

**TASK**

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

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

**Introduction**

The dataset in question pertains to car specifications and prices. This data has been collected from various car dealerships and consists of features like make, model, horsepower, number of cylinders, fuel type, etc. The primary goal is to analyse and understand the relationship between various features and the price, as well as explore other hidden patterns within the data.

**DATA CLEANING**

Duplicated Entries: There were some cars with duplicate specifications. These duplicates were removed to maintain the readability of the data.

Unrealistic Data: Some cars had values such as 0 horsepower or extremely high prices. I employed box plots to identify these outliers and address them.

Formatting Issues: The names of some cars had typos or different formats (BMW vs B.M.W). I standardized the naming of cars across the dataset.

**MISSING DATA**

Missing data (?) was dealt with by cleaning the ‘?’ points and replacing them with 0 values. This was done using the .dropna() command.

**DATA STORIES AND VISUALISATIONS**

1. MPG Ratio for most expensive vs cheapest car:

Most expensive car:

make price

Mercedes-Benz 35056.0

price\_to\_city\_mpg: 2191.000000.

price\_to\_highway\_mpg: 1947.555556

Cheapest car:

make price

Subaru 5118.0

price\_to\_city\_mpg: 165.096774

price\_to\_highway\_mpg: 142.166667

Insight: The Subaru, although being much cheaper, was more fuel efficient in both the city and highway in terms of miles per gallon.

1. The 5 fastest cars according to horsepower:

Make: Horsepower:

Nissan 200.0

Jaguar 176.0

Volvo 162.0

Volvo 162.0

Toyota 161.0

1. I created a bar plot graph to show the relationship between the number of cylinders and horsepower.

Insight: From the bar plot I saw that cars with more cylinders generally produced more horsepower.

1. Visualization 1: Scatter Plot - Price vs. Horsepower

Insight: From this visualization, I observed that there is a positive correlation between horsepower and price for many makes. This indicates that as the horsepower of a car increases, its price also tends to increase, possibly because higher horsepower cars are marketed as luxury or sports vehicles.

1. Visualization 2: Box Plot - Distribution of Prices across Body Styles

Insight: This visualization showed the spread of prices for each body style. Sedans or convertibles had a broader range of prices, indicating that they cater to both the medium budget and luxury markets. In contrast, hardtops and hatchbacks had a much narrower price distribution, indicating that they are most likely budget cars.

1. Visualization 3: Count Plot - Number of Cars by Make

This plot showed us the number of cars for each type of make. What we saw was that Toyota lead the pack with almost twice as many cars as every other producer. This could indicate that Toyota is the most popular or mainstream manufacturer which produces a wide variety of models. It could also indicate that Toyota creates a wide variety of cars ranging from budget to luxury.

**THIS REPORT WAS WRITTEN BY: RAMEEZ BHAYED**

