

# Datawarehouse y Minería de datos (DMD941) G01T

# Proyecto parte 1.

# Integrantes:

- Kevin Alexander Fernández Monge. FM150385

- Patricia Elizabeth Mejia Rivera. MR141857

## **Docente:**

Ing. Karens Medrano.

# Índice

I)	Enunciado del proyecto	
٠,		
II)	Desarrollo de la actividad	4
ı	Parque vehícular	4
ı	Esquelas de infracción de tránsito	7
III)	Recursos.	13
	Publicación de reportes.	13
	Enlace de video de análisis de datos.	13

# I) Enunciado del proyecto.

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

- Esquelas de infracción de transito.
- 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
- Reglas de asociación
- Agrupamiento con k-means
- Arbol de decisión

Por ejemplo: Podria ser cubos y Árbol de decisión, para esquelas y reglas de asociación y agrupamiento con K-means para parque vehicular.

Los recursos son proporcionados en formato csv.

- Esquelas 18102018.csv
- Parque\_vehicular.rar

# II) Desarrollo de la actividad.

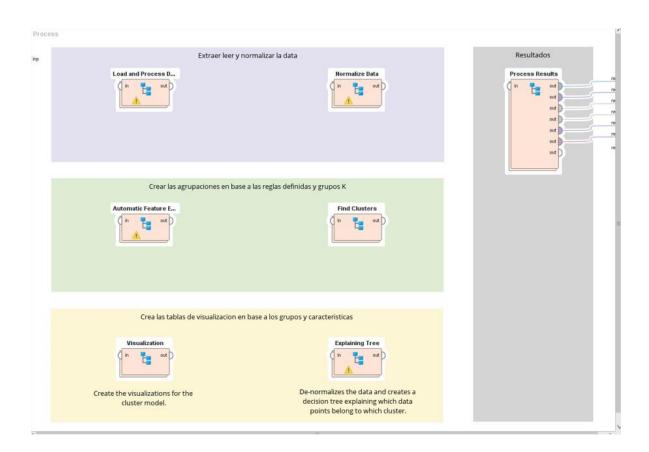
## Parque vehícular.

# • Estrategias utilizadas:

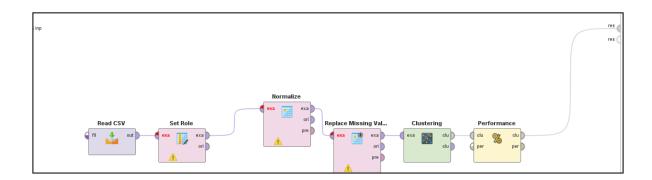
- 1. K Means
- 2. Reglas de asociación

#### • Proceso de análisis:

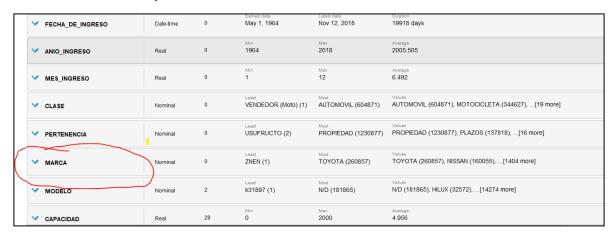
Se realizaron diversos ejercicios de agrupación y se aplicó K means grouping bajo los siguientes parámetros:



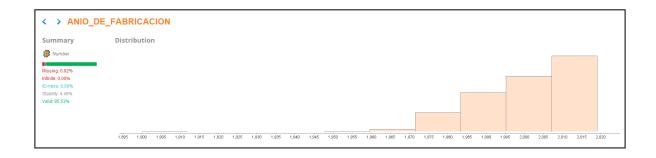
1. Año de ingreso: Con el objetivo de generar grupos que nos permitan visualizar los grupos de años donde ha existido mayor importación de vehículos y las características presentes en cada grupo de años, incluyendo marcas, modelos, tipos de combustible y valor.



**2. Marcas:** Con el objetivo de entender las características del parque vehicular desde la perspectiva de las marcas, incluyendo las características de cada modelo importado.



#### Resultados de análisis



De lo K5 grupos podemos ver que el mayor porcentaje del parque vehicular se fabrica entre el 2010 y 2020 por lo cual tenemos un parque vehicular mayormente moderno, sin embargo, mas del 50% del parque vehicular tiene más de 10 años de antigüedad.

### Algunas de las características por grupo son:

Cluster	ANIO_DE_FABRICACION	COMBUSTIBLE	CONDICION_INGRESO	ESTADO	PROPIETARIO_DEPARTAME	VALOR_DEL_VEHICULO
Cluster 0	1960-1970	Diesel	Vehículo usado	Baja	San Salvador	3285.80
Cluster 1	1970-1985	Diesel	Vehículo usado	Baja	San Salvador	3835.80
Cluster 2	1985-1995	Gasolina	Vehículo usado	Baja	San Salvador	4345.77
Cluster 3	1995-2005	Gasolina	Vehículo nuevo	Alta	San Salvador	7865.80
Cluster 4	2005-2020	Gasolina	Vehículo Nuevo	Alta	San Salvador	9665.80

## Por lo que podemos concluir:

- El precio de los vehículos incremento casi un 30% desde la importación de 1960
- La gasolina es ahora el tipo de combustible predominante sobre todo desde el año 2000.
- La mayor cantidad de propietarios se concentra en San Salvado para todos los grupos

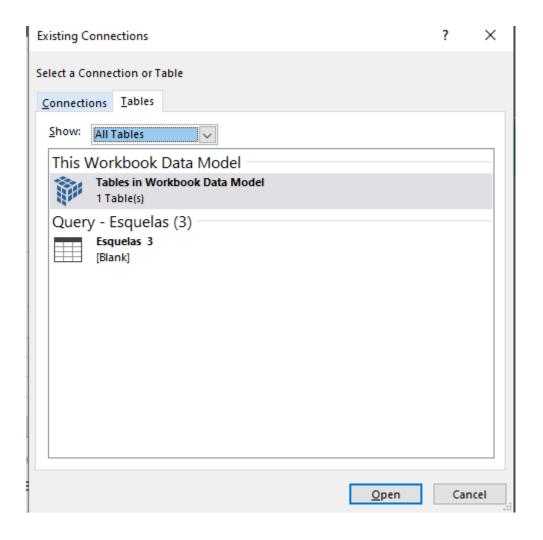
## Esquelas de infracción de tránsito.

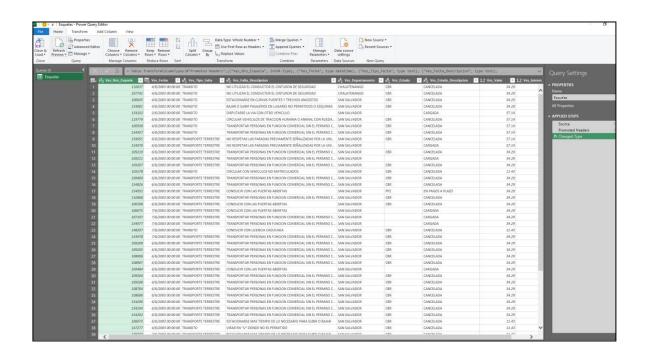
## • Estrategias utilizadas:

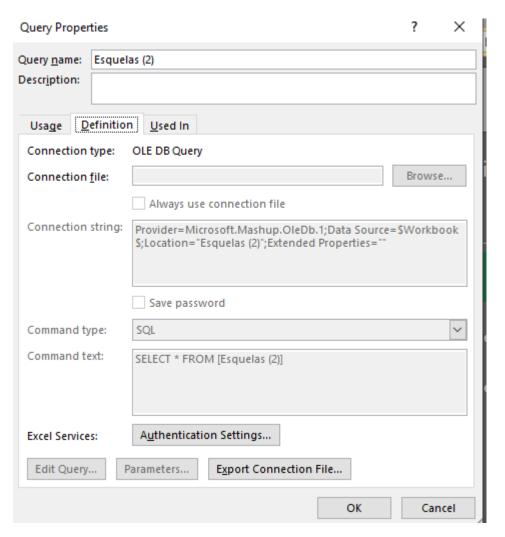
- 1. Cubos OLAP.
- 2. Árbol de decisión.

#### Proceso de análisis:

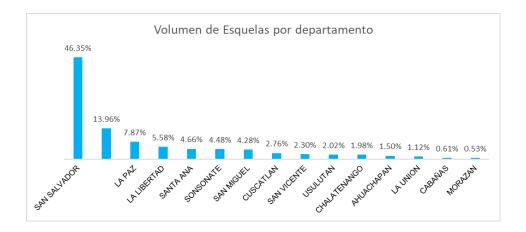
Se realizaron diversos ejercicios mediante la creación del cubo, incluyendo las tablas y medidas necesarias para el análisis de las variables.



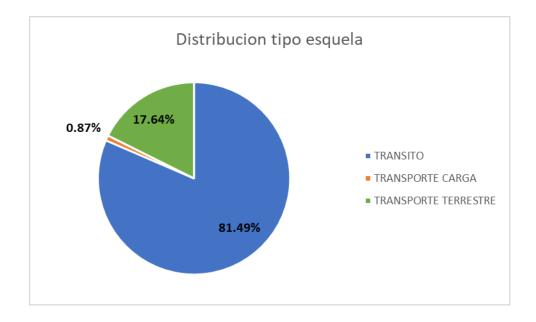




#### Resultados de análisis



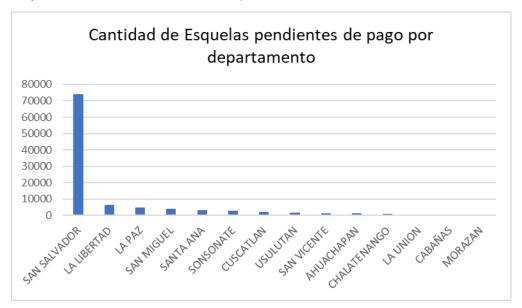
- 1. El 46% de las esquelas se generan en el departamento de San Salvador.
- 2. Existe un 14% que no se categoriza por departamento lo que reduce la calidad de los datos.

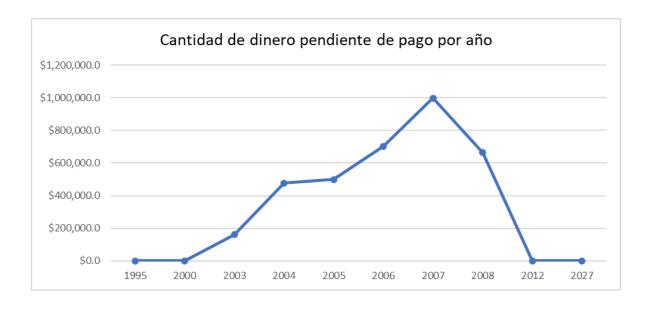


- 3. El 82% de las esquelas son de tránsito.
- 4. Menos del 1% de las esquelas son producidas por transporte de carga.

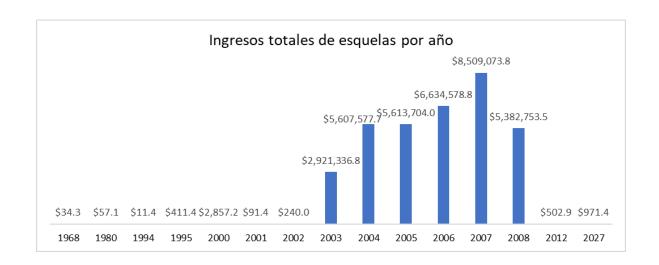


5. El 72% de las Esquelas de San Salvador (departamento con mayor cantidad de esquelas se encuentran canceladas).





- 6. San Salvador tiene más de 70000 esquelas pendientes de pago.
- 7. El año 2007 presenta la mayor cantidad pendiente por pagar.
- 8. Si se colecta la cantidad pendiente se recibiría un ingreso de: \$3,503,819.0



9. 2007 ha sido el año con mayores ingresos por esquelas.



10. De Mayo a Julio tienden a ser los meses con mayor cantidad de esquelas durante el año, por lo tanto podrían ser los meses en donde podrían o deberían hacerse campañas de prevención y concientización del buen manejo.

# III) Recursos.

# Publicación de reportes.

https://app.powerbi.com/links/NwHt1BabzA?ctid=f9afe020-14e8-4555-b638-b98f896aa94b&pbi\_source=linkShare

Enlace de video de análisis de datos.

https://www.youtube.com/watch?v=914dYsIb05Y