

**TASK**

**Exploratory Data Analysis on the Automobiles Data Set**

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

**SUMMARY OF THE DATA SET**

The automobiles data set comprises of 205 rows with values entered for each of the 26 columns including symbolling, normalized-losses, make, fuel-type, aspiration, num-of-doors, body-style, drive-wheels, engine-location, wheel-base, engine-size, fuel-system, bore, stroke, compression-ratio, horsepower, peak-rpm, city-mpg, highway-mpg, price etc.

**DATA CLEANING**

The practice of removing duplicate rows is applied, however the dataset shown the same number of rows from the beginning, hence the dataset had no duplicate rows. Question mark values were replaced with null values.

**MISSING DATA**

The highest number of missing values was found in the normalised losses column. Consequently, the normalised losses column was dropped. Rows with missing values will be excluded in the analysis. Nevertheless, where horsepower missing, peak-rpm was also missing and similarly for bore and stroke.

**DATA STORIES AND VISUALISATIONS**

Value counts per make and null value counts per make showed that all the rows with Renault value for make from the dataset have at least one missing value, half the rows with Isuzu have at least one missing value.

The pie chart shows the representation share of each car make. Alpha Romeo has the biggest share while Volvo has the smallest.

Chart, pie chart

Description automatically generated

The bar chart shows that Jaguar has the highest average price while Chevrolet has the smallest, given the entries in the data set.

Chart, histogram

Description automatically generated

Scatterplot visualisation was performed to analyse whether there is a relationship between the variables with the same number of missing values. No clear correlations between bore and stroke, and between horsepower and peak-rpm.

**Chart, scatter chart

Description automatically generated**

**Chart, scatter chart

Description automatically generated**

Finally, a scatterplot of engine-size and mpg differentiated by colour for city and highway showed that an engine on the highway can carry out more mpg than in the city for the same engine size.

Chart, scatter chart

Description automatically generated

**THIS REPORT WAS WRITTEN BY: ROBERTA DOBRA**

**STUDENT NUMBER: RA23010006214**

