

**TASK**

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

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

**Introduction**

This dataset contains automobile information, including details of automobile makes and details of the cars such as the horsepower, engine size, number of doors.

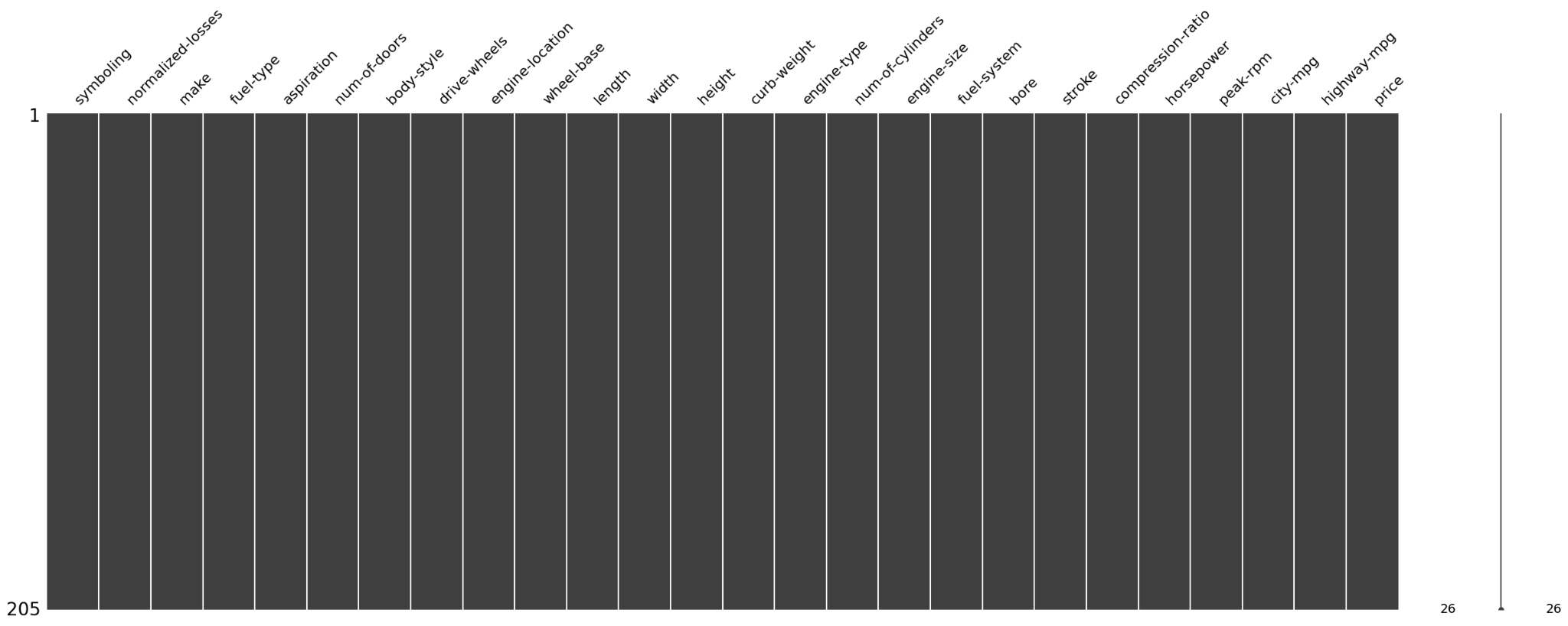
**DATA CLEANING**

Upon loading the dataset, and using .head() to identify columns, from the columns, I identified 2 to be removed that won’t be required, these are aspiration and symboling, using the drop method.

Using dtypes to identify each data type each column is and converting horsepower to a numeric type.

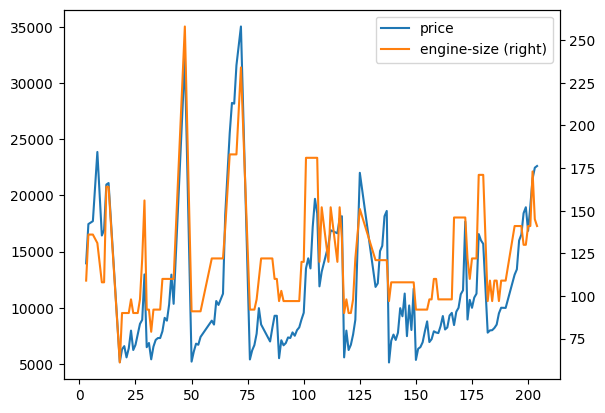
**MISSING DATA**

Using the missingno.matrix function, I have managed to identify that there aren’t any rows after dropping the previous columns and dropping any rows that contain a ‘?’, the below matrix displays after removing the missing data. I decided to remove the data as I felt it wasn’t possible to guess what data should be input into the row.

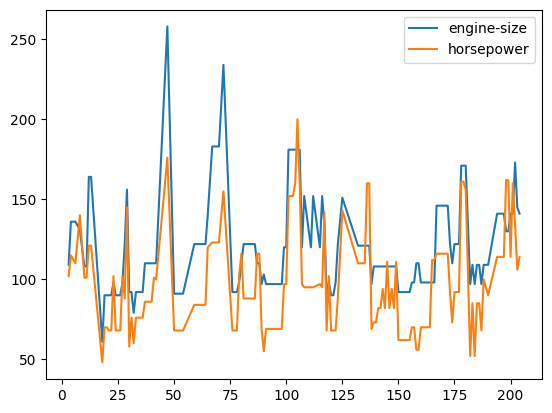


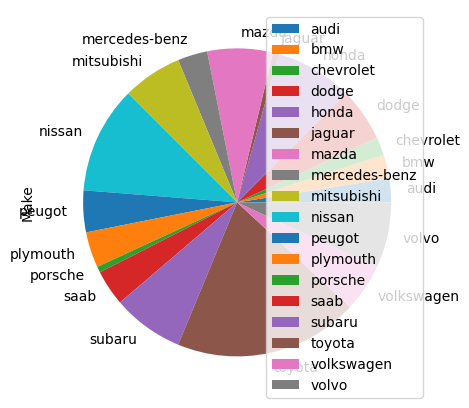
**DATA STORIES AND VISUALISATIONS**

The first visualisation is comparing the price of the automobile to the engine size and whether a higher price means that we get a higher engine size.

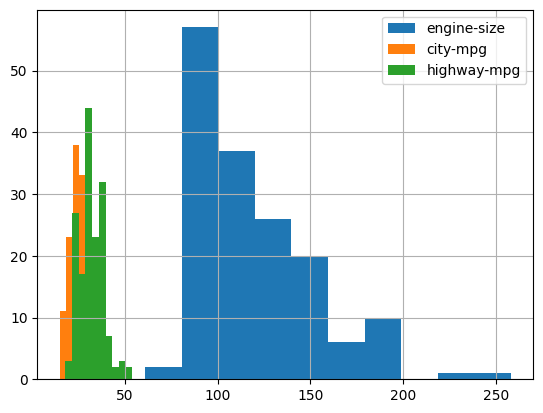
Based on the below graph it shows that we do receive a higher engine size when the price is higher

Based on from the previous question, does a higher price also mean higher horsepower, using the graph below it does show that more horsepower is provided with a higher price.

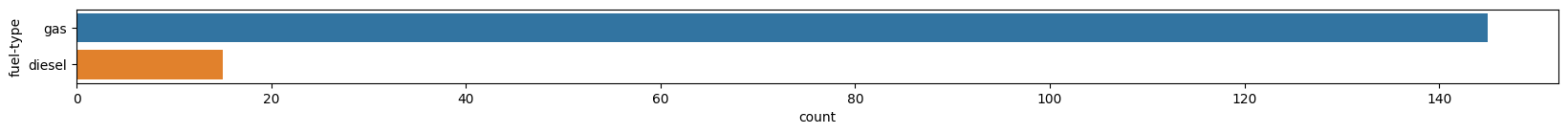
  
Here we can identify that the Toyota automobile brand is the most common



In the below graph, we are comparing the engine sizes to the mpg of city and highway. Here we can see that having a low - mid sized engine obtains a higher mpg within the city / highway.



The last comparison, compared the the difference between the fuel type to identify the most common, here we can see that Gas is the most common fuel type.

****

**THIS REPORT WAS WRITTEN BY : Chris Taylor**

