



DATAWAREHOUSE Y MINERÍA DE DATOS

CATEDRATICO KAREN MEDRANO

FASE 2

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

3 DE DICIEMBRE DEL AÑO 2022

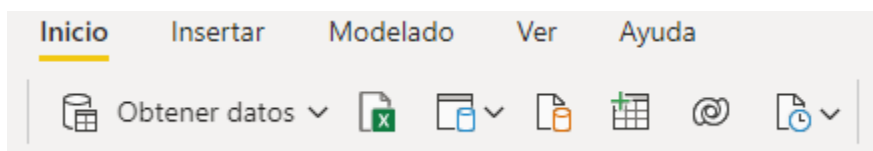
I. OBJETIVOS

- Analizar la información del Ministerio de Obras Públicas (MOP) mediante las estrategias de minería de datos.
- Proporcionar la información y datos mediante K Means y Power BI.

II. MARCO TEORICO

TEMA: USO DE POWER BI

Para introducir el archivo CSV del cual queremos extraer los datos, debemos irnos a inicio y seleccionamos obtener datos, en ese apartado seleccionamos CSV.



Luego nos aparecerá una vista previa de nuestra tabla de datos, ahí podemos hacer cambios de formato o modificaciones, en este caso lo utilizaremos tal y como se nos entregaron los datos.

Esquelas.csv

Origen de archivo: 1252: Europeo occidental (Windows) | Delimitador: Punto y coma | Detección del tipo de datos: Basado en las primeras 200 filas

Ves_Nro_Esquila	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcion	Ves_Departamento
110637	04/06/2003 12:00:00 a. m.	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO
107742	04/06/2003 12:00:00 a. m.	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO
106645	04/06/2003 12:00:00 a. m.	TRANSITO	ESTACIONARSE EN CURVAS PUENTES Y TRECHOS ANGO...	SAN SALVADOR
154061	04/06/2003 12:00:00 a. m.	TRANSITO	BAJAR O SUBIR PASAJEROS EN LUGARES NO PERMITID...	SAN SALVADOR
154102	03/06/2003 12:00:00 a. m.	TRANSITO	DISPUTARSE LA VIA CON OTRO VEHICULO	SAN SALVADOR
119779	06/06/2003 12:00:00 a. m.	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIM...	SAN SALVADOR
106938	04/06/2003 12:00:00 a. m.	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
154457	04/06/2003 12:00:00 a. m.	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
154951	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZAD...	SAN SALVADOR
154976	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZAD...	SAN SALVADOR
109210	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
109211	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
109207	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
105578	09/06/2003 12:00:00 a. m.	TRANSITO	CIRCULAR CON VEHICULOS NO MATRICULADOS	SAN SALVADOR
109469	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
154826	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
154952	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR
110968	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SI...	SAN SALVADOR
106398	06/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR
106075	07/06/2003 12:00:00 a. m.	TRANSPORTE TERRESTRE	CONducir con las puertas abiertas	SAN SALVADOR

Extraer tabla mediante ejemplos | Cargar | Transformar datos | Cancelar

Para las consultas de datos que deseamos realizar necesitamos ejecutar diversos scripts, luego de un análisis estos son los que nosotros hemos creado:

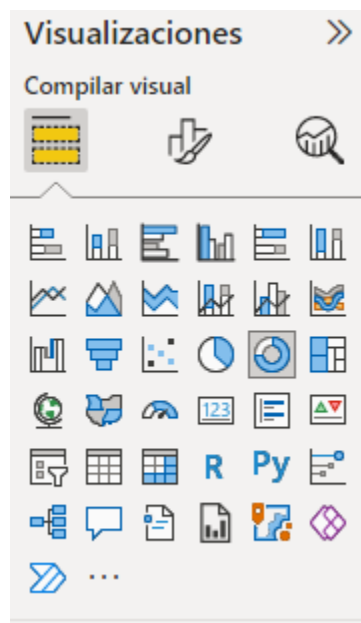
```
Estado = COUNT(Esqueles[Ves_Estado_Descripcion]) //Para obtener las cantidades de cada estado
```

```
Suma total recaudada = SUM(Esqueles[Ves_Valor]) // obtener el dinero total de todas las multas
```

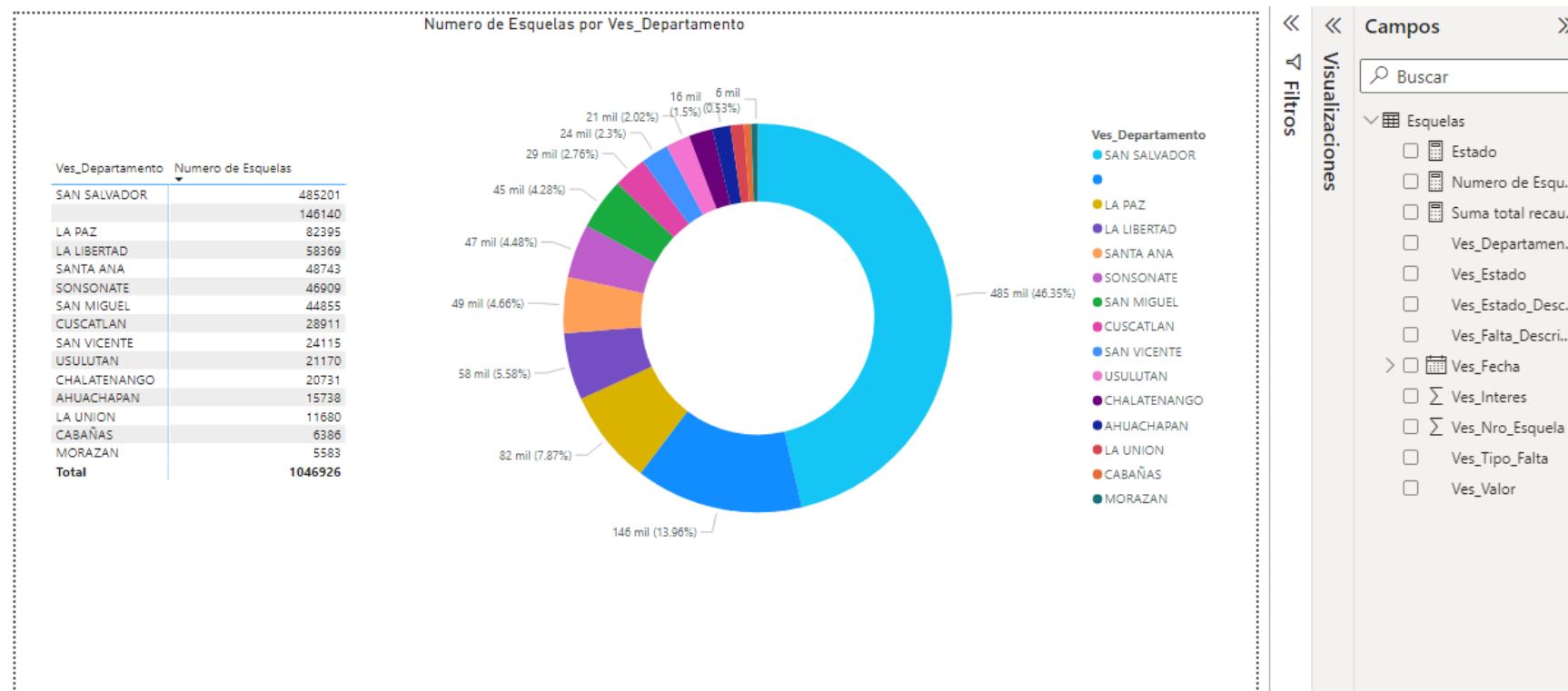
```
Numero de Esqueles = COUNT(Esqueles[Ves_Departamento]) //El numero total de multas impuestas
```

Estos scripts los hemos implementado ya que, si bien Power BI nos implementa ciertas consultas de datos de manera automática, siempre existirán diversas consultas para obtener un “Query” muy específico de algo que deseamos saber.

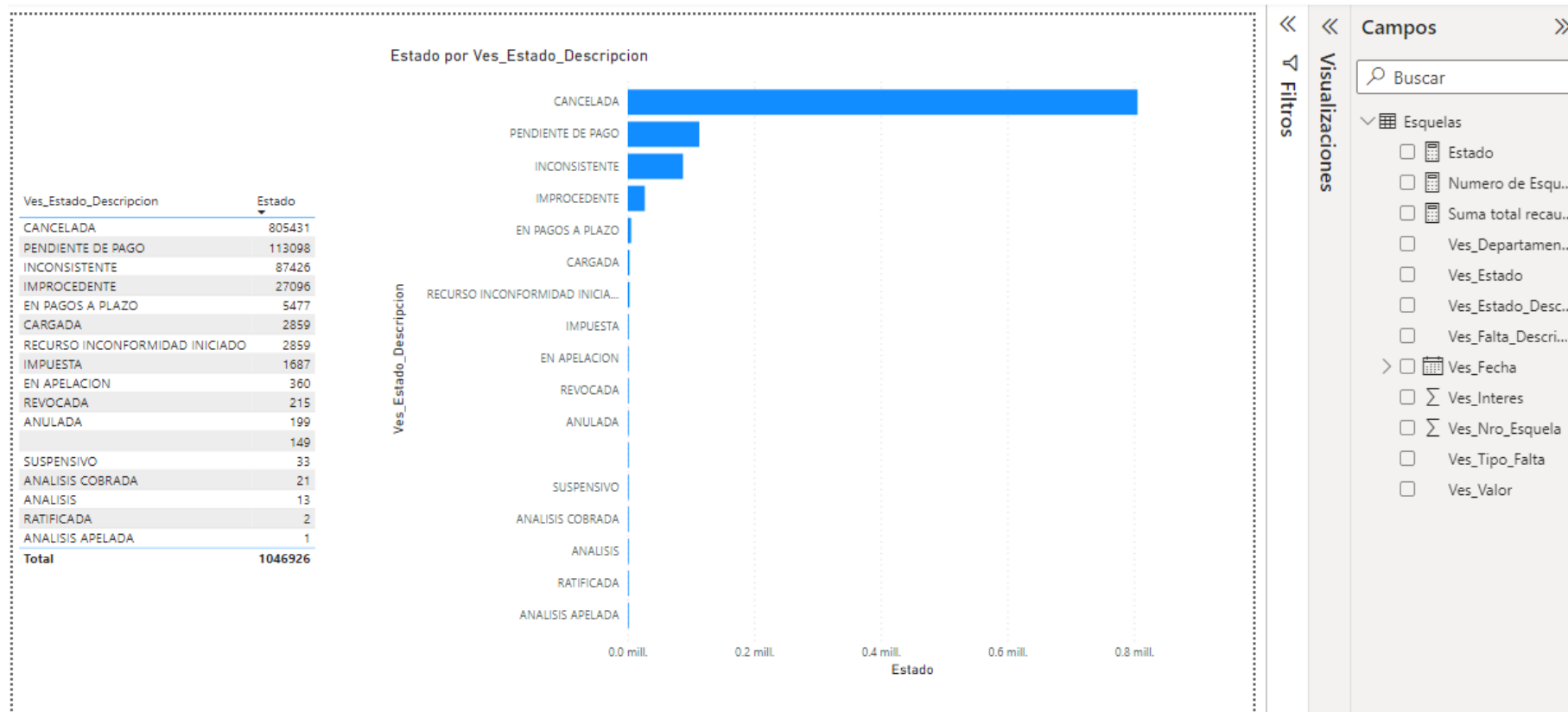
A la hora de mostrar datos, tenemos la pestaña de diseño, donde seleccionaremos el más adecuado según los datos que deseamos mostrar.



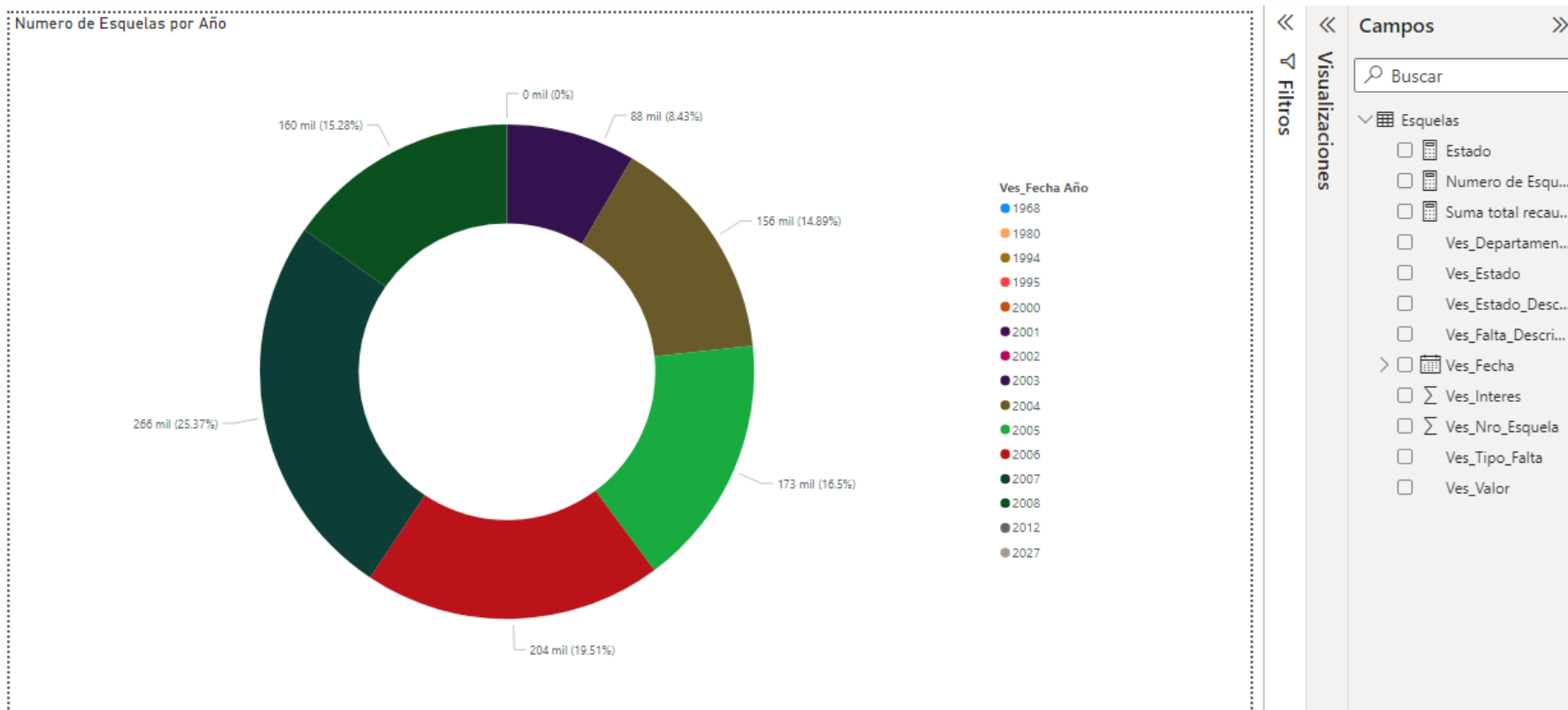
Los primeros datos que mostraremos serán las cantidades de esuelas impuestas en cada departamento del país, siendo San Salvador el departamento que más multas acumula.



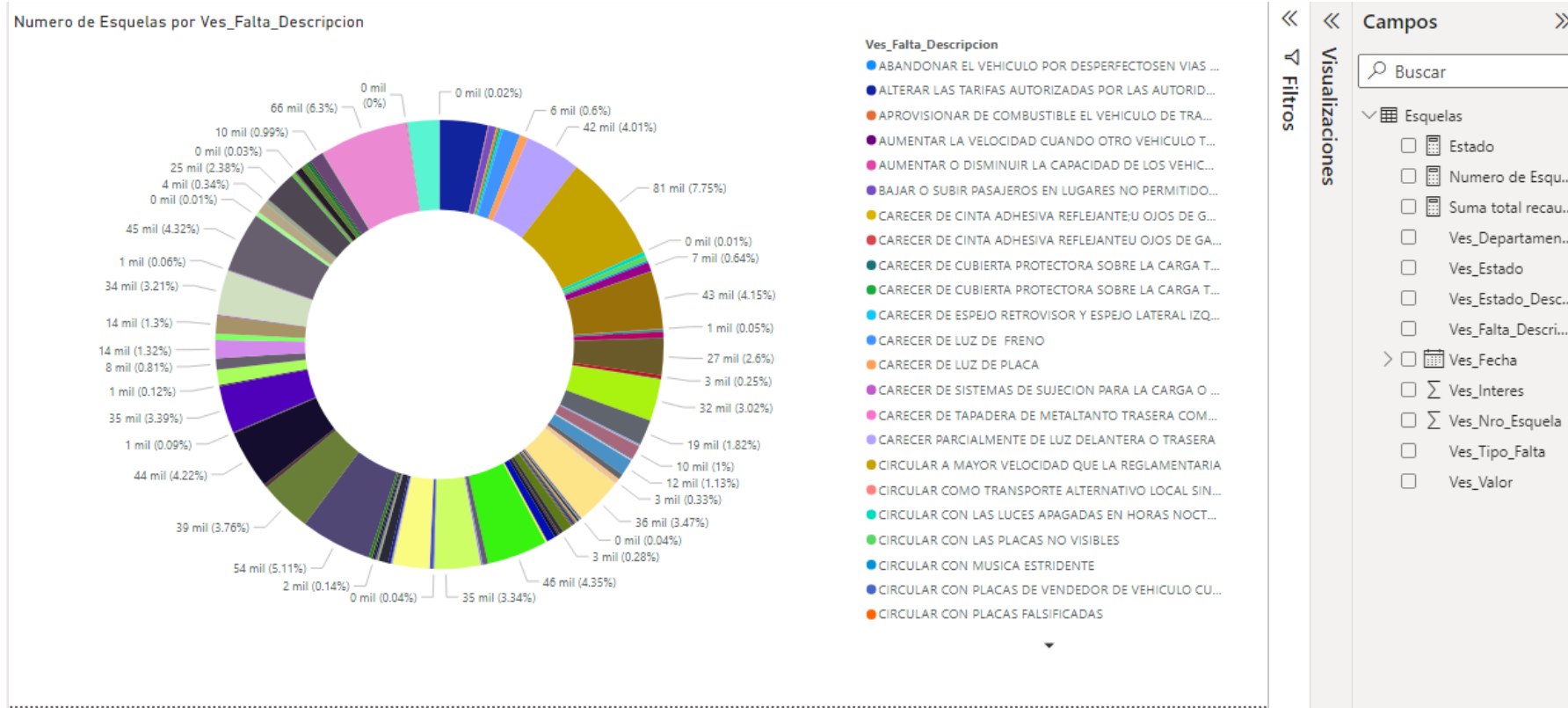
La siguiente consulta la hemos enfocado en conocer los estados de las más de un millón de multas impuestas, como podemos observar, mas de 800 mil multas ya han sido canceladas, representando la mayoría.



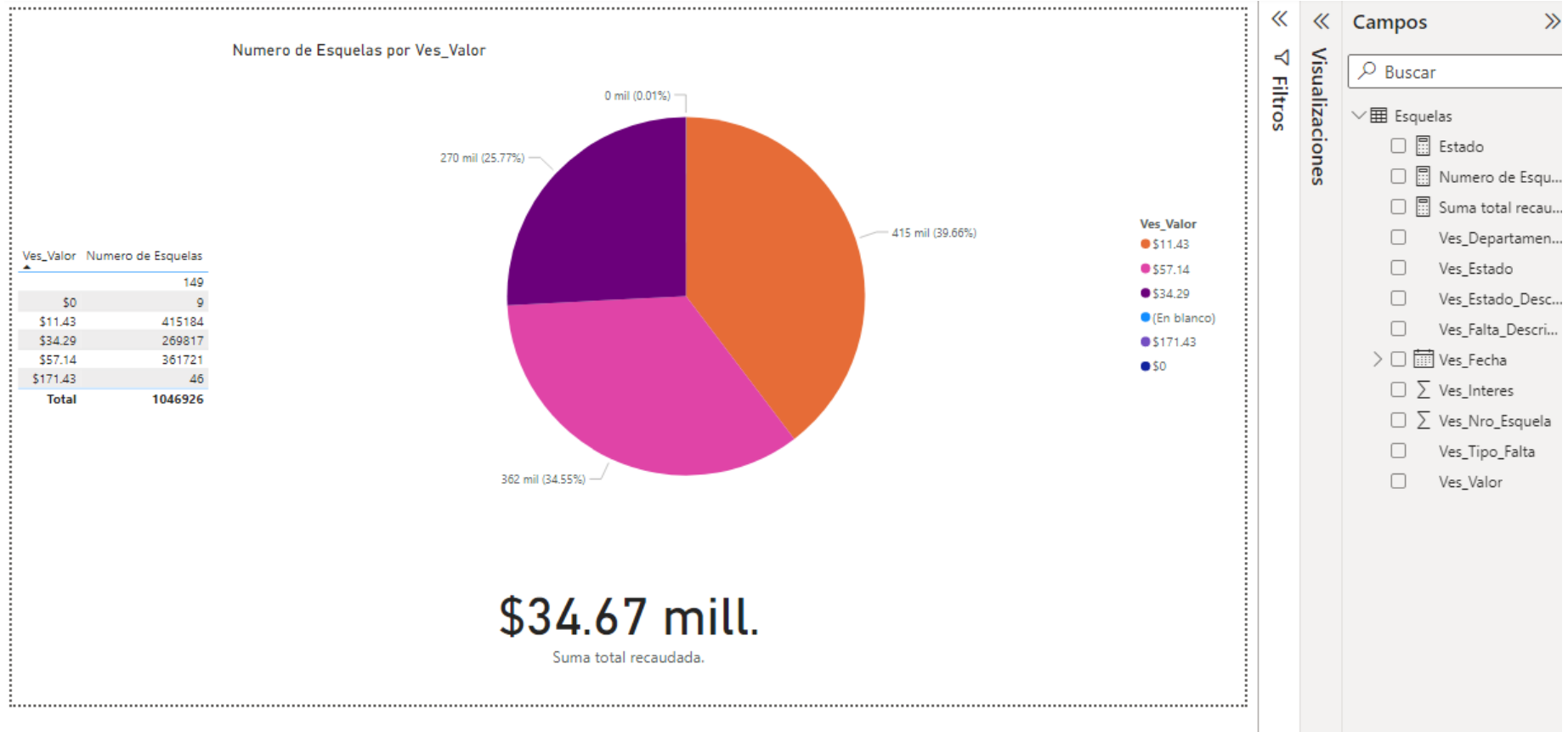
En cuanto al numero de esquelas por años, 2006, 2007 y 2008 encabezan la lista.



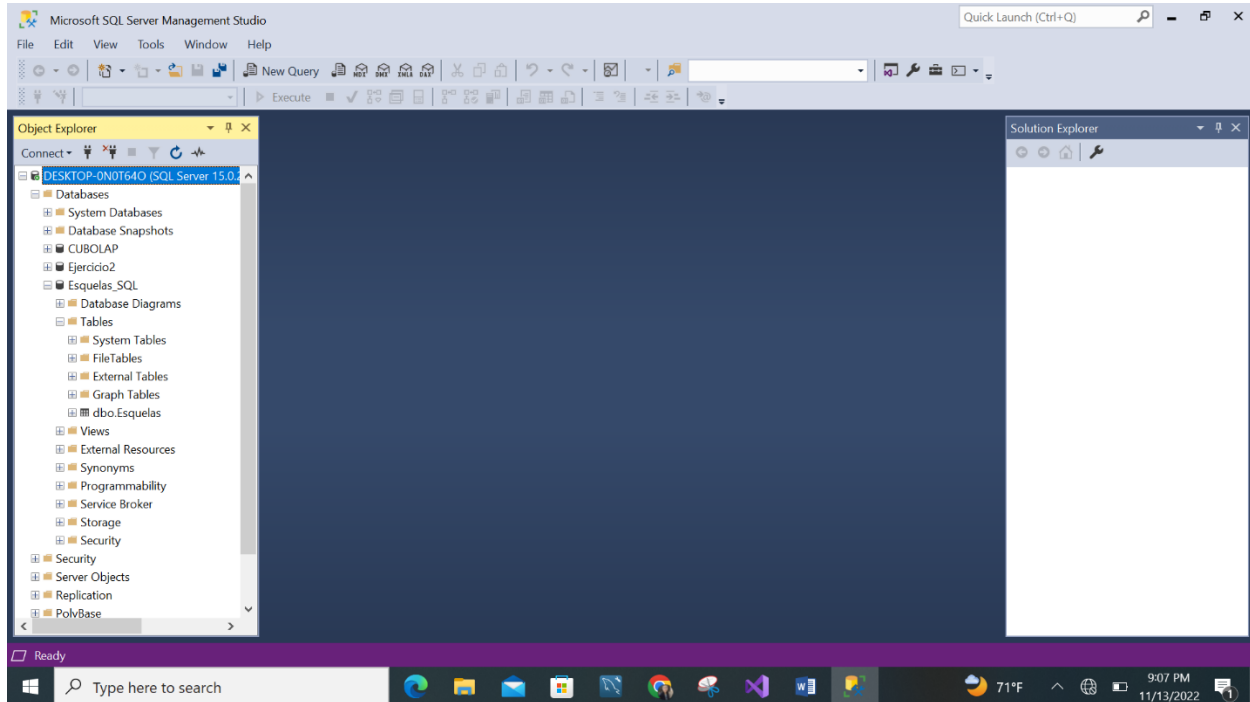
En cuanto a los motivos de las multas, podemos observar que las razones son muy diversas, sin embargo, conducir a una velocidad mayor a la permitida es la razón más frecuente de multas.



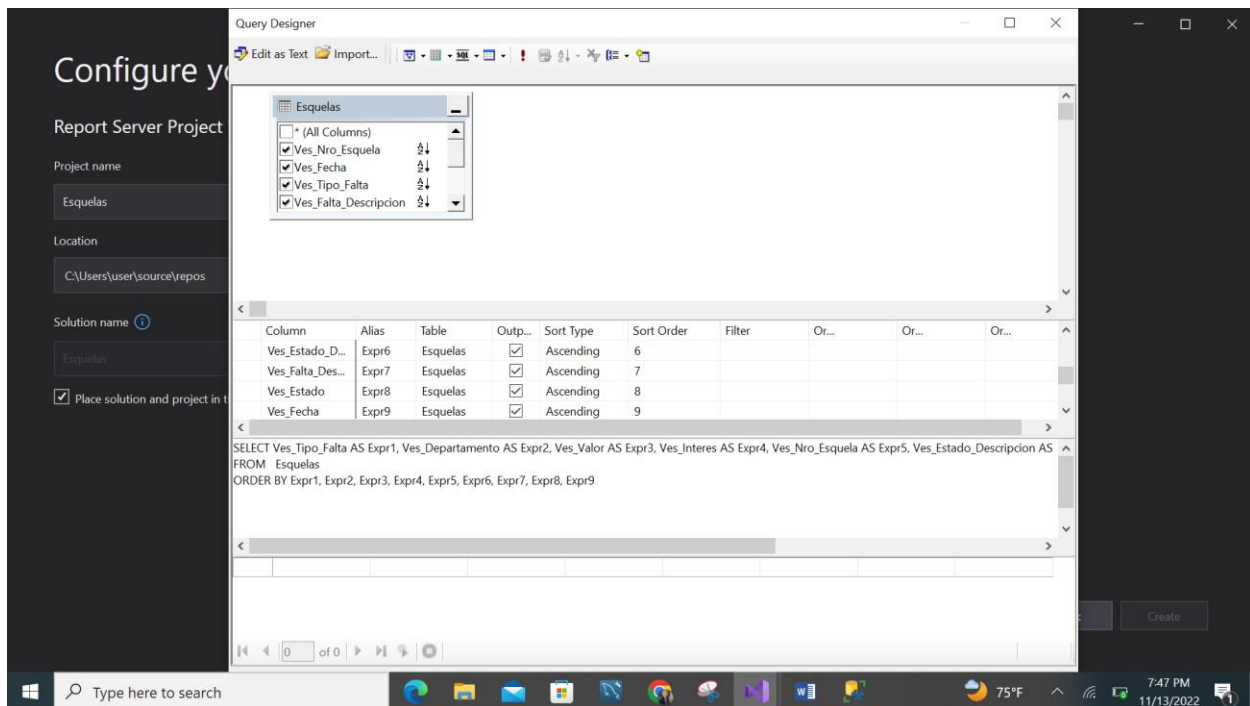
En cuanto a valor monetario, hemos clasificado el grafico de pastel por los montos en que se puede recibir una multa y vemos que la multa mas aplicada es la de \$57.14 y también hemos totalizado el valor de las multas, por lo que obtuvimos que el gobierno ha recaudado más de \$34 millones de dólares en multas.



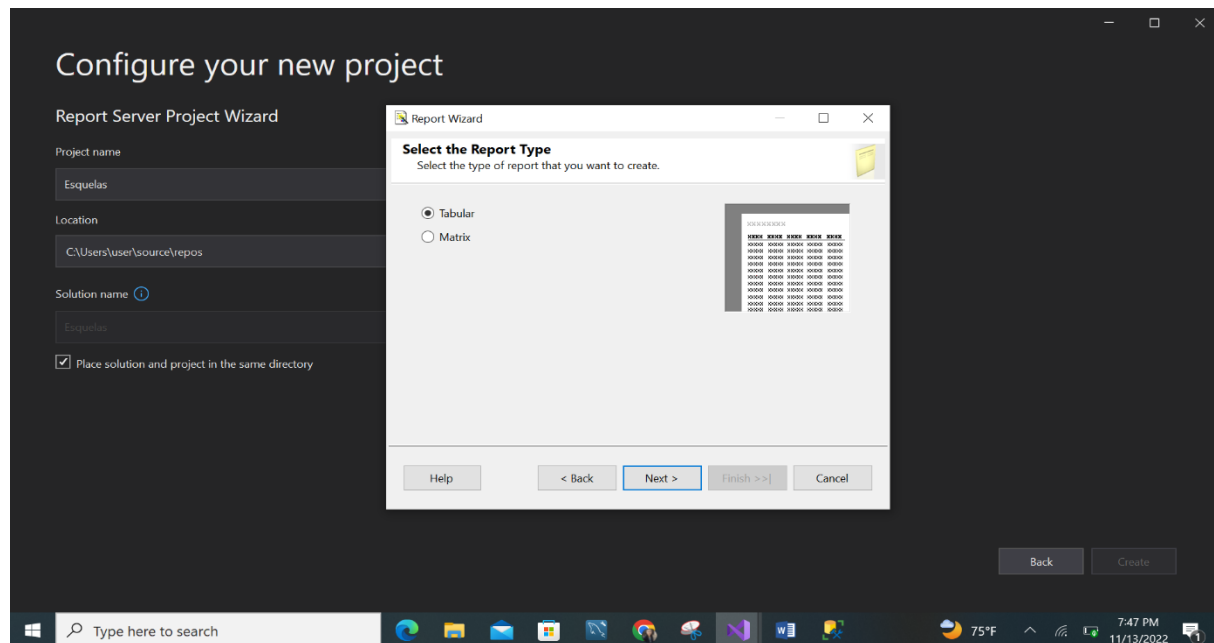
TEMA: REPORTING SERVICES



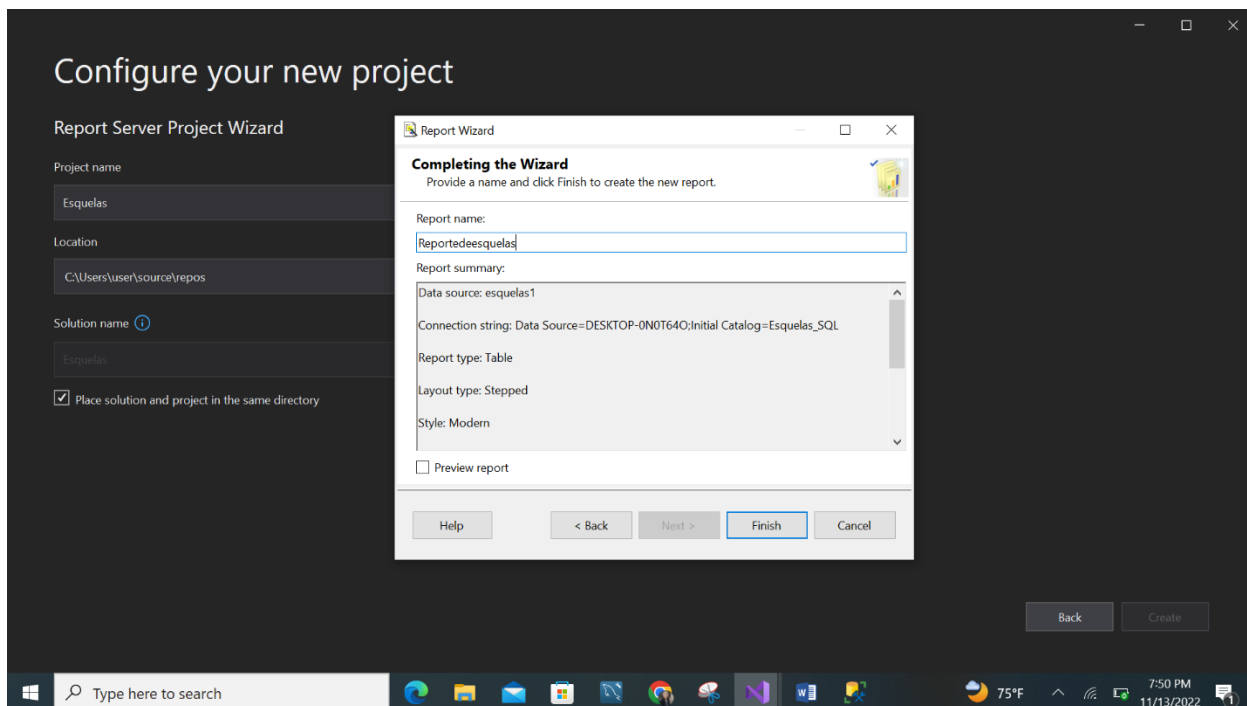
Para poder hacer un mayor análisis y uso de reporting services hemos importado el archivo cvs, el cual nos proporcionara un conocimiento mejor de los datos de Esquelas en el MOP.



Haciendo uso de visual studio creamos un proyecto en el cual diseñamos las características de nuestro reporte, esto para poder organizar los datos e información que se encontraban en SQL Server.



Hacemos una selección del tipo de reporte, en el que tabulamos los datos.



Completamos de organizar nuestro reporte y finalizamos.

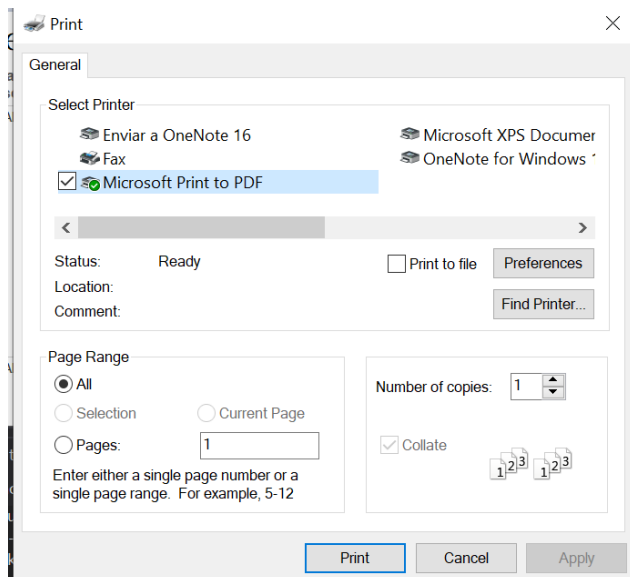
SSIS Toolbox Reportedeesquelas.rdl [Design] 75% Find Next

Design Preview

Reportedeesquelas

falta descripcion	departament o	valor	Interes	Numero estado	falta	estado des	fecha
TRANSITO		11.430000305 1758	0	694 CANCELADA	INCUMPLIR LAS DISPOSICION ES RELATIVAS AL PORCENTAJE MINIMO DE PASO DE LUZ SOLAR EN LOS VIDRIOS POLARIZADO SREGULADOS EN EL ART. 14 -A DE ESTA LEY	CBR	6/12/2006 12:00:00 AM
TRANSITO		11.430000305 1758	0	695 CANCELADA	NO PORTAR LOS TRIANGULOS REFLECTIVO	CBR	6/12/2006 12:00:00 AM

Podemos observar el reporte de esquelas, basado en los datos que se encontraban en nuestro gestor de una manera más gráfica y con alias que ayudan a reconocer los valores internos que se encontraban a modo de dar a conocer la información pertinente.



Y obtener un pdf del reporte, de tal forma que los datos se puedan obtener de una forma más ordenada y presentable.

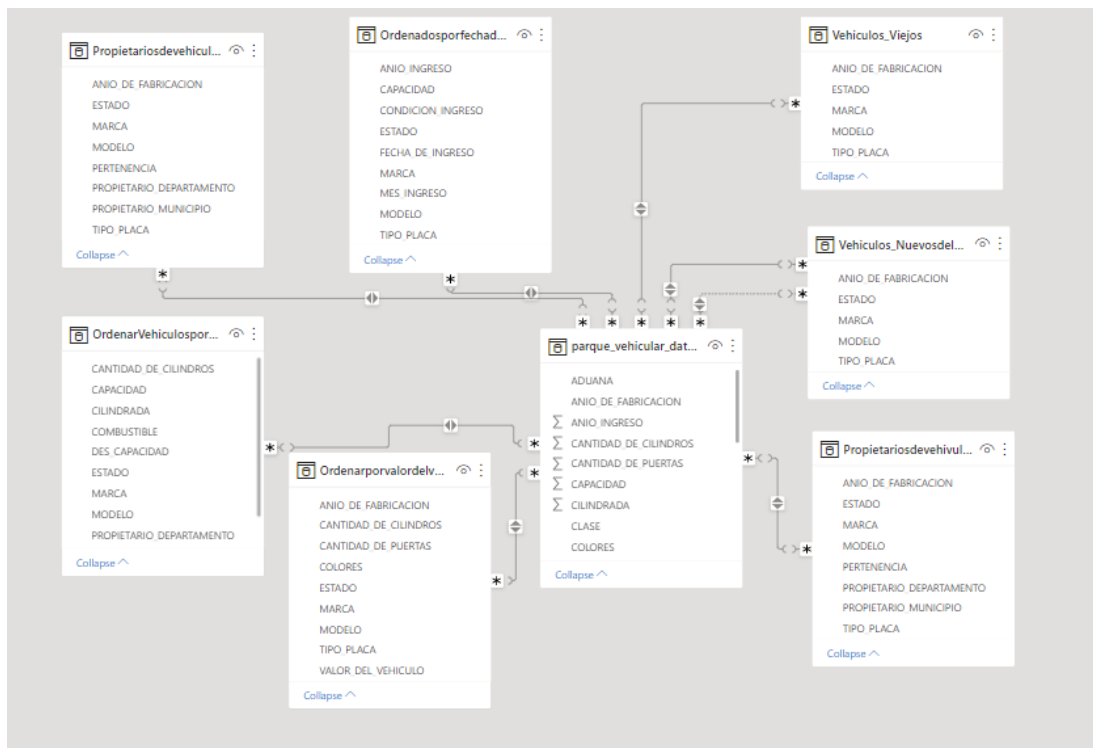
ANALISIS DE PARQUE VEHICULAR

TEMA: CUBO OLAP HACIENDO USO DE POWER BI

- Crearemos las vistas en SQL Server:

```
1 USE OLAP;
2 select*from dbo.parque_vehicular_datos_abiertos_13NOV2018;
3 /*VISTAS PARA CUBO DE OLAP*/
4 USE OLAP;
5 /*VISTA1 AUTOS VIEJOS*/
6 SELECT dbo.parque_vehicular_datos_abiertos_13NOV2018.TIPO_PLACA, dbo.parque_vehicular_datos_abiertos_13NOV2018.ANIO_DE_FABRICACION, dbo.parque_vehicular_datos_abiertos_13NOV2018.ESTADO
7 FROM dbo.parque_vehicular_datos_abiertos_13NOV2018 CROSS JOIN
8     dbo.parque_vehicular_datos_abiertos_13NOV2018 AS parque_vehicular_datos_abiertos_13NOV2018_1
9 WHERE (dbo.parque_vehicular_datos_abiertos_13NOV2018.ANIO_DE_FABRICACION < 2000)
10 /*VISTA2 AUTOS NUEVOS*/
11 SELECT dbo.parque_vehicular_datos_abiertos_13NOV2018.TIPO_PLACA, dbo.parque_vehicular_datos_abiertos_13NOV2018.ANIO_DE_FABRICACION, dbo.parque_vehicular_datos_abiertos_13NOV2018.ESTADO
12 FROM dbo.parque_vehicular_datos_abiertos_13NOV2018 CROSS JOIN
13     dbo.parque_vehicular_datos_abiertos_13NOV2018 AS parque_vehicular_datos_abiertos_13NOV2018_1
14 WHERE (dbo.parque_vehicular_datos_abiertos_13NOV2018.ANIO_DE_FABRICACION > 2000)
15 /*VISTA3 PROPIETARIOS Y AUTOS NUEVOS*/
16 SELECT TIPO_PLACA, ANIO_DE_FABRICACION, MODELO, ESTADO, MARCA, PERTENENCIA, PROPIETARIO_DEPARTAMENTO, PROPIETARIO_MUNICIPIO
17 FROM dbo.parque_vehicular_datos_abiertos_13NOV2018
18 WHERE (ANIO_DE_FABRICACION > 2000)
19 /*VISTA4 PROPIETARIOS Y AUTOS VIEJOS*/
20 SELECT TIPO_PLACA, ANIO_DE_FABRICACION, MODELO, ESTADO, MARCA, PERTENENCIA, PROPIETARIO_DEPARTAMENTO, PROPIETARIO_MUNICIPIO
21 FROM dbo.parque_vehicular_datos_abiertos_13NOV2018
22 WHERE (ANIO_DE_FABRICACION < 2000)
23 /*AUTOS CON SU TIEMPO DE INGRESO EN EL PAIS*/
24 SELECT TOP (100) PERCENT TIPO_PLACA, MODELO, MARCA, CAPACIDAD, CONDICION_INGRESO, ESTADO, ANIO_INGRESO, FECHA_DE_INGRESO, MES_INGRESO
25 FROM dbo.parque_vehicular_datos_abiertos_13NOV2018
26 ORDER BY ANIO_INGRESO;
```

- Luego de cargar el archivo SQL procedemos a hacer el análisis y hacemos la optimización por medio del modelo estrella:



- Organizamos los datos por medio de tablas:

TIPO_PLACA	MODELO	MARCA	CAPACIDAD	CONDICION_INGRESO	ESTADO	ANIO_INGRESO	FECHA_DE_INGRESO	MES_INGRESO
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	17/4/2004 10:31:18	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	21/4/2004 16:15:53	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	21/4/2004 16:23:56	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	5/4/2004 10:30:49	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	5/4/2004 10:32:58	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	13/4/2004 09:54:08	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	30/4/2004 15:04:10	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	30/4/2004 13:42:53	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	24/4/2004 10:44:42	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	19/4/2004 11:08:58	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	19/4/2004 11:10:24	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	16/4/2004 10:44:41	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	27/4/2004 11:47:10	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	30/4/2004 13:44:48	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	22/4/2004 15:13:24	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	7/4/2004 12:27:18	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	7/4/2004 13:56:58	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	7/4/2004 14:00:25	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	16/4/2004 11:07:48	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	30/4/2004 11:28:13	4.00
PARTICULAR	COROLLA	TOYOTA	5.00	VEHICULO USADO	ALTA	2004.00	14/4/2004 15:38:57	4.00

Tabla 1: Ordenados por fecha de ingreso

TIPO_PLACA	MARCA	MODELO	COLORES	CANTIDAD_DE_PUERTAS	CANTIDAD_DE_CILINDROS	ANIO_DE_FABRICACION	VALOR_DEL_VEHICULO	ESTADO
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4991.77	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	6295.46	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	5485.26	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	6726.43	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4437.08	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	6681.12	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4496.83	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	7385.37	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	8292.92	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4177.04	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	5394.47	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	5418.88	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4823.68	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	5459.31	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	2909.75	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	4179.83	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	6235.48	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	7074.36	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	5295.71	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	8891.68	ALTA
PARTICULAR	TOYOTA	COROLLA LE	GRIS	4.00	4.00	2010	3997.37	ALTA

Tabla 2: Ordenadas por valor de Vehículo

TIPO_PLACA	ESTADO	PROPIETARIO_DEPARTAMENTO	PROPIETARIO_MUNICIPIO	CAPACIDAD	MODELO	MARCA	DES_CAPACIDAD	COMBUSTIBLE	CANTIDAD_DE_
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00
PARTICULAR	ALTA	SAN SALVADOR	SAN SALVADOR	5.00	SPECTRA EX	KIA	ASIENTOS	GASOLINA	4.00

Tabla 3: Ordenar vehículos por capacidad

TIPO_PLACA	ANIO_DE_FABRICACION	CILINDRADA	CANTIDAD_DE_CILINDROS	CANTIDAD_DE_PUERTAS	VALOR_DEL_VEHICULO	COLORES	FECHA_DE_IMPORTACION
PARTICULAR	1984	0	0	0	0	ROJO	
PARTICULAR	1979	0	0	0	0	ROJO	
PARTICULAR	1984	0	0	0	0	ROJO	
PARTICULAR	1978	0	0	0	0	ROJO	
PARTICULAR	1982	0	0	0	0	ROJO	
PARTICULAR	1983	0	0	0	0	ROJO	
PARTICULAR	1982	0	0	0	0	ROJO	
PARTICULAR	1986	0	0	0	0	ROJO	
PARTICULAR	1985	0	0	0	0	ROJO	
PARTICULAR	1980	0	0	0	0	ROJO	
PARTICULAR	1978	0	0	0	0	ROJO	
PARTICULAR	1978	0	0	0	0	ROJO	
PARTICULAR	1987	0	0	0	0	ROJO	
PARTICULAR	1985	0	0	0	0	ROJO	
PARTICULAR	1984	0	0	0	0	ROJO	
PARTICULAR	1984	0	0	0	0	ROJO	
PARTICULAR	1980	0	0	0	0	ROJO	
PARTICULAR	1980	0	0	0	0	ROJO	
PARTICULAR	1984	0	0	0	0	ROJO	
PARTICULAR	1981	0	0	0	0	ROJO	

Tabla 4: Parque vehicular datos abiertos

[illegible]

Tabla 5: Propietarios de Vehículos viejos

[illegible]

Tabla 6: Propietarios de vehículos nuevos

TIPO_PLACA ▼	ANIO_DE_FABRICACION ▼	MARCA ▼	MODELO ▼	ESTADO ▼
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA
PARTICULAR	2006	TOYOTA	HILUX	ALTA

Tabla 7: Vehículos nuevos del 2000

TIPO_PLACA ▼	ANIO_DE_FABRICACION ▼	MARCA ▼	MODELO ▼	ESTADO ▼
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA
PARTICULAR	1984	TOYOTA	N/D	ALTA

Tabla 8: Vehículos Viejos del 2000

- Los siguientes gráficos son basados en las vistas y las tablas.

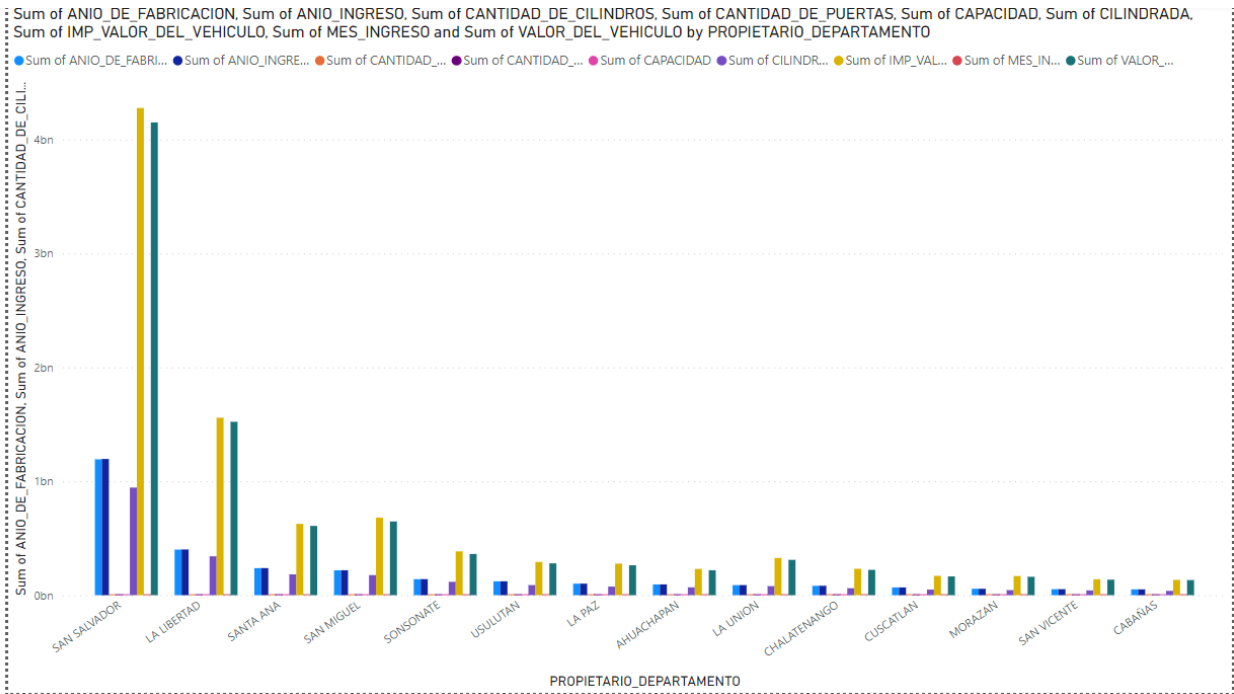


GRAFICO 1: Gráficos por departamentos

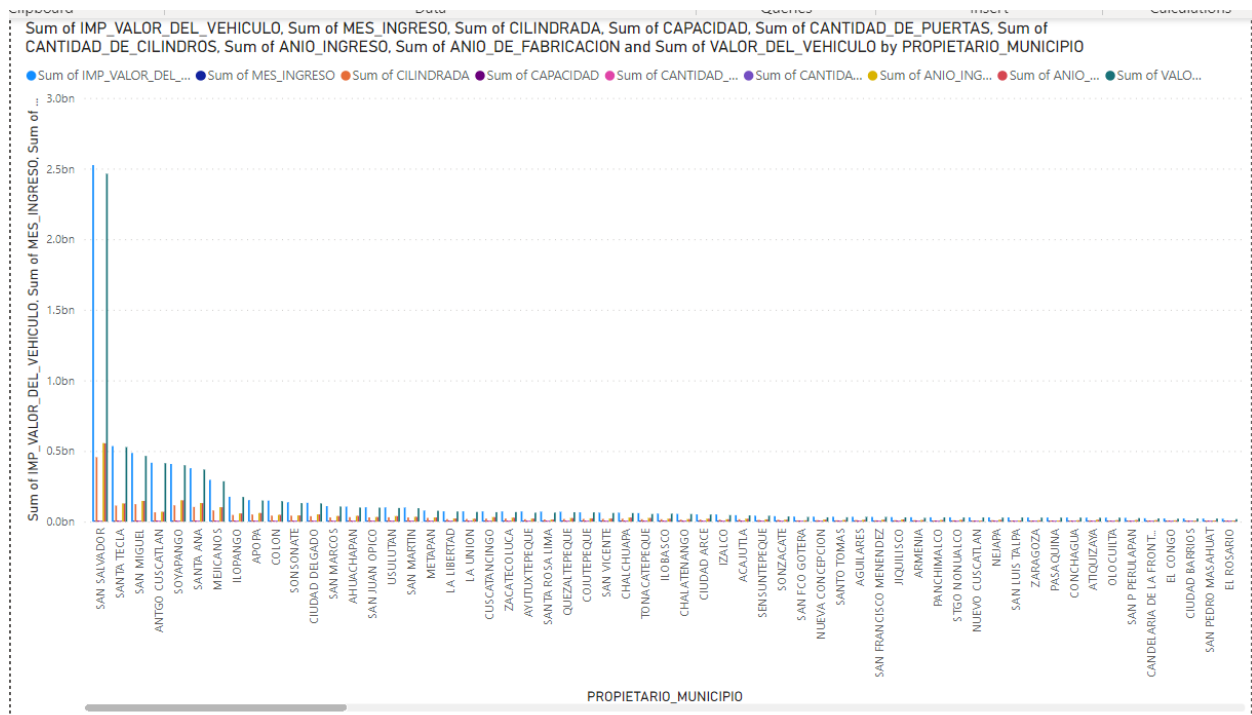


Grafico 2: Gráficos por municipio

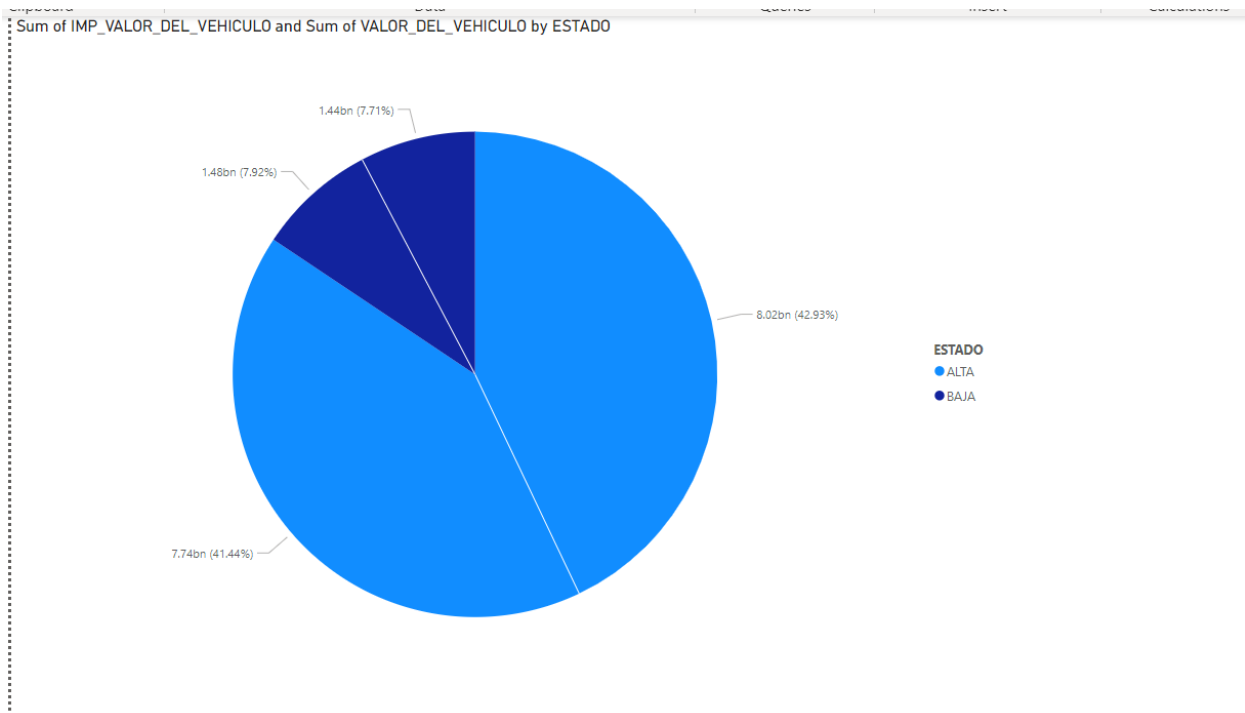


Grafico 3: Valores por estado

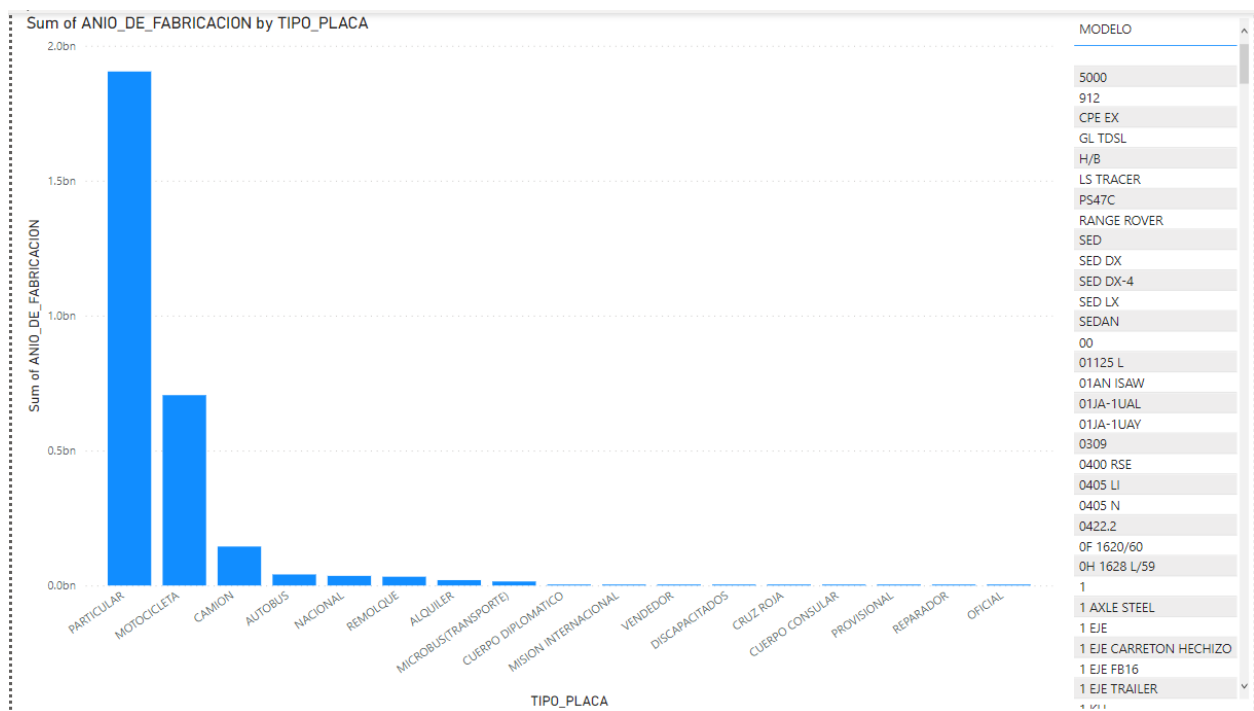


Grafico 4: Vehículos por año

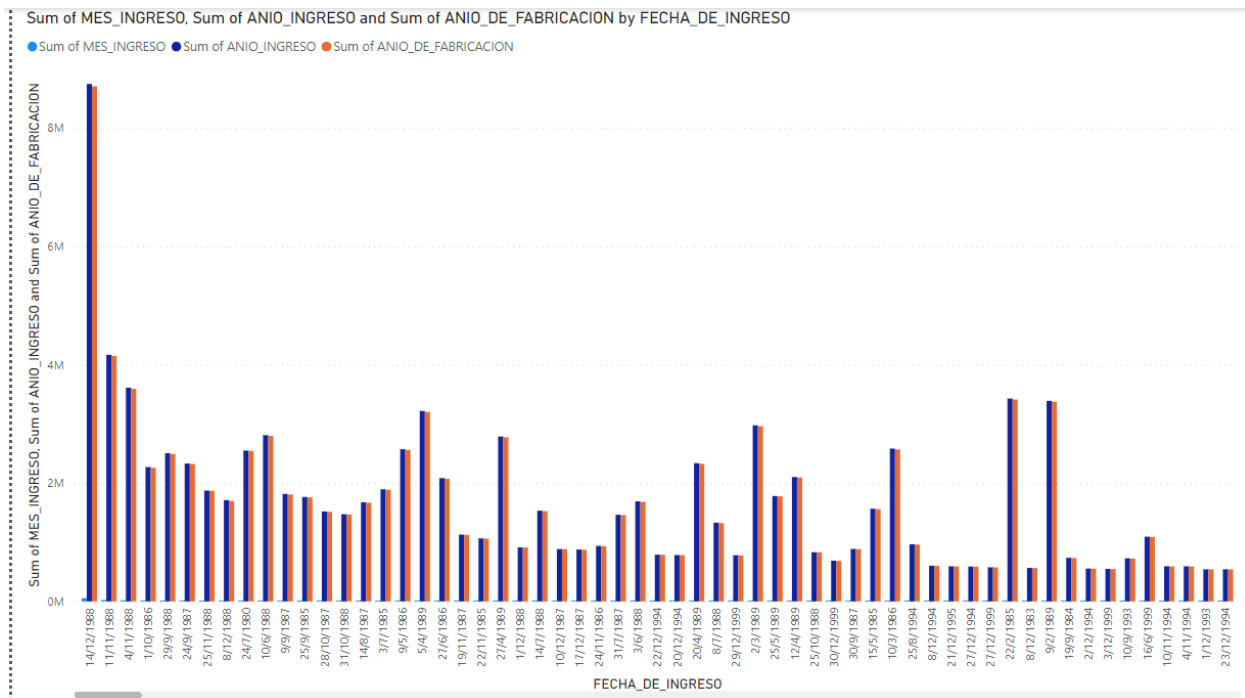


Grafico 5: Valores de egreso

- Impresión del reporte:

The screenshot displays the Power BI Desktop application. On the left, the 'Export' menu is open, showing options: 'Power BI template', 'Export to PDF', 'Recent sources', 'Transform data', 'Refresh', 'New visual', 'Text box', 'More visuals', 'New measure', 'Quick measure', 'Sensitivity', 'Publish', 'Sensitivity', and 'Share'. The main area shows a bar chart visualization of the same data as in Grafico 5. The right-hand pane shows the 'Visualizations' and 'Fields' panes. The 'Visualizations' pane has a 'Build visual' section with a 'Filters' dropdown. The 'Fields' pane shows a list of fields including 'Ordenados por fecha de...', 'Ordenados por valor de veh...', 'Ordenar Vehiculos por ca...', 'parque_vehicular_datos...', 'Propietarios de vehiculo...', 'Propietarios de vehiculo...', 'Vehiculos_Nuevos del 20...', and 'Vehiculos_Viejos'. The bottom status bar shows the date and time as 9:00 PM 12/1/2022.

TEMA: K-MEANS USANDO HERRAMIENTAS DE RAPIDMINER.

Para la realización de análisis K-means podemos hacer uso de diversas herramientas, en nuestro caso usaremos el software Rapidminer y en específico su asistente llamado automodel, en el cual iremos seleccionando la información que representaremos por medio del modelo k-means.

- Para empezar, escogemos la base de datos con la que trabajaremos, en este caso será: `parque_vehicular_datos_abiertos_13NOV2018`

The screenshot shows the 'Auto Model' tab in the Rapidminer interface. The workflow bar at the top has six steps: 'Load Data', 'Select Task', 'Prepare Target', 'Select Inputs', 'Model Types', and 'Results'. The 'Load Data' step is currently selected, indicated by a green circle. Below the workflow bar are buttons for 'RESTART', 'BACK', and 'NEXT'. The main area is titled 'Select Data for a New Model' and displays a tree view of data sources on the left. The 'Local Repository (Local)' is expanded, showing a folder named 'parque_vehicular_datos_abiertos_13NOV2018' which is highlighted. On the right, an 'Information' panel shows details for the selected dataset: Name: 'parque_vehicular_datos_abiertos_13NOV2018', Number of rows: '1,451,100', and Number of columns: '23'. Below this, the 'Attributes / Columns' list includes: TIPO_PLACA, AÑO_DE_FABRICACION, CANTIDAD_DE_CILINDROS, CANTIDAD_DE_PUERTAS, VALOR_DEL_VEHICULO, COLORES, FECHA_DE_IMPORTACION, IMP_VALOR_DEL_VEHICULO, FECHA_DE_INGRESO, AÑO_INGRESO, MES_INGRESO, CLASE, PERTENENCIA, MARCA, MODELO, CAPACIDAD, DES_CAPACIDAD, COMBUSTIBLE, ADUANA, CONDICION_INGRESO, PROPIETARIO_DEPARTAMENTO, PROPIETARIO_MUNICIPIO, ESTADO.

Rapidminer nos da la opción del análisis que deseamos hacer, ya sea: predicciones, agrupamientos o identificación de valores atípicos, para nuestro análisis usaremos los agrupamientos

This screenshot shows the 'Select Task' step in the Rapidminer workflow. The workflow bar at the top has the same six steps as the previous image, but now 'Select Task' is the active step, highlighted with a green circle. The 'NEXT' button is highlighted in green. Below the workflow bar, there are three task options presented in rounded rectangular boxes:

- Predict**: Want to predict the values of a column?
- Clusters**: Want to identify groups in your data?
- Outliers**: Want to detect outliers in your data?

Luego manualmente filtraremos los datos relevantes para nuestros agrupamientos.

Selected: 9 / Total: 23

☒ Select All ☐ Deselect All

Selected	Status ↑	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input checked="" type="checkbox"/>	●		TIPO_PLACA	?	0.00%	65.58%	0.00%	4.50%
<input checked="" type="checkbox"/>	●		ANIO_DE_FABRICACION	?	0.01%	4.11%	0.02%	0.00%
<input type="checkbox"/>	●		CANTIDAD_DE_CILINDROS	?	?	44.36%	14.72%	0.00%
<input type="checkbox"/>	●		CANTIDAD_DE_PUERTAS	?	?	43.06%	39.32%	0.00%
<input checked="" type="checkbox"/>	●		VALOR_DEL_VEHICULO	?	?	11.10%	0.00%	0.00%
<input type="checkbox"/>	●		COLORES	?	0.42%	13.29%	0.01%	15.70%
<input checked="" type="checkbox"/>	●		FECHA_DE_IMPORTACION	?	0.56%	0.13%	15.16%	6.12%
—	—							

Al tener claros los datos que analizaremos, debemos escoger el número de agrupamientos que deseamos, en este caso serán 3 ya que el tamaño de la base de datos es tan extenso que la computadora no soporta un análisis mayor.

Models

☒ **k-Means Clustering**

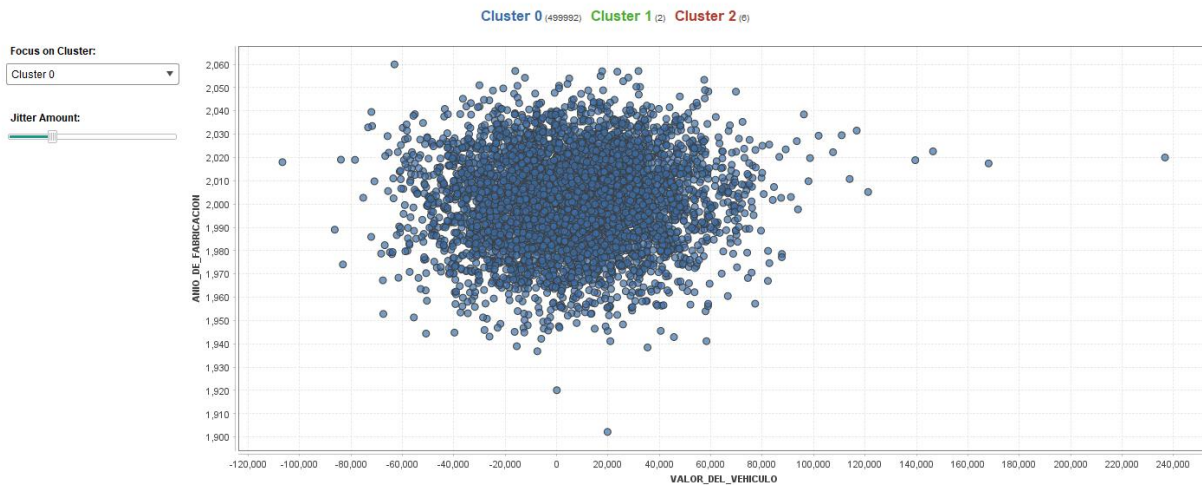
Number of Clusters:

☐ **x-Means Clustering**

Maximal Number of Clusters:

Al finalizar el análisis obtenemos datos relevantes como por ejemplo en nuestro caso analizamos la relación entre el valor promedio de los vehículos y su año de ingreso al país.

k-Means - Scatter Plot



k-Means - Summary

Number of Clusters: 3

Cluster 0

499,992

VALOR_DEL_VEHICULO is on average 0.26% smaller, ANIO_DE_FABRICACION is on average 0.00% larger, ANIO_INGRESO is on average 0.00% larger

Cluster 1

2

VALOR_DEL_VEHICULO is on average 66,392.14% larger, ANIO_INGRESO is on average 25.32% smaller, ANIO_DE_FABRICACION is on average 0.46% smaller

Cluster 2

6

ANIO_DE_FABRICACION is on average 99.04% smaller, VALOR_DEL_VEHICULO is on average 89.68% smaller, ANIO_INGRESO is on average 19.29% smaller

Con esto podemos observar que no necesariamente los vehículos mas nuevos son los mas costosos, ya que los datos arrojan que muchos vehículos antiguos tienen un valor muy elevado.

III. ANTECEDENTES

El análisis y la minería de datos se han convertido en una necesidad para las empresas e instituciones, esto debido a que la toma de decisiones tiene como uno de sus principales factores las predicciones basadas en las consecuencias de las acciones previas, por lo cual, no solamente se trata de llevar un registro de los datos, sino, de saber obtener información de relevancia por medio la interpretación de estos.

IV. SITUACION ACTUAL

El “Ministerio de obras públicas” no cuenta con un análisis de los datos que tienen almacenados, únicamente poseen una base de datos con los registros de las esquelas, dicha base de datos ha sido entregada en un documento CSV el cual consta de una única tabla con diversos campos relacionados a las esquelas impuestas, dicha base consta de más de un millón de registros los cuales deben ser analizados.

La información presentada de un “Parque Vehicular” mediante un archivo CSV el cual consta únicamente de una tabla y de un volumen de archivos grandes, ha sido presentada para ser analizados.

V. METODOLOGIA

A metodología utilizada en el presente proyecto será la de investigación descriptiva, ya que se utilizará dos bases de datos desorganizadas, las cuales serán analizadas por medio de diversas técnicas de minería de datos, con el fin de obtener información relevante las cuales serán presentadas de una manera organizada.

VI. FORMULACION DEL PROBLEMA

La actual base de datos que posee el “Ministerio de obras públicas”, cuenta con registros de las esquelas impuestas sobre el territorio nacional, sin embargo, no tienen la capacidad de analizar dichos datos, por lo que no tienen un funcionamiento más que llevar el registro histórico, es necesario poder analizar la información que poseen ya que de esta manera se pueden tomar medidas de prevención y/o corrección.

En la base de datos presentada de un “Parque Vehicular” no existe una adecuada organización y optimización de los datos, por lo cual se necesita emplear un informe correcto y detallado para presentar un análisis y solución del problema.

VII. FACTIBILIDAD

El análisis de los datos se realizará utilizando diversas herramientas de software y sus respectivas licencias, además se utilizará equipo con características de gamas media, esto debido a que el procesamiento de bases de datos tan extensas obliga a utilizar un poder de procesamiento medio-alto, luego del procesamiento de los datos, se espera que la información obtenida pueda ser utilizada para la creación de nuevas estrategias y planes por parte de la institución.

VIII. JUSTIFICACION

Hemos implementado **estrategias de minería de datos** para poder organizar una cantidad grande de datos y hacer un análisis de las esquelas de tránsito y del parque vehicular, lo cual son temas de los que se desean presentar reportes.

Usando **estrategias de minería de datos**, podemos visualizar los datos de una forma más entendible y gráfica, siendo esta una ayuda para el MOP, ya que podrán valorar los datos y dar una respuesta inmediata a situaciones que se presenten dentro de los resultados, por esa razón se presenta esta documentación.

Haciendo una aplicación correcta de las herramientas podremos conocer la manera correcta de adecuar los datos y beneficiar al parque vehicular, para encontrar respuestas a soluciones que se pueden tener mediante el orden de los datos.

IX. IMPORTANCIA

La importancia de poder hacer uso de estas estrategias es el poder documentar un gran volumen de datos y poder brindar resultados más gráficos, siendo este el uso de Power BI, en el cual hicimos gráficos los resultados, análisis OLAP haciendo uso del programa Power BI, Reporting Services y k-means utilizando Rapidminer , herramientas que facilitan el manejo de grandes cantidades de datos, los cuales sería muy difícil analizar de manera manual , por lo cual, dichas herramientas nos ayudan a hacer la impresión de los datos para que se pueda tener acceso por quienes observaran los análisis realizados.

X. ALCANCES Y LIMITACIONES

Se debe tener equipos y herramientas técnicas adecuadas como así también una excelente gestión a la hora de hacer los procesos de ordenamiento de datos mediante las estrategias de minería, razón por la cual se puede tener una complicación cuando se tenga ya presente los documentos, si estos no han tenido un proceso correcto en el desarrollo del análisis de los datos, por eso se dice que la mayor complicación es que las herramientas son demasiadas complejas.

XI. PLANIFICACION DE RECURSOS

A la hora de planificar se debe de tener presente el volumen, tipo de datos, variables, y todos aquellos aspectos necesarios antes de implementar la herramienta para que se puedan detallar la información y datos de una forma más clara, eso nos dará una mejor referencia de planificación.

XII. RECOMENDACIONES

- Tener presente el formato de archivo para el uso de la herramienta.
- Hacer una investigación de la implementación de la herramienta.
- Tener presente la entrega de los datos físicos. Valorar los datos más importantes que serán expuestos de acuerdo al tipo de reporte que se quiere presentar.