

**Datawarehouse y Minería de datos**

**(DMD941)  G01T**

**Proyecto fase 1**

**Integrantes:**

* Steffany Rocío Leonor Trejo LT172547
* Deysi Guadalupe De Paz Sosa DS150922
* Antonia Francisca Ayala González AG161316

**Docente:**

Ing. Karens Medrano.

Contenido

[1. Análisis Parque vehicular 3](#_Toc103809092)

[2. Análisis Esquelas 6](#_Toc103809093)

# Análisis Parque vehicular

**Estrategias utilizadas:**

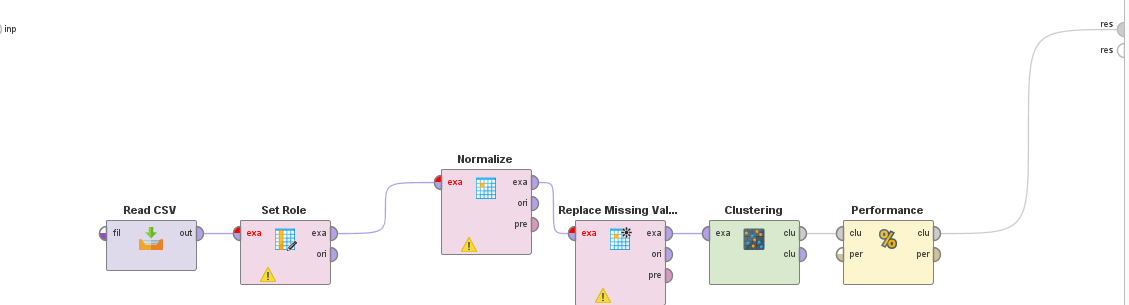
K Means

Reglas de asociación

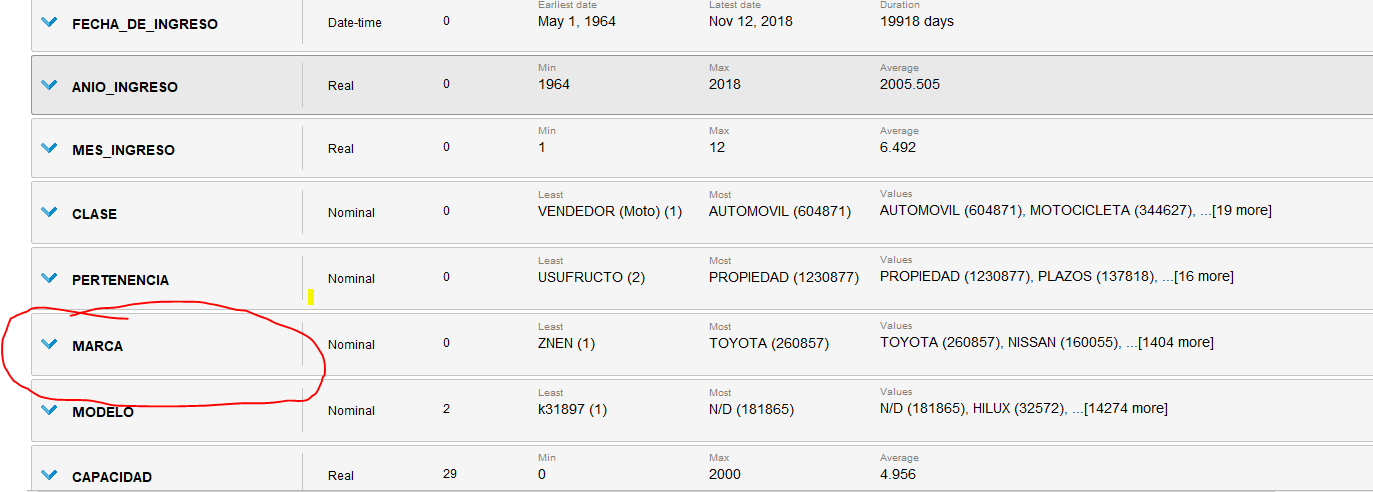
**Proceso de análisis:**

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

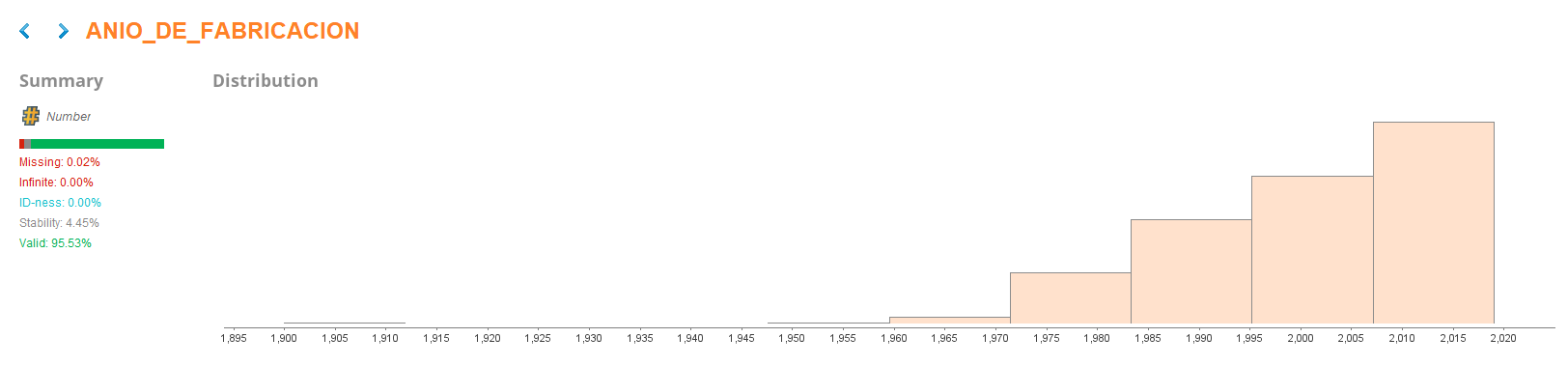
* **Año de ingreso:** Con el objetivo de generar grupos que nos permitan visualizar los grupos de años donde ha existido mayor importación de vehículos y las características presentes en cada grupo de años, incluyendo marcas, modelos, tipos de combustible y valor



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

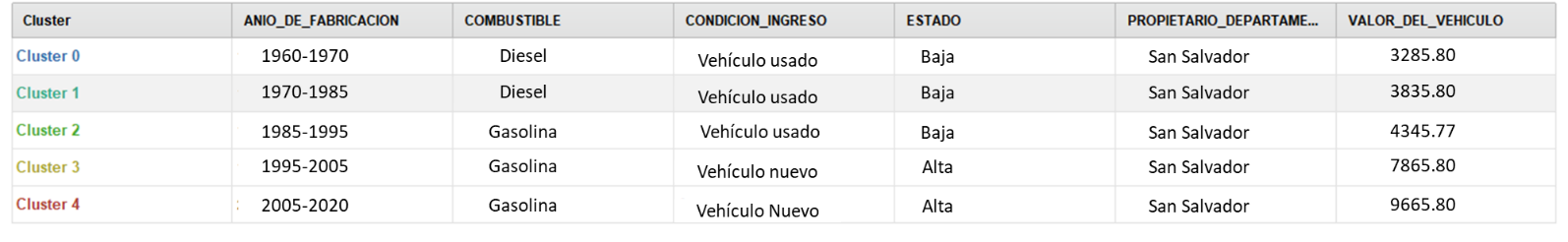


**Resultados de análisis**



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 mas de 10 años de antigüedad.

Algunas de las características por grupo son:



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

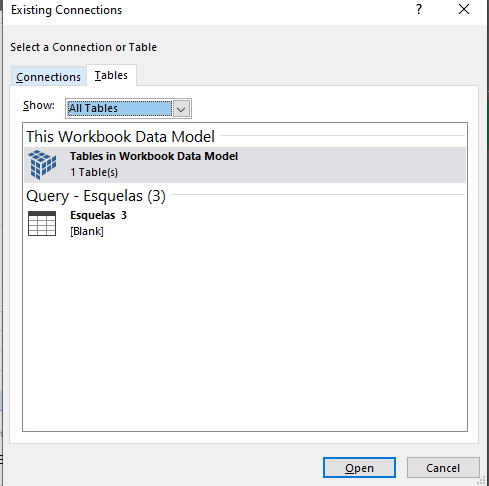
# Análisis Esquelas

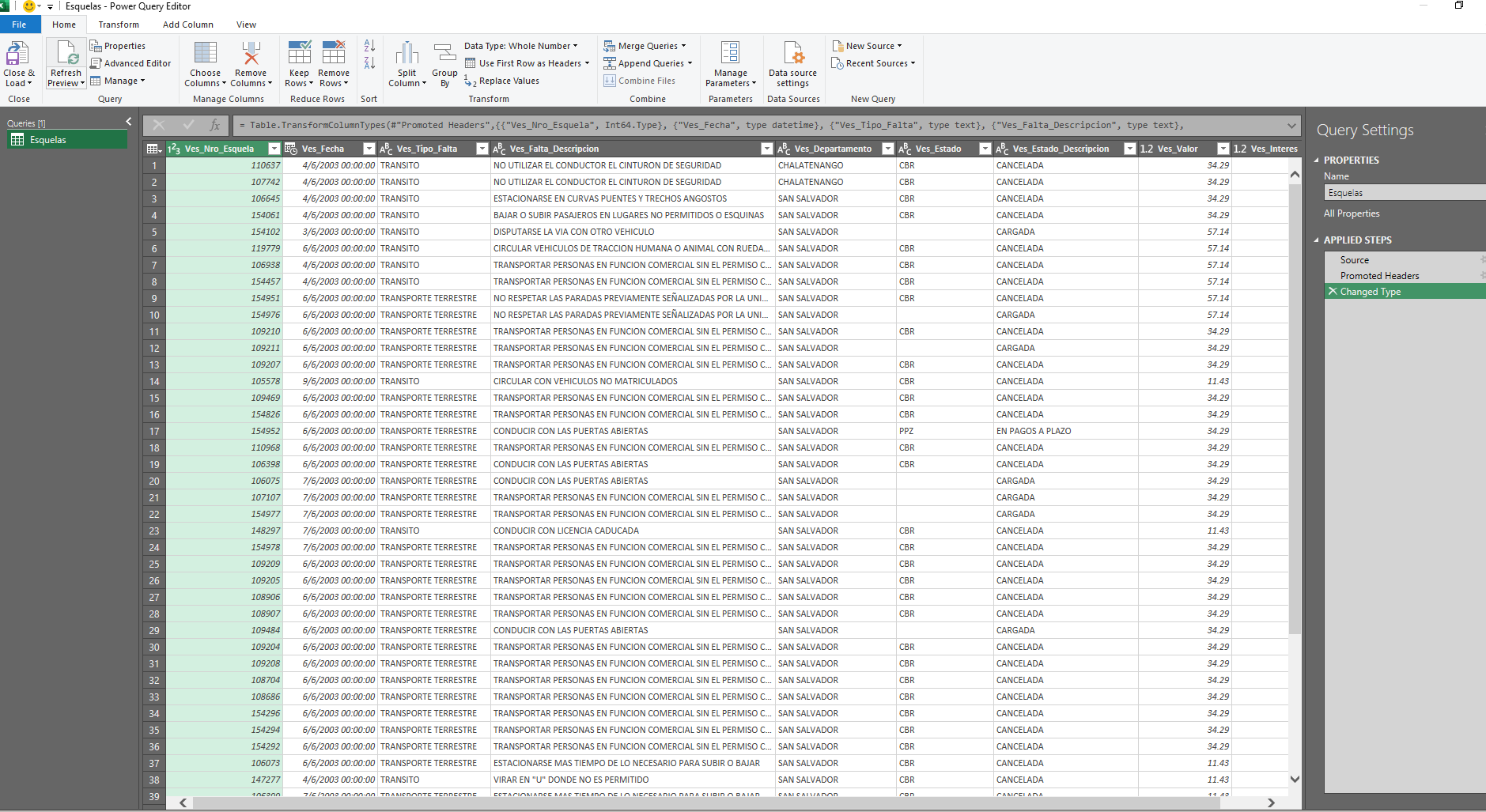
**Estrategias utilizadas:**

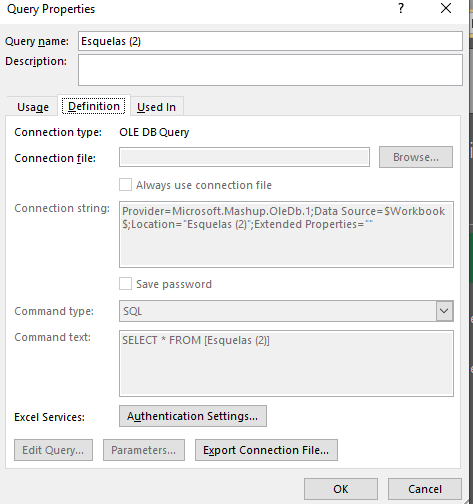
Cubos OLAP

**Proceso de análisis:**

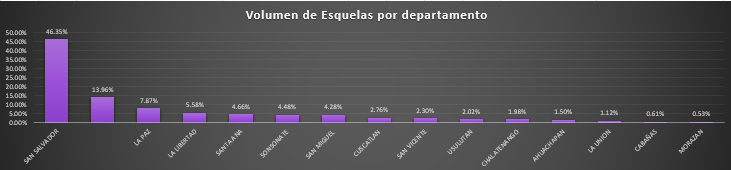
Se realizaron diversos ejercicios mediante la creación del cubo, incluyendo las tablas y medidas necesarias para el análisis de las variables



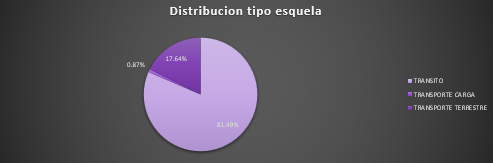




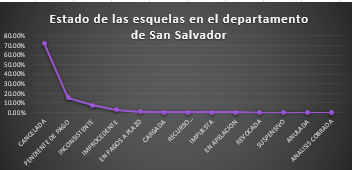
**Resultados de análisis**



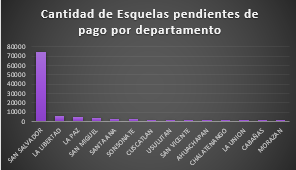
* 46% de las esquelas se generan en el departamento de San Salvador.
* Existe un 14% que no se categoriza por departamento lo que reduce la calidad de los datos.

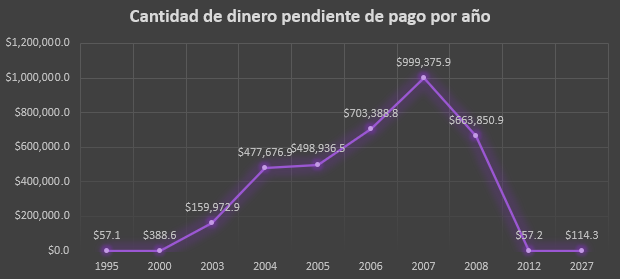


* El 82% de las esquelas son de tránsito.
* Menos del 1% de las esquelas son producidas por transporte de carga.



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

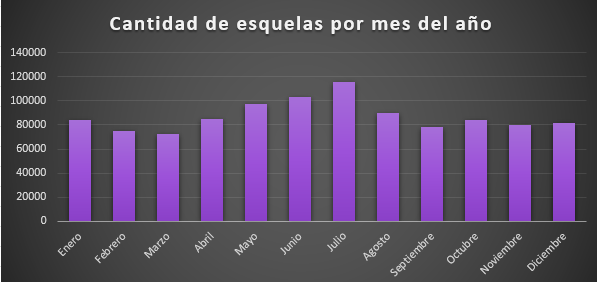




* San Salvador tiene mas de 70000 esquelas pendientes de pago
* 2007 es el año con que presenta la mayor cantidad pendiente por pagar
* Si se colecta la cantidad pendiente se recibiría un ingreso de: **$3,503,819.0**



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



* De mayo a Julio tienden a ser los meses con mayor cantidad de esquelas durante el año, por lo tanto, podrían hacerse campañas de prevención y concientización durante esos meses.

Link al GitHub: