

Universidad del Valle de Guatemala

Minería de Datos

Lynette Garcia Perez

Sección 10



## **Proyecto 1**

### **Proyecto 1 - Consumo de combustible**

Fabián Juárez, 21440

Diego Lemus, 21469

Brian Carrillo, 21108

Diego Alonzo, 20172

Carlos Lopez, 21666

Guatemala, 18 de marzo del 2024

## **Situación Problemática**

La industria automotriz se enfrenta a la problemática de equilibrar la demanda de vehículos con altos estándares de rendimiento y la urgencia de reducir las emisiones de gases de efecto invernadero para combatir el cambio climático. Aunque se presume que los vehículos modernos son más eficientes y ecológicos que sus predecesores, es esencial verificar y cuantificar esta tendencia para orientar las decisiones de consumidores, fabricantes y reguladores. La disponibilidad de modelos de vehículos más limpios y eficientes varía ampliamente, y los consumidores pueden no estar informados o no tener acceso a vehículos que se ajusten a estos estándares debido a barreras como el coste, la percepción de la tecnología o la falta de infraestructura de apoyo.

Los fabricantes de automóviles se encuentran bajo presión para innovar y producir vehículos que cumplan con normativas ambientales cada vez más estrictas, mientras se mantienen competitivos en términos de costos y características deseadas por los consumidores. Además, la ausencia de una tendencia clara hacia la eficiencia en los modelos más recientes podría indicar barreras en la adopción de tecnologías limpias, lo que pone en riesgo los esfuerzos para reducir el impacto ambiental del transporte.

## **Problema científico**

¿Existe una tendencia significativa hacia vehículos más eficientes y menos contaminantes en los modelos más recientes en comparación con los modelos más antiguos?

## **Objetivos**

### **General**

- Describir las tendencias en eficiencia de combustible y emisiones de vehículos para identificar características importantes en los diferentes modelos.

### Específicos

- Determinar una tendencia al agrupamiento y formación de clusters de los diferentes modelos de vehículos, para identificar características específicas que han mostrado mejoras significativas.
- Generar un modelo predictivo capaz de relacionar las características de los vehículos con el consumo de combustible y la emisión de dióxido de carbono que realiza.

## Descripción de los datos

Estas son las variables del primer dataset:

- Year: El año del modelo del vehículo. Indica la fecha de fabricación o el año del modelo al que pertenece el vehículo.
- Make: La marca del vehículo. Representa al fabricante o la compañía que ha construido el vehículo.
- Model: El modelo del vehículo. Refiere a la designación comercial que el fabricante da a una línea de vehículos de una cierta configuración y características.
- Vehicle Class: La clase del vehículo. Categoriza el tipo de vehículo según su tamaño, uso y forma.
- Engine Size: Tamaño del motor del vehículo expresado en litros (L). Se refiere al volumen total de todos los cilindros del motor combinados.
- Cylinders: Número de cilindros del motor. Determina la cantidad de cilindros que tiene el motor.
- Transmission: El tipo de transmisión del vehículo. Puede ser automática (A), manual (M), automática con cambio manual seleccionable (AS), de cambio manual automatizado (AM), o variable continua (AV), junto con el número de velocidades o marchas.
- Fuel: Tipo de combustible que utiliza el vehículo. Puede ser gasolina regular (X), gasolina premium (Z), diésel (D), etanol (E85) o gas natural (N).
- Fuel Consumption: Consumo de combustible del vehículo. Se muestra en litros por cada 100 kilómetros (L/100 km) y se da como un valor combinado basado en un porcentaje ponderado de conducción en ciudad y carretera.

- HWY (L/100 km): Consumo de combustible en carretera. Muestra cuántos litros de combustible utiliza el vehículo para recorrer 100 kilómetros en un entorno de carretera.
- COMB (L/100 km): Consumo de combustible combinado. Es un promedio ponderado del consumo de combustible en ciudad y en carretera.
- COMB (mpg): Consumo de combustible combinado expresado en millas por galón imperial. Es una medida alternativa al consumo de combustible combinado que se utiliza en algunos países.
- Emissions: Emisiones de CO2 estimadas del tubo de escape del vehículo. Se mide en gramos por kilómetro y se basa en el tipo de combustible y el consumo de combustible combinado del vehículo.

### Operaciones de limpieza

Se llevó a cabo una revisión de los datos para ver si existían valores nulos en las diferentes columnas que se utilizarían en el dataset. Se encontró que no se posee ningún valor nulo en el dataset.

```
[ ]
Empty DataFrame
Columns: [YEAR, MAKE, MODEL, VEHICLE CLASS, ENGINE SIZE, CYLINDERS, TRANSMISSION, FUEL, FUEL CONSUMPTION, HWY (L/100 km), COMB (L/100 km), COMB (mpg),
Index: [ ]
0
22556
```

### Análisis exploratorio

Link al repositorio

[https://github.com/FabianJuarez182/Proyecto\\_Analisis\\_Exploratorio\\_MD.git](https://github.com/FabianJuarez182/Proyecto_Analisis_Exploratorio_MD.git)

Link al repositorio

[https://www.canva.com/design/DAF\\_5f8O-OE/hZ8fE6483oGIEmC8MkIXw/edit?utm\\_content=DAF\\_5f8O-OE&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAF_5f8O-OE/hZ8fE6483oGIEmC8MkIXw/edit?utm_content=DAF_5f8O-OE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

## Hallazgos

En el proyecto se identificó una relación directa entre el consumo de combustible y las emisiones de CO<sub>2</sub>, indicando que a mayor consumo de combustible, mayor será la emisión de carbono.

La formación de clusters reveló agrupaciones específicas de vehículos que comparten características similares en términos de eficiencia y emisiones, lo cual es crucial para entender las tendencias y orientar las acciones futuras. Estos hallazgos subrayan la importancia de políticas dirigidas y la innovación tecnológica en la industria automotriz para acelerar la transición hacia un transporte más sostenible.

Se evidenciaron 2 temas clave al momento de llevar a cabo el clustering:

1. Eficiencia en el consumo de combustible por clase de vehículo:

Esta se basa en 2 grupos:

- a. **Grupo con alta eficiencia:** Incluye a las clases 'COMPACT', 'MID-SIZE' y 'SUBCOMPACT', estos presentan los menores consumos promedio de combustible, teniendo así una mayor eficiencia.
- b. **Grupo de Baja eficiencia:** Conformado por 'VAN-PASSENGER', 'SUV-STANDARD', y 'PICKUP TRUCK - STANDARD', estos vehículos muestran consumos de combustible más altos, siendo estos los menos eficientes.

2. Emisiones de CO<sub>2</sub> por clase de vehículo:

Este se basa en 2 grupos:

- a. **Grupo de Bajo Impacto Ambiental:** Vehículos como 'COMPACT', 'MID-SIZE' y 'SUBCOMPACT' con las emisiones más bajas, sugiriendo una mayor amigabilidad ambiental.
- b. **Grupo de Alto Impacto Ambiental:** Incluye a 'PICKUP TRUCK - STANDARD', 'VAN - PASSENGER' y 'SUV - STANDARD', con emisiones promedio más altas, indicando menor eficiencia en términos de emisiones.

## Conclusiones

Respecto al análisis exploratorio y el clustering realizado es posible concluir lo siguiente:

- Existen correlaciones positivas fuertes entre las características técnicas de los autos, y las emisiones de CO<sub>2</sub> producidas. Es posible inducir que estas características pueden llegar a ser buenos predictores para la emisión respectiva de un modelo no registrado.
- En la muestra la mayor parte de los automóviles utilizados son de tamaño normal y grande, por lo que la mayor parte de los puntos atípicos pertenecen a autos de tamaño pequeño. Esto puede llegar a repercutir en los resultados de modelos posteriores.
- Los grupos obtenidos durante el proceso de clustering comparten similitudes en cuanto a mayor o menor emisión de CO<sub>2</sub>, consumo de combustible, tamaño de motor, cantidad de cilindros, etc. Esto refuerza la idea de que las emisiones de CO<sub>2</sub> pueden ser predichas a partir de las características de los diferentes modelos de automóviles.