



U D B V I R T U A L

Universidad Don Bosco, El Salvador

DMD G01T
Ciclo 02 -2022

Proyecto Fase 1

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

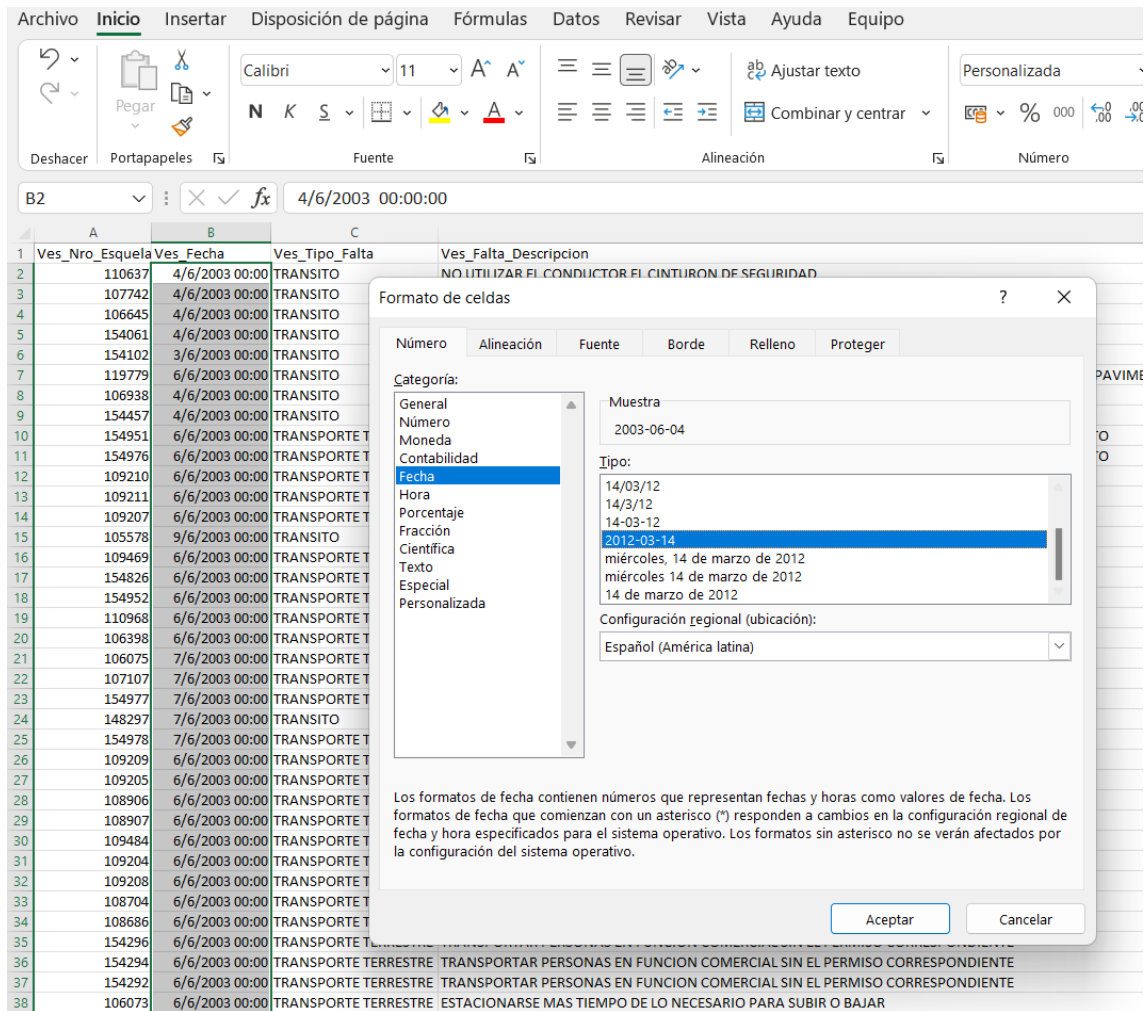


Objetivo

- ✓ 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.

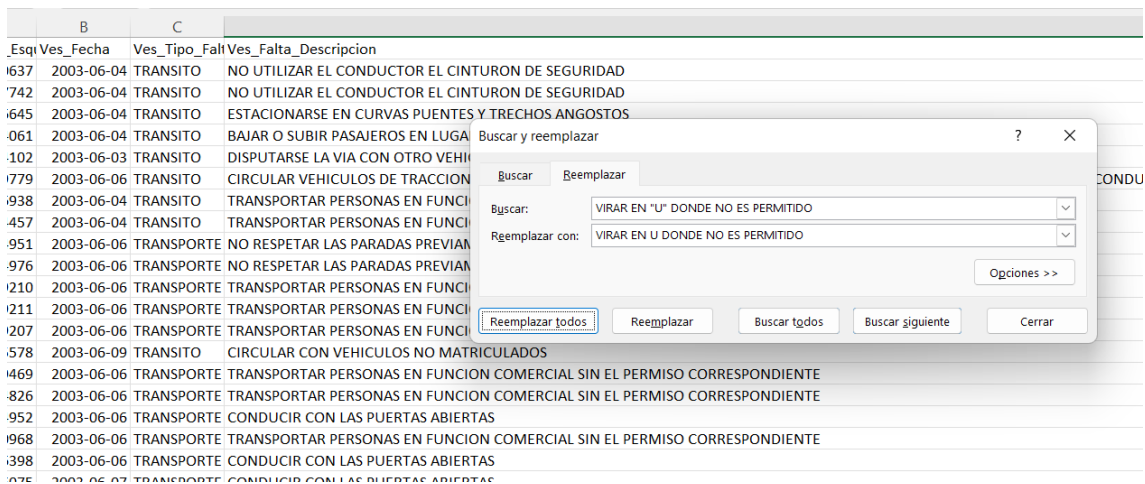
➤ Estrategia de CUBO de OLAP

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

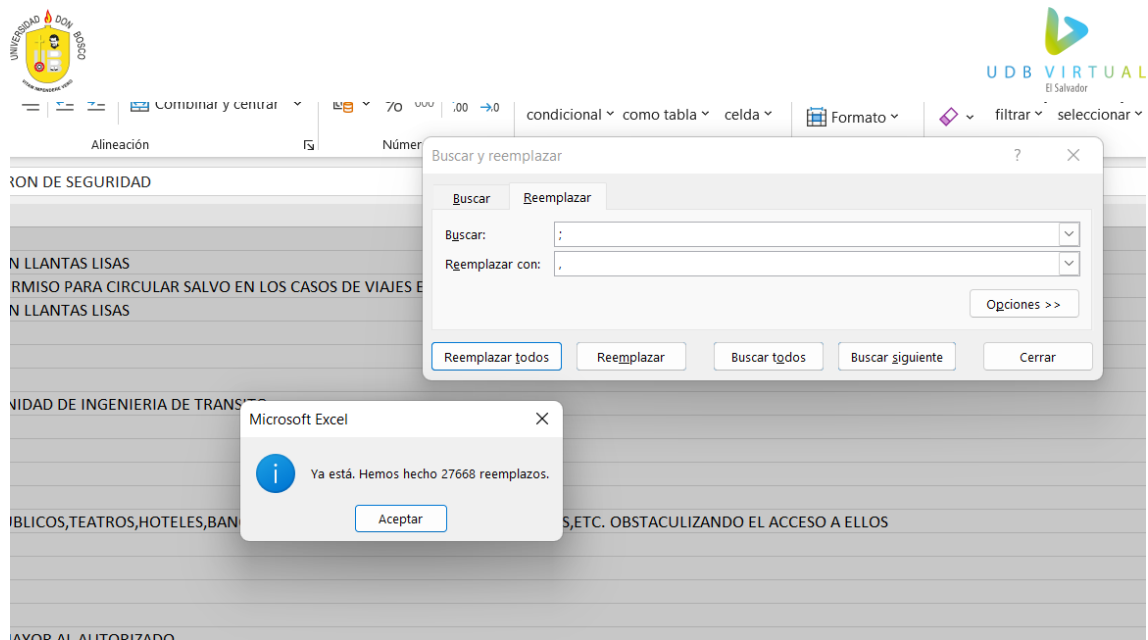


The screenshot shows the Microsoft Excel interface with the 'Formato de celdas' (Format Cells) dialog box open. The dialog box is set to the 'Fecha' (Date) category, and the 'Tipo' (Type) is set to '14-03-12' (dd-mm-yy). The 'Configuración regional' (Regional settings) is set to 'Español (América latina)'. The background shows a data table with columns: Ves_Nro, Esquela, Ves_Fecha, Ves_Tipo, Falta, and Ves_Falta_Descripcion. The data table contains various records, including dates and descriptions of incidents.

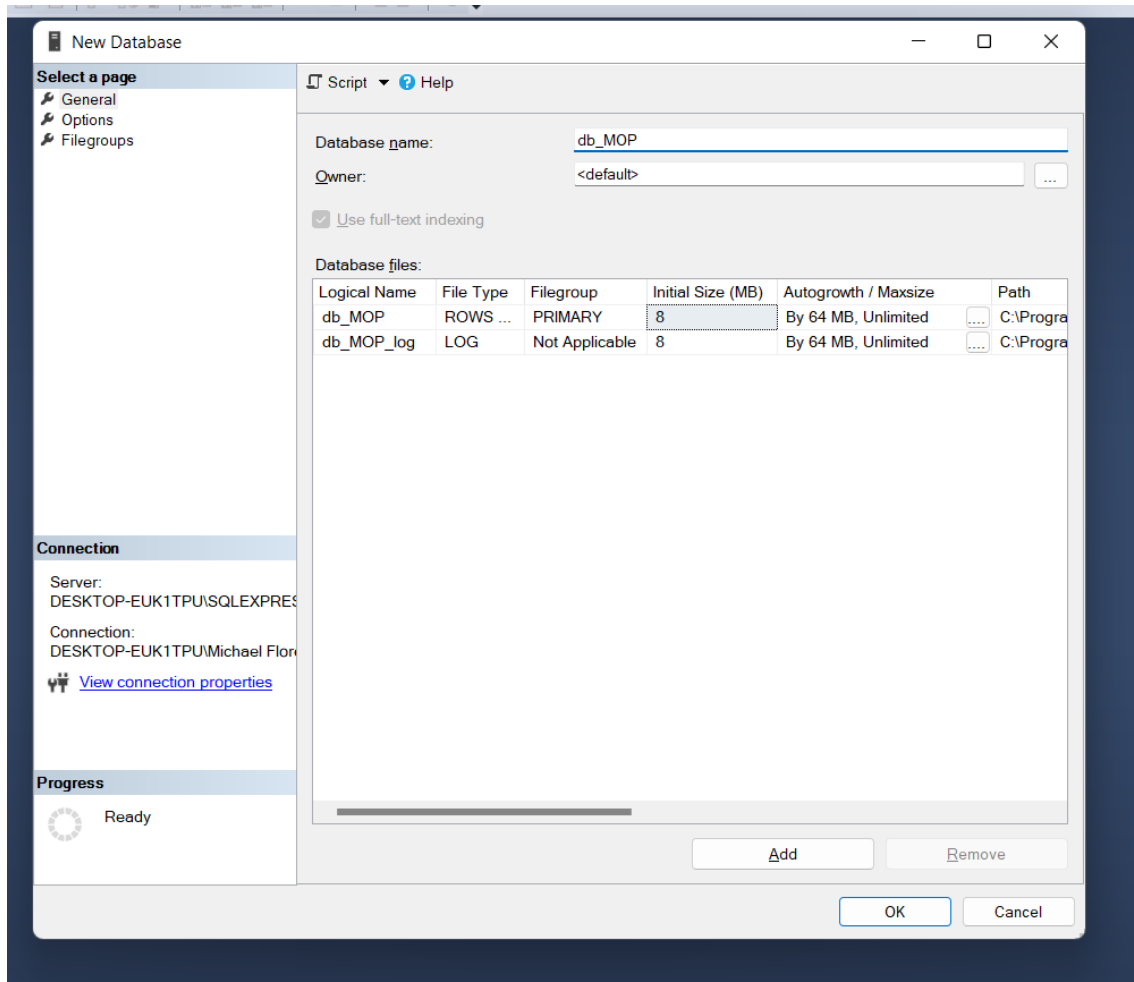
Luego quitaremos valores que nos puedan generar error para insertar la data en SQL Server, por ejemplo, comillas y puntos comas, etc.



The screenshot shows the Microsoft Excel interface with the 'Buscar y reemplazar' (Find and Replace) dialog box open. The dialog box is set to the 'Buscar' (Find) tab, and the 'Buscar' (Find) text is 'VIRAR EN "U" DONDE NO ES PERMITIDO'. The 'Reemplazar con' (Replace with) text is 'VIRAR EN U DONDE NO ES PERMITIDO'. The background shows a data table with columns: Esq_Ves_Fecha, Ves_Tipo, Falta, and Ves_Falta_Descripcion. The data table contains various records, including dates and descriptions of incidents.

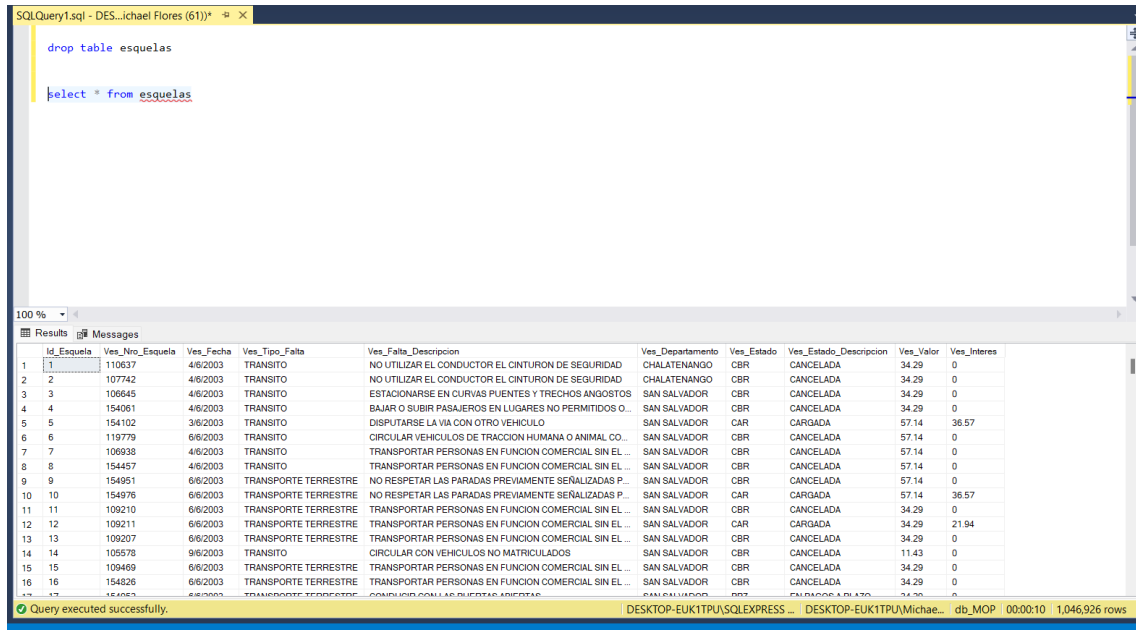


Creamos la base de datos que utilizaremos.





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
```

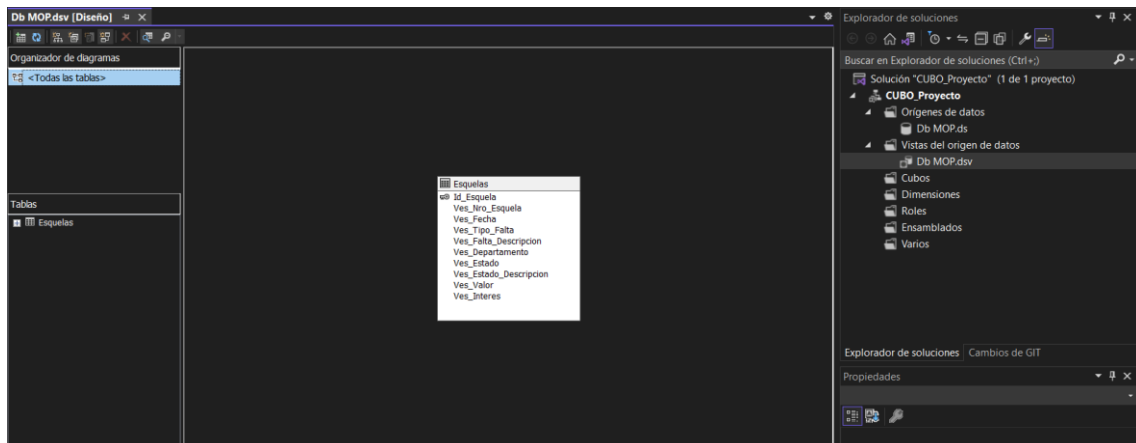
```
select * from esquelas
```

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
 - Origenes 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.

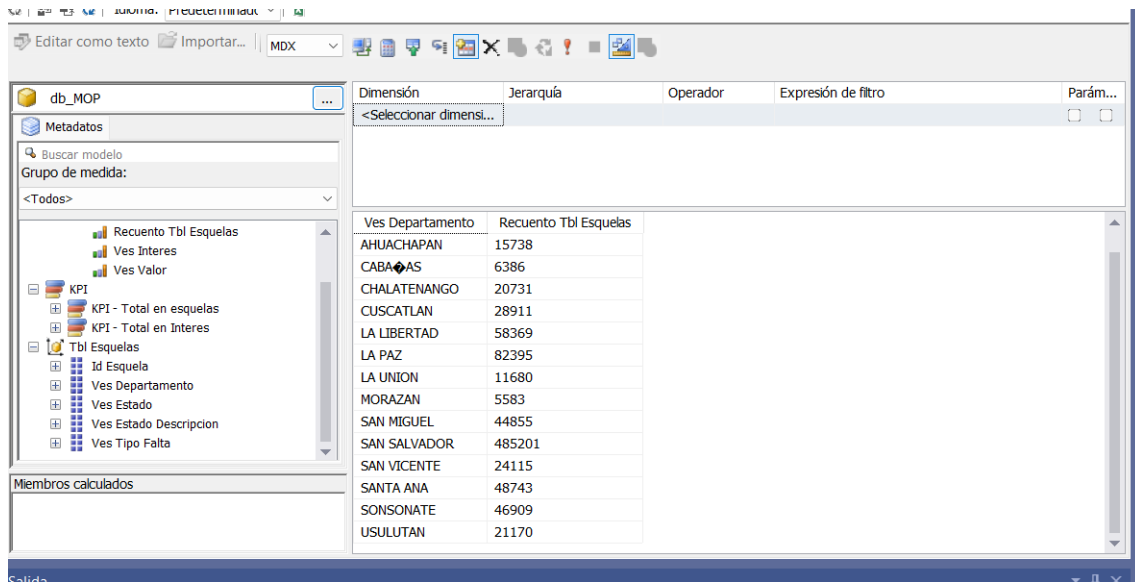
Detener Reprocesar Ver detalles... Copiar

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.

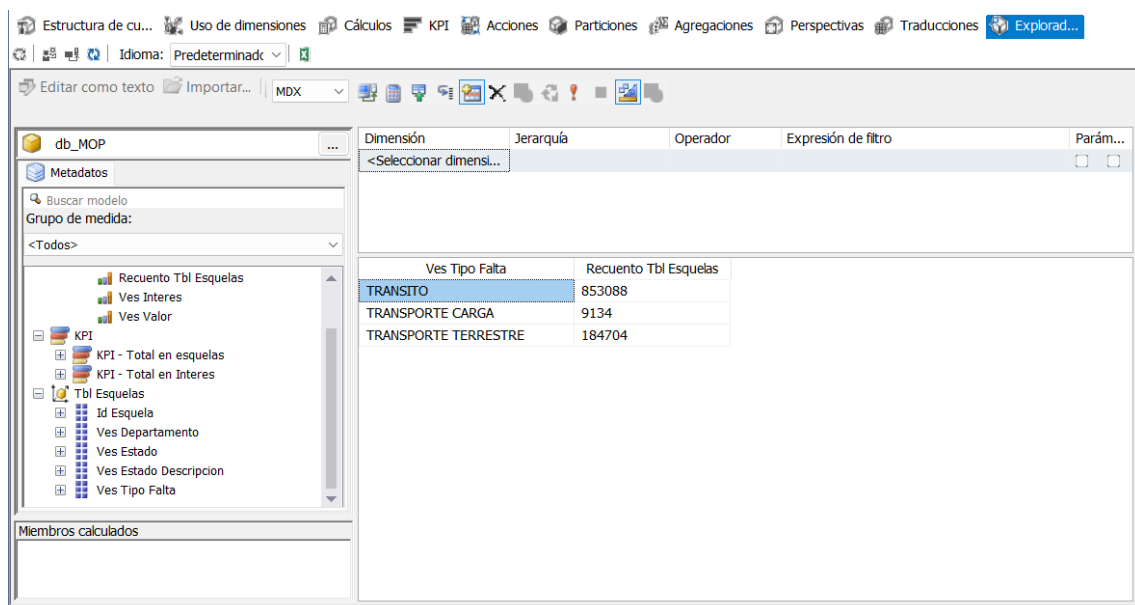


The screenshot shows the OLAP Explorer interface with the following data:

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.

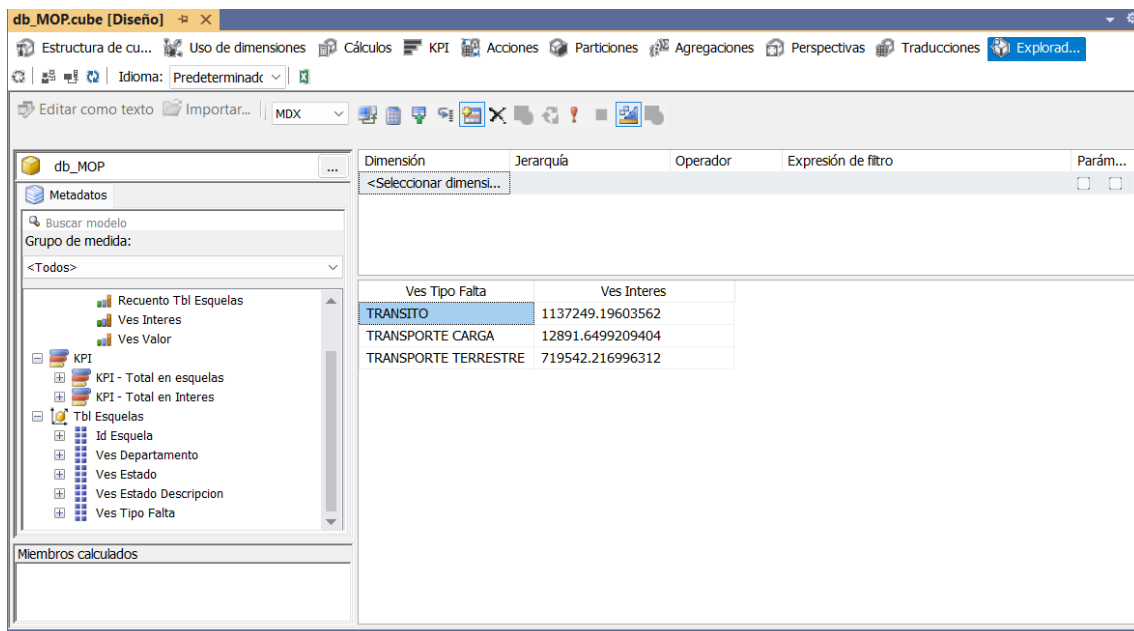


The screenshot shows the OLAP Explorer interface with the following data:

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

3. Tipos que generan más interés

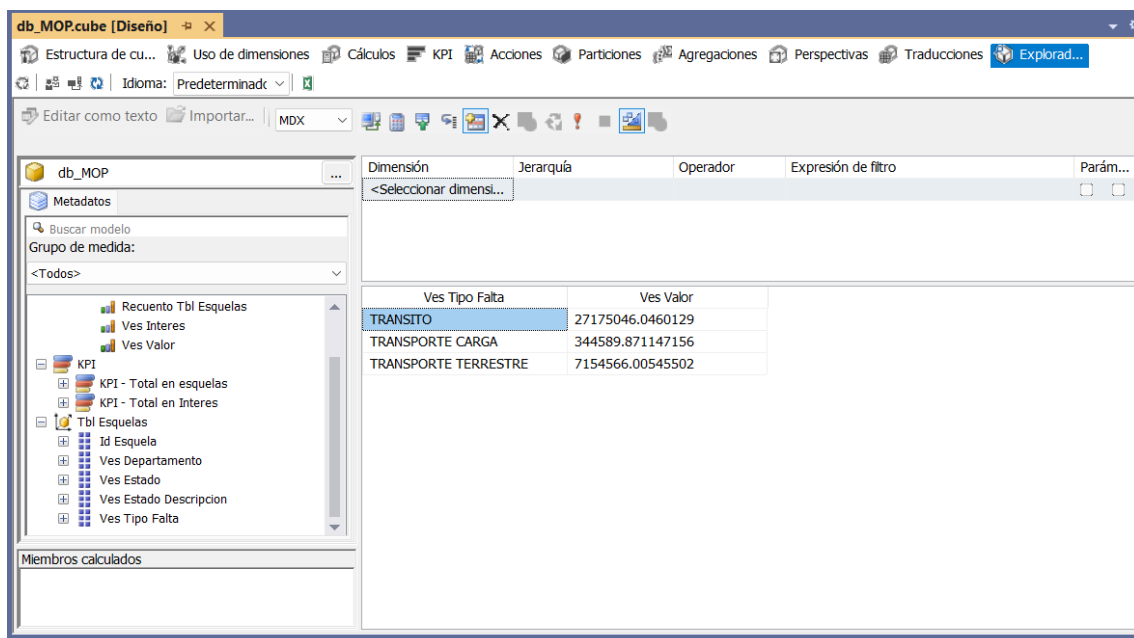
- En conclusión, a mayor valor del interés generado son los que menos pagan sus multas. Las multas de transito son las menos pagadas y esto tiene todo el sentido ya que son las que más se generan.
- Se recomienda crear un programa de facilidades de pago.



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

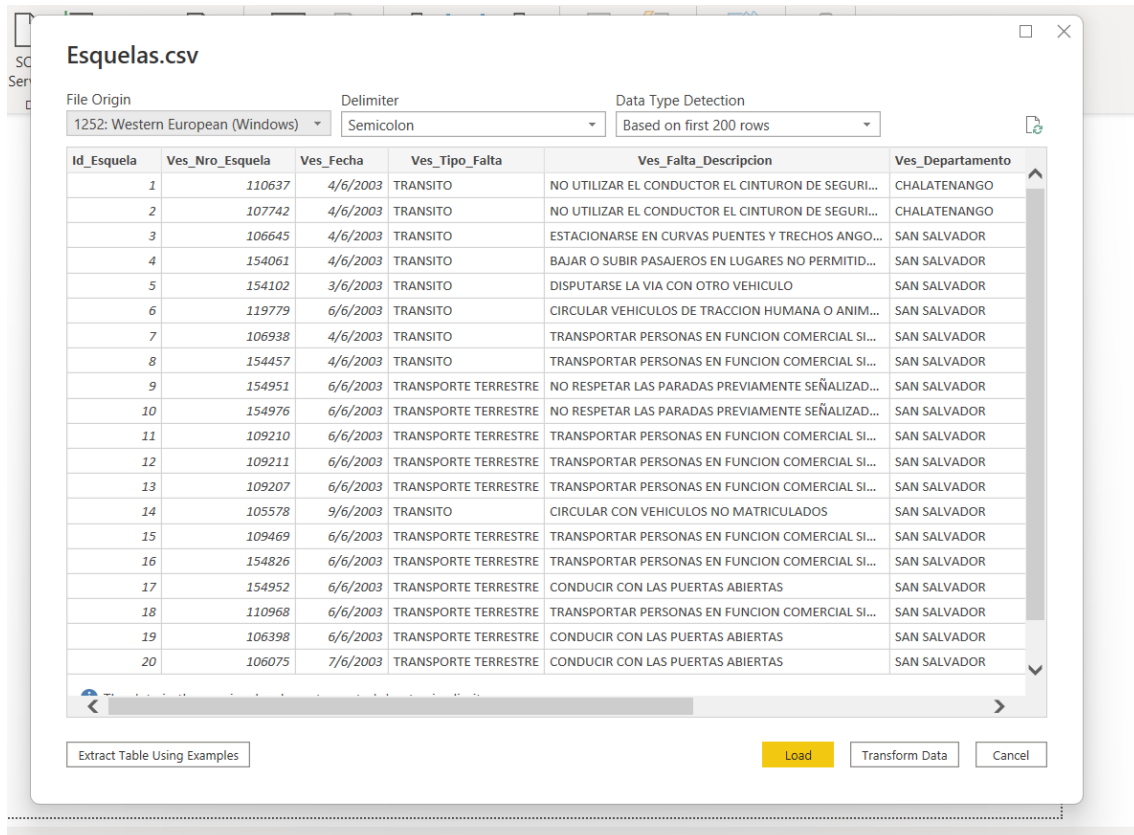
Miembros calculados

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

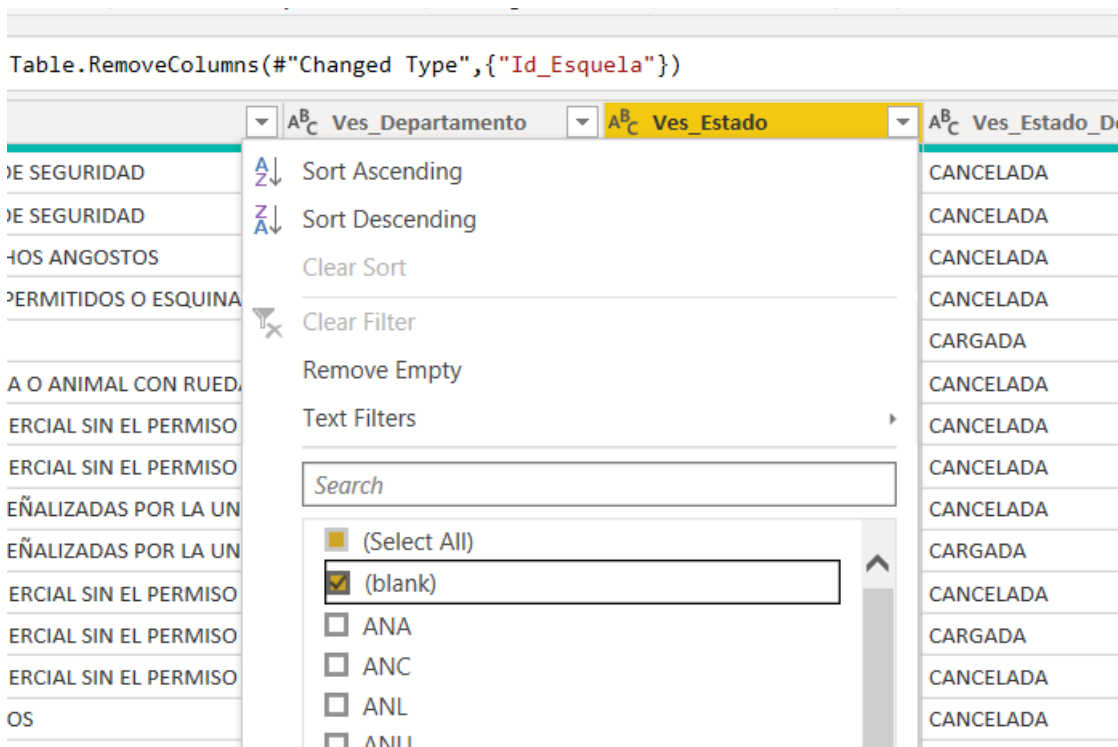
Ves Estado	Ves Estado Descripcion	Recuento Tbl Esquelas
ANA	ANALISIS APELADA	1
ANC	ANALISIS COBRADA	21
ANL	ANALISIS	13
ANU	ANULADA	199
APL	EN APELACION	360
CAR	CARGADA	2859
CBR	CANCELADA	805431
INC	INCONSISTENTE	87426
INI	PENDIENTE DE PAGO	113098
IPR	IMPROCEDENTE	27096
IPT	IMPUESTA	1687
IRI	RECURSO INCONFOR...	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	ICA DE CONDUCIR			Error	7.3
7	ICA DE CONDUCIR			Error	7.3
8	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
9	ICA DE CONDUCIR			Error	7.3
10	ICA DE CONDUCIR			Error	7.3
11	ICA DE CONDUCIR			Error	7.3
12	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
13	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
14	ICA DE CONDUCIR			Error	7.3
15	ICA DE CONDUCIR			Error	7.3
16	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
17	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
18	ICA DE CONDUCIR			Error	7.3
19	LICA PARA COMPETENCIAS AUTOMOVILISTICAS SI...			null	ni
20	ICA DE CONDUCIR			Error	7.3
21	ICA 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

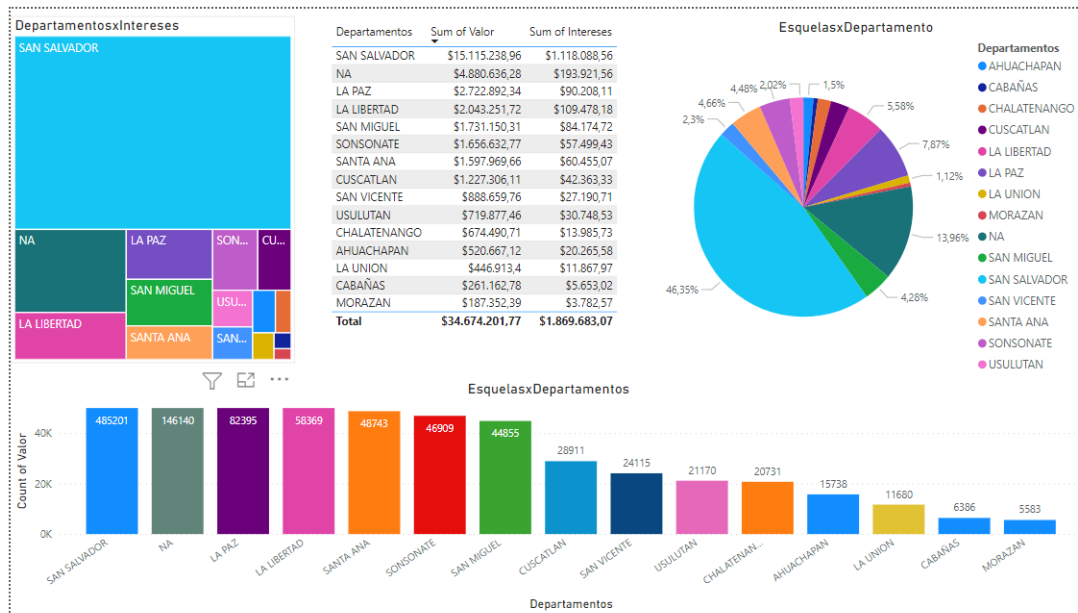
fx = Table.SelectRows("#Removed Columns", each ([Ves_Estado] = ""))

	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

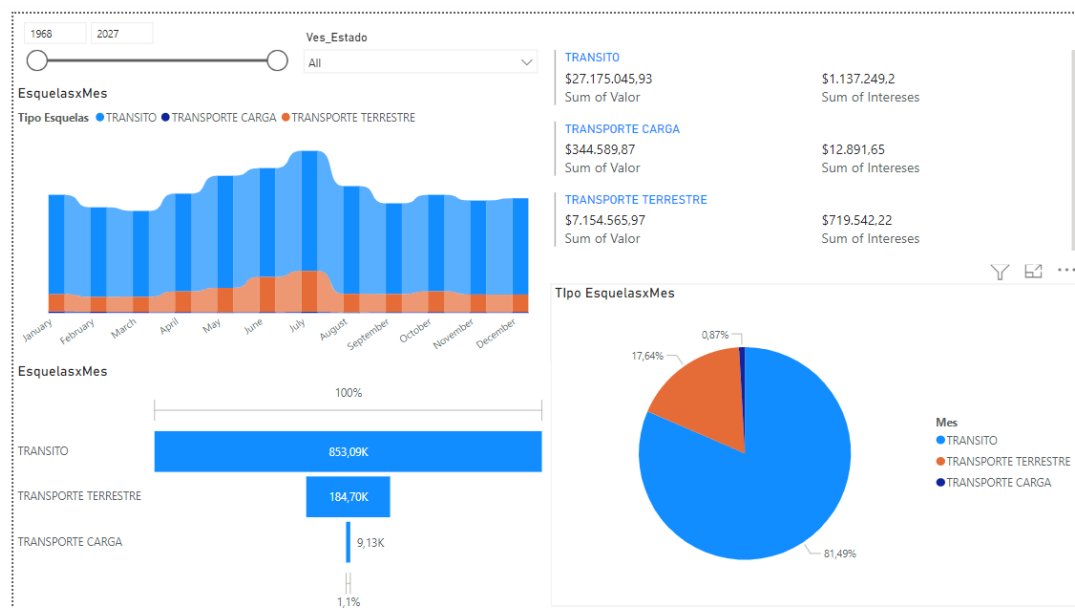
1. Tablero por departamento

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



2. Tablero por tipo de faltas

- La mayoría de las esquelos 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 por que existen tantas inconsistencias en los registros.

