

**TASK**

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

[](http://www.hyperiondev.com/portal/)

**Introduction**

The automobile dataset contains information about cars. The data consist of 25 columns with 205 records. In this dataset, the variable “price” can be regarded as the target. The other variables will be investigated to find the relationship between them and the target variable. EDA can show us how these features affect the market value price of a car.

**DATA CLEANING**

In cleaning this dataset, we start by replacing “?” with NaN value the columns/attributes. The second step was to use the “isnull“ method to display the number of all the missing data points (NaN values) per attribute. To show the missing data points in diagrammatically we used the “ missingno.matrix” method and “heatmap”. The visualization techniques showed that normalized-losses has more missing points than other attributes.

MISSING DATA

The attributes that have missing data points are:

Normalized-losses, num-of-doors, bore, stroke, horsepower, peak-rpm, and price.

The mean method was used to fill the missing data points for normalized-losses, bore, stroke, horsepower, peak-rpm, and prize.

The number of doors was solved by relating the number of doors with body-style attribute. If more hatchback cars have two doors, then any hatchback car with missing num-of-doors is more likely to have the same number of doors.

DATA STORIES AND VISUALIZATIONS

From the automobile dataset, we can try to answer questions such as:

# Which is the most expensive car make?

# How does the engine-size affect the price of a car?

# Body-style in relation to the number of doors

# Which is the most popular make

Visualizations shows that:

# Most cars are four-doors. Sedan has the highest number of four-doors, while most two-doors are hatchbacks.

# The most popular fuel type is a gas, with a percentage of 81.95% . The turbo has 18.05%.

# The scatter plot shows that the engine-size is directly proportional to the price of a car.

# Toyota is the most popular car make, followed by Nissan and Mazda

# Most popular drive-wheels are front wheels.

**THIS REPORT WAS WRITTEN BY : NJABULO MBANJWA**

