



U D B V I R T U A L

Universidad Don Bosco, El Salvador

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.

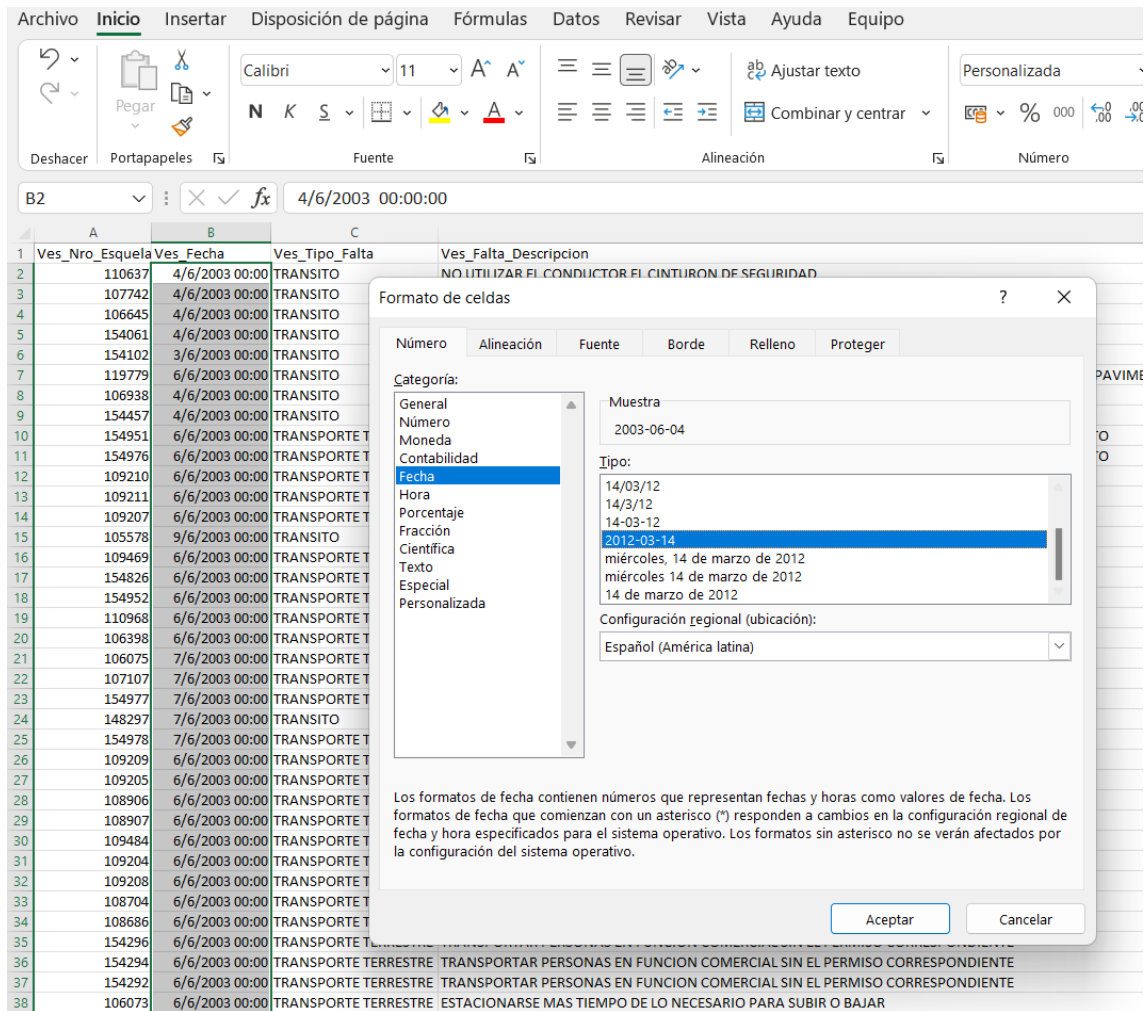
- Esquelas_18102018.csv
- Parque_vehicular.rar

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.

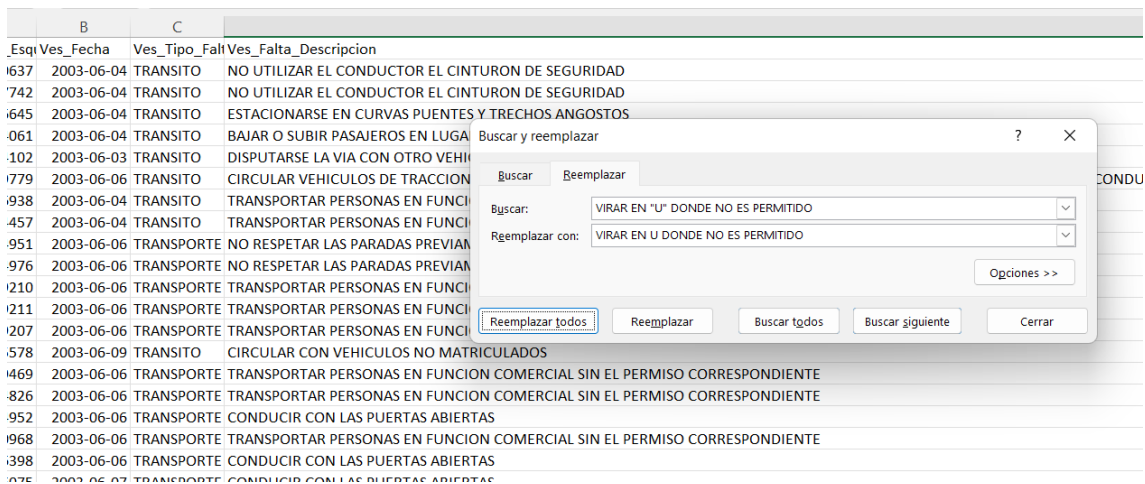
➤ Estrategia de CUBO de OLAP

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

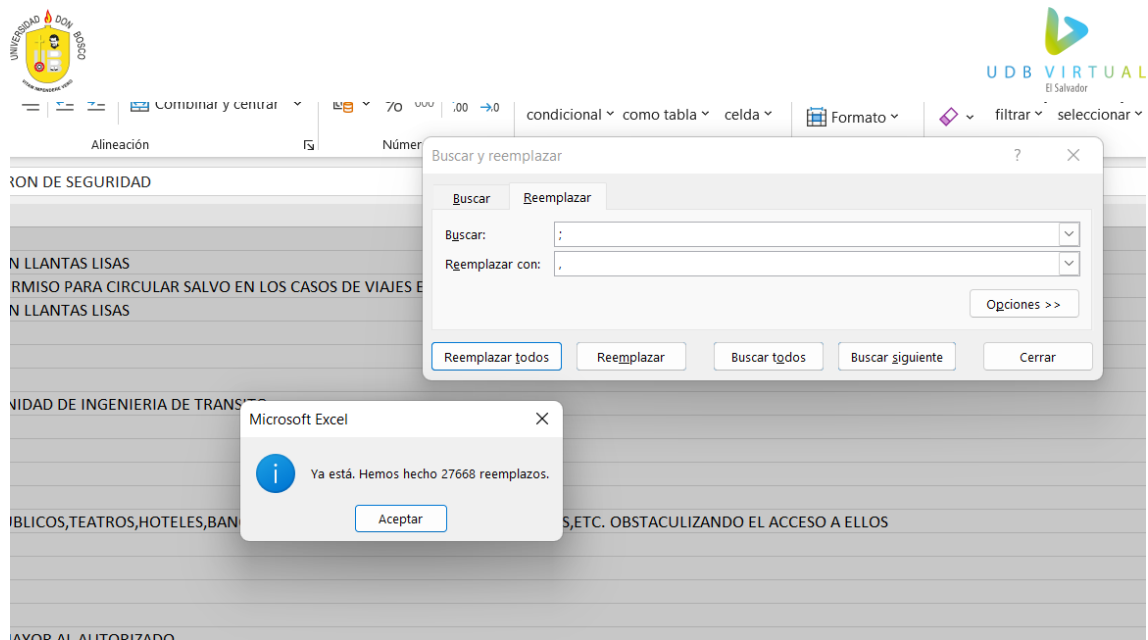


The screenshot shows the Microsoft Excel interface with the 'Formato de celdas' (Format Cells) dialog box open. The 'Fecha' (Date) category is selected in the 'Categoría' (Category) list. The 'Muestra' (Preview) section displays the date '2012-03-14' formatted as 'miércoles, 14 de marzo de 2012'. The 'Configuración regional' (Regional settings) is set to 'Español (América latina)'. The background shows a table with columns 'Ves_Nro', 'Esquela', 'Ves_Fecha', 'Ves_Tipo', 'Falta', and 'Ves_Falta_Descripcion'.

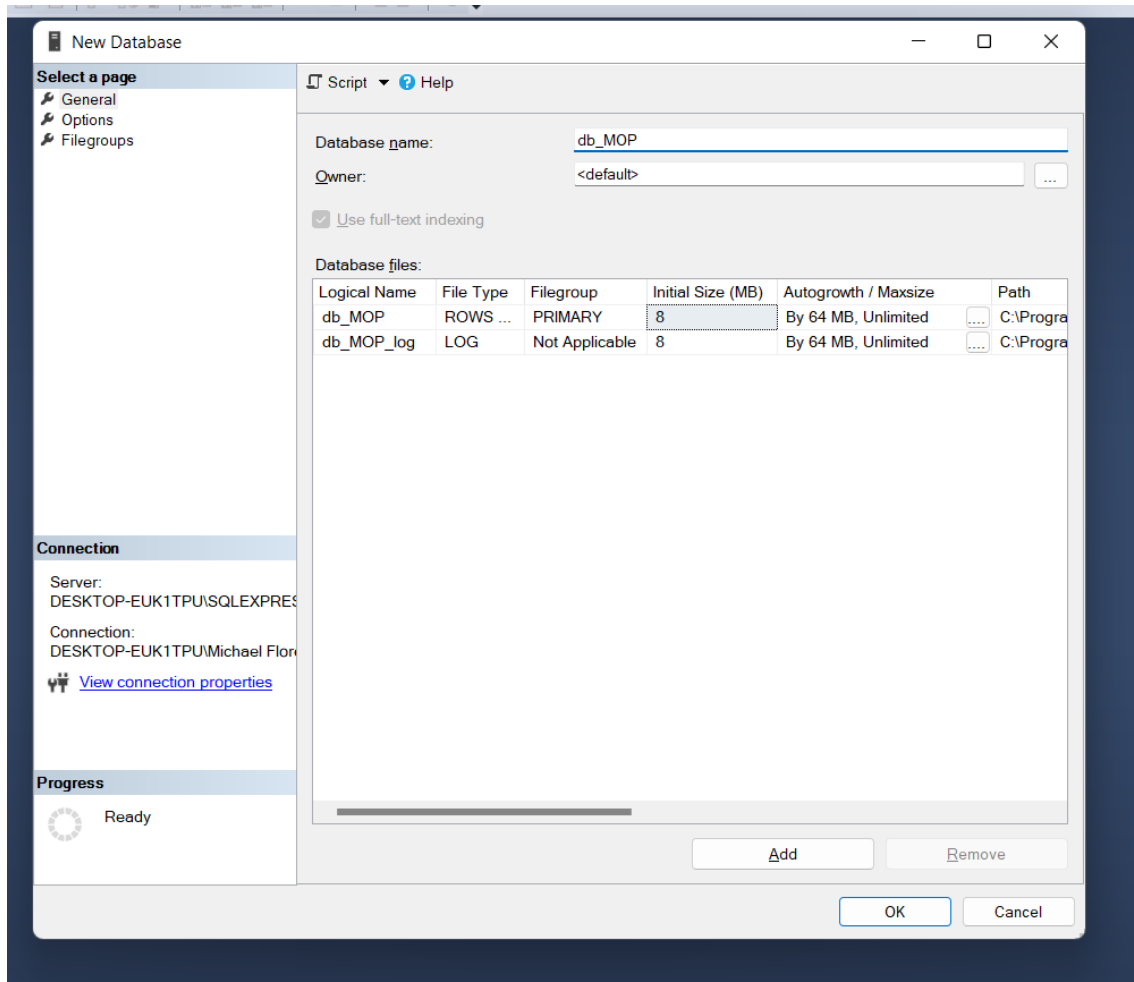
Luego quitaremos valores que nos puedan generar error para insertar la data en SQL Server, por ejemplo, comillas y puntos comas, etc.



The screenshot shows the Microsoft Excel interface with the 'Buscar y reemplazar' (Find and Replace) dialog box open. The 'Reemplazar' (Replace) tab is active. The 'Reemplazar con' (Replace with) field contains the text 'VIRAR EN U DONDE NO ES PERMITIDO'. The 'Buscar' (Find) field contains the text 'VIRAR EN "U" DONDE NO ES PERMITIDO'. The background shows a table with columns 'Ves_Nro', 'Esquela', 'Ves_Fecha', 'Ves_Tipo', 'Falta', and 'Ves_Falta_Descripcion'.

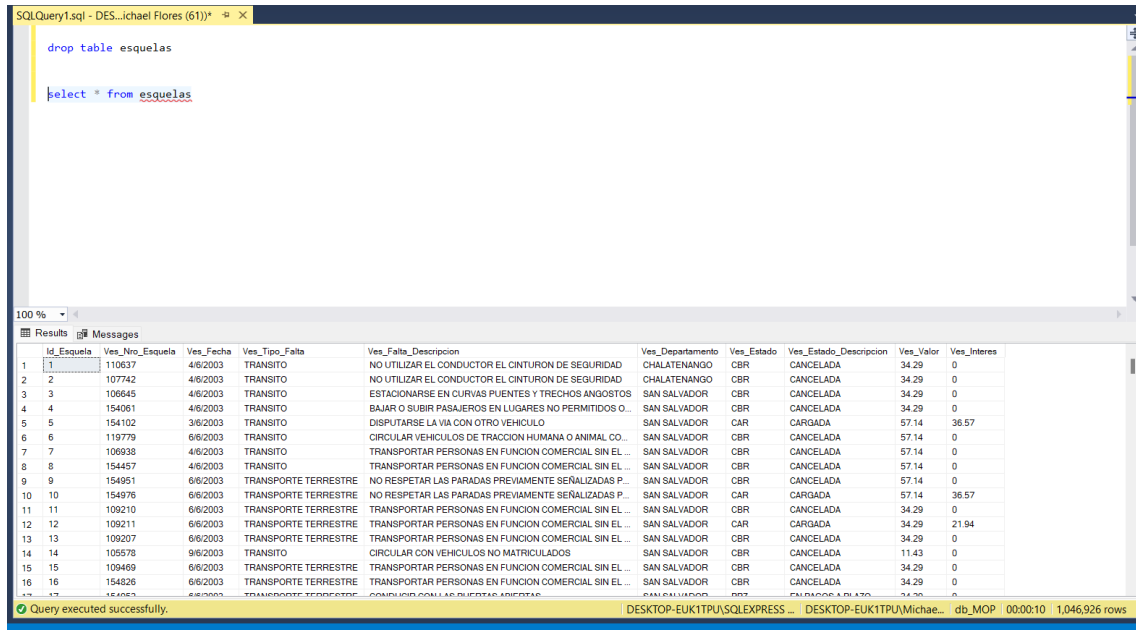


Creamos la base de datos que utilizaremos.





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



SQLQuery1.sql - DES...ichael Flores (61)*

```
drop table esquelas
```

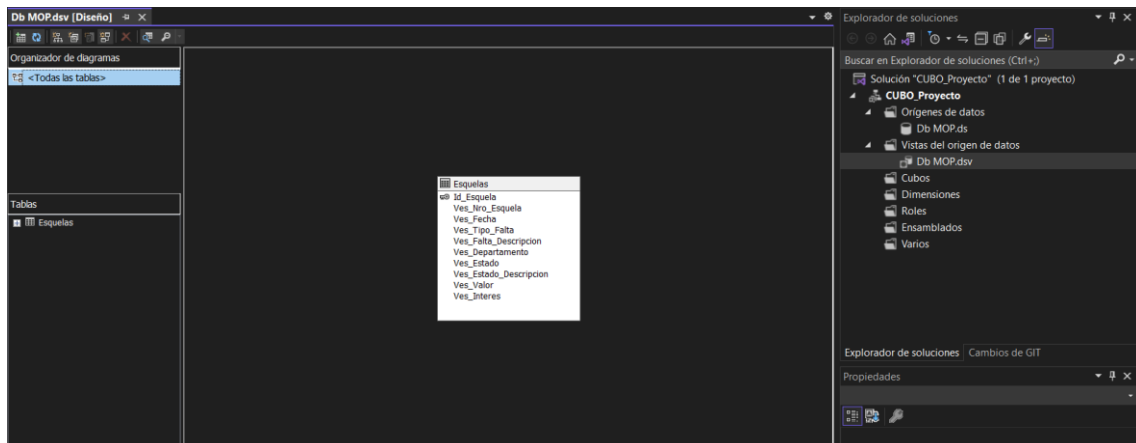
```
select * from esquelas
```

Results

	Id_Esquila	Ves_Nro_Esquila	Ves_Fecha	Ves_Tipo_Falta	Ves_Falta_Descripcion	Ves_Departamento	Ves_Estado	Ves_Estado_Descripcion	Ves_Valor	Ves_Interes
1	1	110637	4/6/2003	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO	CBR	CANCELADA	34.29	0
2	2	107742	4/6/2003	TRANSITO	NO UTILIZAR EL CONDUCTOR EL CINTURON DE SEGURIDAD	CHALATENANGO	CBR	CANCELADA	34.29	0
3	3	106645	4/6/2003	TRANSITO	ESTACIONARSE EN CURVAS PUENTES Y TRECOS ANGOSTOS	SAN SALVADOR	CBR	CANCELADA	34.29	0
4	4	154061	4/6/2003	TRANSITO	BAJAR O SUBIR PASAJEROS EN LUGARES NO PERMITIDOS O...	SAN SALVADOR	CBR	CANCELADA	34.29	0
5	5	154102	3/6/2003	TRANSITO	DISPUTARSE LA VIA CON OTRO VEHICULO	SAN SALVADOR	CAR	CARGADA	57.14	36.57
6	6	119779	6/6/2003	TRANSITO	CIRCULAR VEHICULOS DE TRACCION HUMANA O ANIMAL CO...	SAN SALVADOR	CBR	CANCELADA	57.14	0
7	7	106938	4/6/2003	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	57.14	0
8	8	154457	4/6/2003	TRANSITO	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	57.14	0
9	9	154951	6/6/2003	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZADAS P...	SAN SALVADOR	CBR	CANCELADA	57.14	0
10	10	154976	6/6/2003	TRANSPORTE TERRESTRE	NO RESPETAR LAS PARADAS PREVIAMENTE SEÑALIZADAS P...	SAN SALVADOR	CAR	CARGADA	57.14	36.57
11	11	109210	6/6/2003	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	34.29	0
12	12	109211	6/6/2003	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CAR	CARGADA	34.29	21.94
13	13	109207	6/6/2003	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	34.29	0
14	14	105578	9/6/2003	TRANSITO	CIRCULAR CON VEHICULOS NO MATRICULADOS	SAN SALVADOR	CBR	CANCELADA	11.43	0
15	15	109469	6/6/2003	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	34.29	0
16	16	154826	6/6/2003	TRANSPORTE TERRESTRE	TRANSPORTAR PERSONAS EN FUNCION COMERCIAL SIN EL ...	SAN SALVADOR	CBR	CANCELADA	34.29	0
17	17	154853	6/6/2003	TRANSPORTE TERRESTRE	CONDUCTOR CON LAS PUERTAS ABIERTAS	SAN SALVADOR	CBR	CANCELADA	34.29	0

Query executed successfully.

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



Db MOP.dsv [Diseño]

Organizador de diagramas

<Todas las tablas>

Tablas

Esquelas

Esquelas

- Id_Esquila
- Ves_Nro_Esquila
- Ves_Fecha
- Ves_Tipo_Falta
- Ves_Falta_Descripcion
- Ves_Departamento
- Ves_Estado
- Ves_Estado_Descripcion
- Ves_Valor
- Ves_Interes

Explorador de soluciones

Buscar en Explorador de soluciones (Ctrl+Q)

Solución "CUBO_Proyecto" (1 de 1 proyecto)

- CUBO_Proyecto
 - Orígenes de datos
 - Db MOP.ds
 - Vistas del origen de datos
 - Db MOP.dsv
 - Cubos
 - Dimensiones
 - Roles
 - Ensamblados
 - Varios

Explorador de soluciones | Cambios de GIT

Propiedades

Procedemos a crear el Cubo, las dimensiones respectivas y procesamos

Procesar Cubo - db_MOP

Nombre de objeto	Tipo	Opciones de proceso	Configuración
db_MOP	Cubo	Proceso completo	

Progreso del proceso

Comando

- Procesando Cubo, 'db_MOP' completados.
Hora de inicio: 10/11/2022 14:14:38; Hora de finalización: 10/11/2022 14:17:40; Duración: 0:03:02
- Procesando Grupo de medida, 'Esquelas' completados.
- Procesando Dimensión, 'Esquelas' completados.
Hora de inicio: 10/11/2022 14:14:00; Hora de finalización: 10/11/2022 14:14:38; Duración: 0:00:38
- Procesando Atributo de dimensión, '(All)' completados.
- Procesando Atributo de dimensión, 'Id Esquela' completados. 1046927 filas leídas.
Hora de inicio: 10/11/2022 14:14:00; Hora de finalización: 10/11/2022 14:14:26; Duración: 0:00:25
- Consultas SQL
SELECT DISTINCT [dbo_Esquelas].[Id_Esquela] AS [dbo_EsquelasId_Esquela0_0] FROM [dbo].[Esquelas] AS [dbo_Esquelas]

Estado:

✓ Proceso finalizado correctamente.

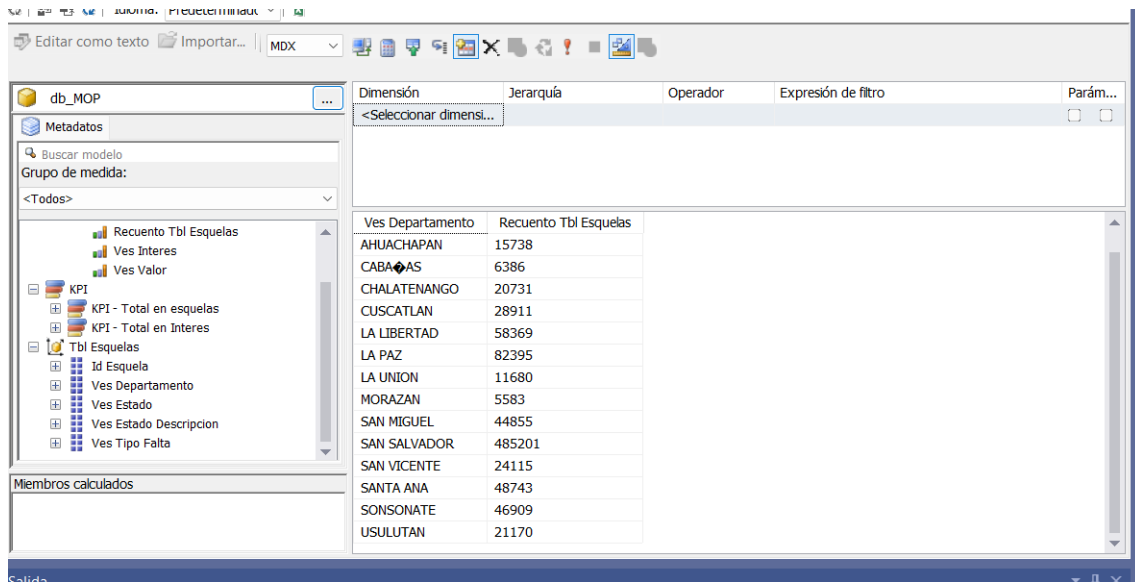
Detener Reprocesar Ver detalles... Copiar

Cerrar Ayuda

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

1. 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.

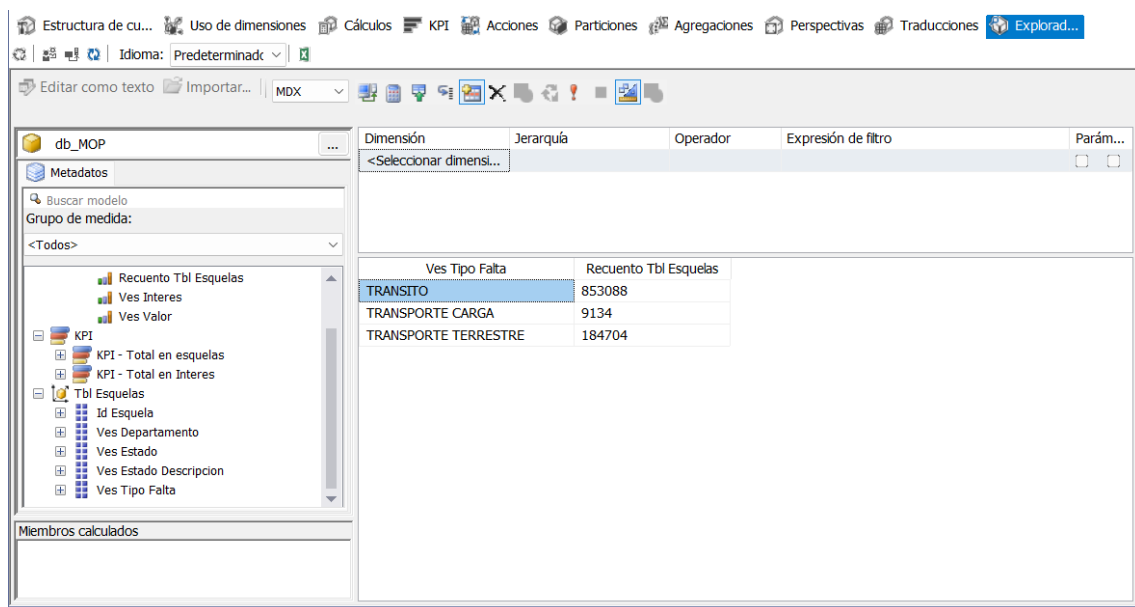


The screenshot shows the MDX Explorer interface with the following data:

Ves Departamento	Recuento Tbl Esquelas
AHUACHAPAN	15738
CABAÑAS	6386
CHALATENANGO	20731
CUSCATLAN	28911
LA LIBERTAD	58369
LA PAZ	82395
LA UNION	11680
MORAZAN	5583
SAN MIGUEL	44855
SAN SALVADOR	485201
SAN VICENTE	24115
SANTA ANA	48743
SONSONATE	46909
USulután	21170

2. 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.

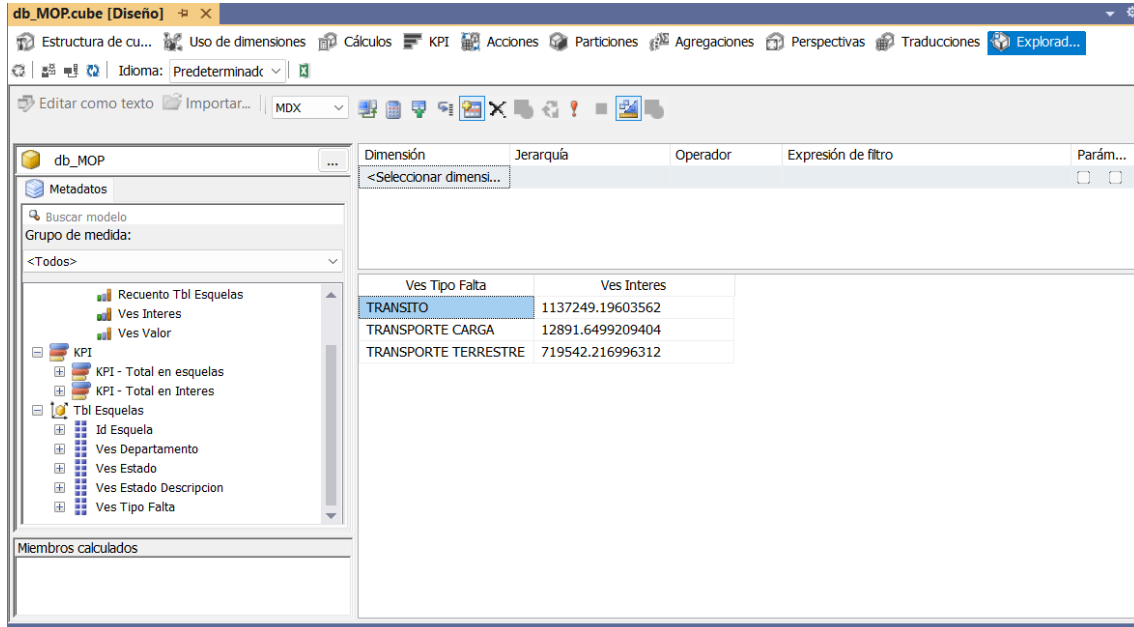


The screenshot shows the MDX Explorer interface with the following data:

Ves Tipo Falta	Recuento Tbl Esquelas
TRANSITO	853088
TRANSPORTE CARGA	9134
TRANSPORTE TERRESTRE	184704

3. Tipos que generan más interés

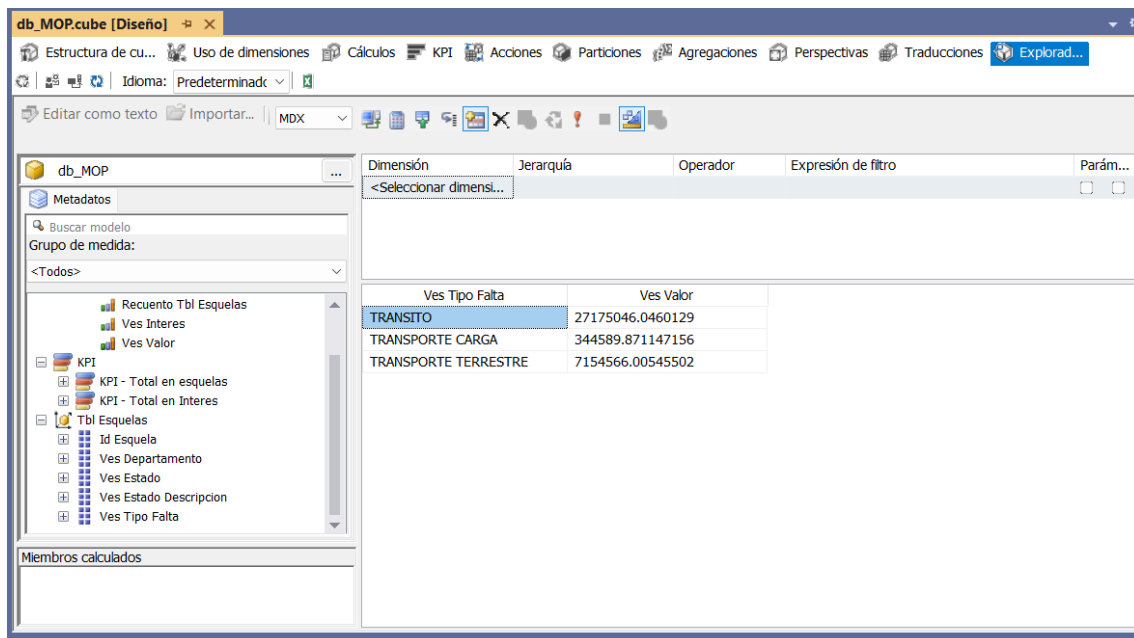
- En conclusión, a mayor valor del interés generado son los que menos pagan sus multas. Las multas de transito son las menos pagadas y esto tiene todo el sentido ya que son las que más se generan.
- Se recomienda crear un programa de facilidades de pago.



Dimensión	Jerarquía	Operador	Expresión de filtro	Parám...
<Seleccionar dimensi...				
Ves Tipo Falta			Ves Interes	
TRANSITO			1137249.19603562	
TRANSPORTE CARGA			12891.6499209404	
TRANSPORTE TERRESTRE			719542.216996312	

4. Tipos de esquila que generan más valor

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



Dimensión	Jerarquía	Operador	Expresión de filtro	Parám...
<Seleccionar dimensi...				
Ves Tipo Falta			Ves Valor	
TRANSITO			27175046.0460129	
TRANSPORTE CARGA			344589.871147156	
TRANSPORTE TERRESTRE			7154566.00545502	

5. Valor e interés generado por departamento

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

db_MOP.cube [Diseño]

Estructura de cu... Uso de dimensiones Cálculos KPI Acciones Particiones Agregaciones Perspectivas Traducciones Explorad...

Idioma: Predeterminad MDX

Editar como texto Importar...

db_MOP

Metadatos

Buscar modelo

Grupo de medida:

<Todos>

Recuento Tbl Esquelas
Ves Interes
Ves Valor

KPI

KPI - Total en esquelas
KPI - Total en Interes

Tbl Esquelas

Id Esquela
Ves Departamento
Ves Estado
Ves Estado Descripcion
Ves Tipo Falta

Miembros calculados

Dimensión	Jerarquía	Operador	Expresión de filtro	Parám...
<Seleccionar dimensl...				

Ves Departamento	Ves Valor	Ves Interes
AHUACHAPAN	520667.12090...	20265.5799365...
CABAAS	261162.77937...	5653.01997947...
CHALATENANGO	674490.71228...	13985.7299656...
CUSCATLAN	1227306.1058...	42363.3297505...
LA LIBERTAD	2043251.7237...	109478.179698...
LA PAZ	2722892.3474...	90208.1095137...
LA UNION	446913.39981...	11867.9699082...
MORAZAN	187352.39081...	3782.57000207...
SAN MIGUEL	1731150.3115...	84174.7196774...
SAN SALVADOR	15115239.070...	1118088.55460...
SAN VICENTE	888659.76177...	27190.7098321...
SANTA ANA	1597969.6657...	60455.0698902...
SONSONATE	1656632.7707...	57499.4298224...
USulután	719877.46311...	30748.5298628...

Salida

6. 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.

db_MOP.cube [Diseño]

Estructura de cu... Uso de dimensiones Cálculos KPI Acciones Particiones Agregaciones Perspectivas Traducciones Explorad...

Idioma: Predeterminad MDX

Editar como texto Importar...

db_MOP

Metadatos

Buscar modelo

Grupo de medida:

<Todos>

db_MOP

Measures

Tbl Esquelas

Recuento Tbl Esquelas
Ves Interes
Ves Valor

KPI

KPI - Total en esquelas
KPI - Total en Interes

Tbl Esquelas

Id Esquela
Ves Departamento
Ves Estado
Ves Estado Descripcion
Ves Tipo Falta

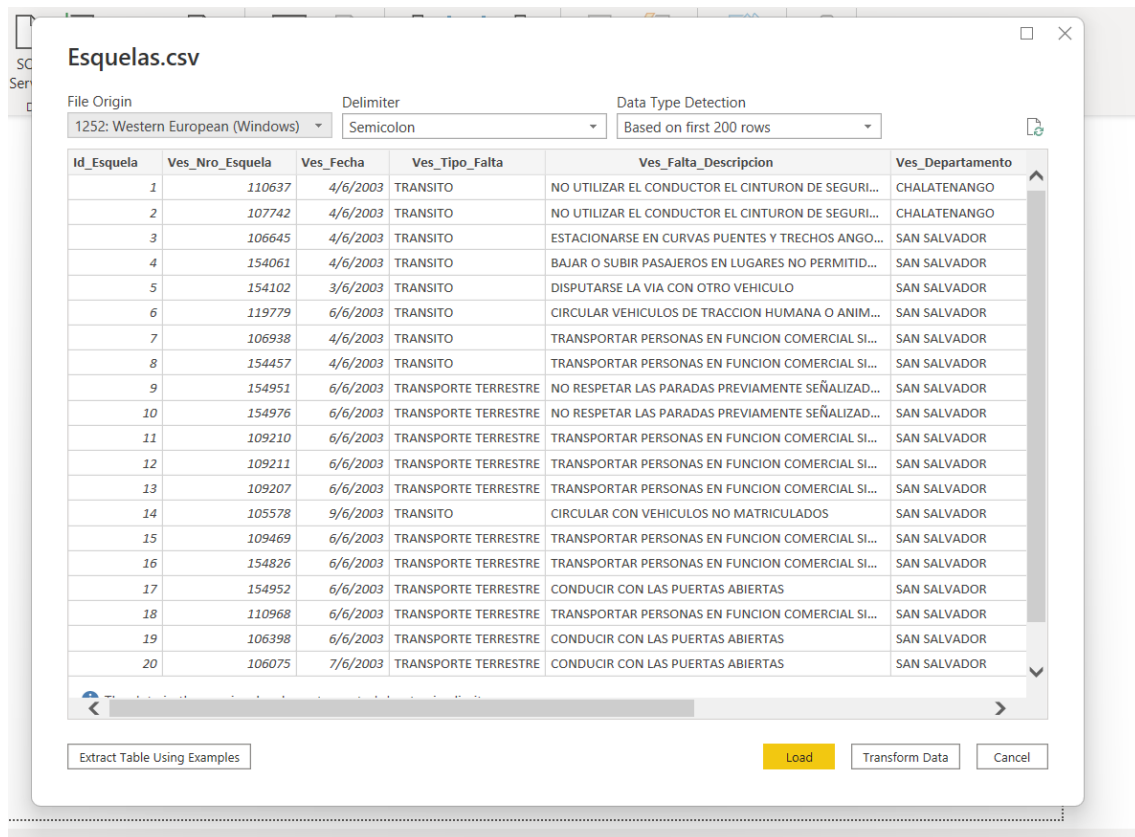
Miembros calculados

Dimensión	Jerarquía	Operador	Expresión de filtro	Parám...
<Seleccionar dimensl...				

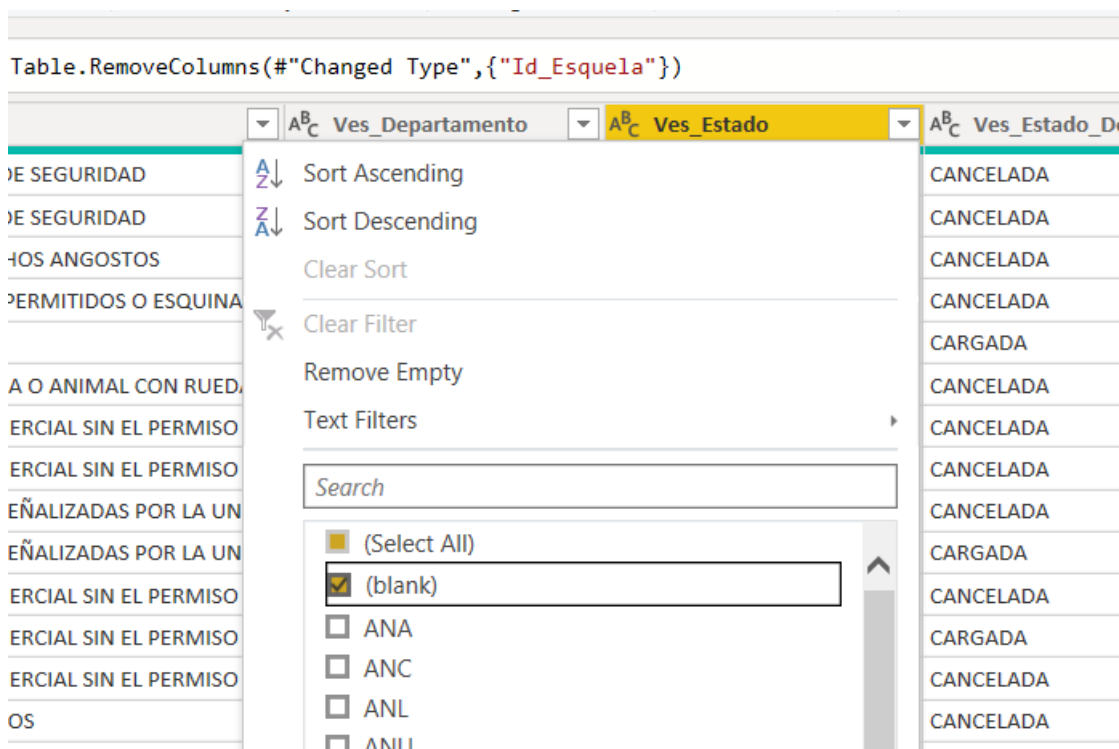
Ves Estado	Ves Estado Descripcion	Recuento Tbl Esquelas
ANA	ANALISIS APELADA	1
ANC	ANALISIS COBRADA	21
ANL	ANALISIS	13
ANU	ANULADA	199
APL	EN APELACION	360
CAR	CARGADA	2859
CBR	CANCELADA	805431
INC	INCONSISTENTE	87426
INI	PENDIENTE DE PAGO	113098
IPR	IMPROCEDENTE	27096
IPT	IMPUESTA	1687
IRI	RECURSO INCONFOR...	2859
PPZ	EN PAGOS A PLAZO	5477
RAT	RATIFICADA	2
REV	REVOCADA	215
SUS	SUSPENSIVO	33

➤ 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.



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



Query

fx = Table.SelectRows("#Removed Columns", each ([Ves_Estado] = ""))

Idioma	A ^B Ves_Departamento	A ^B Ves_Estado	A ^B Ves_Estado_Descripcion	1.2 Ves_Valor	1.2 Ves_Interes
1	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
2	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
3	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
4	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
5	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
6	ICIA DE CONDUCIR			Error	7.3
7	ICIA DE CONDUCIR			Error	7.3
8	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
9	ICIA DE CONDUCIR			Error	7.3
10	ICIA DE CONDUCIR			Error	7.3
11	ICIA DE CONDUCIR			Error	7.3
12	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
13	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
14	ICIA DE CONDUCIR			Error	7.3
15	ICIA DE CONDUCIR			Error	7.3
16	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
17	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
18	ICIA DE CONDUCIR			Error	7.3
19	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
20	ICIA DE CONDUCIR			Error	7.3
21	ICIA DE CONDUCIR			Error	7.3
22	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
23	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
24	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
25					
	null				

sources Parameters Query Manage Columns Reduce Rows Sort Transform Combine

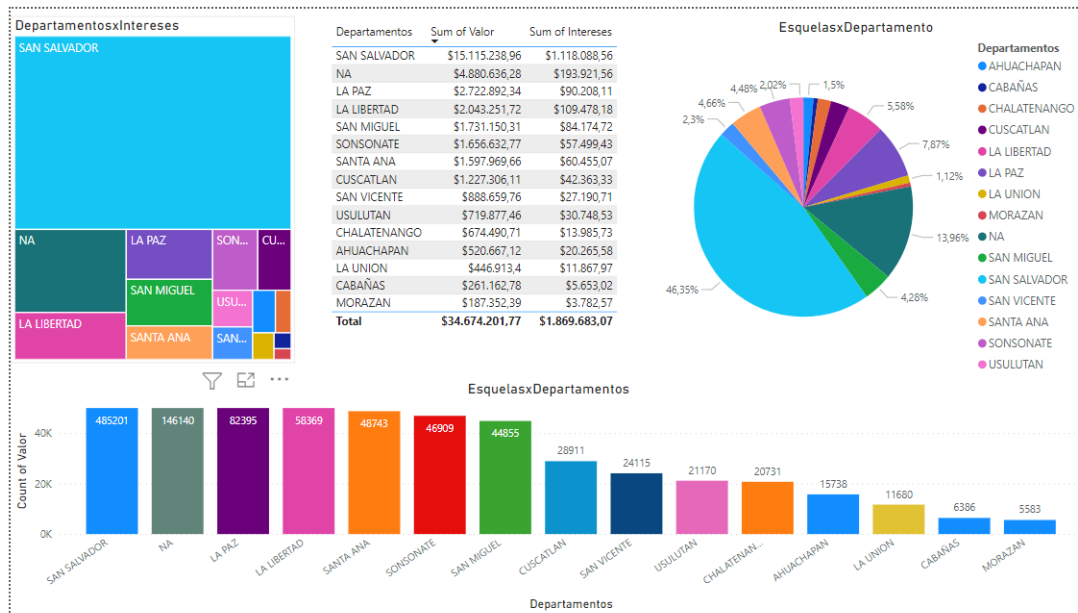
fx = Table.SelectRows("#Removed Columns", each ([Ves_Estado] = ""))

	A ^B Ves_Departamento	A ^B Ves_Estado	A ^B Ves_Estado_Descripcion	1.2 Ves_Valor
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
ONDUCIR				Error
ONDUCIR				Error
A COMPETENCIAS AUTOMOVILISTICAS SI...				ni
ONDUCIR				Error
ONDUCIR				Error

Copy
Remove Columns
Remove Other Columns
Add Column From Examples...
Remove Duplicates
Remove Errors
Replace Values...
Fill
Change Type
Transform

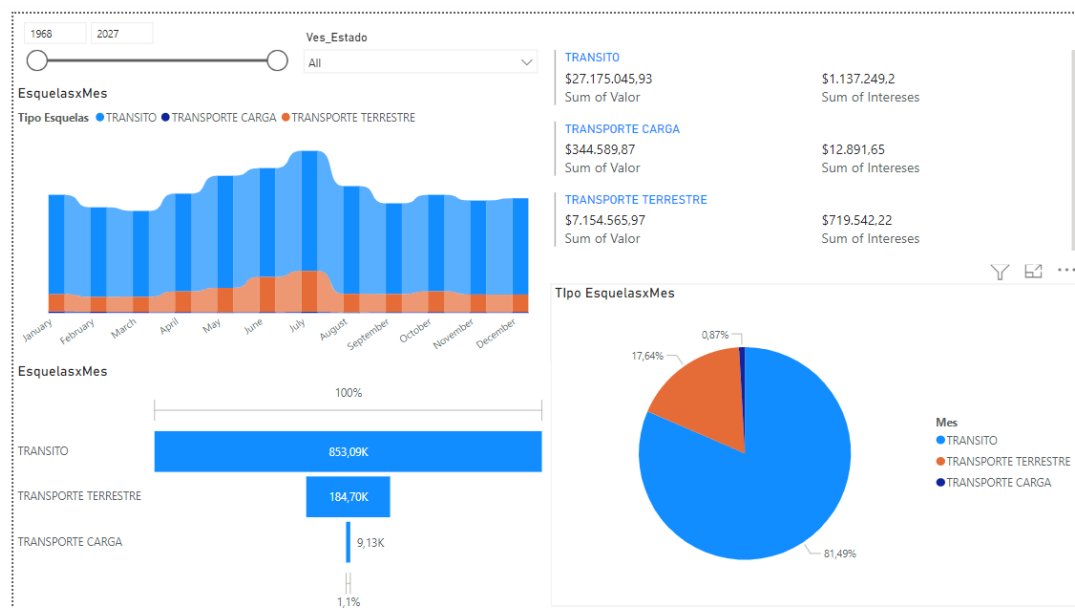
1. Tablero por departamento

- Podemos observar que evidentemente, como en nuestro análisis anterior, a mayor número o porcentaje de esuelas 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 esuelas anuladas.
- Se recomienda poner más atención al departamento para fomentar la educación vial.



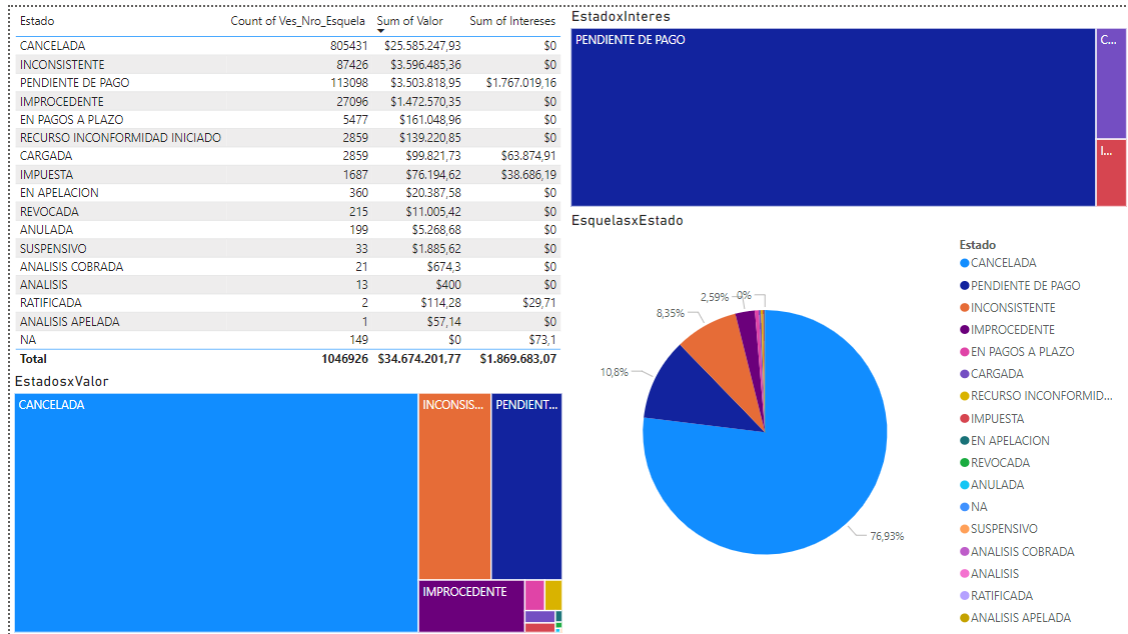
2. Tablero por tipo de faltas

- La mayoría de las esuelas 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



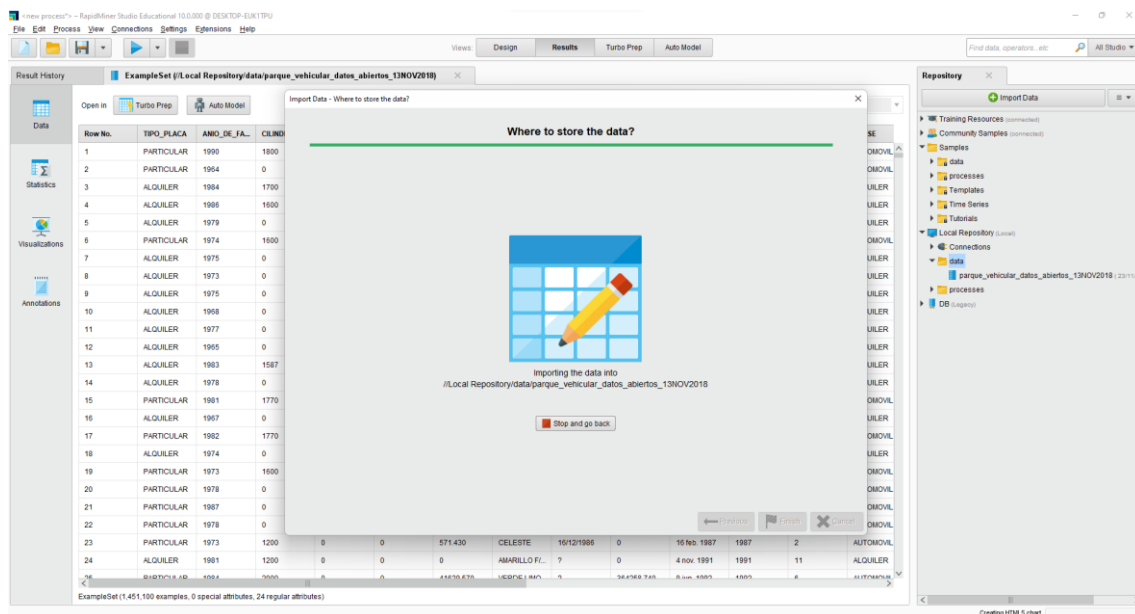
3. Tablero por estado de esquelas

- Como podemos interpretar 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.

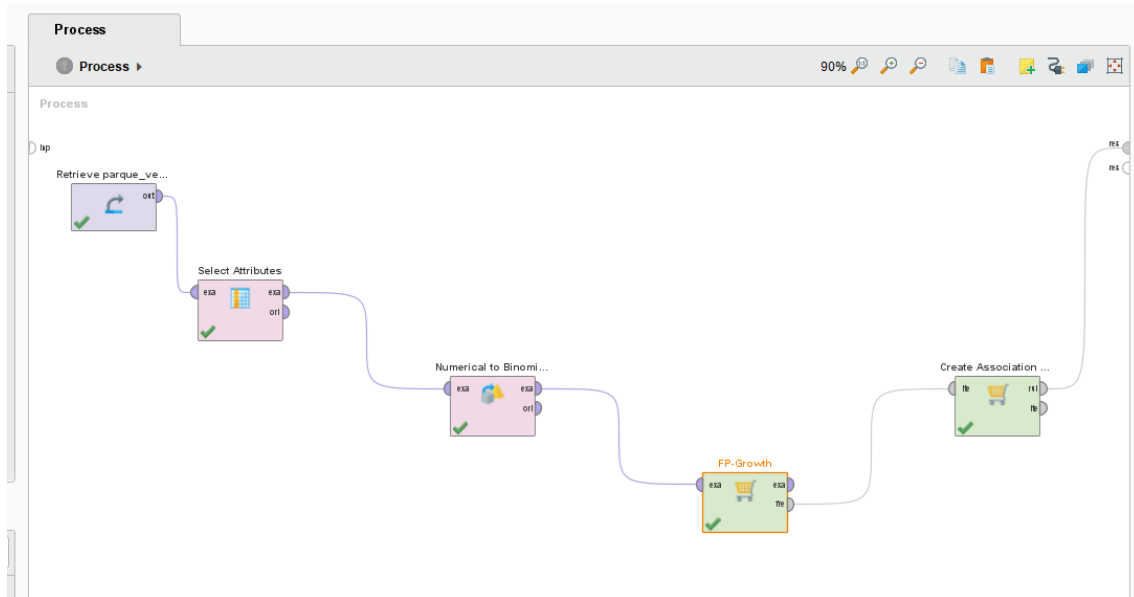


Reglas de asociación – RapidMiner

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



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.



1. 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

Result History | ExampleSet (/Local Repository/data/parque_vehicular_datos_abiertos_13NOV2018) | AssociationRules (Create Association Rules)

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convict...
157	VEHICULO NUEVO	GASOLINA, ASIENTOS, MOTOCICLETA	0.206	0.615	0.903	-0.463	0.127	2.607	1.984
158	VEHICULO NUEVO	GASOLINA, MOTOCICLETA	0.206	0.615	0.904	-0.463	0.127	2.607	1.984
164	VEHICULO NUEVO	ASIENTOS, MOTOCICLETA	0.207	0.619	0.904	-0.462	0.126	2.561	1.989
165	VEHICULO NUEVO	MOTOCICLETA	0.207	0.619	0.904	-0.462	0.126	2.561	1.989
328	ASIENTOS, VEHICULO NUEVO	GASOLINA, MOTOCICLETA	0.206	0.767	0.951	-0.331	0.142	3.250	3.275
336	ASIENTOS, VEHICULO NUEVO	MOTOCICLETA	0.207	0.772	0.952	-0.329	0.142	3.193	3.329
348	GASOLINA, VEHICULO NUEVO	ASIENTOS, MOTOCICLETA	0.206	0.786	0.956	-0.318	0.142	3.252	3.542
349	GASOLINA, VEHICULO NUEVO	MOTOCICLETA	0.206	0.786	0.956	-0.318	0.142	3.252	3.543
369	GASOLINA, ASIENTOS, VEHICULO NUEVO	MOTOCICLETA	0.206	0.811	0.962	-0.301	0.144	3.358	4.023

Result History | AssociationRules (Create Association Rules)

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convict...
431	MOTOCICLETA	GASOLINA, ASIENTOS, VEHICULO NUEVO	0.206	0.851	0.971	-0.278	0.144	3.358	5.001
432	MOTOCICLETA	GASOLINA, VEHICULO NUEVO	0.206	0.851	0.971	-0.278	0.142	3.252	4.948
433	ASIENTOS, MOTOCICLETA	GASOLINA, VEHICULO NUEVO	0.206	0.851	0.971	-0.278	0.142	3.252	4.948
446	MOTOCICLETA	ASIENTOS, VEHICULO NUEVO	0.207	0.856	0.972	-0.276	0.142	3.193	5.086
447	MOTOCICLETA	VEHICULO NUEVO	0.207	0.856	0.972	-0.276	0.126	2.561	4.628
448	ASIENTOS, MOTOCICLETA	VEHICULO NUEVO	0.207	0.856	0.972	-0.276	0.126	2.561	4.628
468	GASOLINA, MOTOCICLETA	ASIENTOS, VEHICULO NUEVO	0.206	0.871	0.975	-0.266	0.142	3.250	5.692
469	GASOLINA, ASIENTOS, MOTOCICLETA	VEHICULO NUEVO	0.206	0.872	0.975	-0.266	0.127	2.607	5.180
470	GASOLINA, MOTOCICLETA	VEHICULO NUEVO	0.206	0.872	0.975	-0.266	0.127	2.607	5.180

- 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

Result History | ExampleSet (/Local Repository/data/parque_vehicular_datos_abiertos_13NOV2018) | AssociationRules (Create Association Rules)

Show rules matching: all of these conclusions: PROPIEDAD, GASOLINA, ASIENTOS, VEHICULO USADO, PARTICULAR, AUTOMOVIL, SAN SALVADOR, VEHICULO NUEVO, MOTOCICLETA

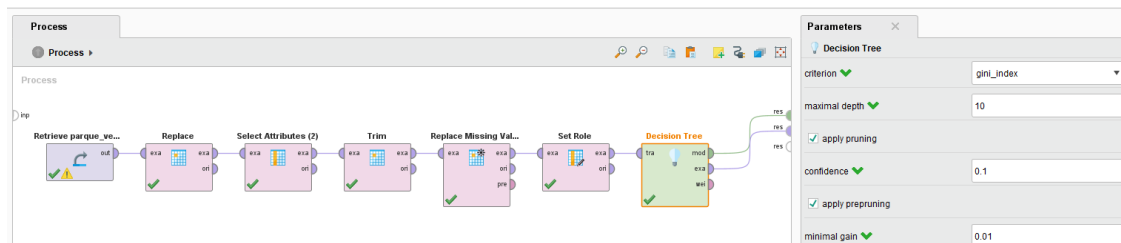
No.	Premises	Conclusion	Support	Confiden...	LaPlace	Gain	p-s	Lift
694	GASOLINA, ASIENTOS, PARTICULAR	AUTOMOVIL	0.400	0.986	0.996	-0.411	0.231	2.366
693	PROPIEDAD, GASOLINA, ASIENTOS, PARTICULAR	AUTOMOVIL	0.365	0.986	0.996	-0.375	0.211	2.366
692	GASOLINA, ASIENTOS, PARTICULAR, SAN SALVADOR	AUTOMOVIL	0.200	0.986	0.998	-0.206	0.116	2.365
689	PROPIEDAD, GASOLINA, ASIENTOS, VEHICULO USADO, PARTICULAR	AUTOMOVIL	0.332	0.985	0.996	-0.343	0.192	2.363
687	GASOLINA, ASIENTOS, VEHICULO USADO, PARTICULAR	AUTOMOVIL	0.353	0.985	0.996	-0.364	0.204	2.363
646	PROPIEDAD, ASIENTOS, VEHICULO USADO, PARTICULAR	AUTOMOVIL	0.340	0.962	0.990	-0.367	0.193	2.309
642	ASIENTOS, VEHICULO USADO, PARTICULAR	AUTOMOVIL	0.361	0.961	0.989	-0.390	0.204	2.305
641	ASIENTOS, PARTICULAR, SAN SALVADOR	AUTOMOVIL	0.206	0.959	0.993	-0.223	0.116	2.301
634	PROPIEDAD, ASIENTOS, PARTICULAR	AUTOMOVIL	0.377	0.957	0.988	-0.410	0.213	2.296
631	ASIENTOS, PARTICULAR	AUTOMOVIL	0.412	0.953	0.986	-0.453	0.232	2.285
618	PROPIEDAD, ASIENTOS, VEHICULO USADO, PARTICULAR	GASOLINA, AUTOMOVIL	0.332	0.941	0.984	-0.374	0.190	2.334
611	ASIENTOS, VEHICULO USADO, PARTICULAR	GASOLINA, AUTOMOVIL	0.353	0.939	0.983	-0.399	0.201	2.331
601	ASIENTOS, PARTICULAR, SAN SALVADOR	GASOLINA, AUTOMOVIL	0.200	0.934	0.988	-0.229	0.114	2.318
593	GASOLINA, ASIENTOS, VEHICULO USADO, PARTICULAR	PROPIEDAD, AUTOMOVIL	0.332	0.928	0.981	-0.384	0.196	2.443
592	PROPIEDAD, ASIENTOS, PARTICULAR	GASOLINA, AUTOMOVIL	0.365	0.928	0.980	-0.422	0.207	2.303
585	ASIENTOS, PARTICULAR	GASOLINA, AUTOMOVIL	0.400	0.923	0.977	-0.466	0.225	2.290
542	ASIENTOS, VEHICULO USADO, PARTICULAR	PROPIEDAD, AUTOMOVIL	0.340	0.905	0.974	-0.411	0.197	2.384
525	GASOLINA, ASIENTOS, PARTICULAR	PROPIEDAD, AUTOMOVIL	0.365	0.901	0.972	-0.445	0.211	2.373
520	PROPIEDAD, GASOLINA, ASIENTOS, PARTICULAR	VEHICULO USADO, AUTOMOVIL	0.332	0.898	0.972	-0.408	0.196	2.465
514	PROPIEDAD, GASOLINA, ASIENTOS, VEHICULO USADO	AUTOMOVIL	0.334	0.894	0.971	-0.413	0.178	2.145
513	GASOLINA, ASIENTOS, VEHICULO USADO	AUTOMOVIL	0.355	0.892	0.969	-0.441	0.189	2.140
512	PROPIEDAD, GASOLINA, ASIENTOS, VEHICULO USADO	PARTICULAR, AUTOMOVIL	0.332	0.890	0.970	-0.415	0.178	2.156
511	GASOLINA, ASIENTOS, VEHICULO USADO	PARTICULAR, AUTOMOVIL	0.353	0.886	0.967	-0.444	0.188	2.147
507	ASIENTOS, VEHICULO USADO, PARTICULAR	PROPIEDAD, GASOLINA, AUTOMOVIL	0.332	0.885	0.969	-0.419	0.194	2.408
467	GASOLINA, ASIENTOS, PARTICULAR	VEHICULO USADO, AUTOMOVIL	0.353	0.871	0.963	-0.467	0.205	2.391

Min. Criterion: confidence
Min. Criterion Value:

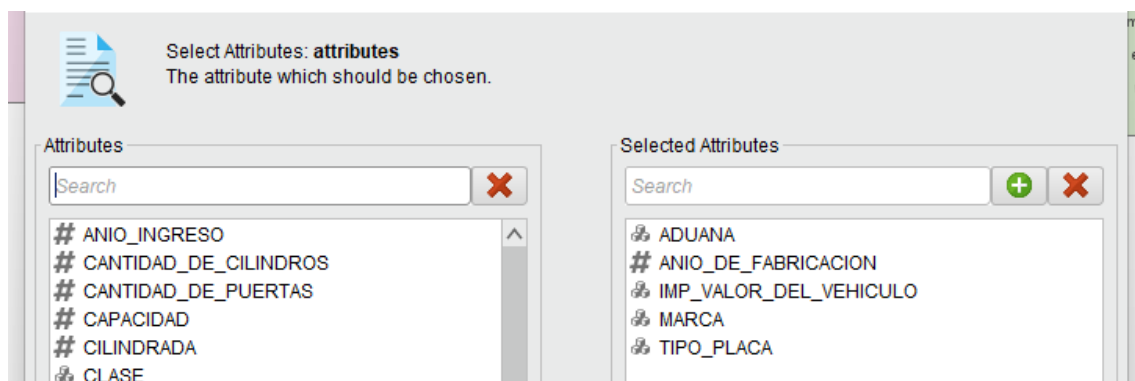
Árbol de decisiones - RapidMiner

2. 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



Ejecutamos nuestra primera vez con los siguientes atributos

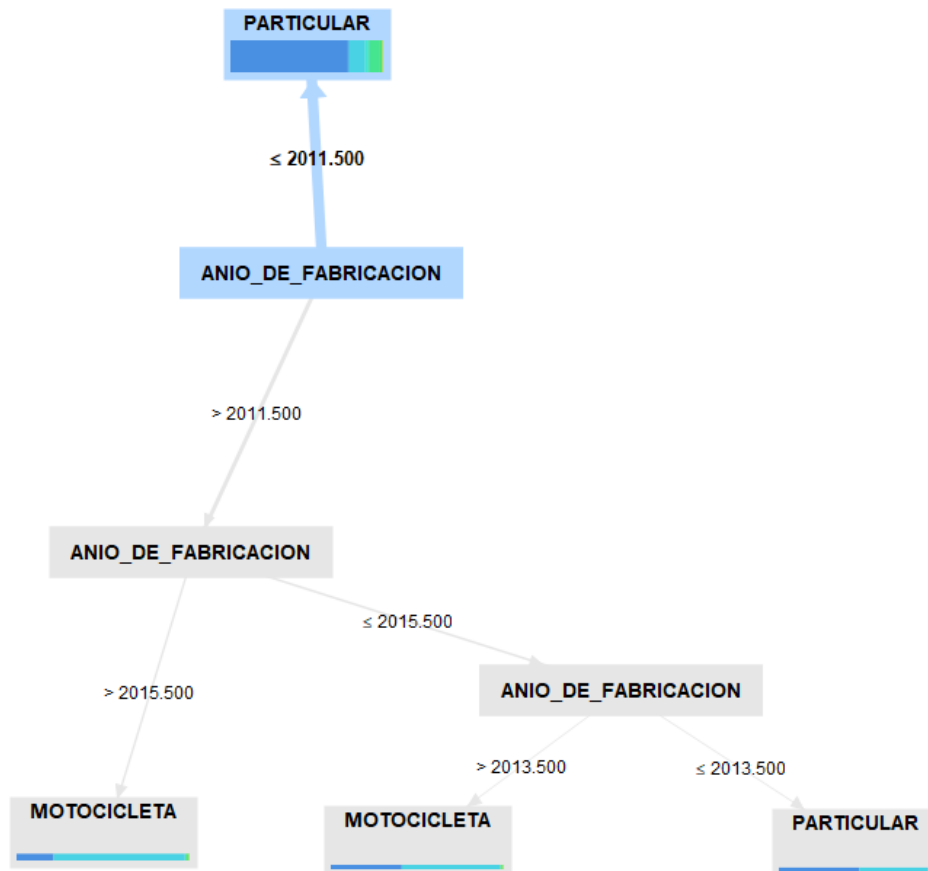


Nuestro dataset de muestreo quedaría de la siguiente manera

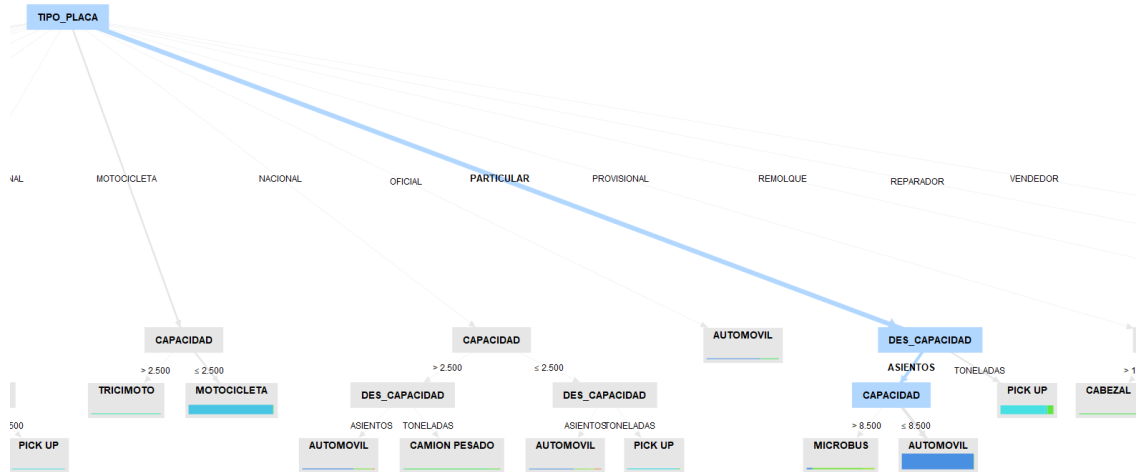
Result History					
ExampleSet (Set Role)					
Tree (Decision Tree)					
Open in Turbo Prep Auto Model					
Row No.	TIPO_PLACA	ADUANA	IMP_VALOR_DEL_VEHICULO	MARCA	ANIO_DE_FABRICACION
1	PARTICULAR	SAN BARTOL...	4094.56	NISSAN	1990
2	PARTICULAR	MARITIMA DE...	0.00	TOYOTA	1964
3	ALQUILER	MARITIMA DE...	0.00	TOYOTA	1984
4	ALQUILER	MARITIMA DE...	0.00	HYUNDAI	1986
5	ALQUILER	MARITIMA DE...	0.00	DATSUN	1979
6	PARTICULAR	MARITIMA DE...	0.00	DATSUN	1974
7	ALQUILER	MARITIMA DE...	0.00	TOYOTA	1975
8	ALQUILER	MARITIMA DE...	0.00	TOYOTA	1973
9	ALQUILER	MARITIMA DE...	0.00	TOYOTA	1975
10	ALQUILER	MARITIMA DE...	0.00	TOYOTA	1968
11	ALQUILER	MARITIMA DE...	0.00	DATSUN	1977
12	ALQUILER	MARITIMA DE...	0.00	PEUGEOT	1965
13	ALQUILER	SAN BARTOL...	4183.76	TOYOTA	1983
14	ALQUILER	MARITIMA DE...	0.00	SUBARU	1978
15	PARTICULAR	SAN BARTOL...	4442.74	TOYOTA	1994

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

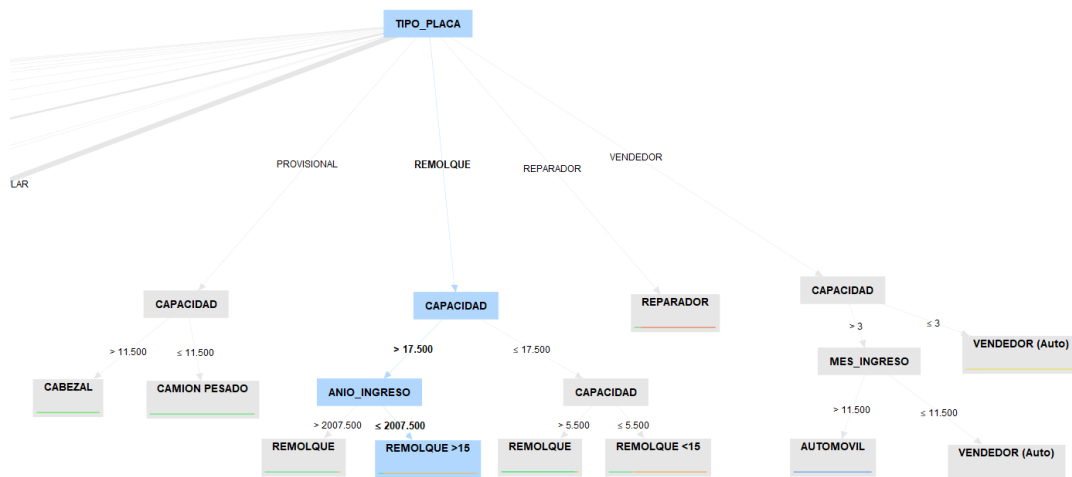
- La mayoría de las motocicletas fueron fabricadas antes del año 2015, partiendo de base que tipo de placas de estas particulares



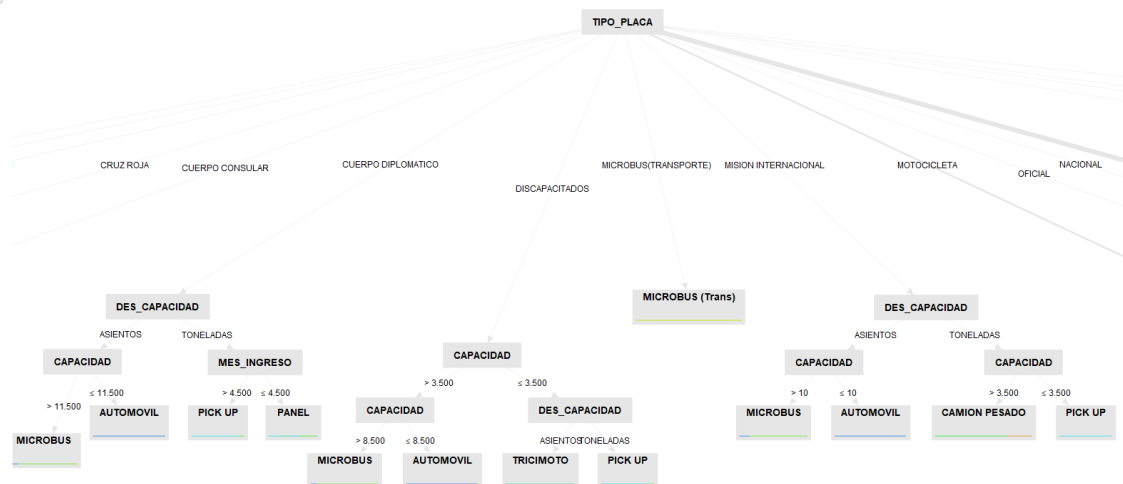
- De acuerdo con la capacidad podemos definir que si es medida en toneladas se trata pickups y si es en base a asientos podemos encontrar a la capacidad que pueden ser automóviles o microbuses, todo esto bajo el tipo de placa que es particular.
- Si es bajo la rama de motocicleta se puede definir que la capacidad del tanque aumenta o baja según sea tricótopo 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



- 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)



- 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

