

**TASK**

**Exploratory Data Analysis on the Automobile Data Set**

[](https://www.hyperiondev.com/)

**Introduction**

Automobile data set containing 26 columns and 206 rows of information. I imported varies libraries to assist my analysis and loaded the automobile.txt file into Jupyter.

**DATA CLEANING**

In summary I carried out the following:

* I identified columns that had a ? and replaced with the mean average
* I converted data types to integers
* I converted non- numeric to null

**MISSING DATA**

There was no missing data in this data frame

**DATA STORIES AND VISUALISATIONS**

In summary I carried out the following:

Doors and body style

* I looked at the number of doors in the data frame and found 4 doors was the most popular with around 60% and 40% for two doors
* I did a comparison of body style, number of doors and loses. My analysis revealed that normalised losses are distributed across different body style but the two door cars has more number of losses than the four door cars in this data frame.
* I produced a pie chart to visualise vehicle body styles across the whole dataset and found sedan was most popular with 46.3% and second was hatchback with 34.5%.

Wheels

* I analysed the wheel transmission data and found forward drive vehicles were the most popular
* I analysed the wheel transmission data vs city MPG and found forward drive vehicles were the most popular
* I analysed the wheel transmission data vs highway MPG and again found forward drive vehicles were the most popular
* I analysed the wheel transmission data vs price and found rear wheel drive vehicles were the most expensive to buy

Manufacturers

* I analysed the number of cars produced by manufacturers and found Toyota the highest number of vehicles followed by Nissan.
* I analysed the manufacturers vs price and found Mercedes Benz is the most expensive car and the least expensive is Chevrolet. The manufacturer with have the highest list price are Porsche, Jaguar, BMW and Mercedes. The manufacturer with the least list price are Plymouth, Honda, dodge and Chevrolet.
* I analysed the manufacturers vs fuel type and toyota has the highest produced gas fuel type cars and 4th highest diesel produced cars. Peugot produces the most diesel cars closely followed by volkswagon and mercedes-benz. There is no mention of hybrid or electric fuel types which would suggest this data frame is at least 10 years old from current day (early 2023).
* I analysed the manufacturers vs aspiration and found that both mitsubishi and peugot manufacturers have the highest counts for std and turbo aspirations in their vehicles.

Engines

* I analysed the number of cylinders in a pie chart and found four cylinders was most popular with 77.3% of vehicles in this data frame.
* I also created a seaborn relplot to visualise the engine size and number of cylinders.

**THIS REPORT WAS WRITTEN BY: LOUISE RANDALL**

