

Universidad Don Bosco



Materia:

Datawarehouse y Minería de
Datos

Docente:

Karens Medrano

Actividad:

Proyecto de cátedra

Estudiantes:

René Saúl Jovel Calderón - JC211517

Debora Elizabeth Rosales Angel - RA211941

Enlace de Github:

<https://github.com/saulcalderon/DMD-proyecto>

Objetivos

Objetivo General:

- Generar un informe para la presentación al Gobierno de la República de El Salvador utilizando estrategias de minería de datos para el análisis de la información del tema: **Esquelas de infracción de tránsito.**

Objetivo Específico:

- Demostrar el uso de la aplicación POWER BI, representando la información proporcionada por el MOP visualizada en gráficos.
- Determinar el comportamiento general de los datos proporcionados, en este caso de las esquelas para encontrar patrones de interés.

Problema

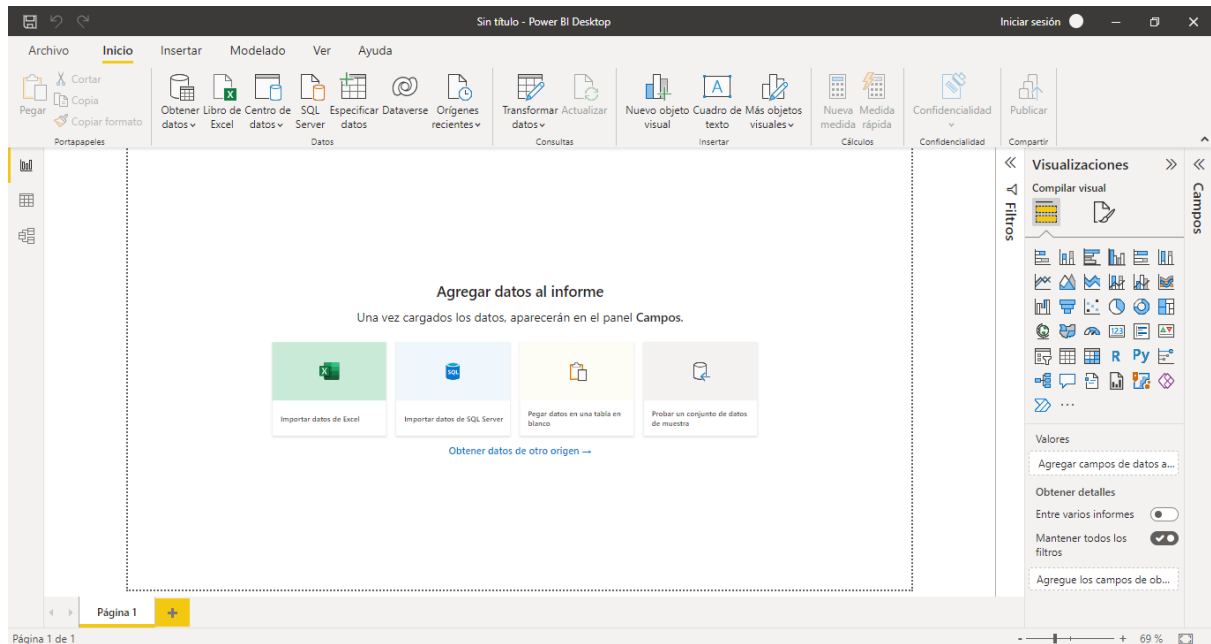
Presentación de información al gobierno central de esquelas desde un punto estadístico/informativo.

Dada la información proporcionada por el Ministerio de Obras Públicas (MOP), se recomienda realizar diferentes análisis de los datos para buscar posibles soluciones para el futuro. Con estos datos también podemos determinar un comportamiento general de las esquelas registradas. Los datos son agrupados mediante la frecuencia de cada uno de los atributos o campos relevantes en el documento proporcionado.

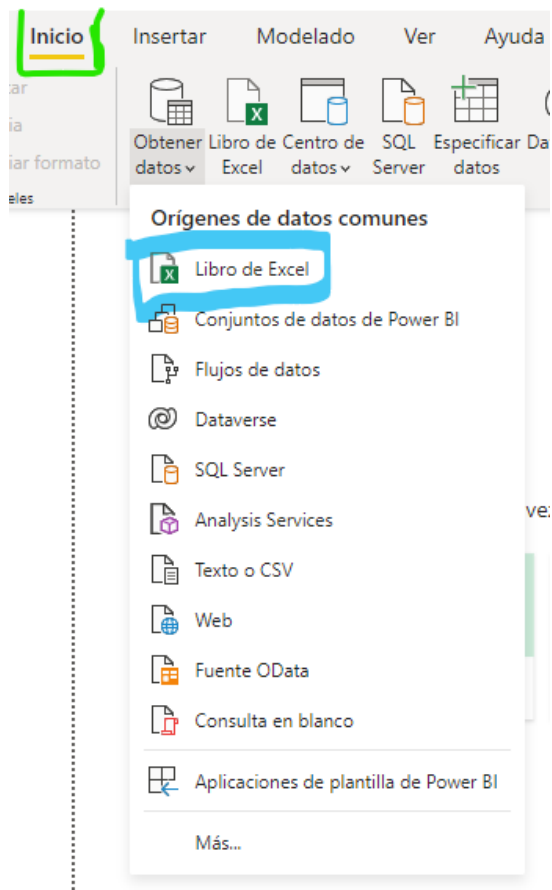
Power BI

La primera estrategia de minería de datos para esuelas usamos **POWER BI**:

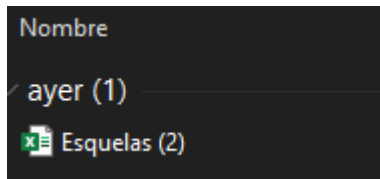
Como primer paso se abre POWER BI



Dar clic en INICIO, opción OBTENER DATOS clic en Libro de Excel



Se agrega el documento deseado, en nuestro caso el libro de ESQUELAS

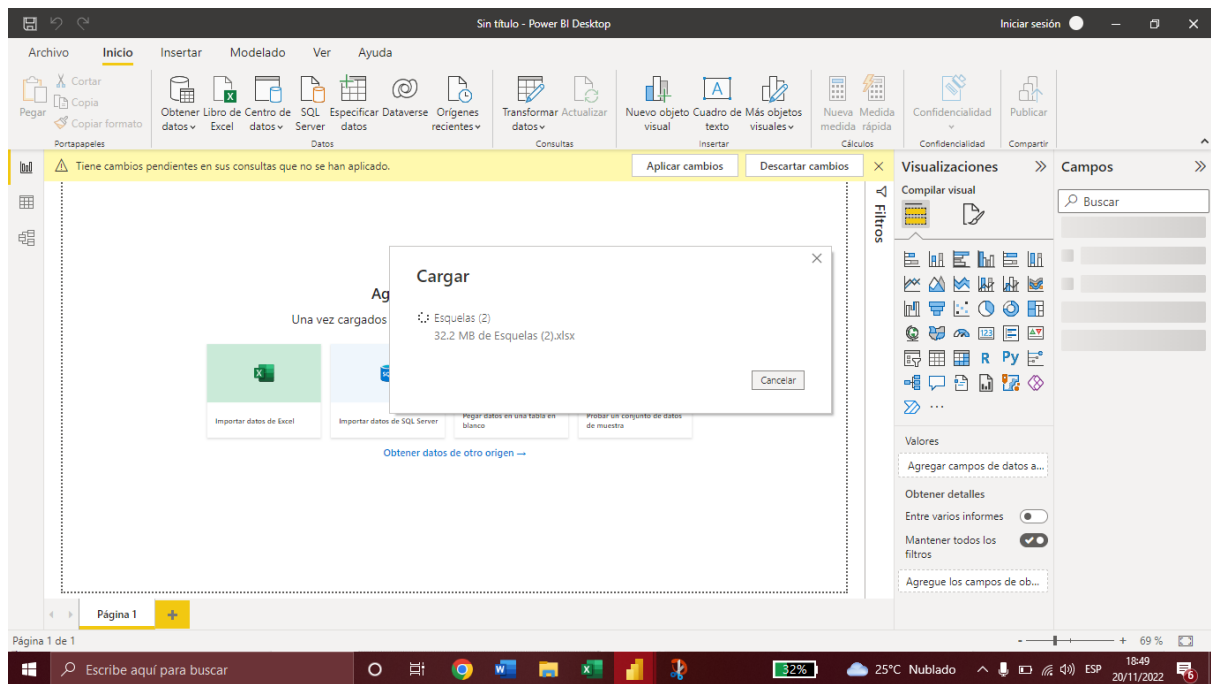


Le damos clic en el cuadro que aparece cerca del nombre del archivo para poder seleccionar el documento y le damos CARGAR

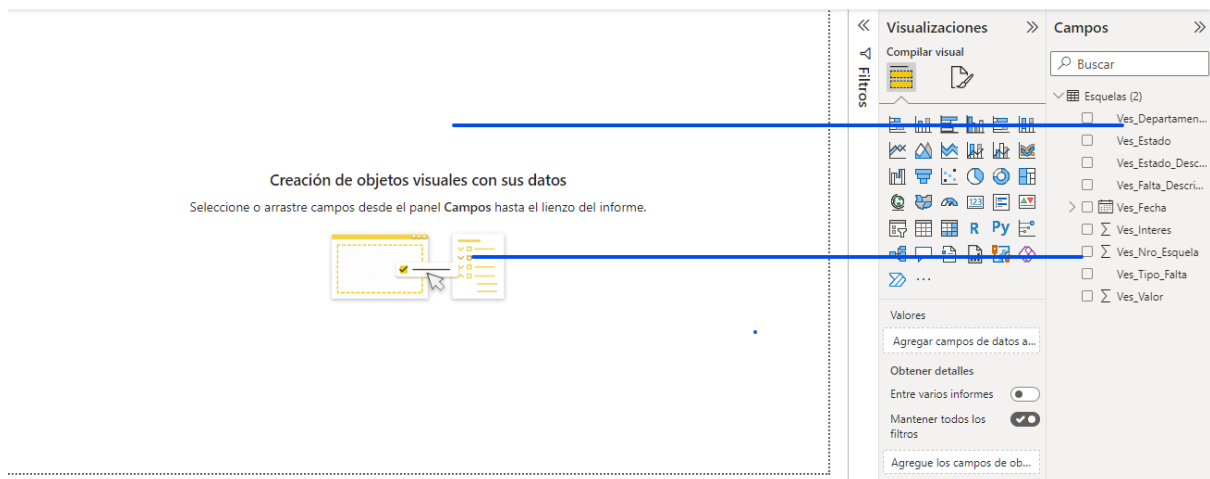
The 'Navegador' window is shown with a search bar and a list of files. 'Esquelas (2).xlsx [1]' is expanded, showing 'Esquelas (2)' with a checkmark and a table icon. To the right, a table titled 'Esquelas (2)' is displayed. The table has four columns: 'Ves_Nro_Esqueda', 'Ves_Fecha', 'Ves_Tipo_Falta', and 'Ves_Falta_Descripcio'. Below the table, there are three buttons: 'Cargar' (highlighted with a blue circle), 'Transformar datos', and 'Cancelar'.

Ves_Nro_Esqueda	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcio
110637	4/6/2003 00:00:00	TRANSITO	NO UTILIZAR EL CC
107742	4/6/2003 00:00:00	TRANSITO	NO UTILIZAR EL CC
106645	4/6/2003 00:00:00	TRANSITO	ESTACIONARSE EN
154061	4/6/2003 00:00:00	TRANSITO	BAJAR O SUBIR PA
154102	3/6/2003 00:00:00	TRANSITO	DISPUTARSE LA VI
119779	6/6/2003 00:00:00	TRANSITO	CIRCULAR VEHICU
106938	4/6/2003 00:00:00	TRANSITO	TRANSPORTAR PEI
154457	4/6/2003 00:00:00	TRANSITO	TRANSPORTAR PEI
154951	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	NO RESPETAR LAS
154976	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	NO RESPETAR LAS
109210	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
109211	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
109207	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
105578	9/6/2003 00:00:00	TRANSITO	CIRCULAR CON VE
109469	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
154826	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
154952	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	CONducir con LA
110968	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
106398	6/6/2003 00:00:00	TRANSPORTE TERRESTRE	CONducir con LA
106075	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	CONducir con LA
107107	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
154977	7/6/2003 00:00:00	TRANSPORTE TERRESTRE	TRANSPORTAR PEI
148297	7/6/2003 00:00:00	TRANSITO	CONducir con LI

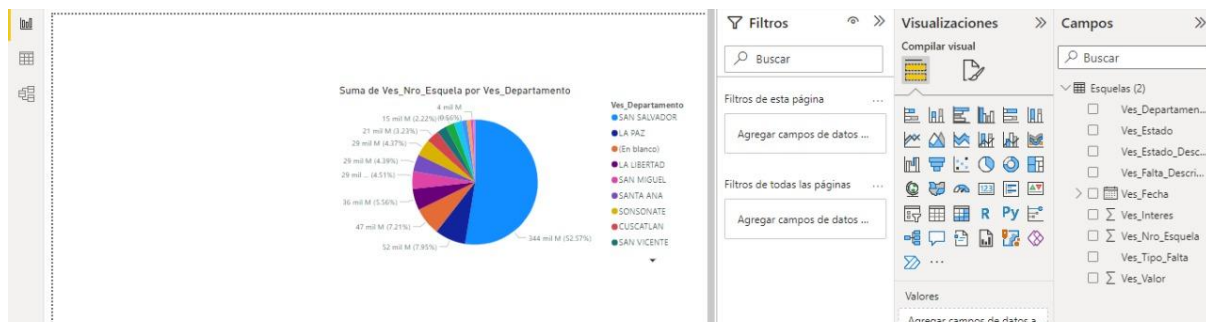
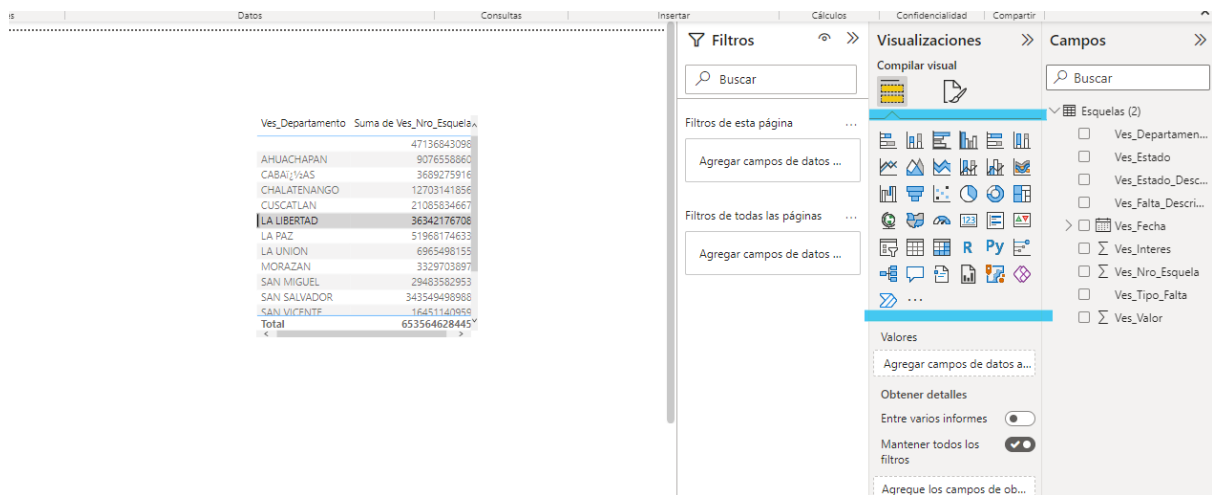
Esperamos a que cargue.



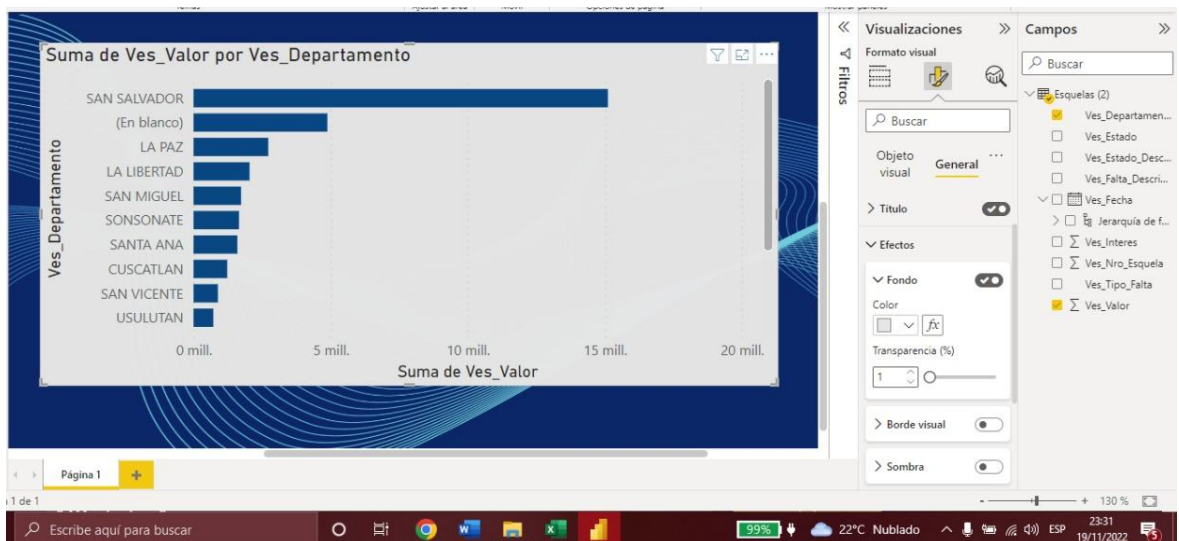
Arrastramos los campos que deseamos para poder crear la gráfica.



Nos aparecerá una tabla como esta y lo siguiente es elegir el gráfico que desea.



Como primer informe gráfico tenemos la comparación entre número de esquelas por departamento, esto nos ayuda a llevar un mejor informe para saber qué departamento tiene el número más alto de esquelas.



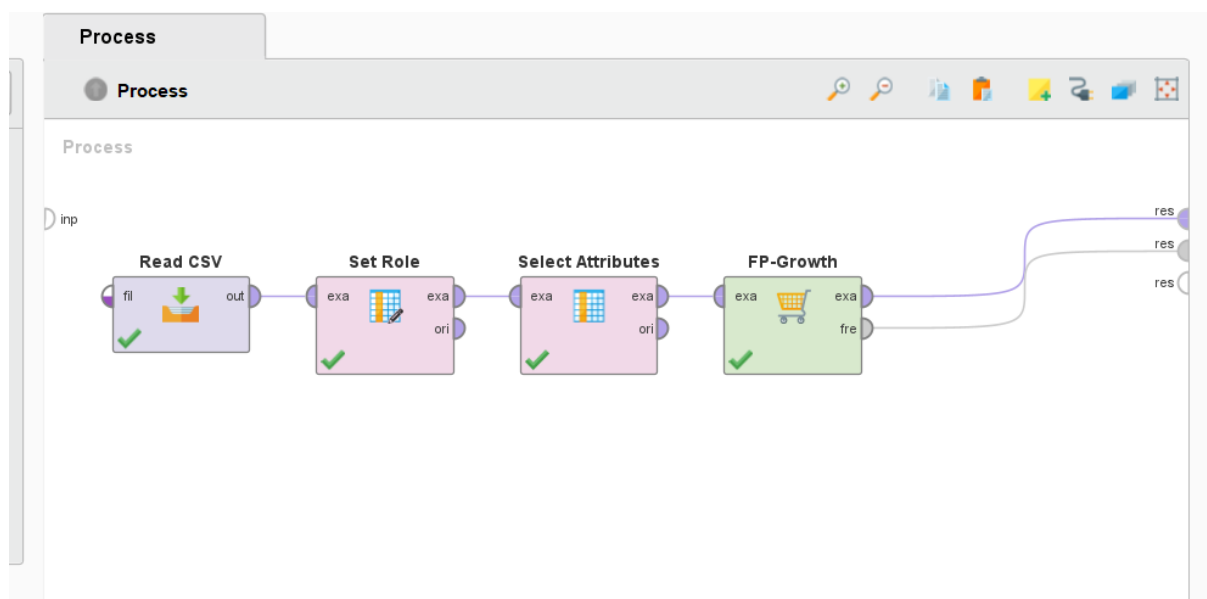
En el segundo informe gráfico se muestra la suma del valor de las esquelas por departamento, este muestra la cantidad de dinero que se obtiene por las esquelas realizadas y en qué departamento es el que más recibe dinero por dichas esquelas.

Reglas de asociación

Para el segundo análisis de comportamiento general se utilizó una estrategia minería de datos llamada **Reglas de Asociación**.

La creación de este modelo consta de 3 partes:

- Configuración de los datos a analizar (bases de datos, csv, etc.).
- Selección de atributos/campos relevantes para el análisis.
- Procesamiento de la información en base a la frecuencia de los datos.



Los resultados son los siguientes:

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.815	TRANSITO			
1	0.769	CANCELADA			
1	0.769	CBR			
1	0.463	SAN SALVADOR			
2	0.652	TRANSITO	CANCELADA		
2	0.652	TRANSITO	CBR		
2	0.349	TRANSITO	SAN SALVADOR		
2	0.769	CANCELADA	CBR		
2	0.334	CANCELADA	SAN SALVADOR		
2	0.334	CBR	SAN SALVADOR		
3	0.652	TRANSITO	CANCELADA	CBR	
3	0.267	TRANSITO	CANCELADA	SAN SALVADOR	
3	0.267	TRANSITO	CBR	SAN SALVADOR	
3	0.334	CANCELADA	CBR	SAN SALVADOR	
4	0.267	TRANSITO	CANCELADA	CBR	SAN SALVADOR

De este resultado, podemos mencionar comportamientos importantes:

1. La mayoría de las esquelas fueron del tipo **TRÁNSITO**, con un 81% de menciones en la muestra de datos.
2. El 77% de las esquelas fueron correctamente **CANCELADAS**.
3. Un 35% de las esquelas del tipo **TRÁNSITO** ocurrieron en el departamento de **SAN SALVADOR**.

Recomendaciones

Teniendo en cuenta los análisis anteriores creados por medio de dos diferentes estrategias de minería de datos, se proponen las siguientes recomendaciones:

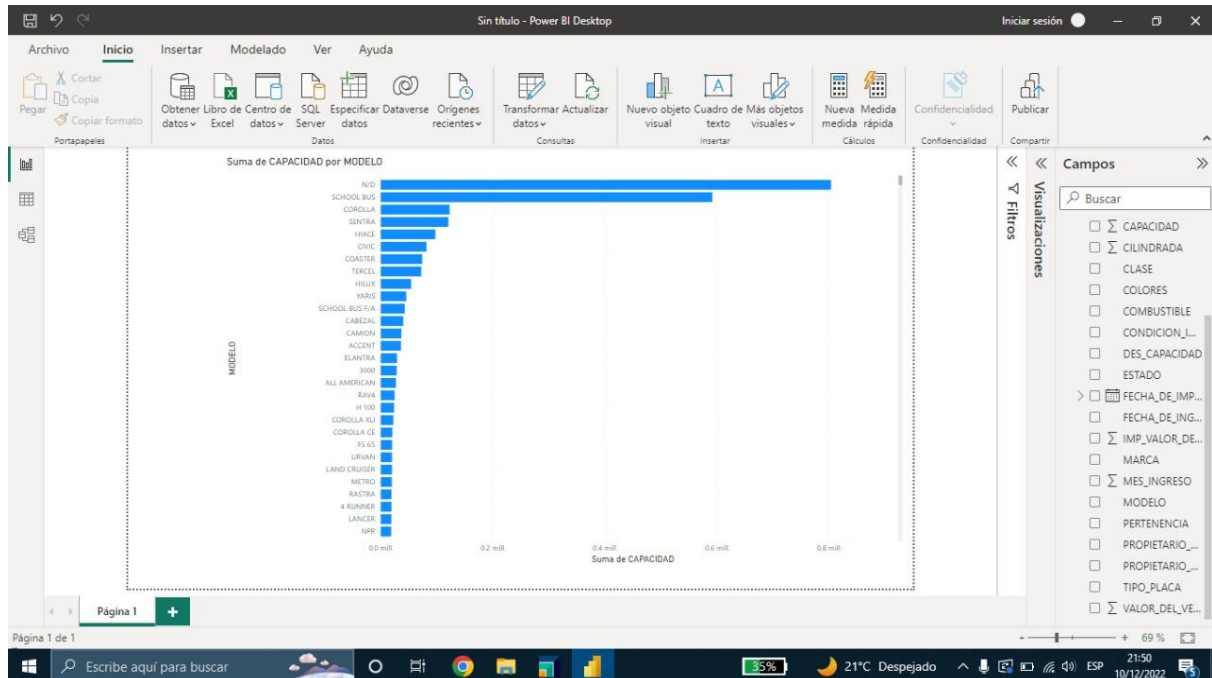
1. El departamento de San Salvador, por ser el más transitado y habitado, debe ser la prioridad para el MOP en controles vehiculares para promover en los automovilistas el cumplimiento de la ley y así evitar accidentes..
2. Mayor intervención policial para los autos particulares, donde se verifique el cumplimiento de las leyes de tránsito.

3. Una forma de contrarrestar la gran cantidad de esquelas, es incrementar la cantidad a pagar por cada tipo de esquila. De esta manera, al ser una multa más grande y costosa, el automovilista tendrá que analizar con mayor cuidado sus acciones para no cometer ninguna infracción.

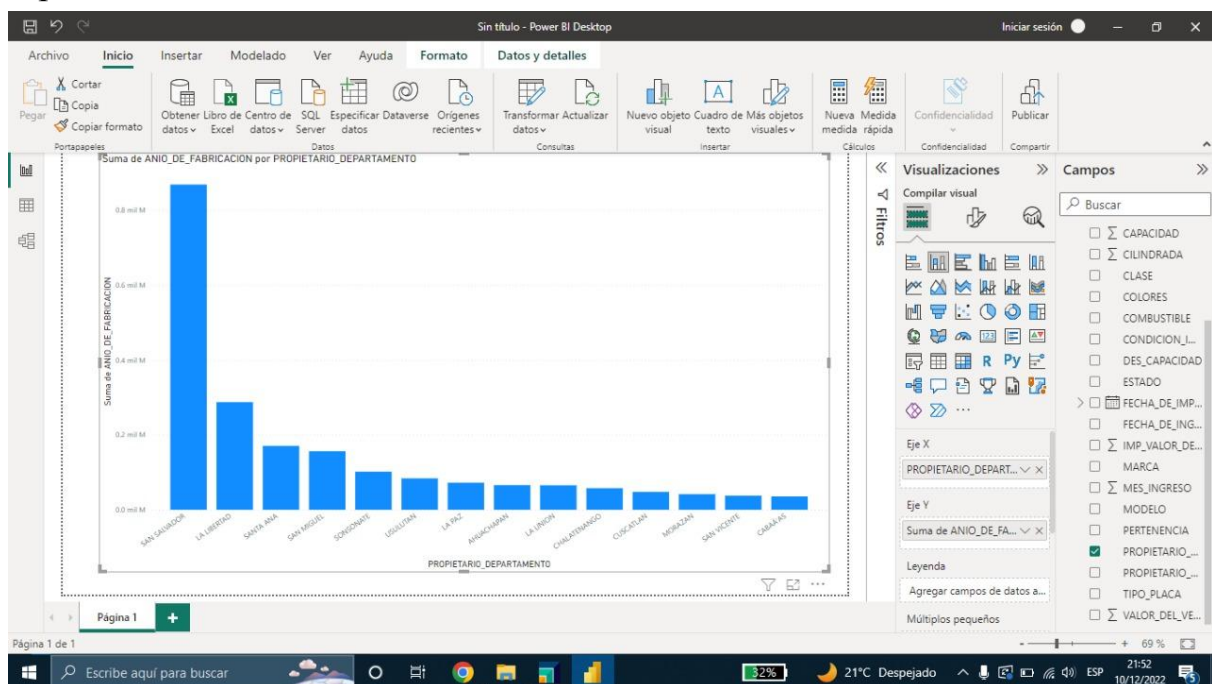
Parqueo Vehicular

Informes vehiculares.

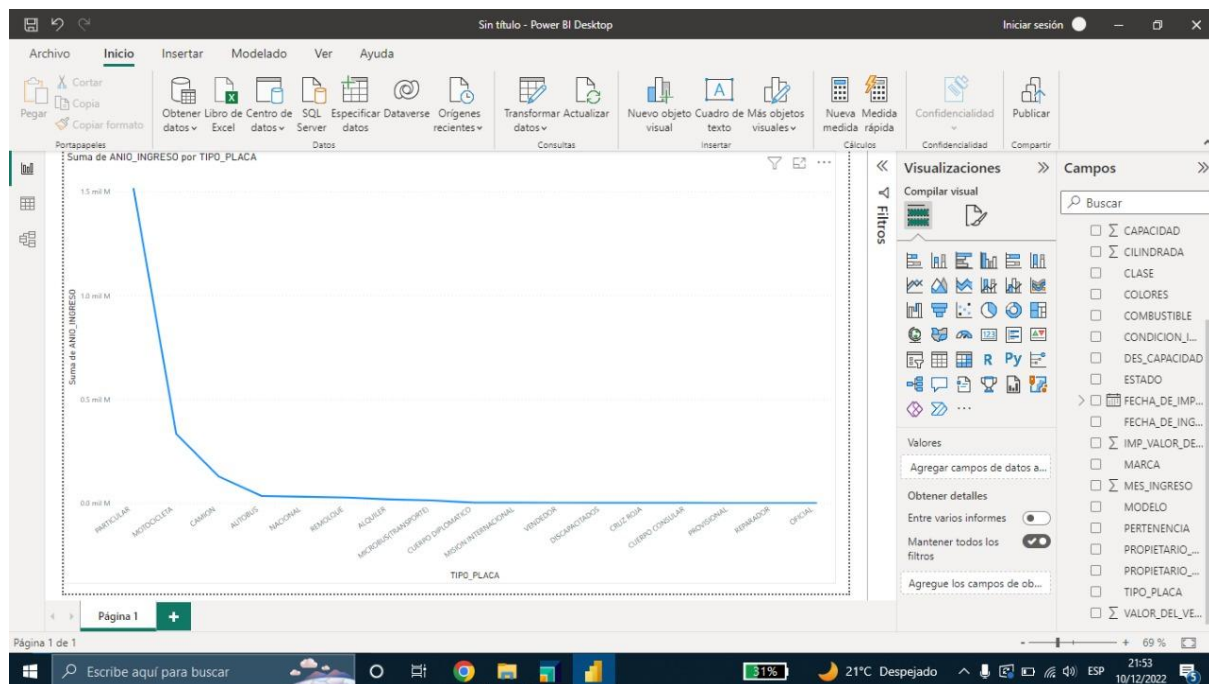
En el primer informe se muestran la capacidad del vehículo VS el modelo.



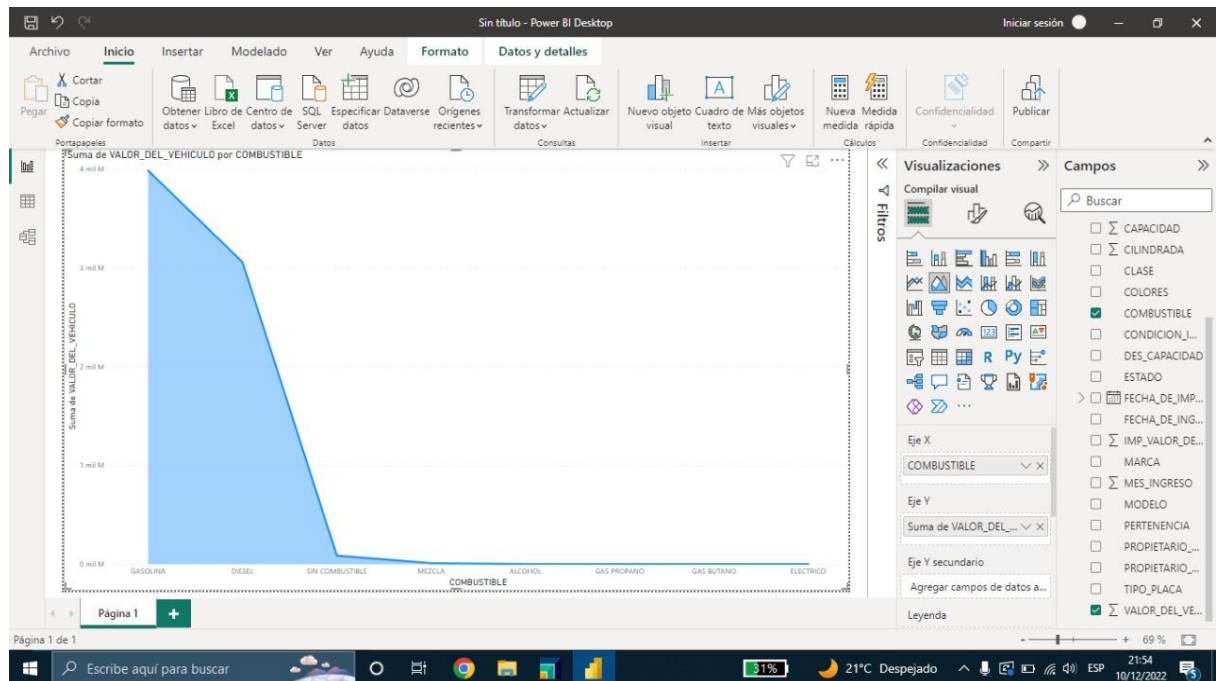
Aquí se muestra el resultado entre el año de fabricación VS el propietario por departamento.



En este informe se muestra el año de ingreso VS el tipo de placa.



Informe del valor del vehículo VS el tipo de combustible que utiliza.



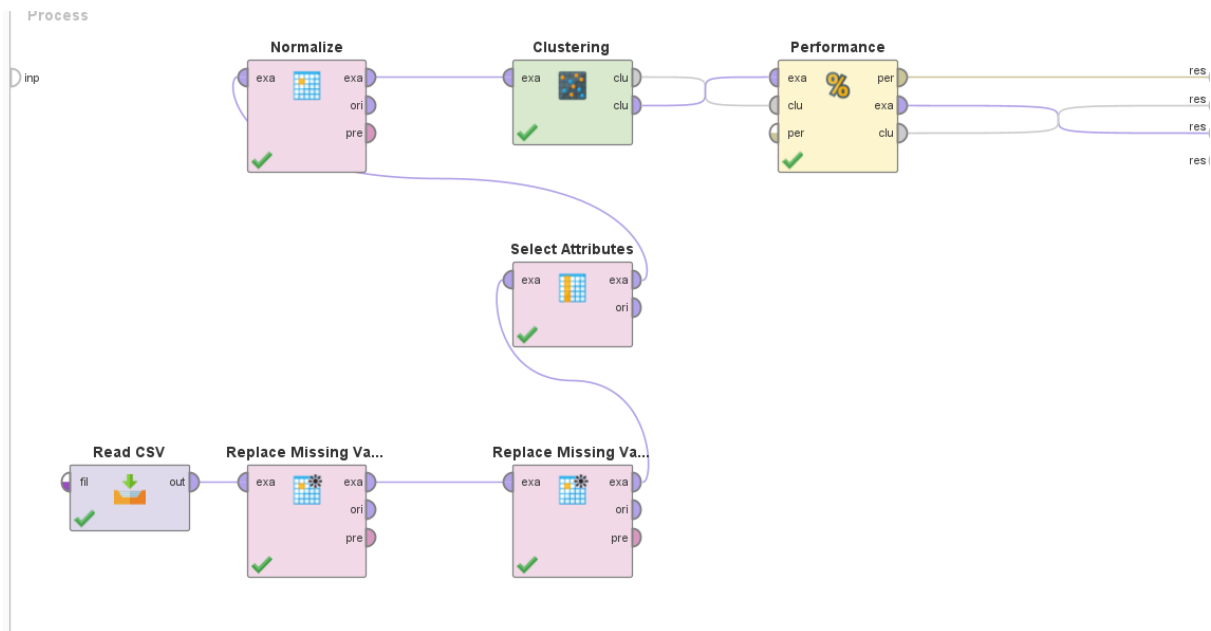
Se muestra la cantidad de cilindros VS el color del vehículo.

Análisis usando estrategias de minería de datos

Agrupamiento con K-means

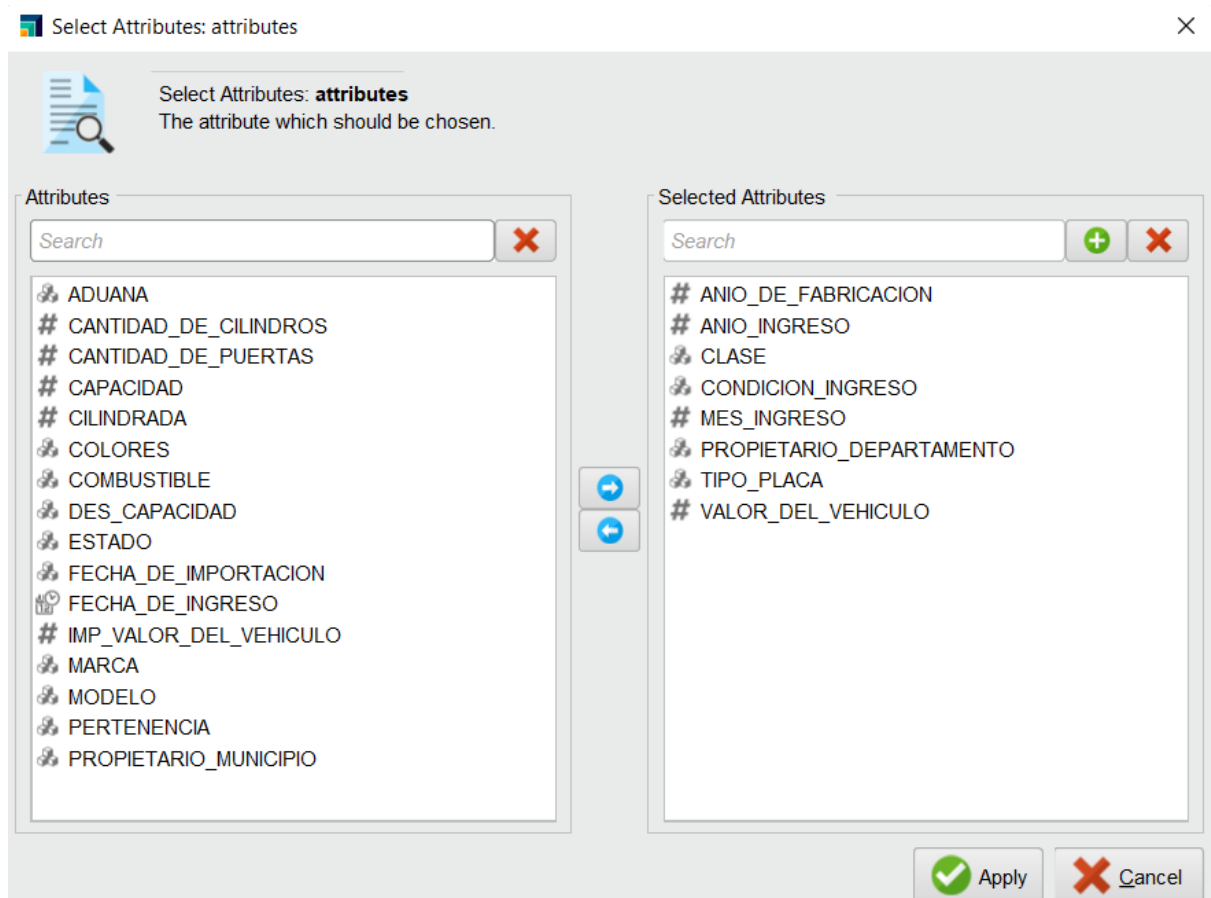
Para el primer análisis, se utiliza la estrategia de Agrupamiento con K-means, para la detección de patrones en la información.

El proceso propuesto:

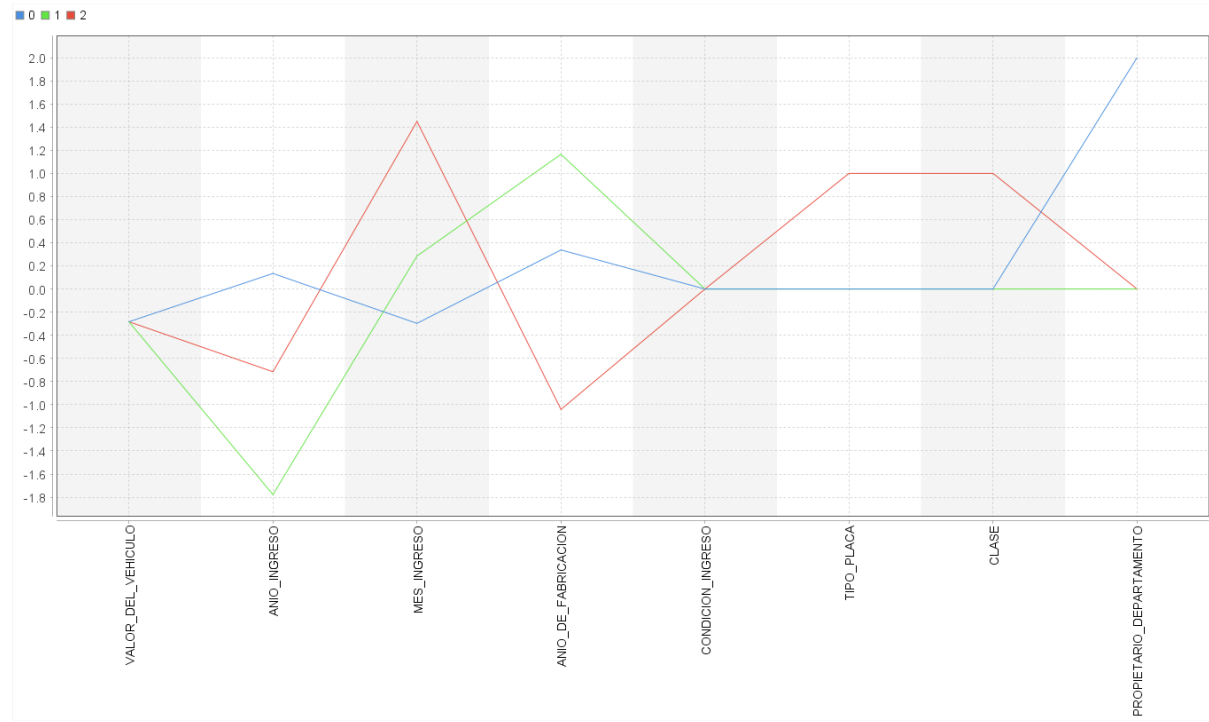


Donde tomamos la información, realizando un paso de limpieza, selección y normalización de la información para que pueda ser procesado por el algoritmo.

Los atributos utilizados para realizar este análisis fueron:



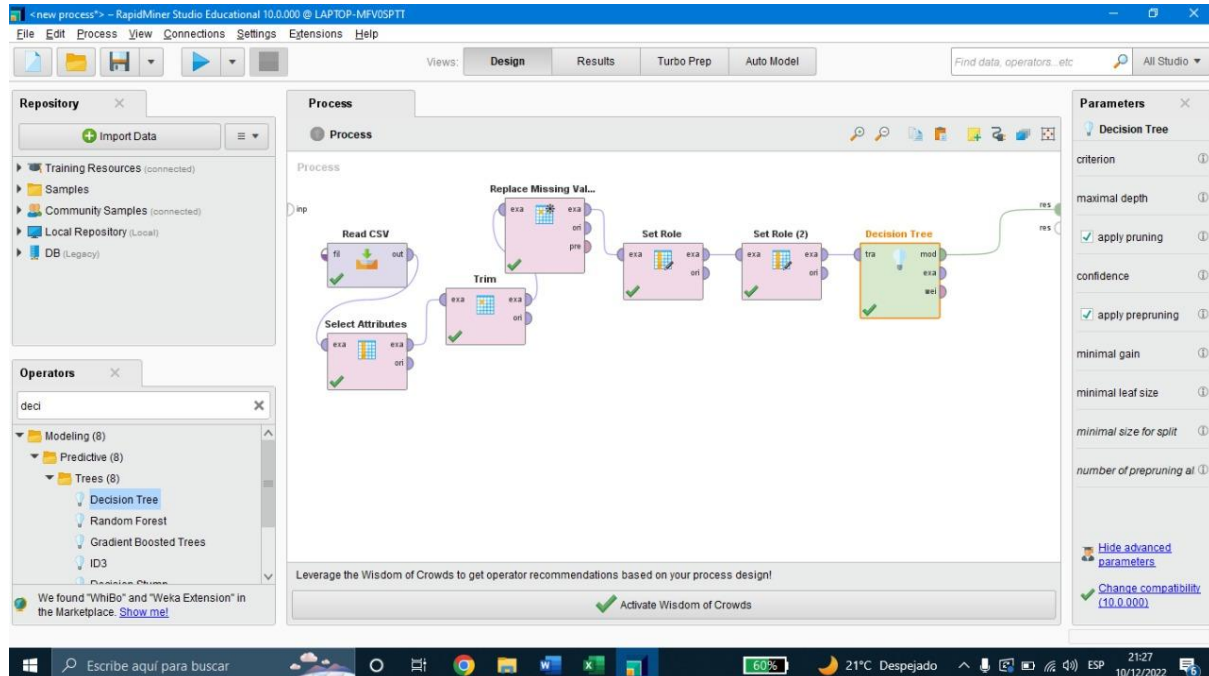
La distribución de los atributos fue el siguiente:



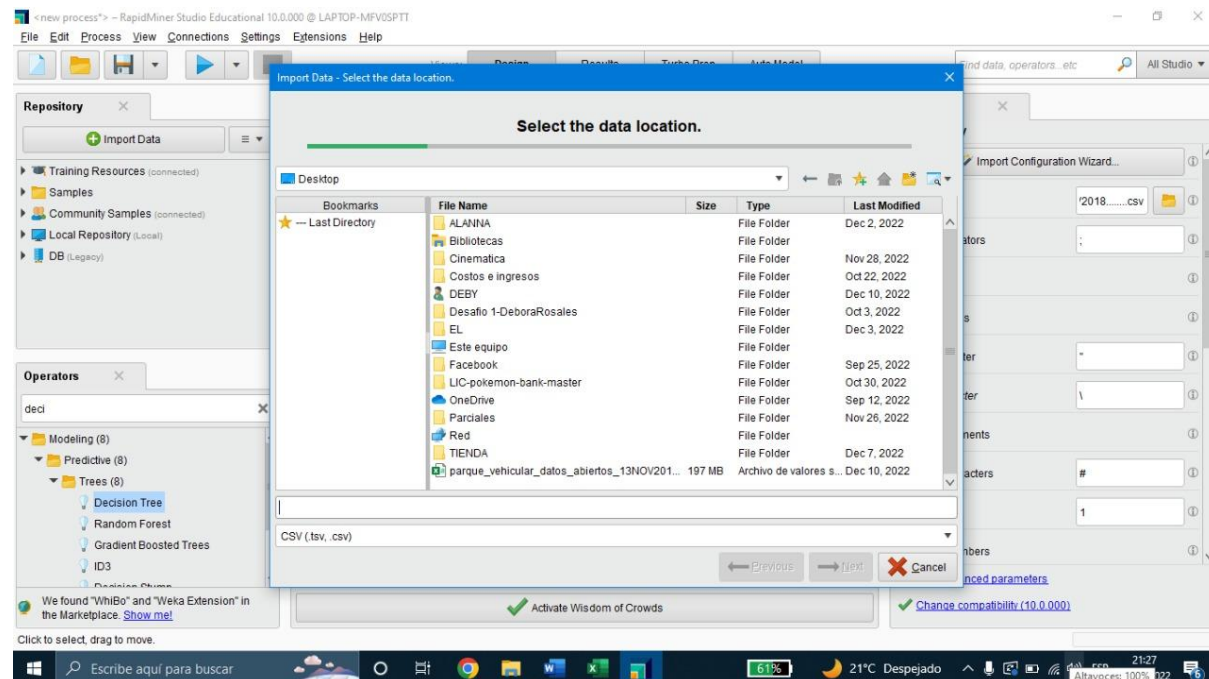
El análisis del primer cluster, al tomar poca información como se define, no hay mucha variación de la información, al contrario con los análisis del cluster n° 2 y n°3, donde se muestra que a mayor toma de datos, mayor es la variación en la información proporcionada.

ÁRBOL DE DECISIONES

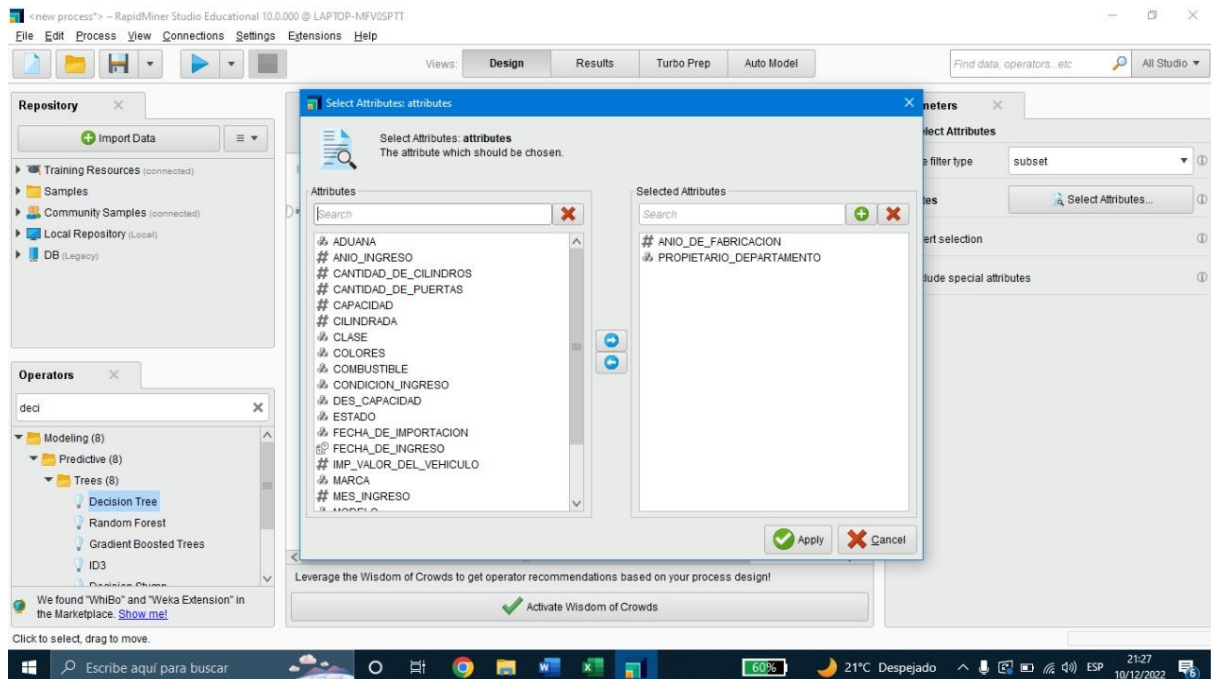
Para empezar nuestro análisis insertamos los operadores, READ CSV, SELECT ATTRIBUTE, TRIM, REPLACE MISSING, SET ROL, DECISION TREE



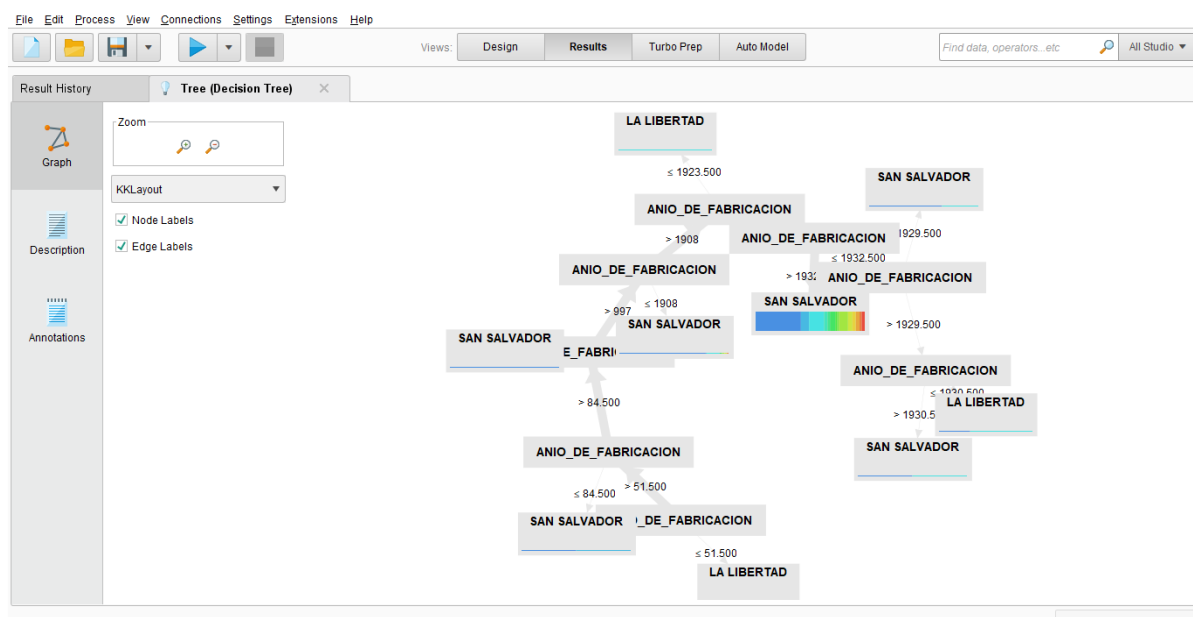
Se agrega el documento a analizar en Read csv.



Se seleccionan los atributos, en este caso se seleccionó el año de fabricación, y el departamento al que pertenece el propietario, para mostrar la relación que poseen estos dos atributos.



Se obtiene el árbol de decisiones. Este árbol muestra la relación del año de fabricación del vehículo y el departamento al que pertenece, un caso en particular para determinar y mostrar información de los vehículos con más años dentro del país, y en el departamento donde proviene el propietario de este.



new process - RapidMiner Studio Educational 10.0.000 @ LAPTOP-MBV0SP1T

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators, etc. All Studio

Result History Tree (Decision Tree)

Graph

Description

Annotations

Tree

```
ANIO_DE_FABRICACION > 51.500
|
| ANIO_DE_FABRICACION > 84.500
| |
| | ANIO_DE_FABRICACION > 997
| | |
| | | ANIO_DE_FABRICACION > 1908
| | | |
| | | | ANIO_DE_FABRICACION > 1923.500
| | | | |
| | | | | ANIO_DE_FABRICACION > 1932.500: SAN SALVADOR (SAN SALVADOR=434968, SAN MIGUEL=78299, LA LIBERTAD=143961, AHUACHAPAN=33049, CUSCATLAN=231
| | | | | ANIO_DE_FABRICACION ≤ 1932.500
| | | | | |
| | | | | | ANIO_DE_FABRICACION > 1929.500
| | | | | | |
| | | | | | | ANIO_DE_FABRICACION > 1930.500: SAN SALVADOR (SAN SALVADOR=1, SAN MIGUEL=0, LA LIBERTAD=1, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=
| | | | | | | ANIO_DE_FABRICACION ≤ 1930.500: LA LIBERTAD (SAN SALVADOR=2, SAN MIGUEL=0, LA LIBERTAD=4, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=
| | | | | | | ANIO_DE_FABRICACION ≤ 1929.500: SAN SALVADOR (SAN SALVADOR=6, SAN MIGUEL=0, LA LIBERTAD=3, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=0,
| | | | | | | ANIO_DE_FABRICACION ≤ 1923.500: LA LIBERTAD (SAN SALVADOR=0, SAN MIGUEL=0, LA LIBERTAD=2, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=0, SAN VICEN
| | | | | | | ANIO_DE_FABRICACION ≤ 1908: SAN SALVADOR (SAN SALVADOR=186, SAN MIGUEL=2, LA LIBERTAD=28, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=5, SAN VICENTE=0
| | | | | | | ANIO_DE_FABRICACION ≤ 997: SAN SALVADOR (SAN SALVADOR=2, SAN MIGUEL=0, LA LIBERTAD=0, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=0, SAN VICENTE=0, CABAÑ
| | | | | | | ANIO_DE_FABRICACION ≤ 84.500: SAN SALVADOR (SAN SALVADOR=1, SAN MIGUEL=1, LA LIBERTAD=0, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=0, SAN VICENTE=0, CABAÑ
| | | | | | | ANIO_DE_FABRICACION ≤ 51.500: LA LIBERTAD (SAN SALVADOR=2, SAN MIGUEL=0, LA LIBERTAD=4, AHUACHAPAN=0, CUSCATLAN=0, SONSONATE=0, SAN VICENTE=0, CABAÑAS=0,
|
```

55% 21°C Despejado 21:33 10/12/2022