Texto

Descripción generada automáticamente

# DMD G01T

**Ciclo 02 -2022**

**Proyecto Fase 2**

Alumno

|  |  |  |
| --- | --- | --- |
| Apellidos | Nombres | Carné |
| Flores Henríquez | John Michael | FH120965 |
|  |  |  |
|  |  |  |
|  |  |  |

# Enunciado del proyecto

El ministerio de obras públicas(MOP), ha recolectado mucha información sobre dos temas de suma importancia, pero no sabe cómo presentar los informes al gobierno central para el cierre del año 2018.

* Esquelas de infracción de tránsito.
* Parque vehicular.

Por tal motivo a contratado un grupo de consultores para que analicen la información y proporcione un documento con el análisis que realizaron, solo tiene una restricción el análisis deben de ocupar dos estrategias por tema de minería de datos, puede ser cualquier de las descritas. Se le pide elaborar dos estrategias de minería de datos, por archivo CSV, en total seria 4 análisis:

* Cubos OLAP
* Reporting Services
* Power BI
* Reglas de asociación
* Agrupamiento con k-means
* Arbol de decisión

Por ejemplo: Podría ser cubos y Power BI, para esquelas y reglas de asociación y árbol de decisión para parque vehicular.

Los recursos son proporcionados en formato csv.

* Esquelas\_18102018.csv
* Parque\_vehicular.rar

# Objetivo

* Crear un análisis de inteligencia de negocios utilizando 4 estrategias para obtener diferentes puntos de vista de la problemática y lograr de manera eficiente otorgar recomendaciones en base a los resultados obtenido a nuestro cliente.

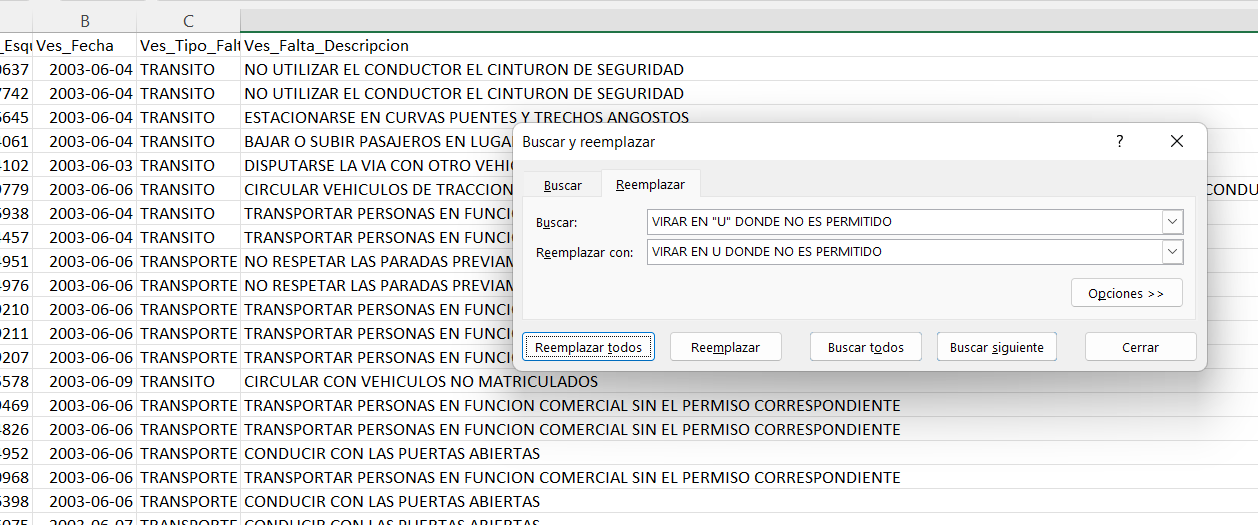
# Estrategia de CUBO de OLAP

Para la ejecución de la estrategia formatearemos y limpiaremos un poco los datos, comenzando por dar formato a las fechas

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Luego quitaremos valores que nos puedan generar error para insertar la data en SQL Server, por ejemplo, comillas y puntos comas, etc.



Interfaz de usuario gráfica, Texto, Aplicación

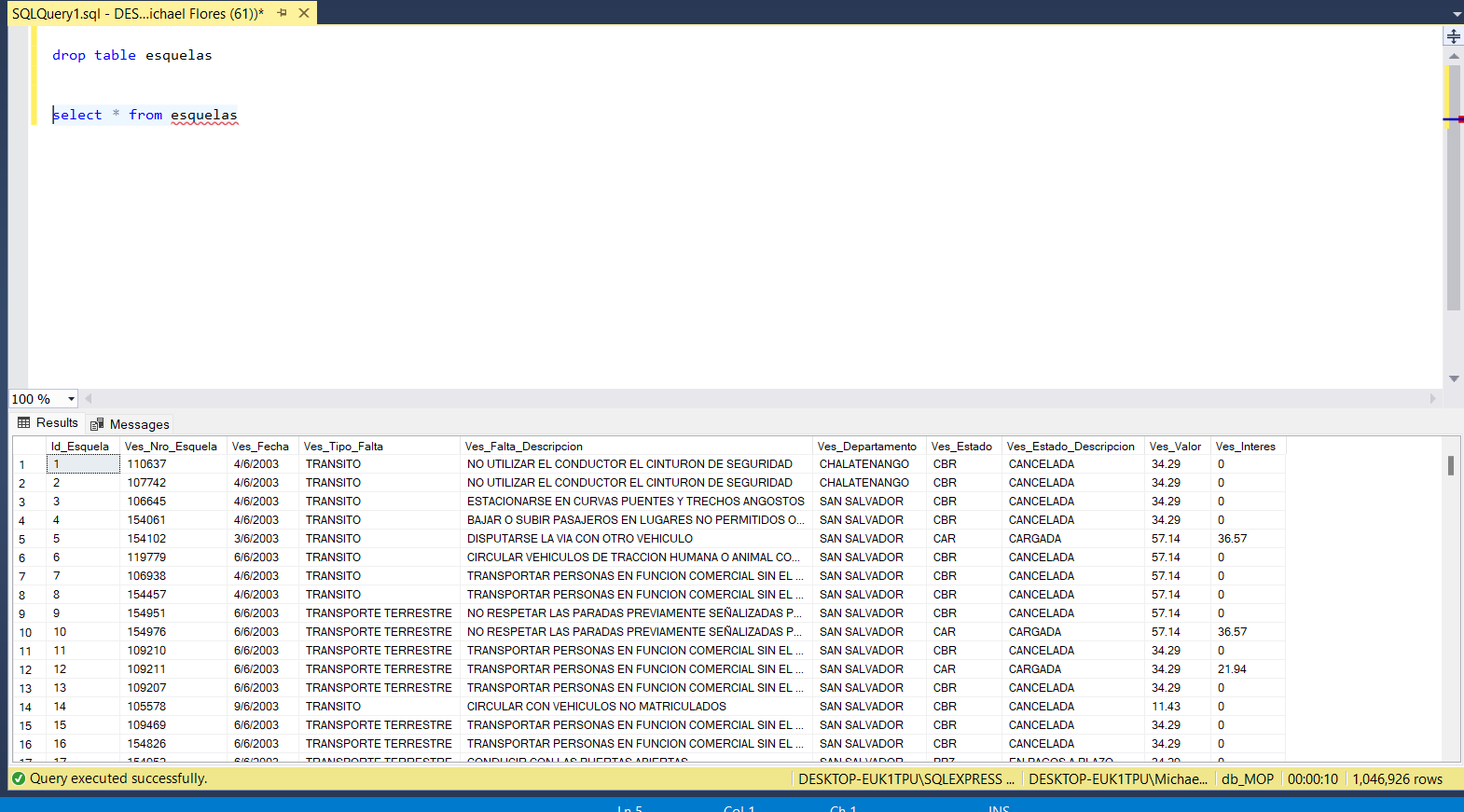
Descripción generada automáticamente

Creamos la base de datos que utilizaremos.

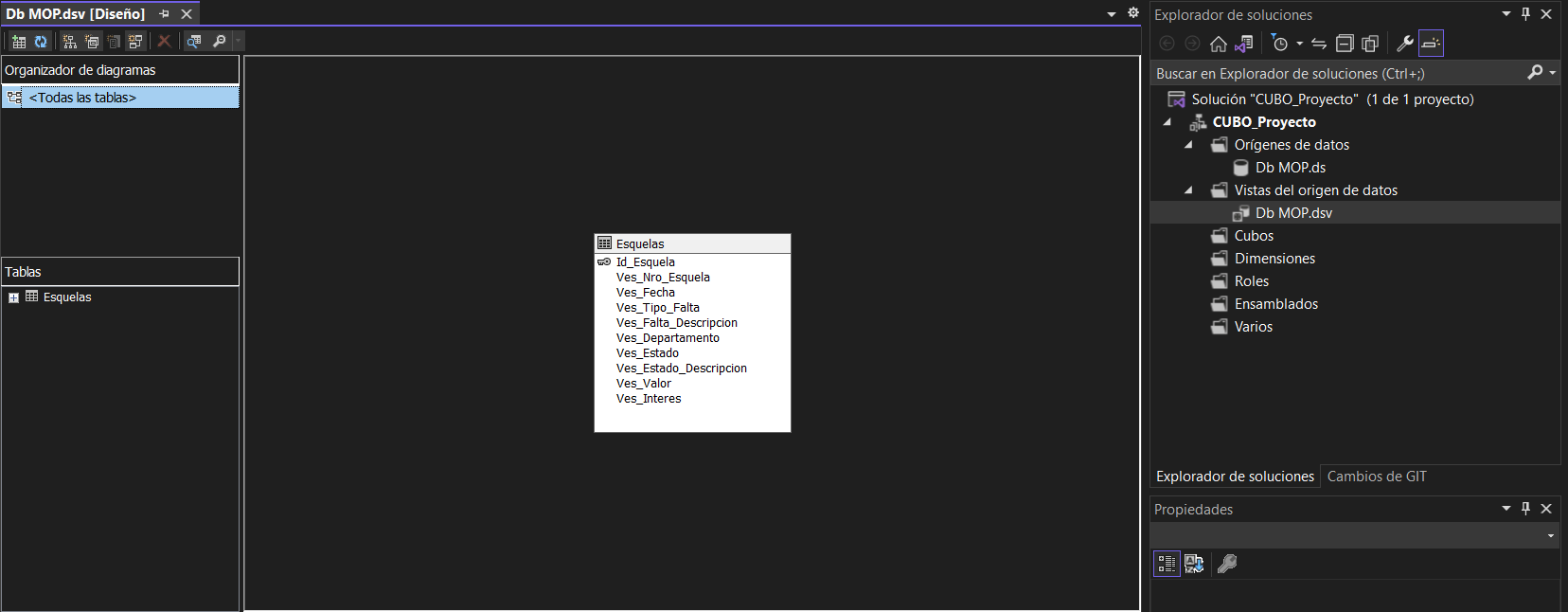
Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Ingresamos la data por medio del Wizard de SQL y consultamos para verificar que todo está en orden



Procedemos a crear un proyecto de Analysis Service, creando la conexión a nuestro origen de datos y luego la vista respectiva.



Procedemos a crear el Cubo, las dimensiones respectivas y procesamos

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Una vez procesado procedemos al explorador del cubo y ejecutamos nuestra primera consulta:

## Esquelas por departamento

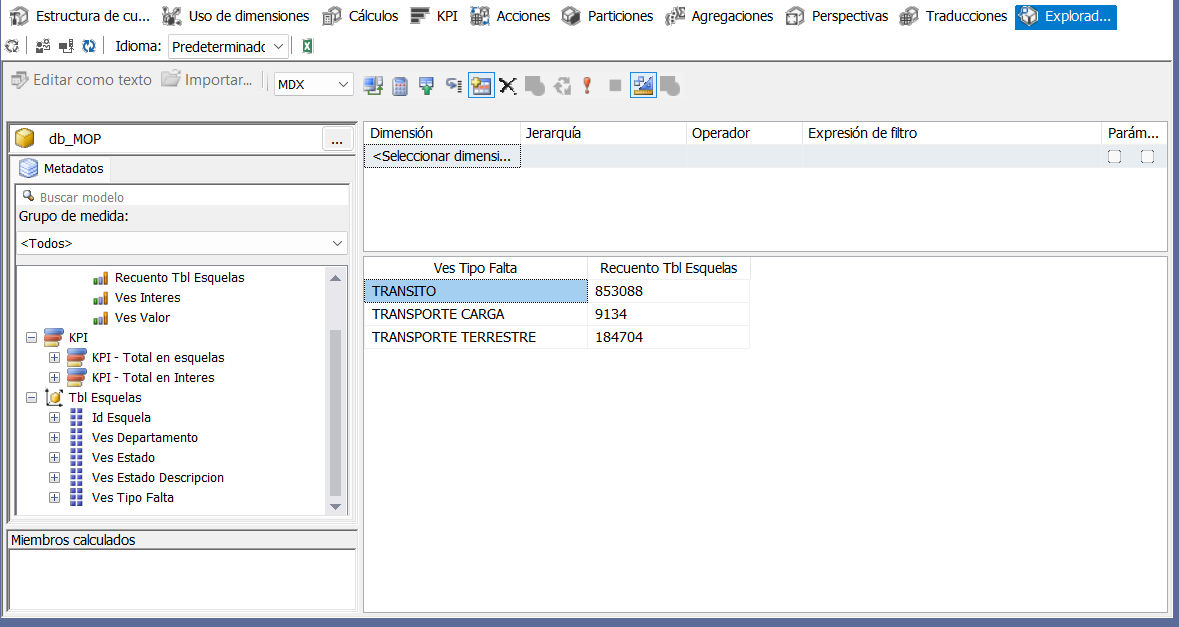
* En la cual podemos concluir que San Salvador es el departamento con más esquelas de todo el país.
* Se recomienda poner más atención a las campañas para la prevención y cuidado en el manejo en este departamento.

Interfaz de usuario gráfica

Descripción generada automáticamente

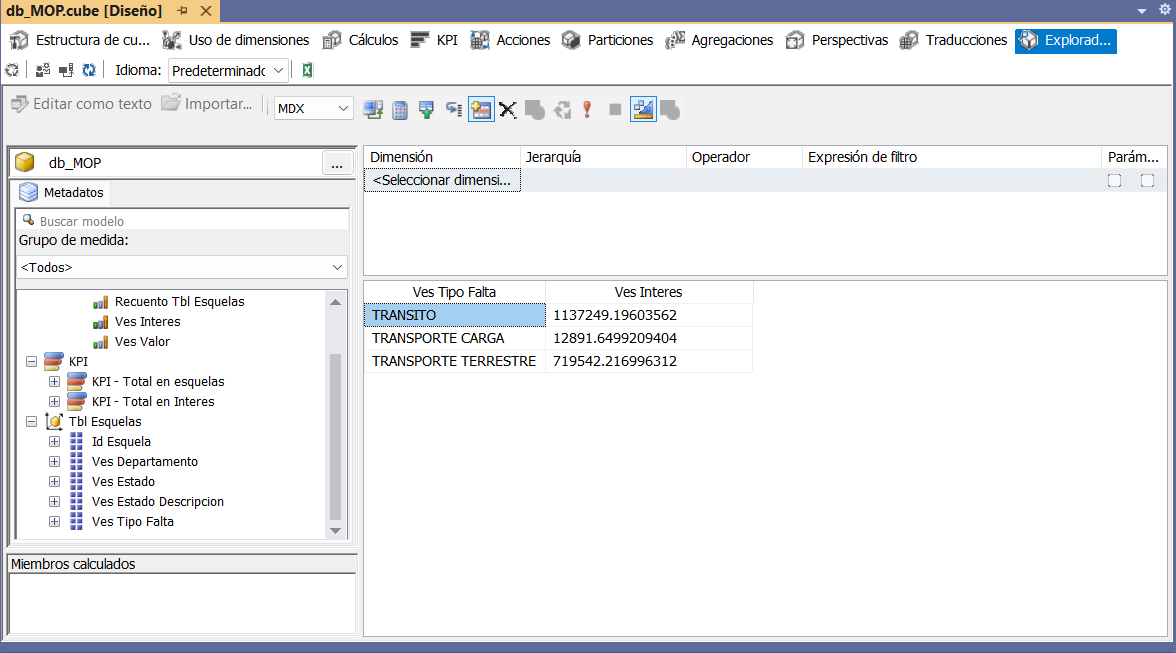
## Esquelas por tipo

* Se concluye que la mayor cantidad de esquelas son de tránsito.
* se recomienda que las campañas de concientización deberían ser orientadas a este público.



## Tipos que generan más interés

* En conclusión, a mayor valor del interés generado son los que menos pagan sus multas. Las multas de transito son las menos pagadas y esto tiene todo el sentido ya que son las que más se generan.
* Se recomienda crear un programa de facilidades de pago.



## Tipos de esquela que generan más valor

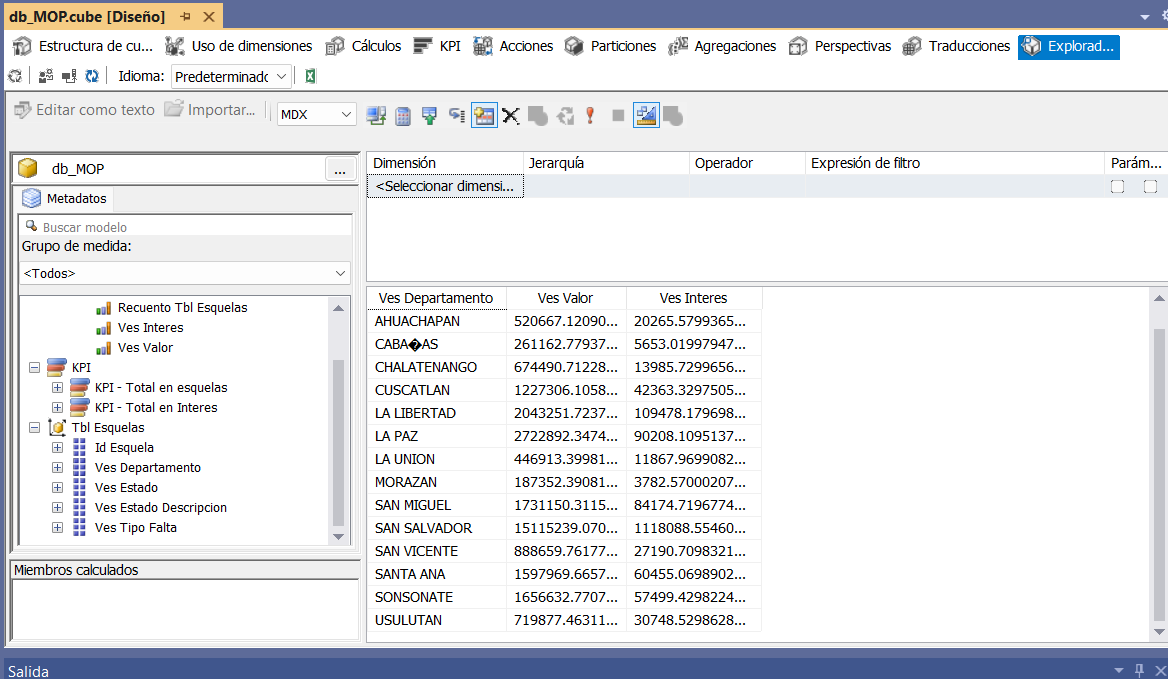
* Las esquelas de transito al ser más también generan más valor monetario.

Interfaz de usuario gráfica, Texto, Correo electrónico

Descripción generada automáticamente

## Valor e interés generado por departamento

* San Salvador es el departamento que más genera interés y valor en cuanto a las esquelas.



## Esquelas por estado

* Podemos concluir que la mayoría de las esquelas son canceladas, por lo que debe haber una inconsistencia en el registro de estas.
* Se recomienda revisar el sistema de registro de esquelas o investigar el motivo de todas estas cancelaciones.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

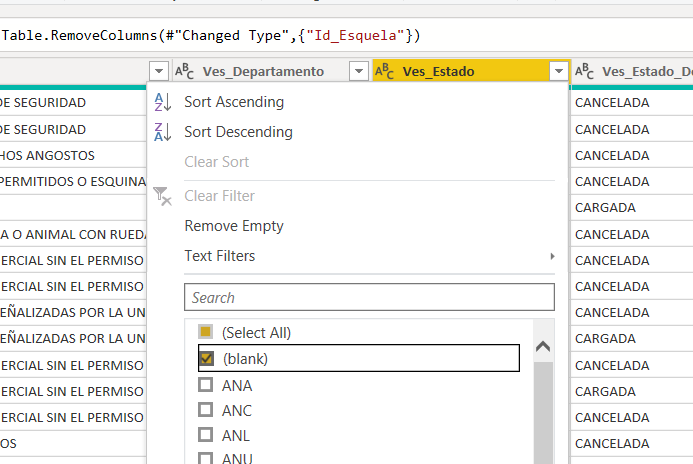
# Análisis con Power BI

Cargamos el documento CSV a PBI o nos conectamos directamente a la base de datos que se creó para este análisis.

Interfaz de usuario gráfica, Texto, Aplicación, Tabla

Descripción generada automáticamente

Transformamos nuestra data quitando columnas o posibles datos nulos, o columnas que no necesitamos para nuestro análisis, en este caso rellenamos los espacios nulos con NA



Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

## Tablero por departamento

* Podemos observar que evidentemente, como en nuestro análisis anterior, a mayor número o porcentaje de esquelas mayor es la cantidad de valor e interés generado por el departamento de San Salvador.
* Existen muchos registros sin el registro del departamento, por lo cual se debe investigar a que se debe esta inconsistencia, aunque lo más posible es que sean las esquelas anuladas.
* Se recomienda poner más atención al departamento para fomentar la educación vial.

Interfaz de usuario gráfica, Gráfico, Aplicación, Excel

Descripción generada automáticamente

## Tablero por tipo de faltas

* La mayoría de las esquelas son por falta en las reglas de tránsito
* La mayor cantidad de faltas se dan en el mes de Julio.
* Se recomienda ser más riguroso con las faltas de transito

Gráfico

Descripción generada automáticamente

## Tablero por estado de esquelas

* Como podemos interpretas existen una gran cantidad de esquelas canceladas o con inconsistencias.
* Las pendientes de pago suman la mayor cantidad de intereses
* Se recomienda gestionar el pago de las pendientes e investigar porque existen tantas inconsistencias en los registros.

Gráfico, Gráfico circular

Descripción generada automáticamente

# Reglas de asociación – RapidMiner

Cargamos la data del CSV dentro del aplicativo para que podamos iniciar a crear nuestro flujo de las reglas de asociación.

Interfaz de usuario gráfica, Aplicación, Word, Excel

Descripción generada automáticamente

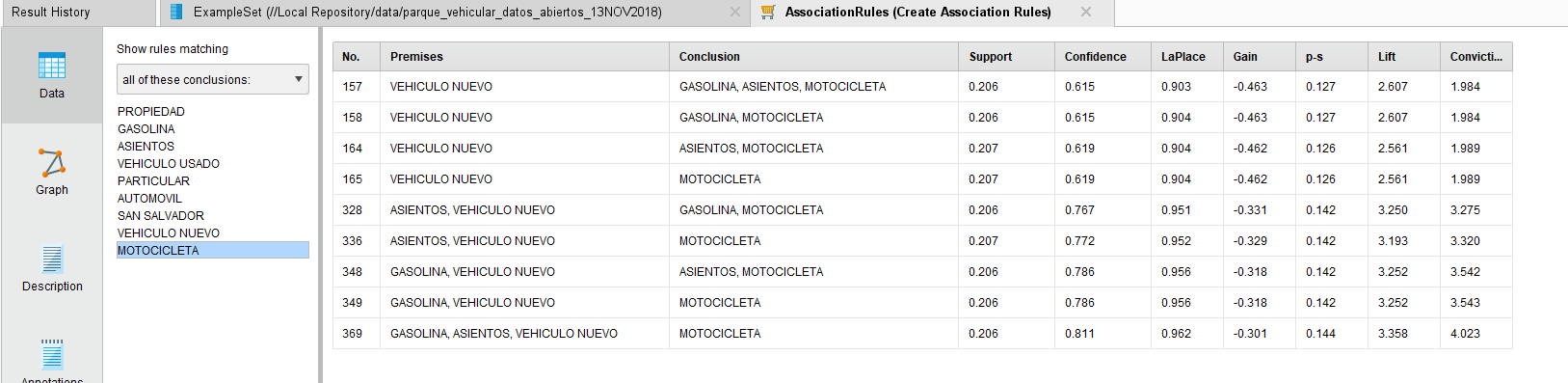
Creamos el flujo una vez cargada la data, en el cual vamos a recuperar los datos, seleccionar los atributos que, convertir los datos en el caso que vengan numéricos y vamos a crear las reglas de asociación.

Diagrama

Descripción generada automáticamente

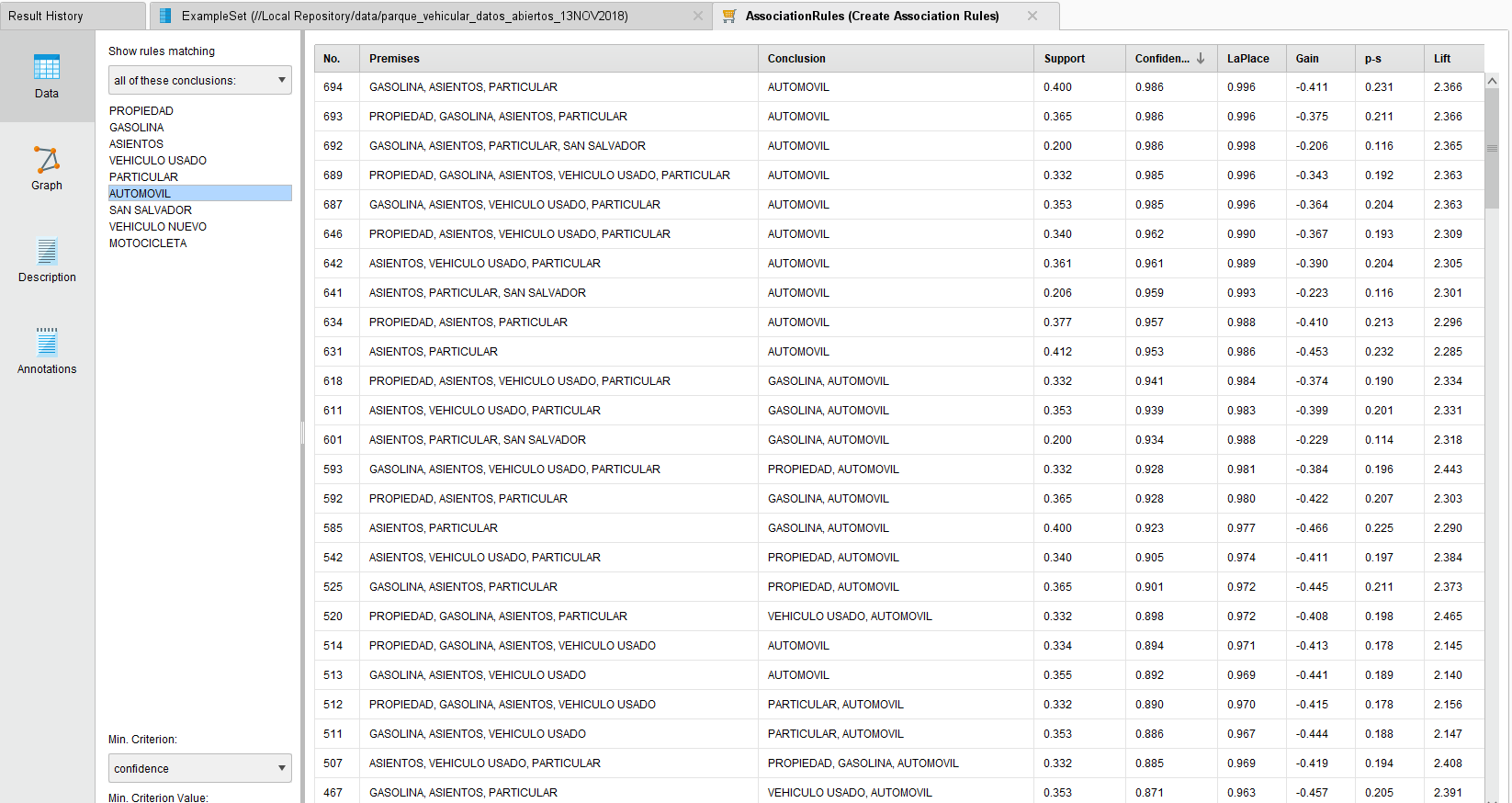
## Resultados de reglas de asociación

* La mayoría de los vehículos nuevos están asociados a motocicletas y el tipo de combustible que suelen utilizar es gasolina





* Como era de esperarse también, los automóviles están fuertemente ligados a que son medios de transporte con asientos, de uso particular y también que utilizan gasolina como combustible principal



# Árbol de decisiones - RapidMiner

## Resultados Árbol de decisión

Para comenzar realizamos cambios en nuestro flujo, reemplazando valores nulos por “NA”, seleccionando atributos específicos para poder ir intercambiando análisis, quitamos espacios en blanco y un rol “Label” para cada uno de nuestros atributos, al final el criterio que estaremos utilizando sería el Gini\_Index

Imagen que contiene Gráfico de cajas y bigotes

Descripción generada automáticamente

Ejecutamos nuestra primera vez con los siguientes atributos

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Nuestro dataset de muestreo quedaría de la siguiente manera

Interfaz de usuario gráfica

Descripción generada automáticamente

En árbol nos quedaría de la siguiente manera en el cual podemos concluir que

* La mayoría de las motocicletas fueron fabricadas antes del año 2015, partiendo de base que tipo de placas de estas particulares

Diagrama

Descripción generada automáticamente

* De acuerdo con la capacidad podemos definir que si es medida en toneladas se trata pickups y si es en base a asientos podemos encontrar a la capacidad que pueden ser automóviles o microbuses, todo esto bajo el tipo de placa que es particular.
* Si es bajo la rama de motocicleta se puede definir que la capacidad del tanque aumenta o baja según sea tricótomo o motocicleta como tal.
* Bajo la rama de tipo de placa nacional podemos encontrar que según la capacidad verifica si efectivamente es un automóvil o un camión pesado, por ejemplo, de recolección de basura.
* En la rama oficial podemos definir que solo se encuentran automóviles

Escala de tiempo

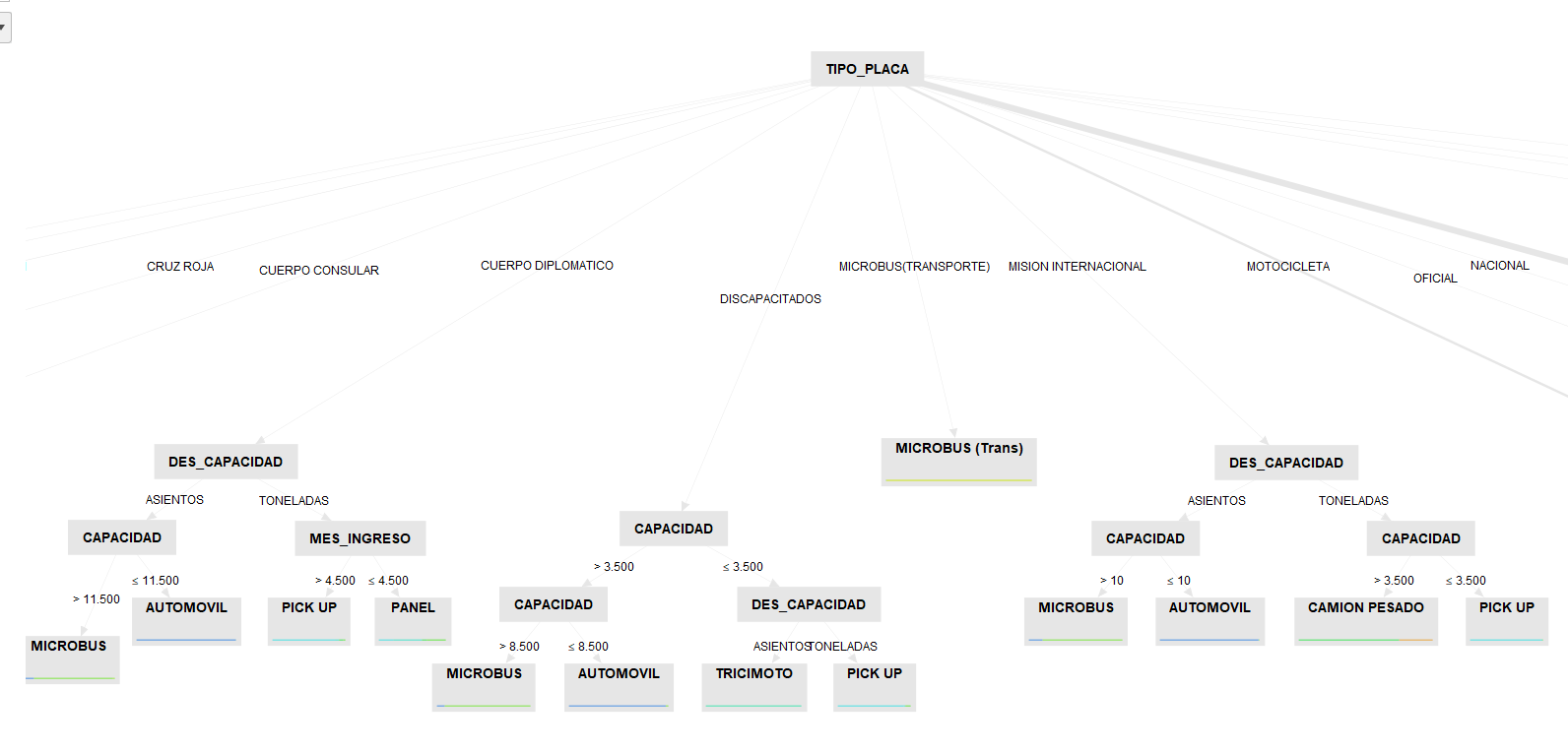
Descripción generada automáticamente

* En la rama provisional tenemos camiones pesados y cabezales, seleccionados de acuerdo con la capacidad de estos vehículos.
* En la calidad de remolque podemos notar que tenemos dos variantes, por año de ingreso, en la cual se calcula que los vehículos remolques a partir del año 2008 y del 2007 hacia atrás son remolques de otro tipo según la data.
* En el tipo de placa calificada como vendedor donde según el mes de ingreso son calificados, el automóvil se recibió en diciembre y en los meses anteriores son de tipo vendedor(auto)

Escala de tiempo

Descripción generada automáticamente

* En la rama de misión internacional, tenemos un análisis de capacidad en el cual podemos calificar también a los automóviles, microbuses, camiones pesados y pickups
* En el tipo de placas de discapacitados podemos encontrar microbuses, automóviles, tricótomos y pickups
* Entre los vehículos pertenecientes al cuerpo diplomático podemos encontrar microbuses y automóviles. En cuanto a los pickups y paneles que ingresaron al país entre mayo y diciembre y los paneles de enero a abril.



* Para terminar, podemos definir que el cuerpo consular solo ingreso al país automóviles
* Bajo el tipo de placa de cruz roja tenemos un ordenamiento por valor del vehículo en el que los pickups que ingresaron fueron mayor a $9990.730 y los camiones livianos menores o iguales a esta misma cantidad
* En calidad de camiones tenemos cabezales que ingresaron con una capacidad mayor de 12.50 y camiones pesados con una capacidad menor o igual a 12.50

Escala de tiempo

Descripción generada automáticamente con confianza baja