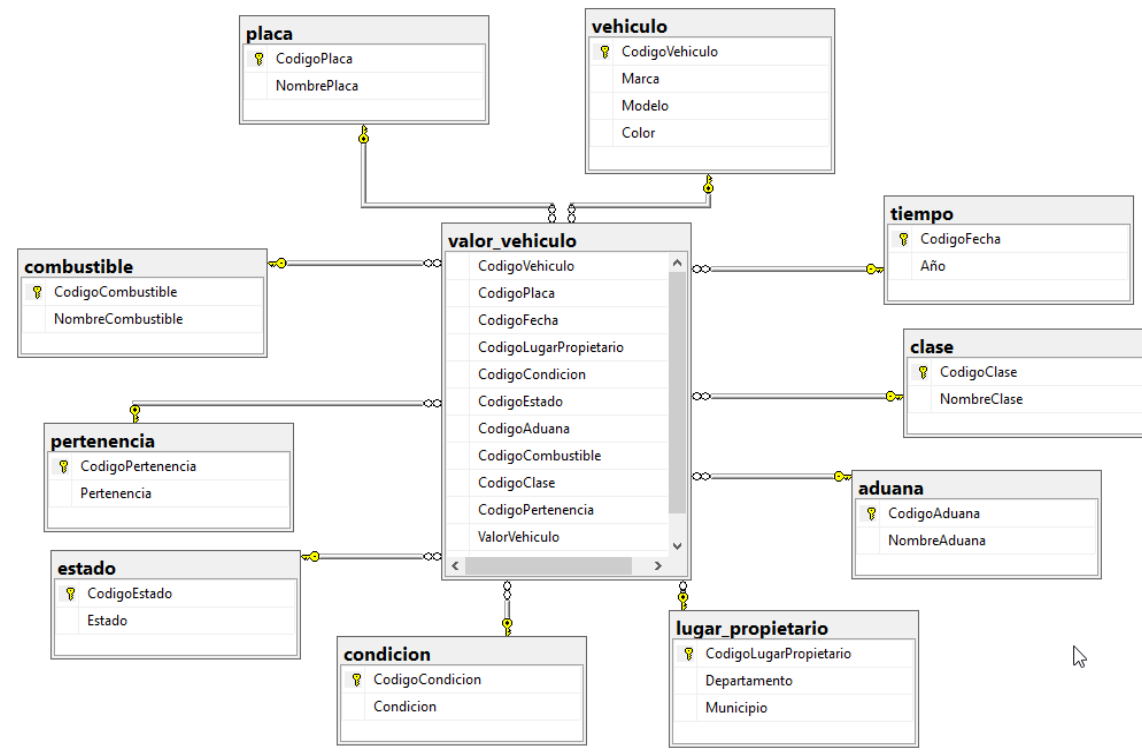
Para poder ayudar al Ministerio de Obras Públicas (MOP) , de la manera más efectiva, para así presentar informes al gobierno central para el cierre del año 2018, nos hemos apoyado de en las estrategias de minería de datos como lo son tanto el Cubo OLAP como los gráficos elaborados en la herramienta de Power Bi.

1. Preliminarmente, decidimos convertir el archivo: parque\_vehicular.csv en formato .xlsx, para una mayor comodidad a la hora de hacer el proceso ETL, sumado a eso, decidimos usar una muestra de 10,00 datos, debido a las limitantes de nuestros equipos de procesar el 1 millón de datos que originalmente tiene el archivo proporcionado. Sin embargo, consideramos que la muestra es lo suficientemente grande como para obtener resultado con fines prácticos de la actividad.

A1																					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Codigo	TIPO_PLA	ANIO_DE	CILINDRA	CANTIDAD	CANTIDAD	VALOR_D	COLORES	FECHA_DE	IMP_VALC	FECHA_DE	ANIO_ING	MES_ING	CLASE	PERTENEN	MARCA	MODELO	CAPACIDA	DES_CAPA	COMBUST	ADUA
2	1	PARTICUL	1990	1800	0	4	4094.56	AMARILLC	#####	4094.56	#####	1994	11	AUTOMOV	PROPIEDA	NISSAN	N/D	5	ASIENTOS	GASOLINA	SAN B
3	2	PARTICUL	1964	0	0	0	0	AMARILLO		0	#####	1989	2	AUTOMOV	PROPIEDA	TOYOTA	TIARA	5	ASIENTOS	GASOLINA	
4	3	ALQUILER	1984	1700	0	0	0	AMARILLO	F/C B/N	0	#####	1985	3	ALQUILER	PROPIEDA	TOYOTA	CORONA	5	ASIENTOS	GASOLINA	
5	4	ALQUILER	1986	1600	0	0	0	AMARILLO	F/C B/N	0	#####	1988	2	ALQUILER	PROPIEDA	HYUNDAI	STELLAR	5	ASIENTOS	GASOLINA	
6	5	ALQUILER	1979	0	0	0	0	AMARILLO		0	#####	1982	10	ALQUILER	PROPIEDA	DATSUN	210	5	ASIENTOS	GASOLINA	
7	6	PARTICUL	1974	1600	0	0	0	AMARILLO		0	#####	1982	10	AUTOMOV	PROPIEDA	DATSUN	1600	5	ASIENTOS	DIESEL	
8	7	ALQUILER	1975	0	0	0	0	AMARILLO		0	#####	1984	5	ALQUILER	PROPIEDA	TOYOTA	CORONA	5	ASIENTOS	GASOLINA	
9	8	ALQUILER	1973	0	0	0	0	AMARILLO	F/C B/N	0	3/6/1988	1988	6	ALQUILER	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	
10	9	ALQUILER	1975	0	0	0	0	AMARILLO		0	4/2/1988	1988	2	ALQUILER	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	
11	10	ALQUILER	1968	0	0	0	800	AMARILLC	#####	0	#####	1983	8	ALQUILER	PROPIEDA	TOYOTA	N/D	5	ASIENTOS	GASOLINA	
12	11	ALQUILER	1977	0	0	0	0	AMARILLO		0	#####	1984	6	ALQUILER	PROPIEDA	DATSUN	120 Y	5	ASIENTOS	GASOLINA	
13	12	ALQUILER	1965	0	0	0	0	AMARILLO		0	#####	1982	10	ALQUILER	PROPIEDA	PEUGEOT	404	5	ASIENTOS	GASOLINA	
14	13	ALQUILER	1983	1587	0	4	4183.76	AMARILLC	#####	4183.76	#####	1995	3	ALQUILER	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	SAN B
15	14	ALQUILER	1978	0	0	0	0	AMARILLO		0	#####	1989	4	ALQUILER	PROPIEDA	SUBARU	N/D	5	ASIENTOS	GASOLINA	
16	15	PARTICUL	1981	1770	0	0	4413.71	AMARILLC	#####	4413.71	#####	1993	10	AUTOMOV	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	SAN B
17	16	ALQUILER	1967	0	0	0	0	AMARILLO		0	#####	1982	10	ALQUILER	PROPIEDA	TOYOTA	CORONA	5	ASIENTOS	GASOLINA	
18	17	PARTICUL	1982	1770	0	4	4619.44	AMARILLC	#####	4619.44	#####	1994	5	AUTOMOV	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	SAN B
19	18	ALQUILER	1974	0	0	0	0	AMARILLO		0	#####	1978	5	ALQUILER	PROPIEDA	MAZDA	808	5	ASIENTOS	GASOLINA	
20	19	PARTICUL	1973	1600	0	0	0	AMARILLO		0	#####	1993	2	AUTOMOV	PROPIEDA	TOYOTA	CORONA	5	ASIENTOS	GASOLINA	
21	20	PARTICUL	1978	0	0	0	0	AMARILLO		0	2/2/1989	1989	2	AUTOMOV	PROPIEDA	TOYOTA	COROLLA	5	ASIENTOS	GASOLINA	
22	21	PARTICUL	1987	0	0	0	0	AMARILLO		0	#####	1989	2	AUTOMOV	PROPIEDA	TOYOTA	CORONA	5	ASIENTOS	GASOLINA	
23	22	PARTICUL	1978	0	0	0	0	AMARILLO		0	#####	1986	7	AUTOMOV	PROPIEDA	MAZDA	N/D	5	ASIENTOS	GASOLINA	

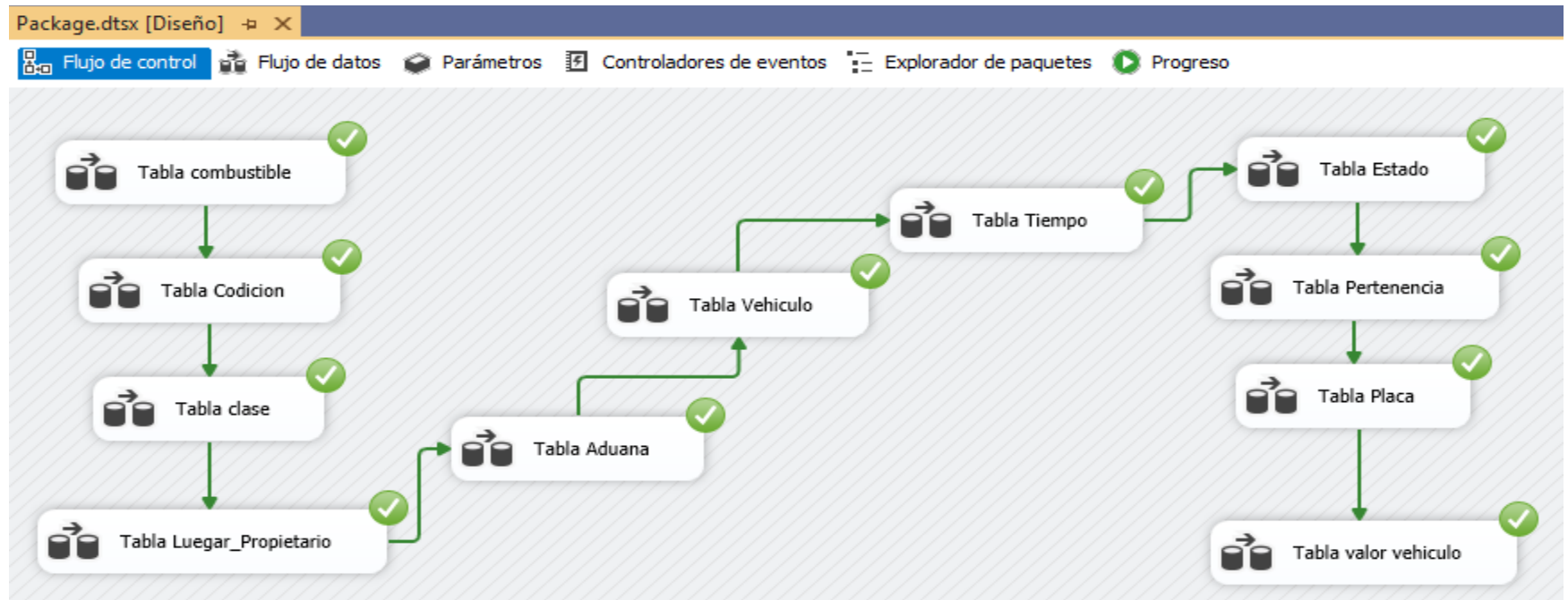
Nota: Cabe mencionar que se agrego una columna de código, utilizando una de las funciones de correlativo de Excel.

2. Como segundo paso, vimos la necesidad de crear una base de datos disponibles. Considerando, claro, los campos más relevantes en nuestra investigación y además de desechar los que no consideramos interesantes, como lo son: año de fabricación del vehículo, cilindrada, cantidad de cilindros, número de puertas, y fecha de importación entre otros. Podemos verlo en más detalle en el diagrama tipo estrella resultante:

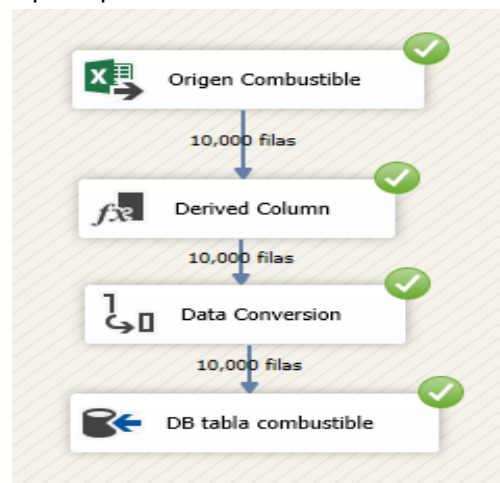


Nota: Considerando la gran cantidad de dimensiones resultantes, hemos omitido una parte en el cubo. Cabe tener en cuenta que se ha buscado englobar todas las características de los vehículos que ingresan al parque vehicular, y hay que tener en cuenta que para la dimensión de tiempo solo se tomo el año de ingreso como dato de tiempo. El código del modelo dimensional se adjunta en este mismo repositorio.

3. El tercer paso fue la transferencia de los datos a la, recién creada, base de datos tipo estrella. En esta imagen se muestra la estructura y resultado del proceso del ETL:



Nota: Para todos los data Flow que aparecen en la imagen se ha seguido una estructura similar, se derivó una columna para el código usando las primeras letras de la característica que representa cada tabla.



4. Resultado tenemos la base de datos repleta con los datos necesario, omitiendo los datos nulos para poder garantizar la calidad, en la imagen podemos observar la tabla de valor\_vehiculo, la que será nuestra tabla de hechos:

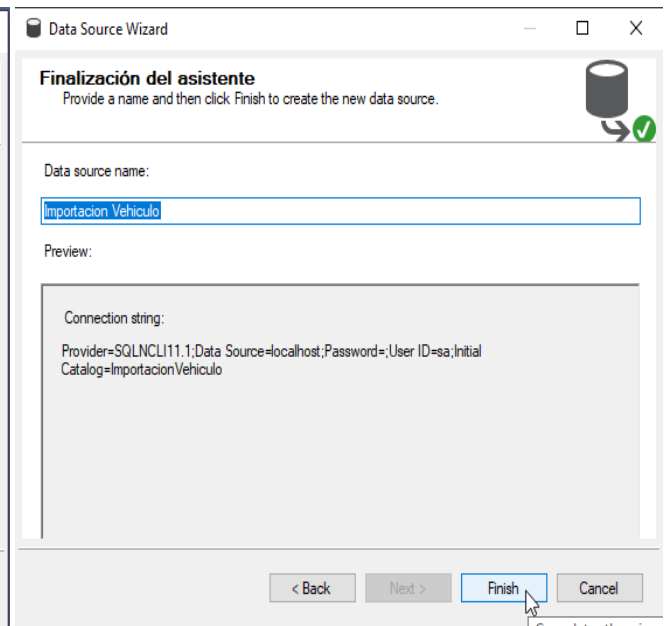
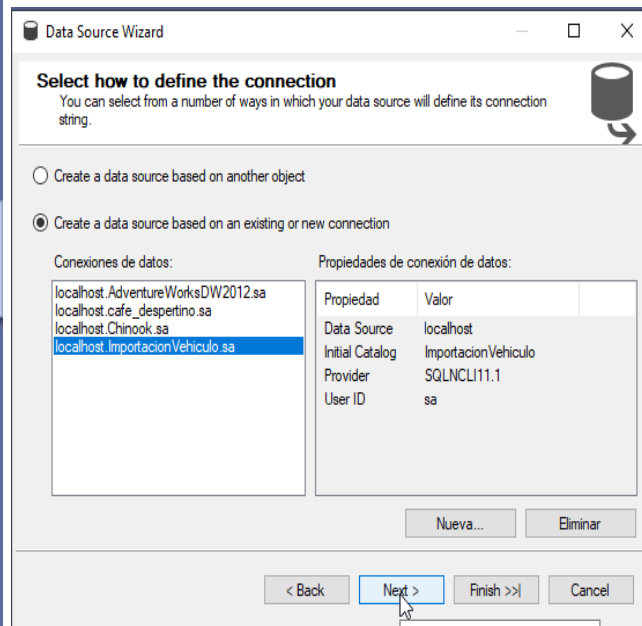
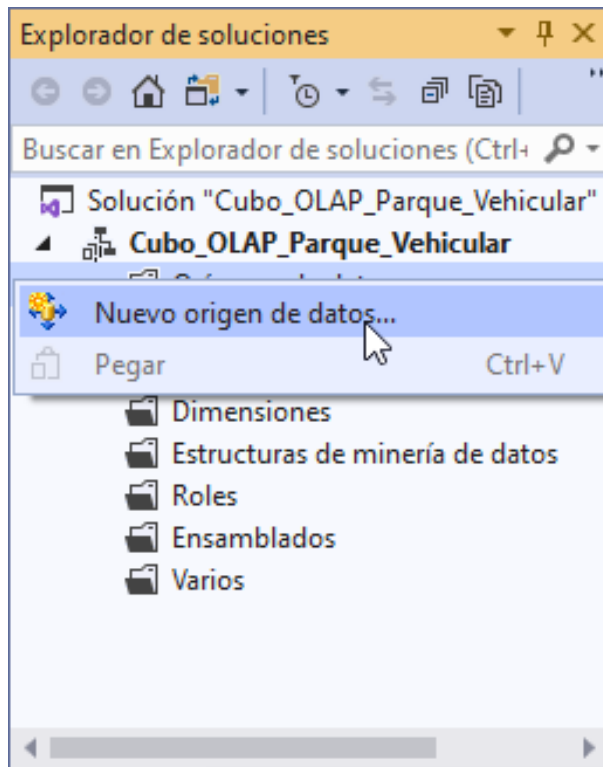
```
use ImportacionVehiculo
select * from valor_vehiculo
```

	CodigoVehiculo	CodigoPlaca	CodigoFecha	CodigoLugarPropietario	CodigoCondicion	CodigoEstado	CodigoAduana	CodigoCombustible	CodigoClase	CodigoPertenencia	ValorVehiculo	Impu
1	1	PAR	94	TONSAL	US	AL	SAN B	GAS	AUT	PRO	4094.56	4094
2	13	ALQ	95	CUSSAL	US	BA	SAN B	GAS	ALQ	PRO	4183.76	4183
3	15	PAR	93	SANSAL	US	BA	SAN B	GAS	AUT	PRO	4413.71	4413
4	17	PAR	94	SANSAL	US	BA	SAN B	GAS	AUT	PRO	4619.44	4619
5	31	PAR	97	SANMIG	US	AL	LA UN	GAS	AUT	PRO	3278.62	3278
6	34	PAR	97	SANVIC	US	AL	SAN B	GAS	AUT	PRO	4200.86	4200
7	43	PAR	96	SANSAL	US	BA	SAN B	GAS	AUT	PRO	3901.96	3901
8	45	ALQ	97	SOYSAL	US	BA	MARIT	GAS	ALQ	PRO	9153.51	9153
9	62	ALQ	95	SOYSAL	US	BA	SAN B	GAS	ALQ	PRO	4271.45	4271
10	65	MOT	12	META A	NU	AL	MARIT	GAS	MOT	PRO	1136.10	1136

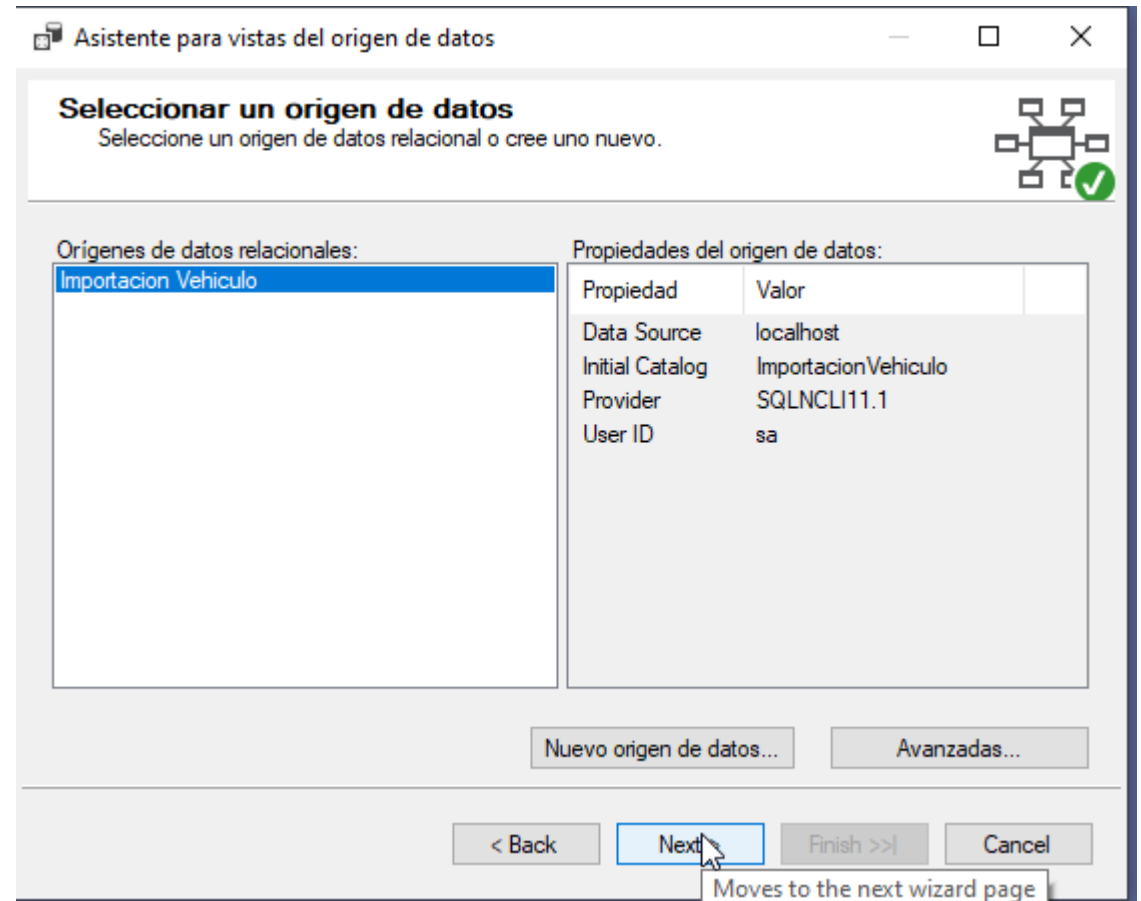
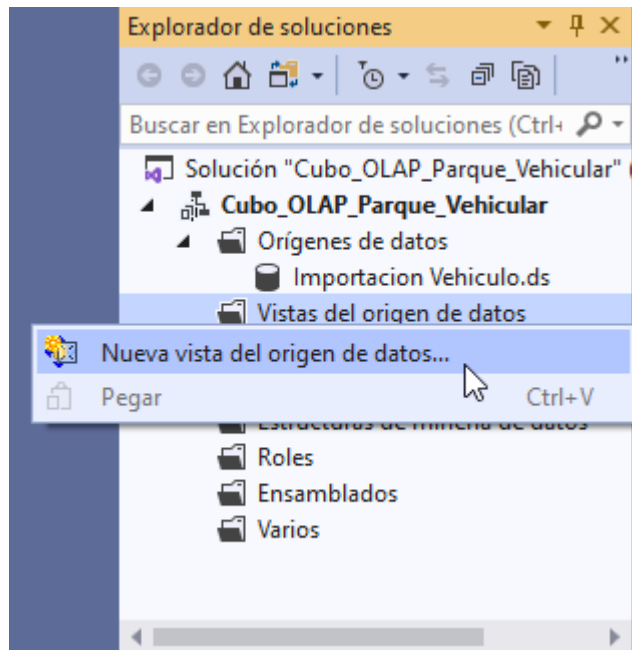
Query executed successfully. | LAPTOP-KQSF4D5I (15.0 RTM) | sa (53) | ImportacionVehiculo | 00:00:00 | 2,905 rows

Nota: Es importante notar que los datos resultantes son un total 2904 filas, lo que significa que un poco menos de un tercio de los datos cumple con el estándar requerido para realizar el análisis.

5. Para continuar la investigación, proseguimos a aplicar la primera técnica de minería de datos listada para este ejemplo, la cual constituye al Cubo OLAP.
- 5.1 Crear origen de datos:** Seleccionamos Nuevo Origen de datos, dando click derecho en la carpeta de Origen de datos, en la parte de la solución del proyecto. Proseguimos a seleccionar la base de datos en donde tenemos almacenados los datos, y por último damos en finalizar.



## 5.2 Creación de Vista de origen de datos: Seleccionamos el origen de datos anteriormente declarado en el paso anterior.



En esta parte seleccionamos las partes que posteriormente se convertirán en las dimensiones de nuestro Cubo. Como anteriormente se mencionó se decidió remover la tabla: placa, pertenencia y estado. Se consideró que no era información necesaria para el análisis buscado.

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
sysdiagrams (dbo)	Tabla
placa (dbo)	Tabla
pertenencia (dbo)	Tabla
estado (dbo)	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Objetos incluidos:

Nombre	Tipo
lugar_propietario (dbo)	Tabla
vehiculo (dbo)	Tabla
valor_vehiculo (dbo)	Tabla
tiempo (dbo)	Tabla
combustible (dbo)	Tabla
clase (dbo)	Tabla
aduana (dbo)	Tabla
condicion (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:

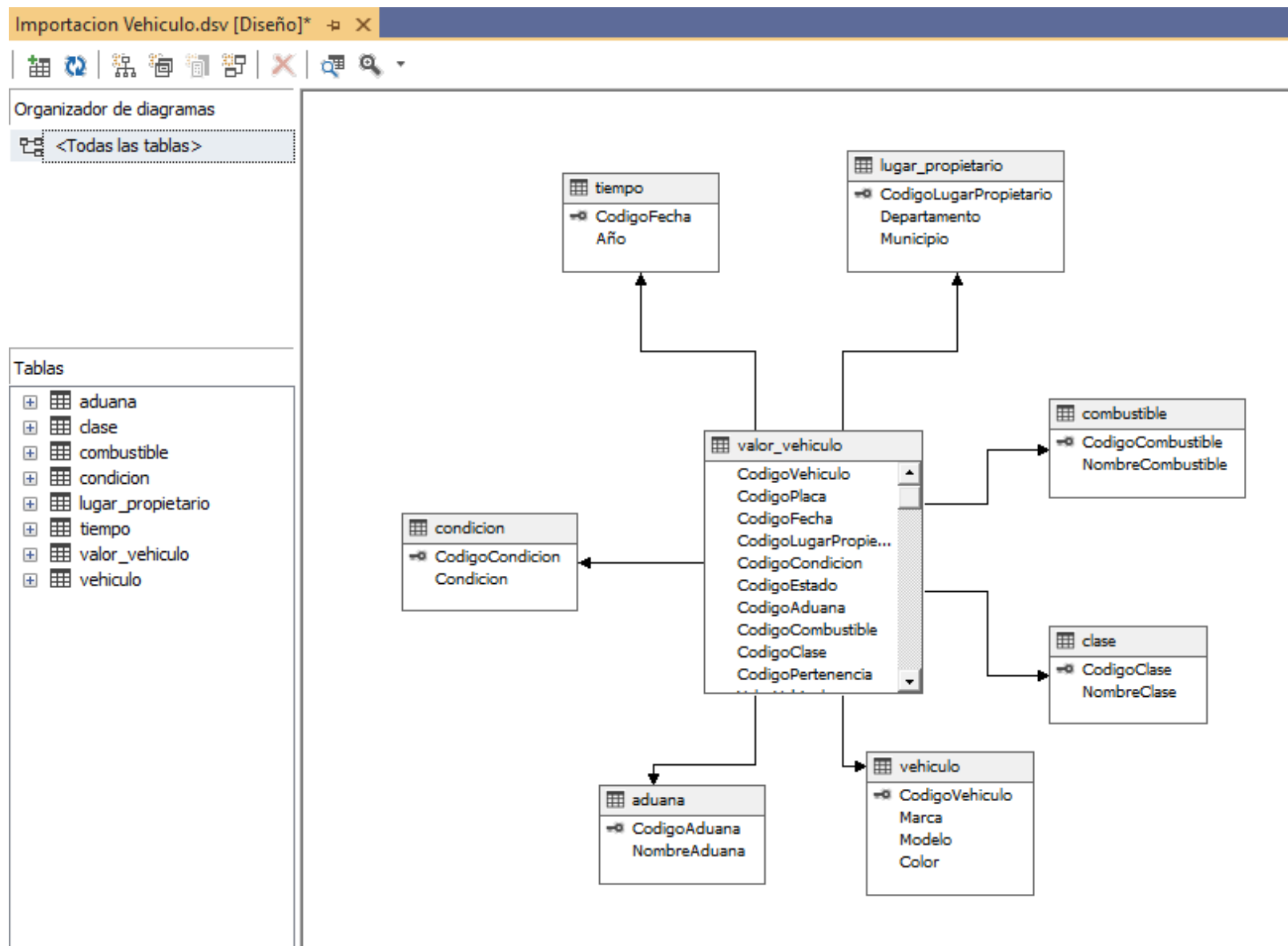
Vista previa:

- Importacion Vehiculo
  - lugar\_propietario (dbo)
  - vehiculo (dbo)
  - valor\_vehiculo (dbo)
  - tiempo (dbo)
  - combustible (dbo)
  - clase (dbo)
  - aduana (dbo)
  - condicion (dbo)

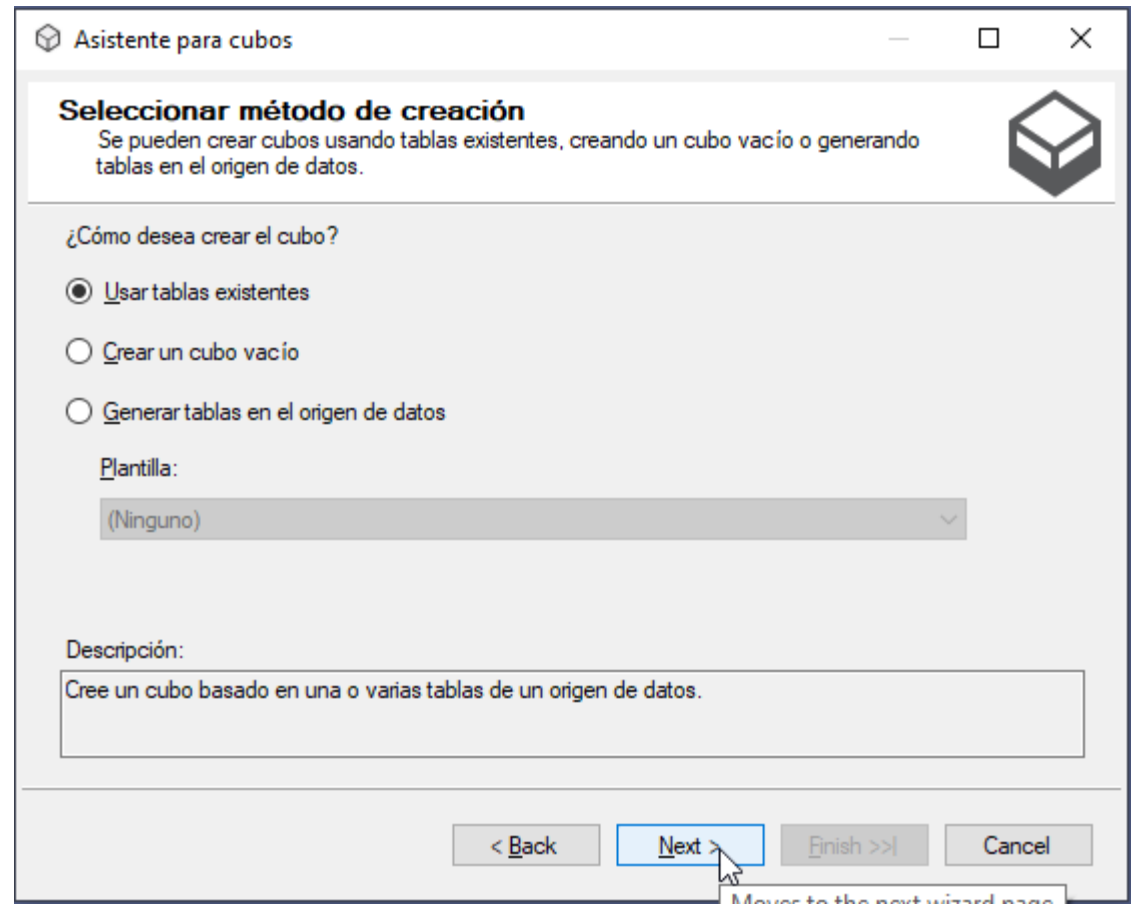
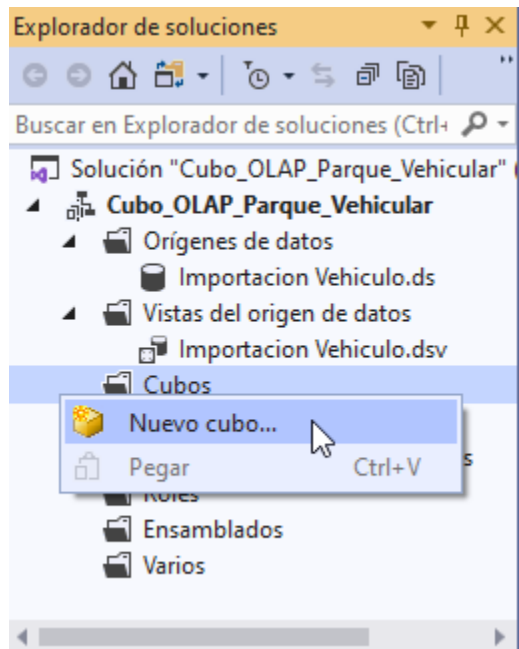
< Back   Next >   Finalizar   Cancel



Como resultado podemos apreciar la vista de origen de datos:



**5.3 Creación del Cubo:** Damos click derecho en la carpeta Cubos y nos aparecerá la ventana de Seleccionar método de creación y nos aseguraremos de marcar Usar tablas existentes, lo que nos permitirá usar las tablas contenidas en la parte de vistas de origen de datos.



En esta parte en donde seleccionamos la dimensión de hechos de nuestro cubo, el cual seleccionaremos la tabla valor\_vehiculo, debido a que almacena el valor en monetario de cada clase de vehículo que visita el parque vehicular.

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:  
Importacion Vehiculo

Tablas de grupo de medida:

- ☐ lugar\_propietario
- ☐ vehiculo
- ☒ valor\_vehiculo
- ☐ tiempo
- ☐ combustible
- ☐ clase
- ☐ aduana
- ☐ condicion

Sugerir

< Back   Next >   Finish >>   Cancel

Asistente para cubos

### Seleccionar medidas

Seleccione las medidas que desea incluir en el cubo.

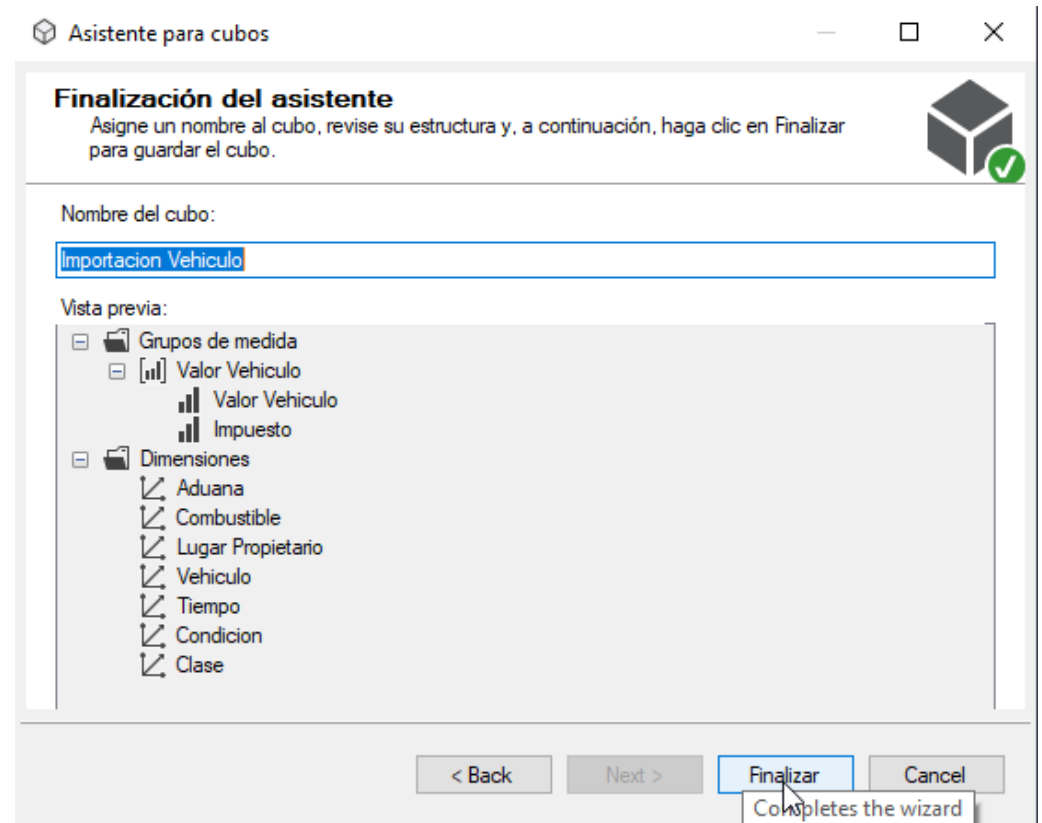
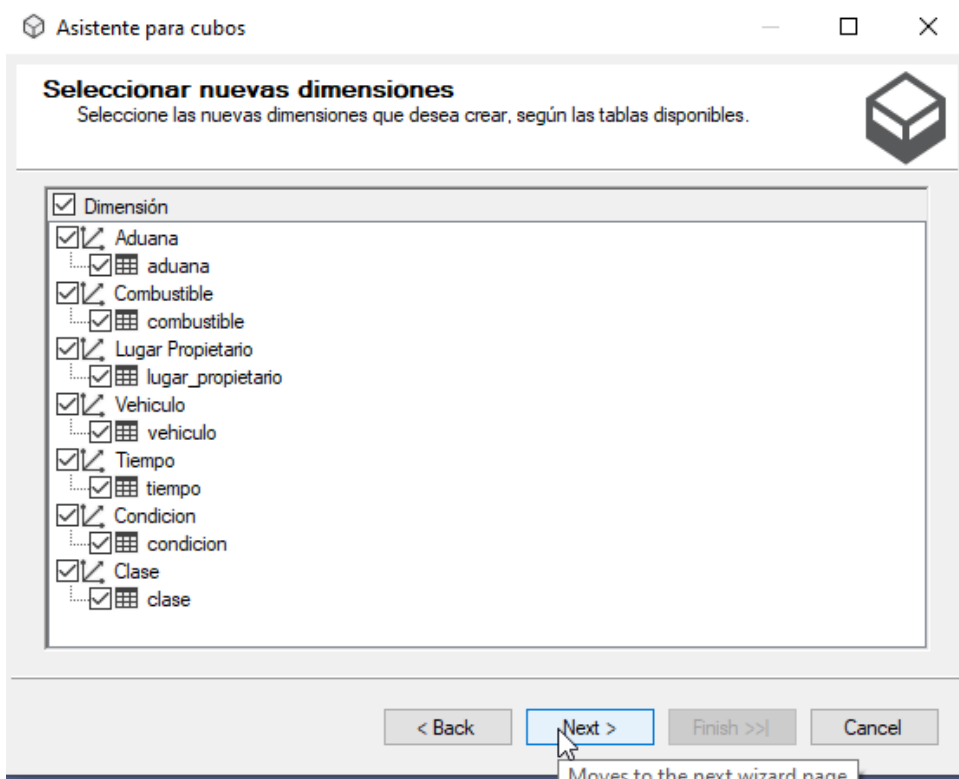
Medida

- ☒ Valor Vehiculo
- ☒ Valor Vehiculo
- ☒ Impuesto
- ☐ Recuento Valor Vehiculo

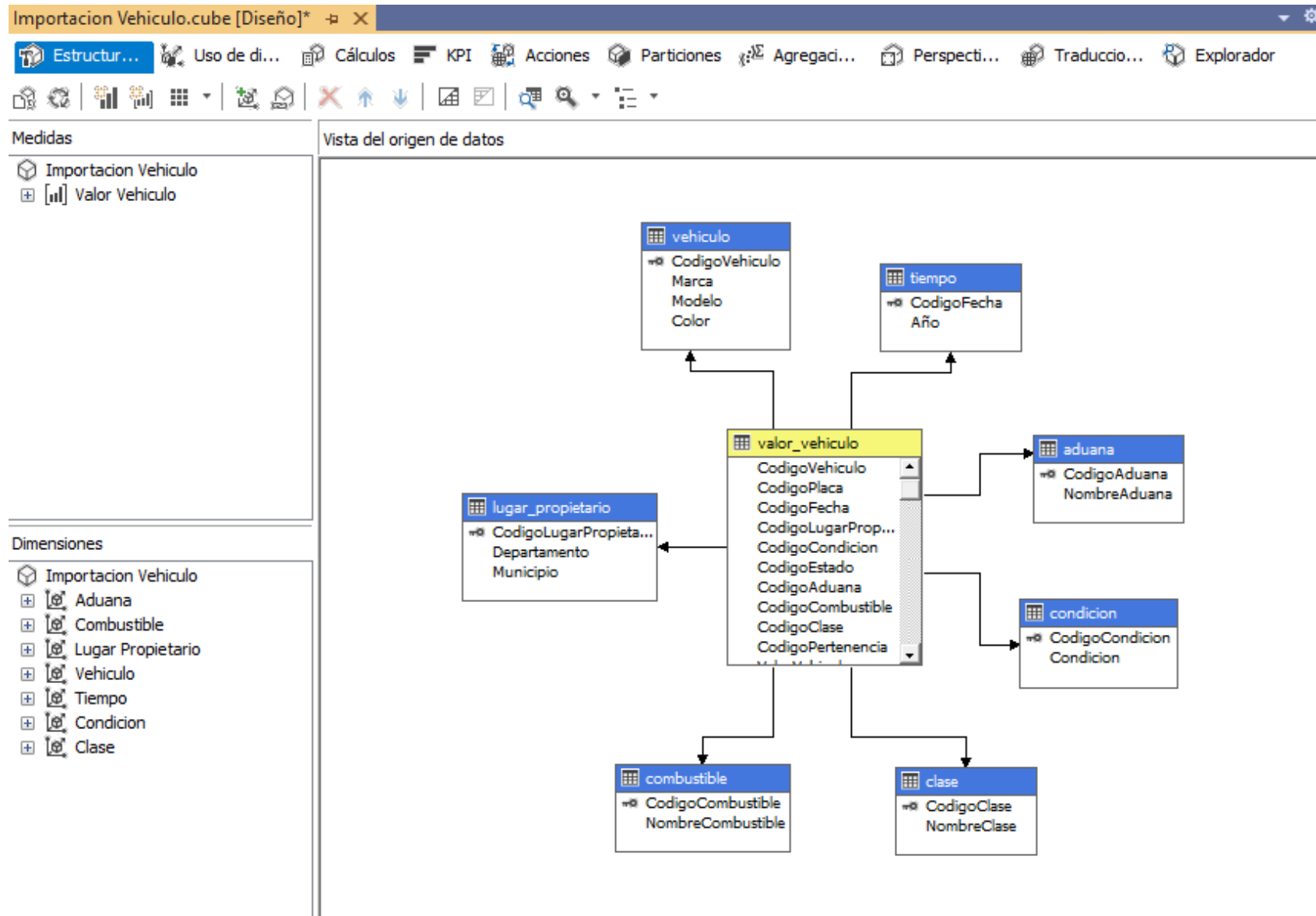
< Back   Next >   Finish >>   Cancel

Moves to the next wizard page

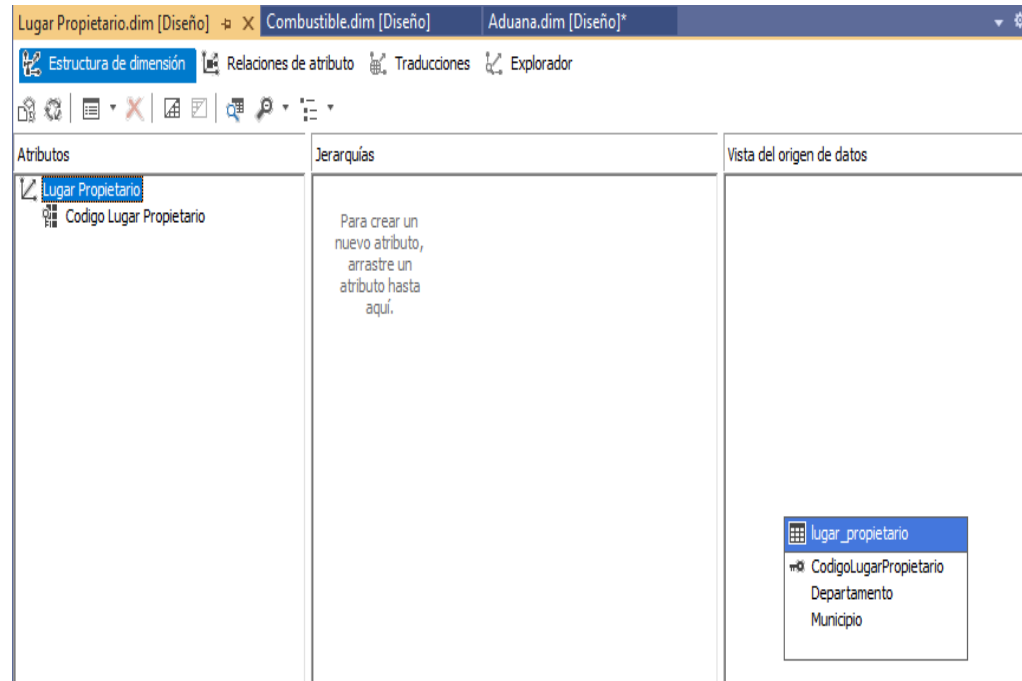
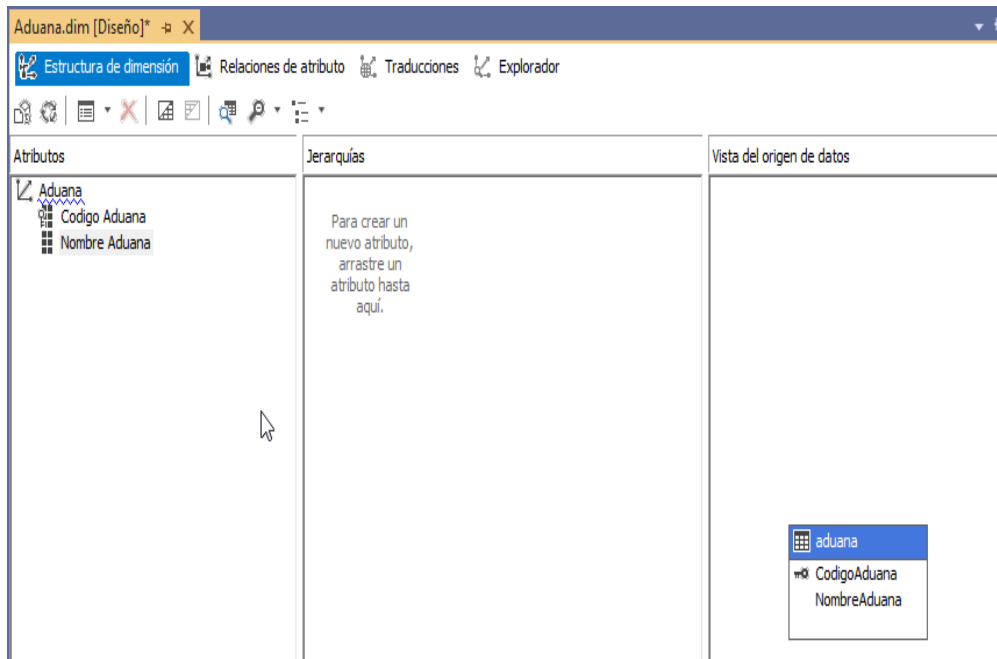
Las siguientes imágenes muestran como seleccionamos las dimensiones que conforman el cubo. Por último se muestra como se finaliza el cubo.



Podemos ver los resultados de manera grafica en esta imagen:

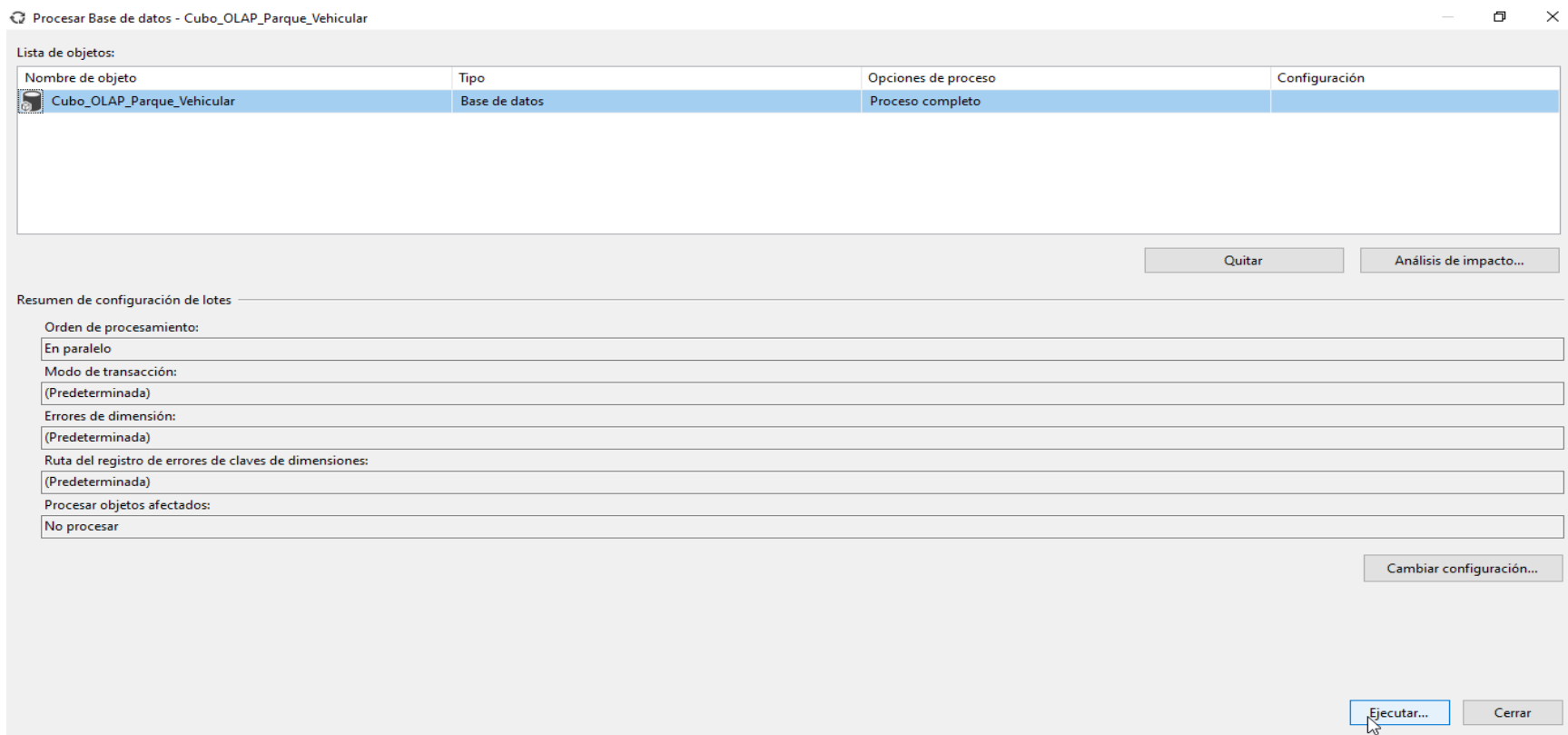
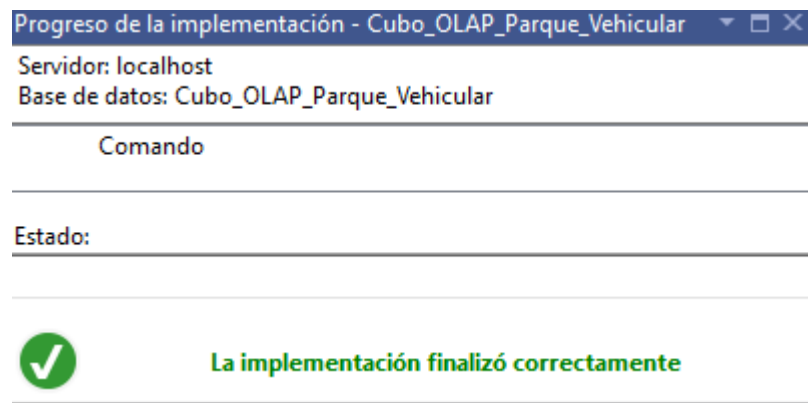
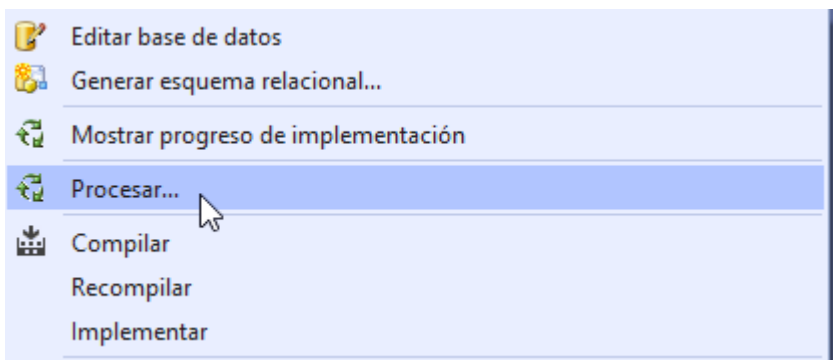


**Ahora solo resta agregar los campos necesarios a las dimensiones correspondientes:**



**Nota:** las imágenes son ejemplos de como se realizo el proceso para las demás dimensiones existentes en el Cubo.

Procesamos el cubo y podemos proseguir a analizar los datos por medio del examinador del cubo: Damos clic derecho sobre el nombre del proyecto y luego clic en procesar..., esperamos el mensaje que nos indica que la implementación finalizo correctamente. Se nos presentará de procesar, en la cual damos a ejecutar.





En esta parte podemos ver el resultado del proceso y posteriormente el examinador del cubo en donde haremos nuestras consultas, para posteriormente mostrarlas de manera grafica en el programa de Microsoft Power Bi.

The image shows two overlapping windows from a Business Intelligence tool. The left window, titled 'Progreso del proceso', displays a list of completed tasks under the 'Comando' (Command) section. The tasks include processing the data base, dimensions (Aduana, Clase, Combustible, Condicion), the cube ('Importacion Vehiculo'), and a measure group ('Valor Vehiculo'). The status at the bottom indicates 'Proceso finalizado correctamente.' (Process completed successfully). Buttons for 'Detener', 'Reprocesar', 'Ver detalles...', 'Copiar', 'Cerrar', and 'Ayuda' are visible.

The right window, titled 'Importacion Vehiculo.cube [Diseño]', shows the cube design interface. It includes a toolbar with options like 'Estructur...', 'Usar de di...', 'Cálculos', 'KPI', 'Acciones', 'Particiones', 'Agregaci...', 'Perspecti...', 'Traduccio...', and 'Explorador'. The main area is divided into sections: 'Metadatos' (Metadata) with a search bar and a list of measures (Valor Vehiculo, Impuesto, Valor Vehiculo); 'KPI' (KPI); 'Aduana' (Customs) with fields like 'Codigo Aduana' and 'Nombre Aduana'; 'Clase' (Class) with fields like 'Codigo Clase' and 'Nombre Clase'; 'Combustible' (Fuel); and 'Condicion' (Condition). Below these is the 'Miembros calculados' (Calculated Members) section. On the right, there are tabs for 'Dimensión', 'Jerarquía', 'Operador', and 'Expresión de filtro', with a dropdown menu for '<Seleccionar dimensi...'.

Las consultas que hemos seleccionado para el análisis son las siguientes:

Marca	Condicion	Valor Vehiculo	Impuesto
MITSUBISHI	VEHICULO USADO	117489.76	117489.76
NISSAN	VEHICULO NUEVO	413007.78	413007.78
NISSAN	VEHICULO USADO	897797	1186345.12
OPEL	VEHICULO USADO	0	0
PENNINE	VEHICULO USADO	0	0
PLYMOUTH	VEHICULO USADO	3709.25	3709.25
PONTIAC	VEHICULO USADO	12853.81	12853.81
ROMAN	VEHICULO USADO	0	0

**Nota:** Esta consulta contiene la Marca, Condición del vehículo con su valor y su impuesto en dólares. Podemos analizar posteriormente en Power Bi la diferencia de porcentaje que hay entre los vehículos nuevos y usados. Lo que será de suma importancia para el estado en la identificación de posibles incumplimientos de las leyes de tránsito.

Nombre Clase	Departamento	Valor Vehiculo	Impuesto
ALQUILER	CUSCATLAN	0	0
ALQUILER	LA LIBERTAD	50141.57	50141.57
ALQUILER	LA PAZ	13609.68	13609.68
ALQUILER	LA UNION	15444.79	15444.79
ALQUILER	SAN MIGUEL	119729.54	119729.54
ALQUILER	SAN SALVADOR	695183.470...	698239.69
ALQUILER	SANTA ANA	11997.74	11997.74
ALQUILER	SONSONATE	66836.99	66836.99

**Nota:** En esta consulta podemos ver el nombre de la clase del vehículo, como por ejemplo si es propio o es de alquiler, el departamento de procedencia. Nos puede ayudar a identificar el lugar de donde residen los vehículos. Con este análisis podemos proponer nuevas sedes de parques vehicular para usuarios que llegan desde lugares alejados del presente parque vehicular.

Año	Nombre Clase	Valor Vehiculo	Impuesto
2012	PICK UP	1171149.67	1171149.67
2012	REMOLQUE	65791.99	65791.99
2012	TRICIMOTO	5946.84	5946.84
2013	ALQUILER	5311	5311
2013	AUTOMOVIL	845969.08	845969.08
2013	CABEZAL	18164.75	18164.75
2013	CAMION PESADO	365513.75	365513.75
2013	MOTOCICLETA	164693.46	164693.46

**Nota:** esta consulta contiene el año del primer ingreso al parque vehicular, y la respectiva clase del vehículo. Con estos datos históricos podemos analizar que vehículos han sido más populares o comunes en determinadas épocas. Con esto podemos adaptarnos para ofrecer mejores servicios dependiendo del tipo de vehículo que ingrese al parque.

Marca	Modelo	Color	Valor Vehi...	Impuesto
BLUE BIRD	SCHOOL BUS	BLANCO VERDE C/F R/B	3507.93	3507.93
BLUE BIRD	SCHOOL BUS	BLANCO/AZUL	3891.61	3891.61
BLUE BIRD	SCHOOL BUS	BLANCO/AZUL C/F R/B	62421.55	62421.55
BLUE BIRD	SCHOOL BUS	BLANCO/CELESTE C/F R/B	3507.93	3507.93
BLUE BIRD	SCHOOL BUS	BLANCO/MORADO C/F R/B	3507.93	3507.93
BLUE BIRD	SCHOOL BUS	BLANCO/ROJO	6071.74	6071.74
BLUE BIRD	SCHOOL BUS	BLANCO/ROJO C/F R/B	3823.35	3823.35
BLUE BIRD	SCHOOL BUS	BLANCO/ROJO/VERDE C/...	4850.81	4850.81

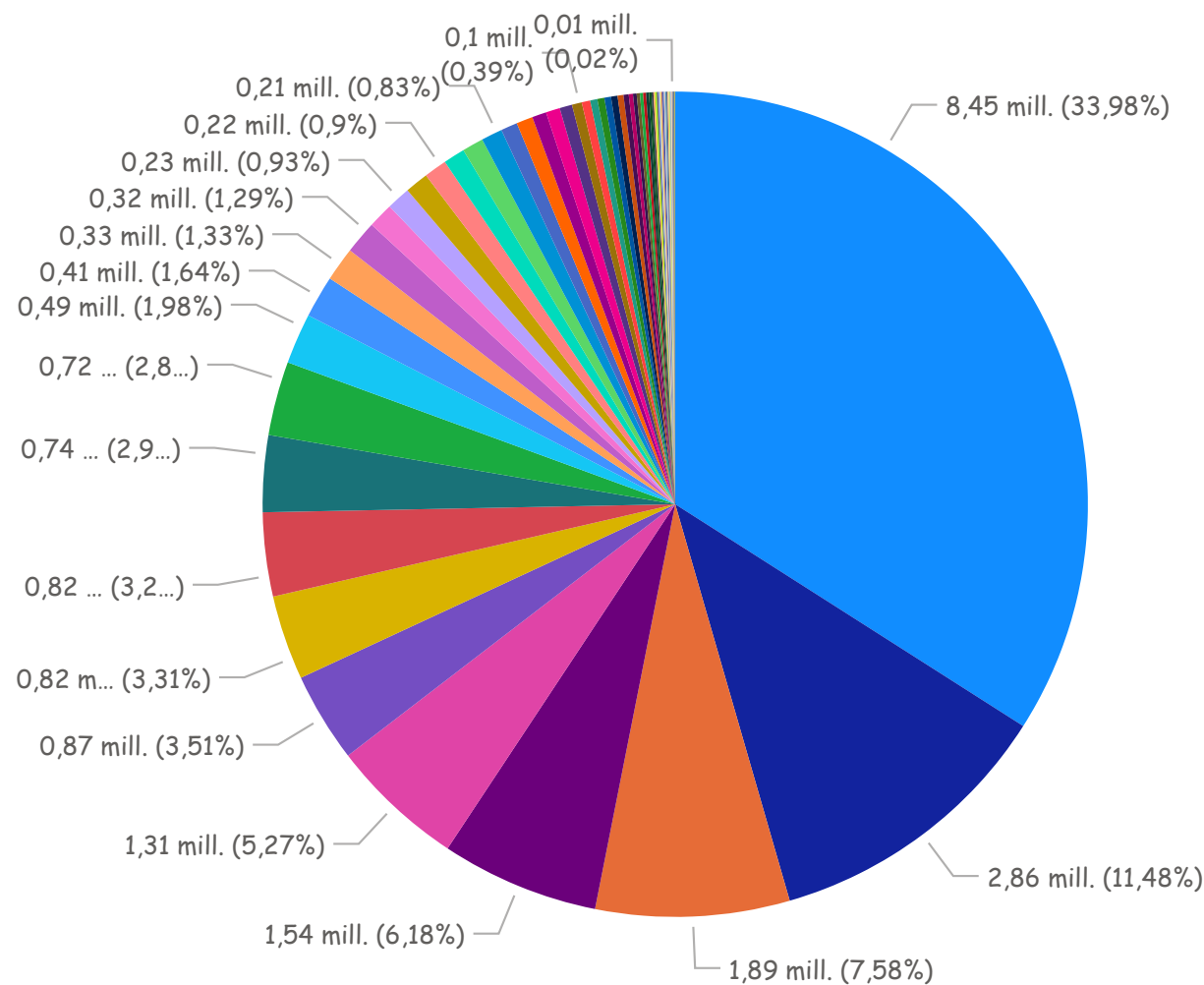
**Nota:** En esta consulta podemos observar las características generales de todos los vehículos será de gran ayuda para el ministerio, debido a que podrá comparar los datos con las tarjetas de circulación, esto con el objetivo de mejorar el sistema de multas por el incumplimiento de las leyes de tránsito.

Nombre Clase	Nombre Combustible	Nombre Aduana	Valor Vehiculo	Impuesto
ALQUILER	DIESEL	AEREA COMALAPA	11014.49	11014.49
ALQUILER	DIESEL	ANGUIATU	0	0
ALQUILER	DIESEL	LA UNION	4721.01	4721.01
ALQUILER	DIESEL	SAN BARTOLO (SAN SAL...	68691.13	68691.13
ALQUILER	DIESEL	SANTA ANA	3874.51	3874.51
ALQUILER	GASOLINA	AEREA COMALAPA	0	0
ALQUILER	GASOLINA	ANGUIATU	0	0
ALQUILER	GASOLINA	EL AMATILLO	3474.4	3474.4

**Nota:** la última consulta que presentamos es la que contiene el nombre de la clase, el tipo de combustible del vehículo y el nombre de la aduana de procedencia. No está más decir que los datos recabados pueden ayudar a impedir que los vehículos tengan orígenes no deseados.

**Los resultados obtenidos en Power bi se adjuntan a continuación:**

Valor por cada vehiculo



- Marca
- MERCEDES BENZ
  - TOYOTA
  - BLUE BIRD
  - INTERNATIONAL
  - NISSAN
  - ISUZU
  - HYUNDAI
  - CAIO
  - KIA
  - FORD
  - MITSUBISHI
  - CHEVROLET
  - HONDA
  - DINA
  - SCANIA
  - GMC
  - DATSUN
  - MASA
  - MAUCO
  - BMW
  - MAZDA
  - VOLKSWAGEN
  - JEEP
  - DAEWOO
  - FREIGHTLINER

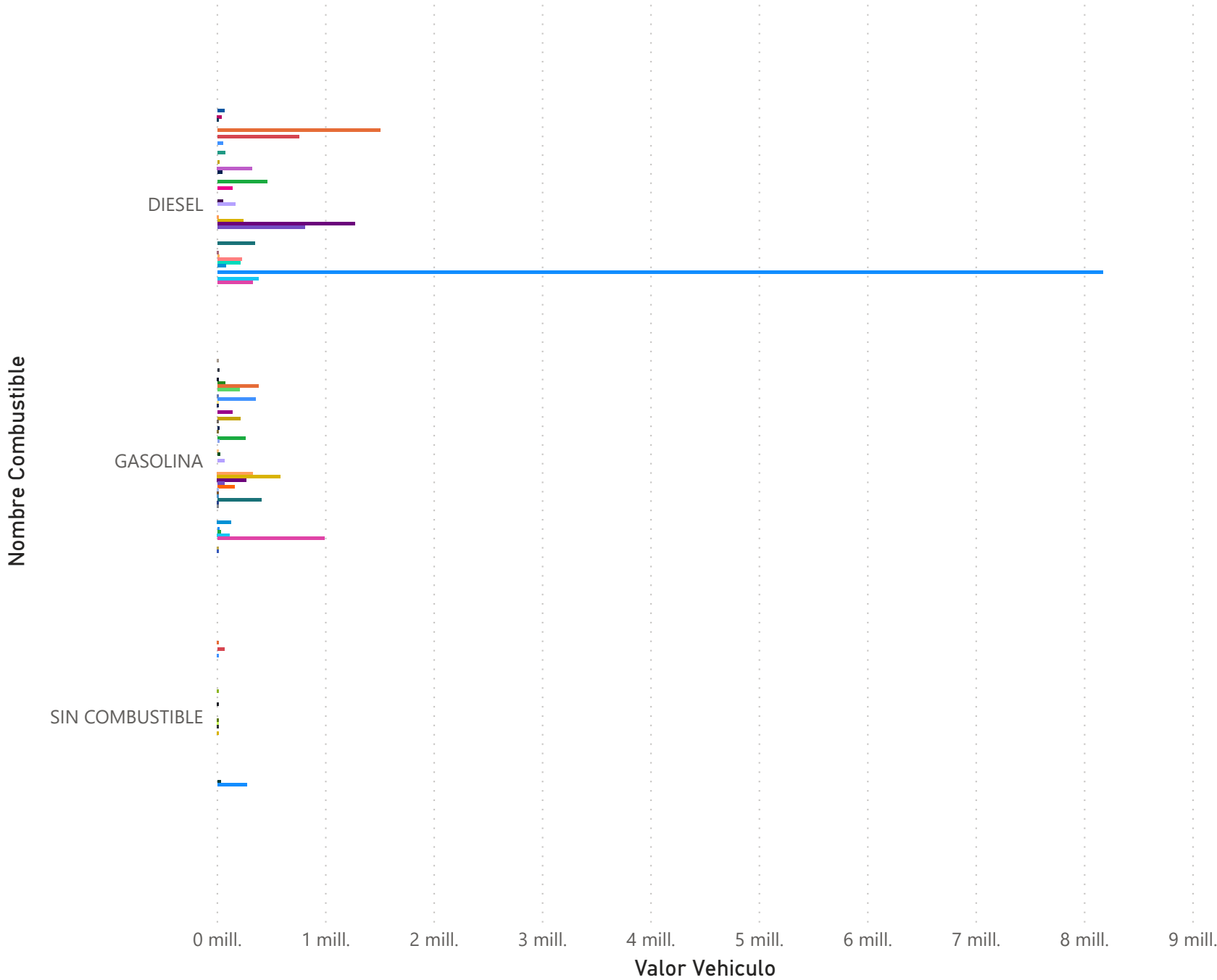
Año

- 1975
- 1976
- 1977
- 1978
- 1979
- 1980
- 1981
- 1982
- 1983
- 1984
- 1985
- 1986
- 1987
- 1988
- 1989
- 1990
- 1991
- 1992
- 1993

Marca	Condicion	Valor Vehiculo
ACURA	VEHICULO USADO	4.950,53
ASHOK LEYLAND	VEHICULO USADO	0,00
ASIA	VEHICULO USADO	64.445,60
ASIA HERO	VEHICULO NUEVO	18.251,28
ASIA SMITH	VEHICULO USADO	40.744,08
AUTOCAR	VEHICULO USADO	3.058,68
AVA	VEHICULO USADO	541,43
BAJAJ	VEHICULO NUEVO	62.309,26
BAJAJ	VEHICULO USADO	5.946,84
BLUE BIRD	VEHICULO USADO	1.886.184,13
BMW	VEHICULO NUEVO	157.555,04
BMW	VEHICULO USADO	49.552,46
CAIO	VEHICULO USADO	815.469,88
CFMOTO	VEHICULO USADO	3.785,09
Total		24.881.335,74

Tipo de gasolina por vehiculo

Marca ACURA ASHOK LE... ASIA ASIA HERO ASIA S... AUTOCAR AVA BAJAJ BLUE BIRD BMW CAIO



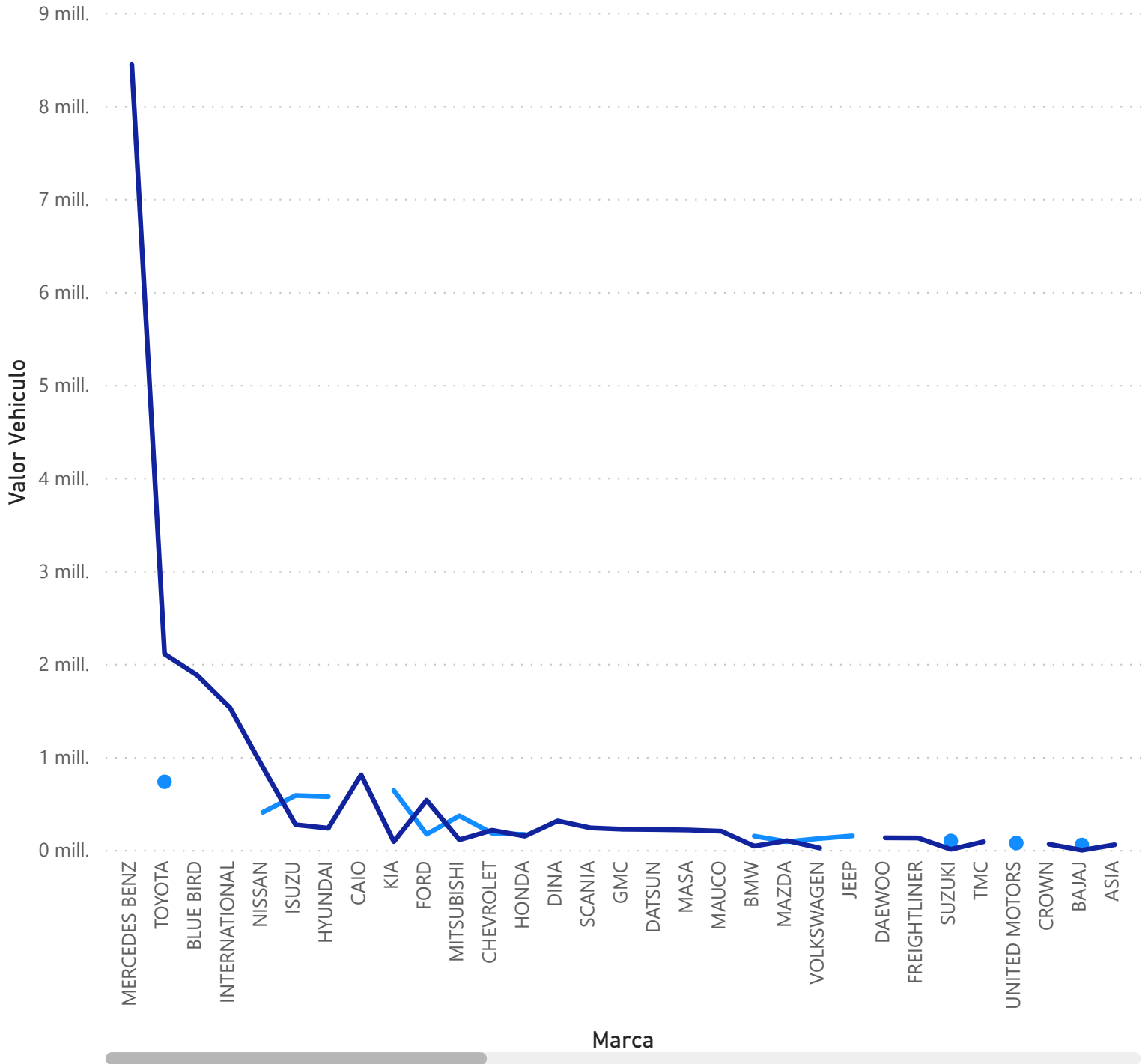
Año

- 1975
- 1976
- 1977
- 1978
- 1979
- 1980
- 1981
- 1982
- 1983
- 1984
- 1985
- 1986
- 1987
- 1988
- 1989
- 1990
- 1991
- 1992

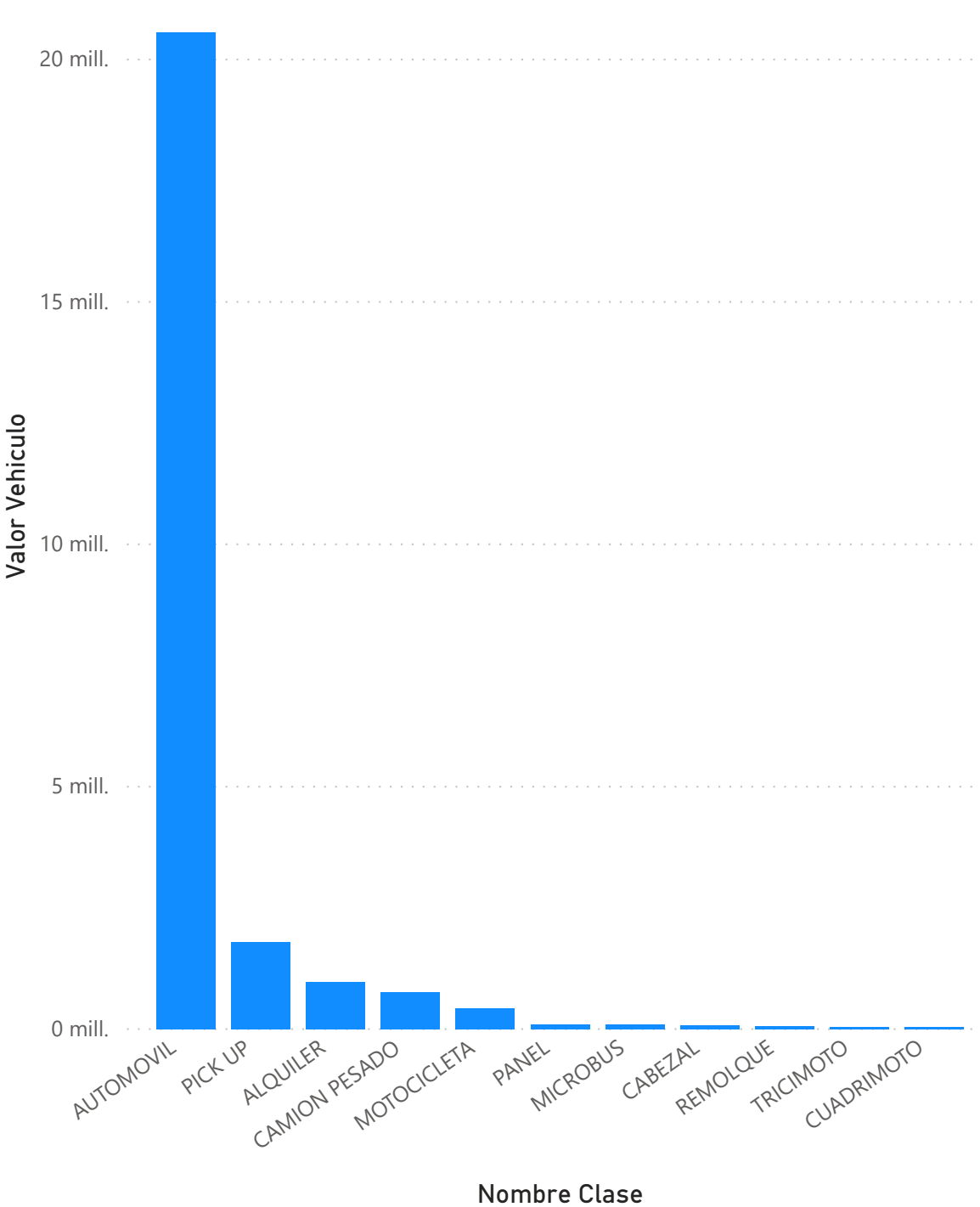
Marca	Nombre Combustible
ACURA	GASOLINA
ASHOK LEYLAND	DIESEL
ASIA	DIESEL
ASIA HERO	GASOLINA
ASIA SMITH	DIESEL
AUTOCAR	DIESEL
AVA	GASOLINA
BAJAJ	GASOLINA
BLUE BIRD	DIESEL
BLUE BIRD	GASOLINA
BLUE BIRD	SIN COMBUSTIBLE
BMW	GASOLINA
CAIO	DIESEL
CAIO	SIN COMBUSTIBLE
CFMOTO	GASOLINA
CHEVROLET	DIESEL

Valor Vehiculo por Marca y Condicion

Condicion ● VEHICULO NUEVO ● VEHICULO USADO

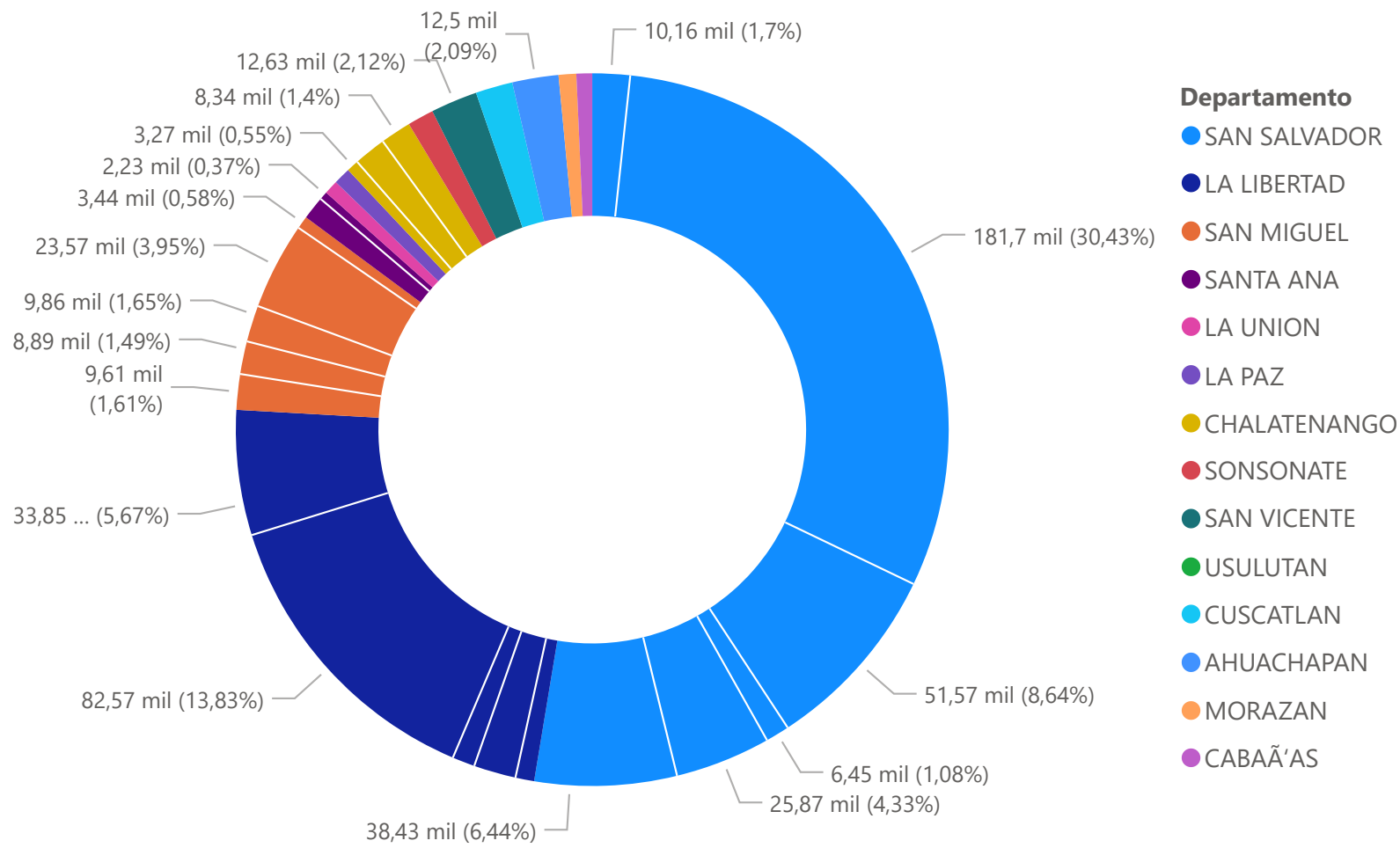


Valor Vehiculo por Nombre Clase





Cantidad de vehiculos por departamentos



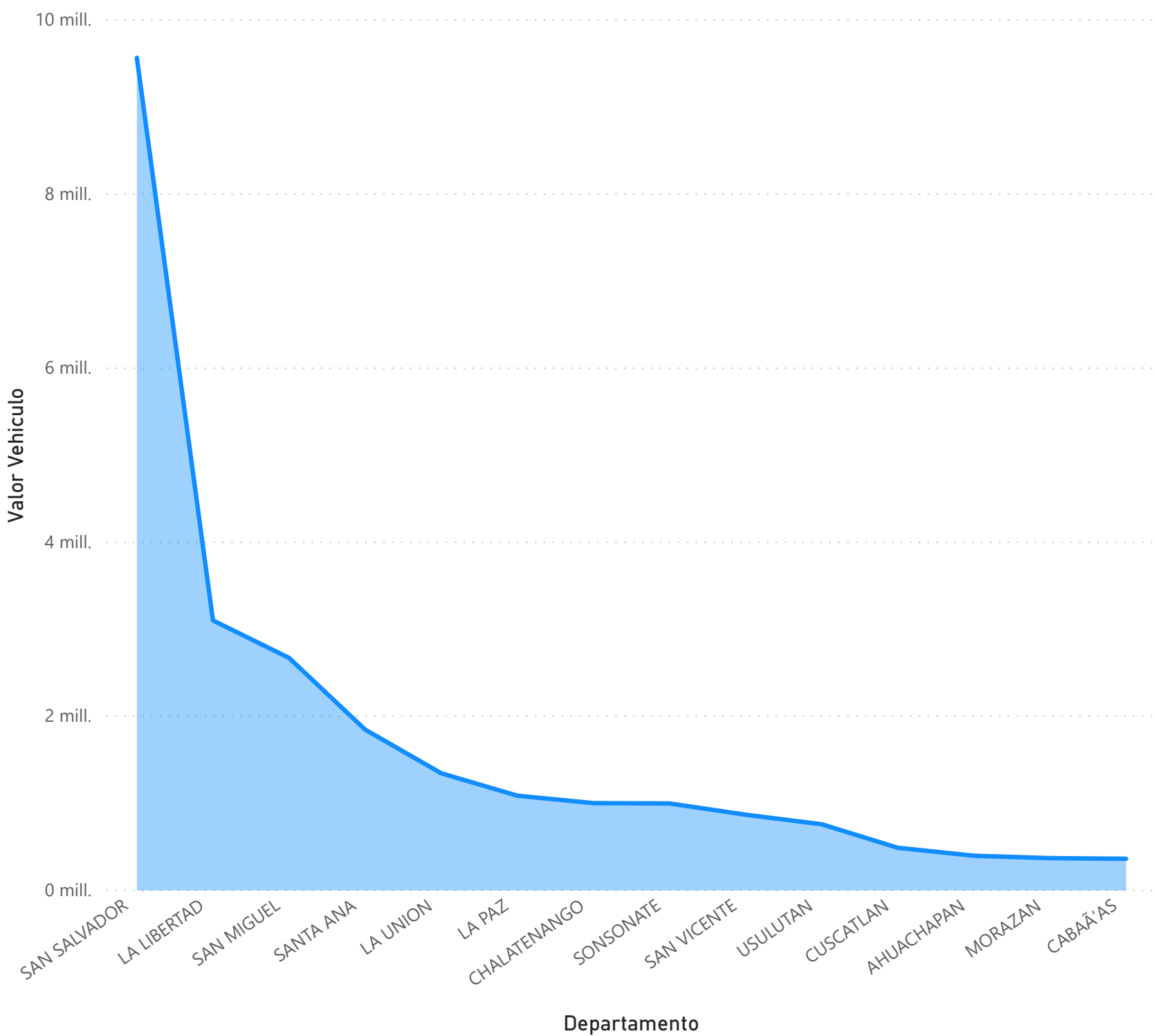
Aduanas

Nombre Aduana

- ☐ AEREA COMALAPA
- ☐ AGDOSIA - ALMACENES GENERALES DEPOSITO DE OCCIDENTE
- ☐ ALDESA - ALMACENADORA DE DESARROLLO
- ☐ ALMACENADORA NEJAPA, S.A DE C.V
- ☐ ANGUIATU
- ☐ BODESA - BODEGAS GENERALES DE DEPOSITO
- ☐ DEPOSITO TEMPORAL LA ROCA
- ☐ EL AMATILLO
- ☐ EL COCO
- ☐ EL POY
- ☐ FERIA INTERNACIONAL
- ☐ LA HACHADURA
- ☐ LA UNION
- ☐ MARITIMA DE ACAJUTLA
- ☐ SAN BARTOLO (SAN SALVADOR)
- ☐ SANTA ANA

Marca	Modelo	Color
ACURA	TSX	ROJO
AM GENERAL	N/D	BLANCO ROJO
AM GENERAL	N/D	BLANCO/ROJO C/F R/B
ASHOK LEYLAND	EVK 55	BLANCO AZUL C/F R/B
ASHOK LEYLAND	EVK 55	BLANCO ROJO
ASHOK LEYLAND	EVK 55	ROJO/BLANCO C/F R/B
ASHOK LEYLAND	N/D	AZUL C/ FJAS ANARANJADO
ASHOK LEYLAND	N/D	BLANCO AZUL
ASHOK LEYLAND	N/D	BLANCO AZUL DIST COM
ASHOK LEYLAND	N/D	BLANCO ROJO
ASHOK LEYLAND	N/D	BLANCO ROJO VERDE
ASHOK LEYLAND	N/D	BLANCO VERDE
ASHOK LEYLAND	N/D	BLANCO VERDE DORADO
ASHOK LEYLAND	N/D	BLANCO/AZUL C/F R/B
ASHOK LEYLAND	N/D	BLANCO/ROJO C/F R/B
ASHOK LEYLAND	N/D	ROJO BLANCO

Valor ingreso vehicular



Año

- 1975
- 1976
- 1977
- 1978
- 1979
- 1980
- 1981
- 1982
- 1983
- 1984
- 1985
- 1986
- 1987
- 1988
- 1989
- 1990
- 1991
- 1992
- 1993

Marca	Modelo	Color	Departamento
ACURA	TSX	ROJO	SAN MIGUEL
ASHOK LEYLAND	N/D	ROJO/BLANCO C/F R/B	SAN SALVADOR
ASIA	N/D	AZUL/BLANCO C/F R/B	SAN SALVADOR
ASIA	N/D	BLANCO VERDE C/F V/B	LA PAZ
ASIA	N/D	ROJO/BLANCO C/F R/B	SAN SALVADOR
ASIA	N/D	VERDE/BLANCO C/F V/B	CUSCATLÁN
ASIA	SCHOOL BUS	VERDE/BLANCO C/F V/B	SAN VICENTE
ASIA HERO	CARGO 125	AZUL MULTICOLOR	SAN SALVADOR
ASIA HERO	CARGO 125	AZUL MULTICOLOR	SAN VICENTE
ASIA HERO	CARGO 125	AZUL MULTICOLOR	USULUTAN
ASIA HERO	CARGO 125	GRIS MULTICOLOR	LA PAZ
ASIA HERO	CARGO 125	GRIS MULTICOLOR	SAN MIGUEL
ASIA HERO	CARGO 125	GRIS MULTICOLOR	SAN SALVADOR
ASIA HERO	CARGO 125	GRIS MULTICOLOR	SAN SALVADOR
ASIA HERO	CARGO 125	GRIS MULTICOLOR	SAN SALVADOR

**Conclusión:** tomando como base los resultados de ambos análisis, presentados en este documento, el Ministerio de Obras Públicas (MOP) puede presentar buenos informes para el cierre del año 2018. Considerando haber abordado los aspectos que serán de mucha utilidad en la rubro de la seguridad vial y la administración del parque vehicular.