## Indian Institute of Information Technology Surat

****

# Lab Report on

# Machine Learning (CS 601) Practical

**Submitted by**

### [RAHUL KUMAR SINGH] (UI21CS44)

**Course Faculty**

### Dr. Pradeep Kumar Roy

### Dr. Rajesh K. Ahir

## Department of Computer Science and Engineering

## Indian Institute of Information Technology Surat

## Gujarat-394190, India

**Jan-2024**

## Lab No: 2

**Aim:**

To perform exploratory data analysis on the attached dataset

**Description:**

Perform the Exploratory Data Analysis (EDA) by considering the following tasks. Use the attached dataset for the same.

1. Check for Duplication

2. Missing Values Calculation

3. Data Reduction (Some columns or variables can be dropped if they do not add value to our analysis.)

4. Feature Engineering

5. Creating Features

6. Data Cleaning/Wrangling

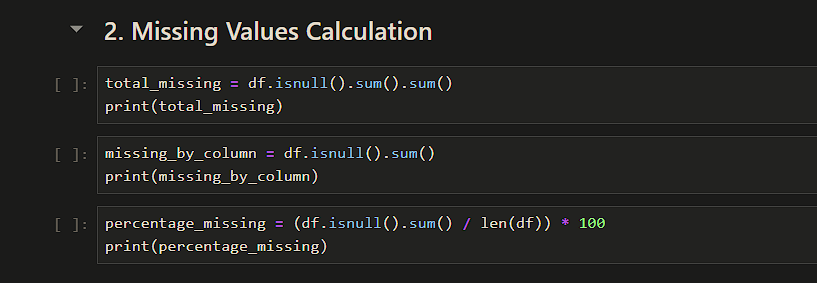
7. Statistics Summary (Count, Mean, Standard Deviation, median, mode, minimum value, maximum value, range, standard deviation)

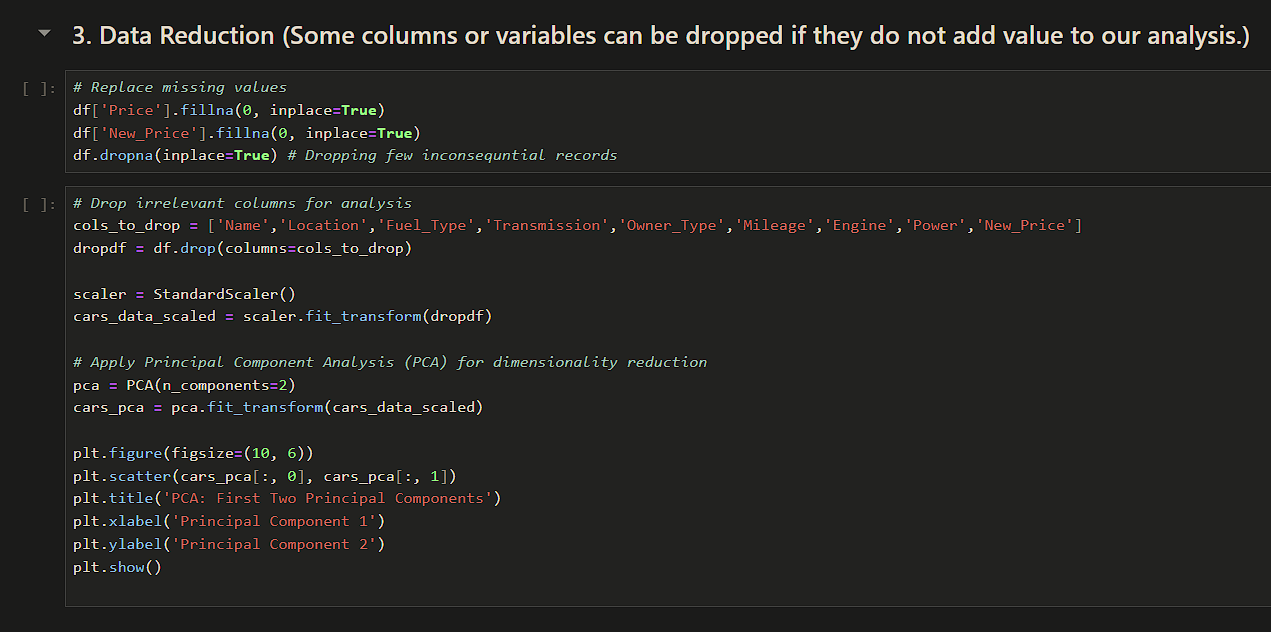
8. Analyzing/visualizing the dataset by taking one variable at a time

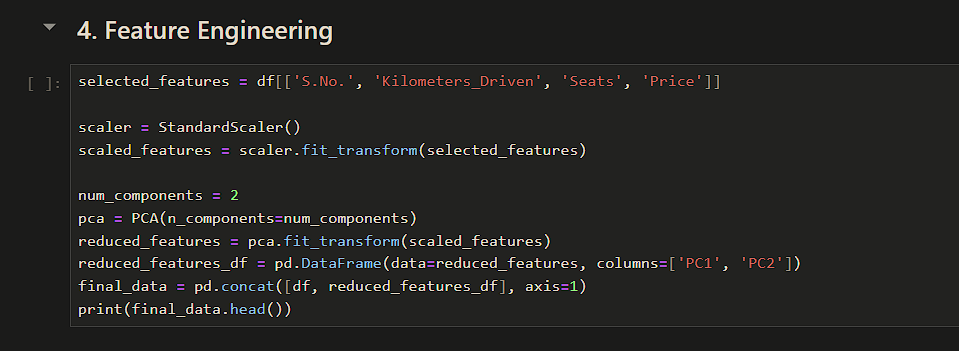
9. Data Transformation

## Source Code:

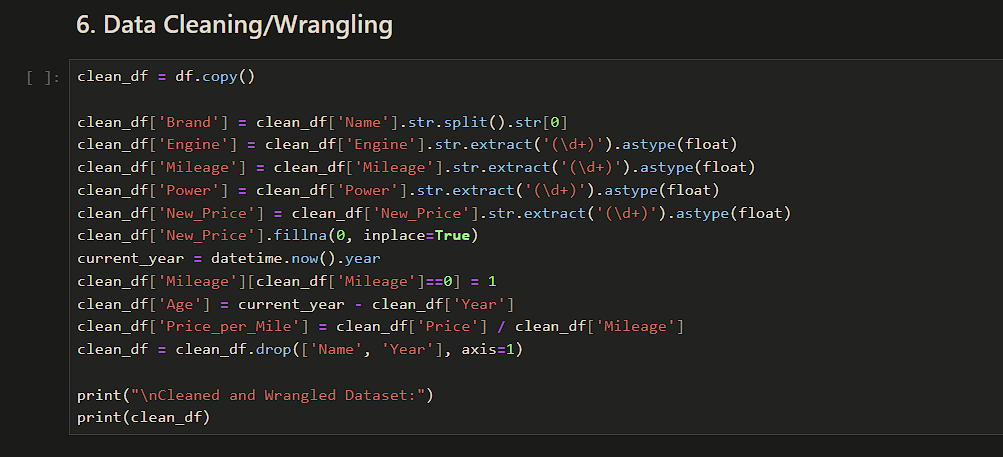
****

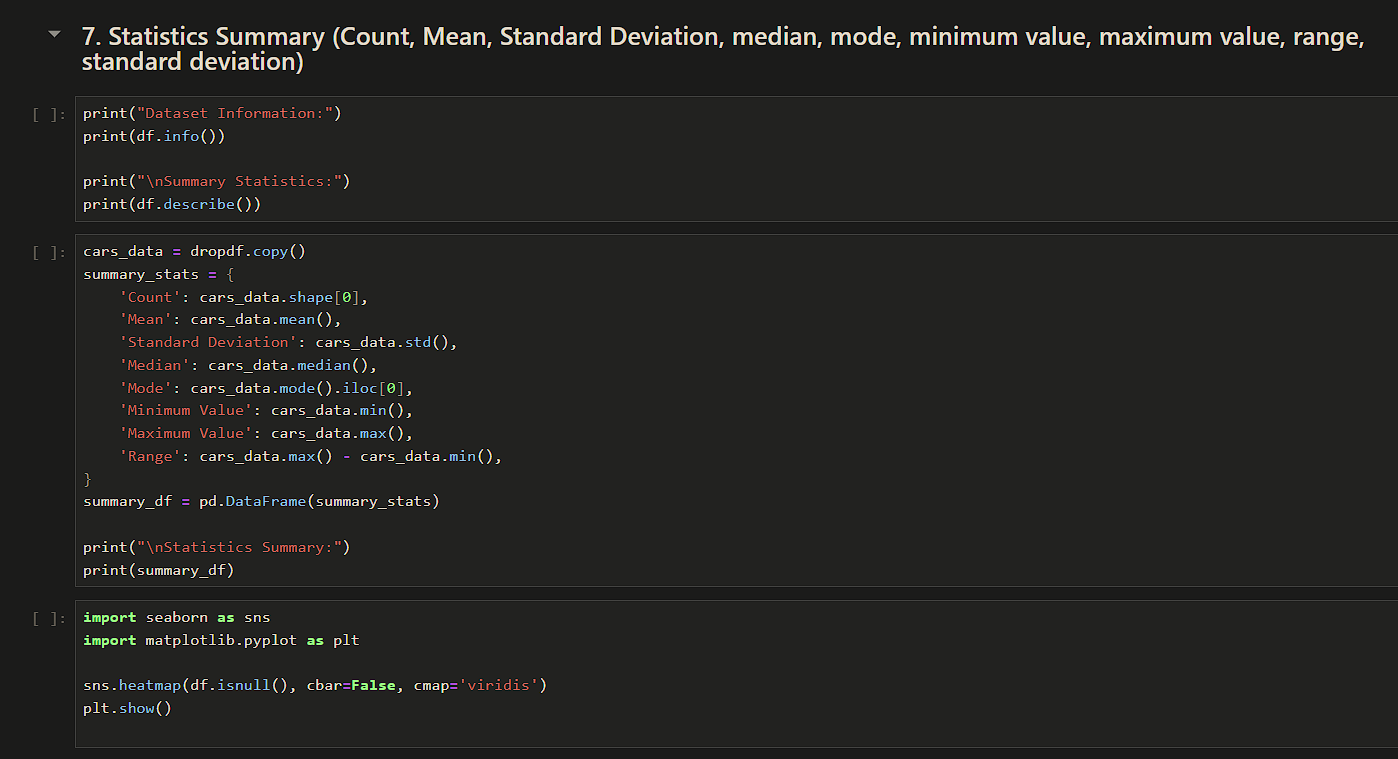
****

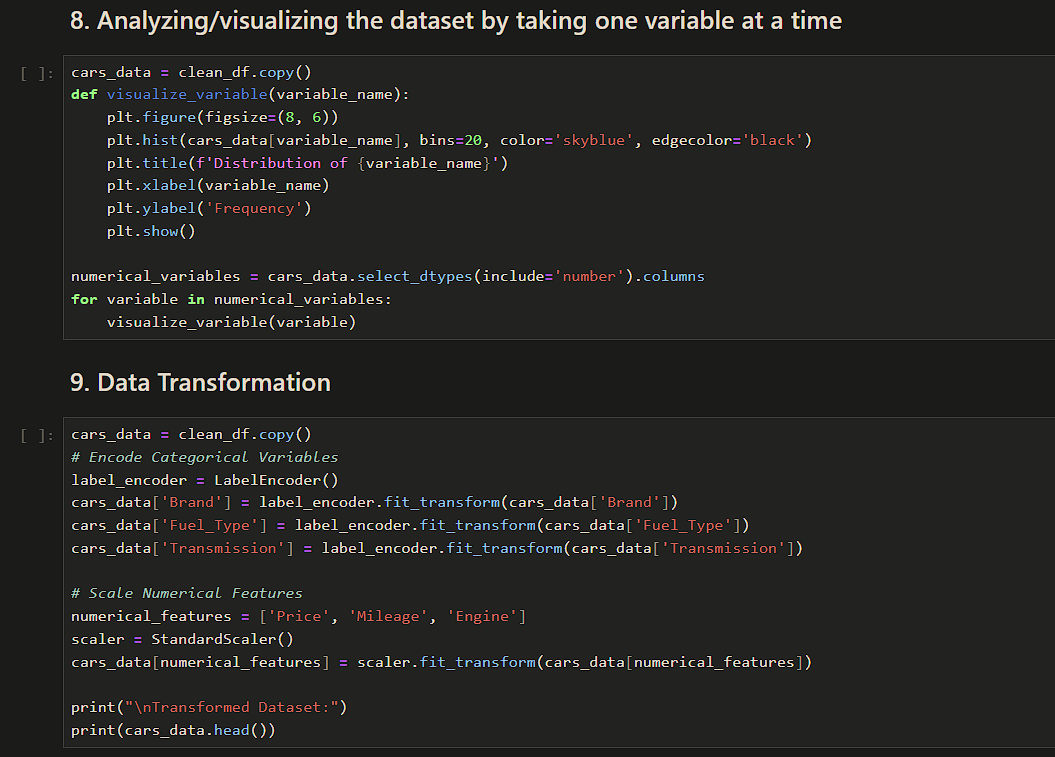
****

****

****

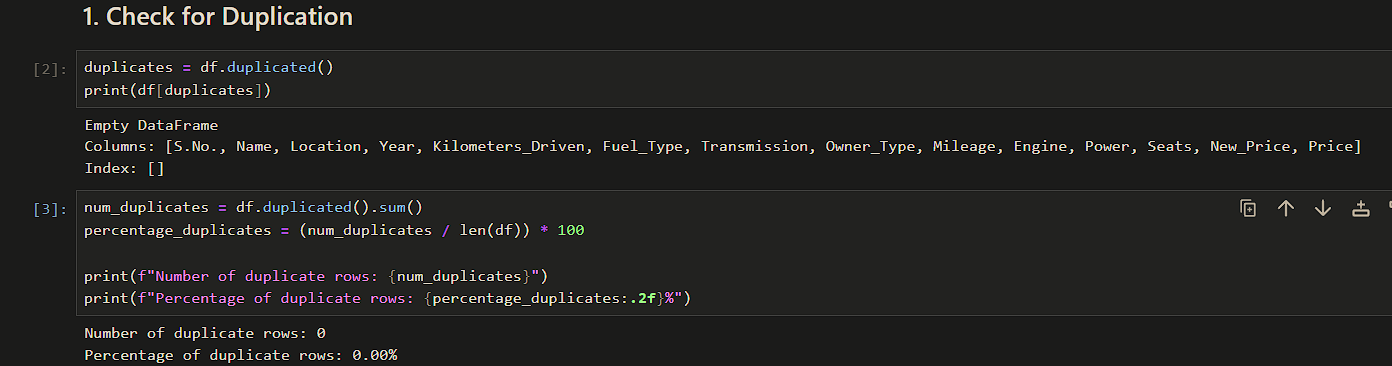
****

****

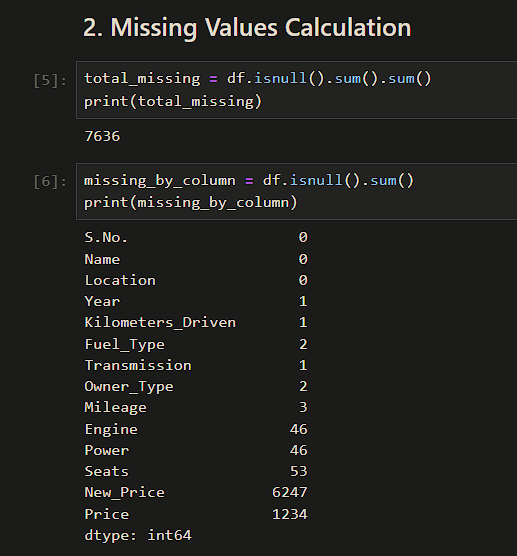
****

## Output:

**1. Check for Duplication**

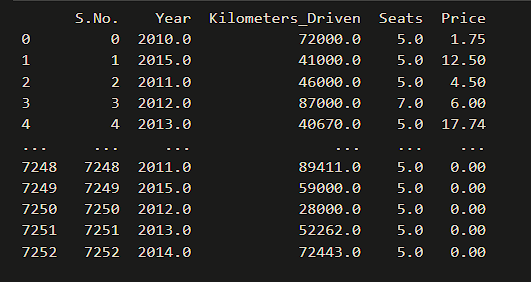
****

**2. Missing Values Calculation**

****

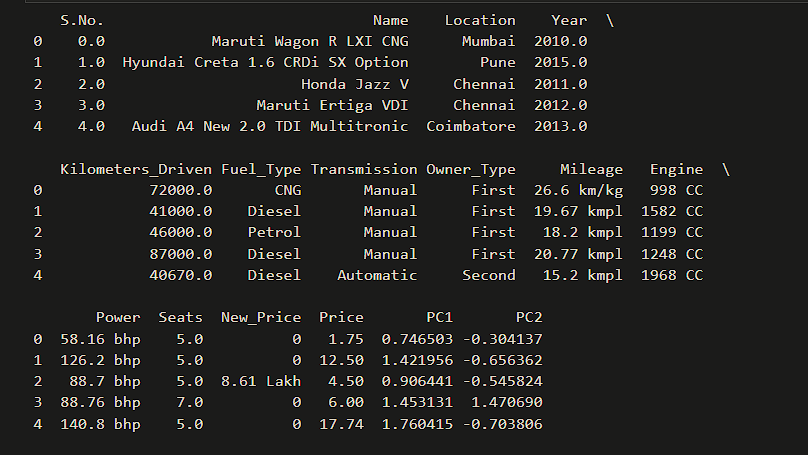
****

**3. Data Reduction (Some columns or variables can be dropped if they do not add value to our analysis.)**

****

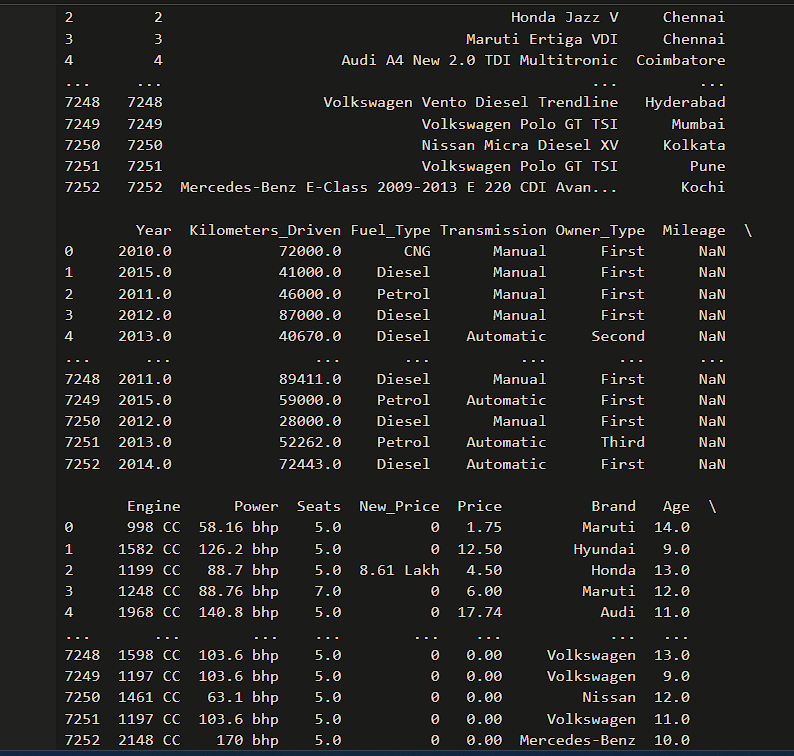
****

**4. Feature Engineering**

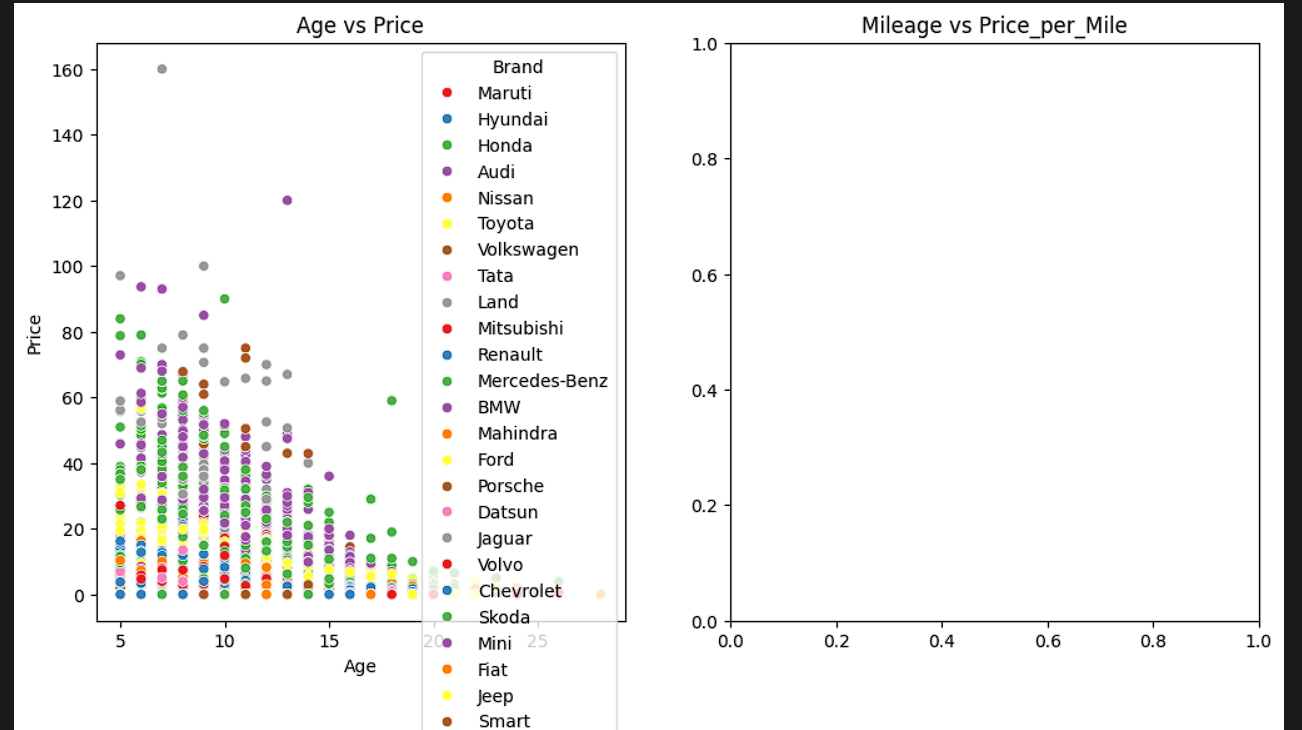
****

****

**5. Creating Features**

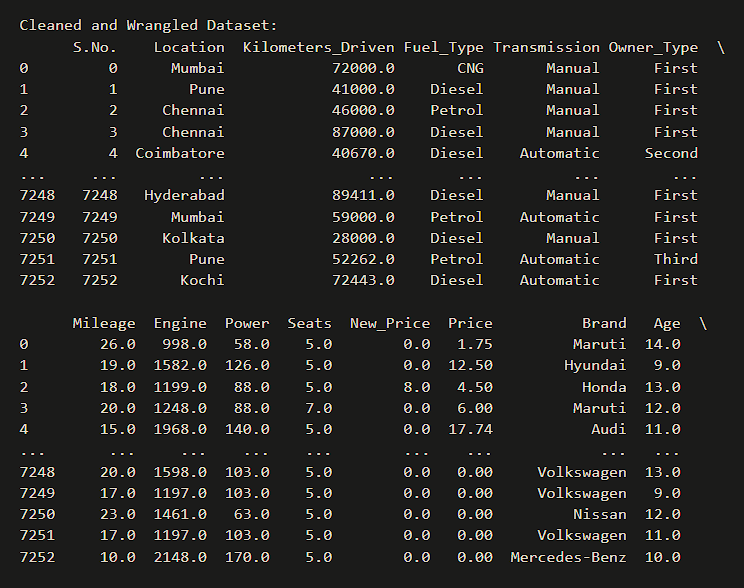
****

****

****

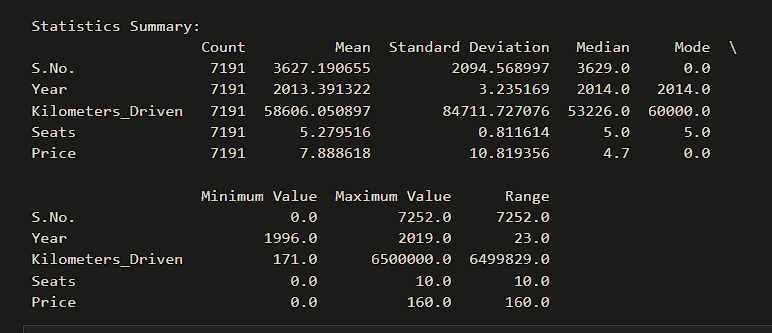
****

**6. Data Cleaning/Wrangling**

****

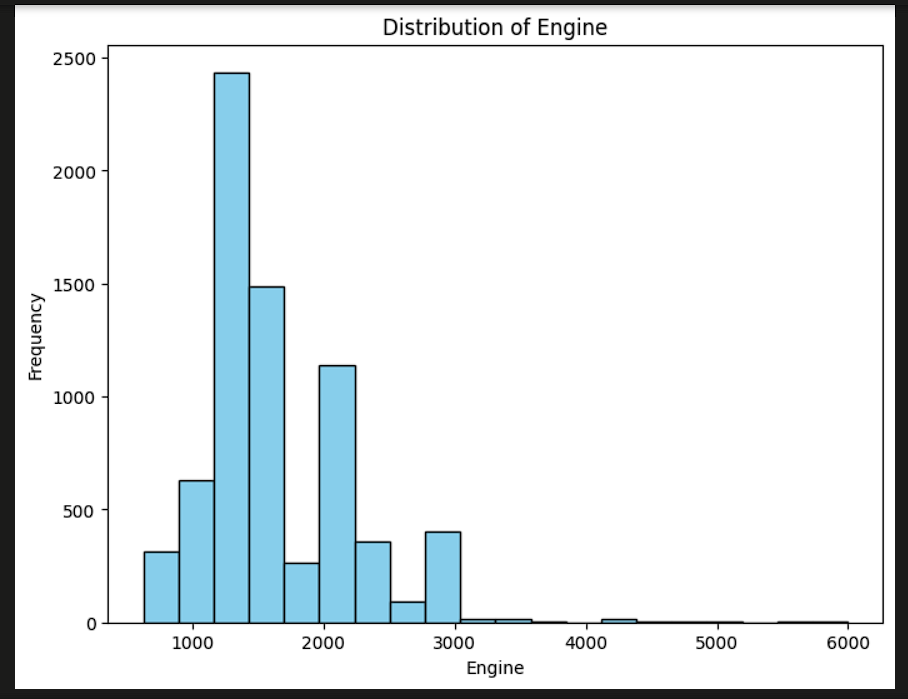
****

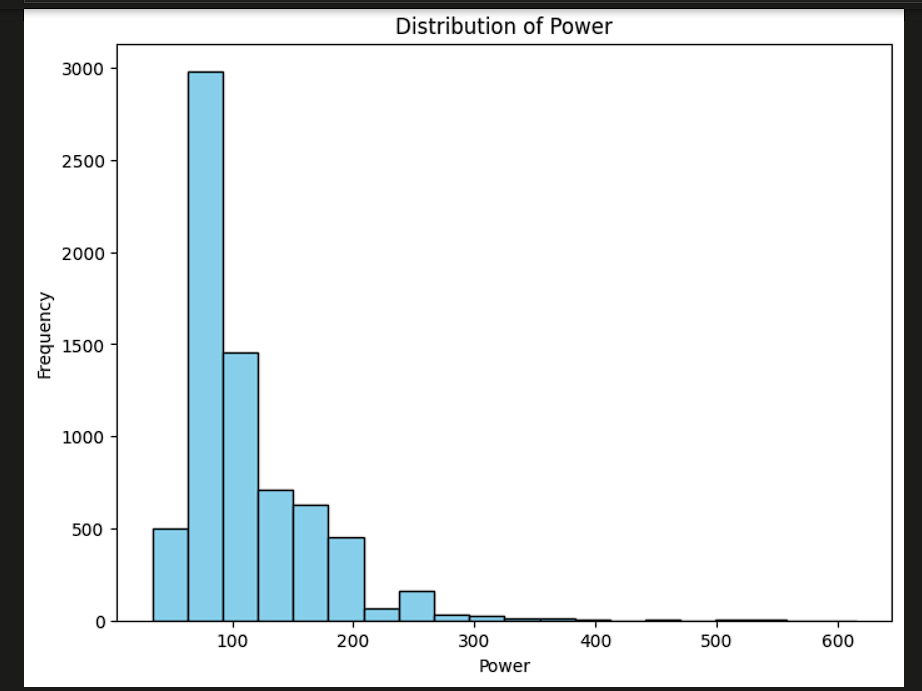
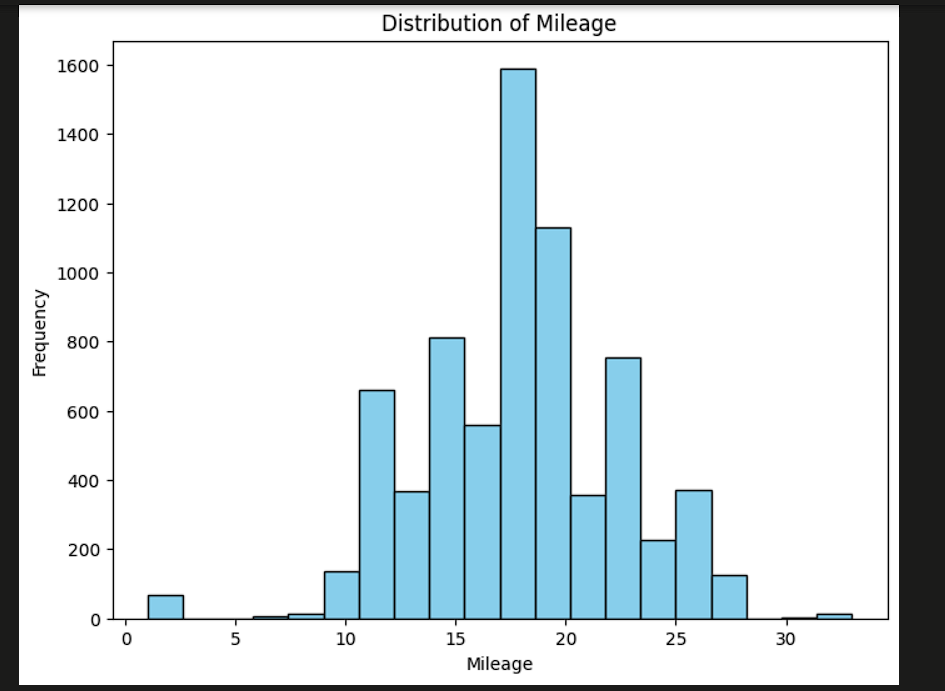
**7. Statistics Summary (Count, Mean, Standard Deviation, median, mode, minimum value, maximum value, range, standard deviation)**

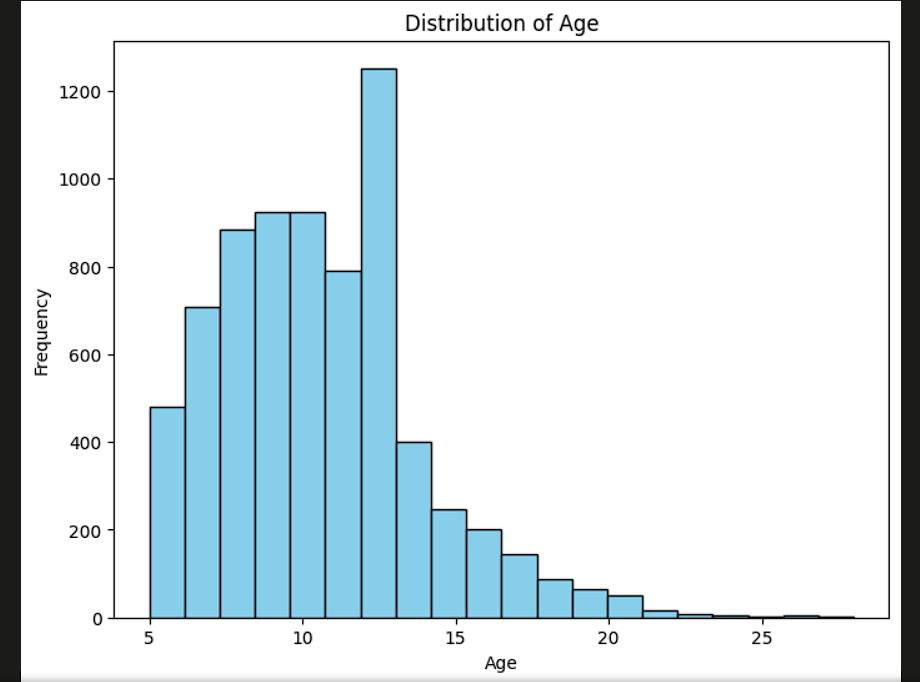
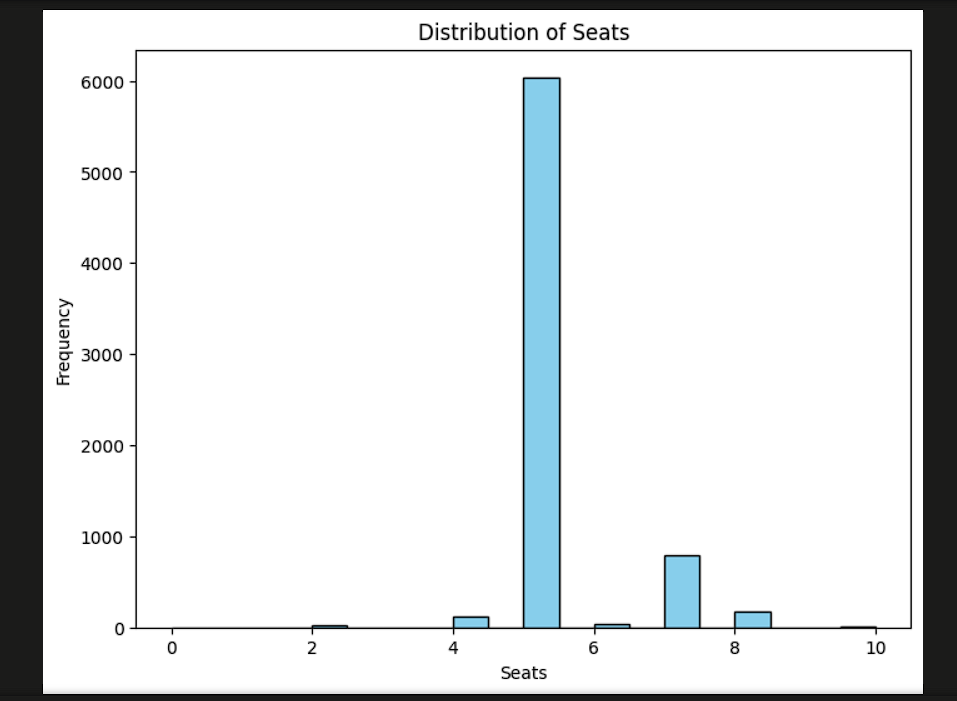
****

****

**8. Analyzing/visualizing the dataset by taking one variable at a time**

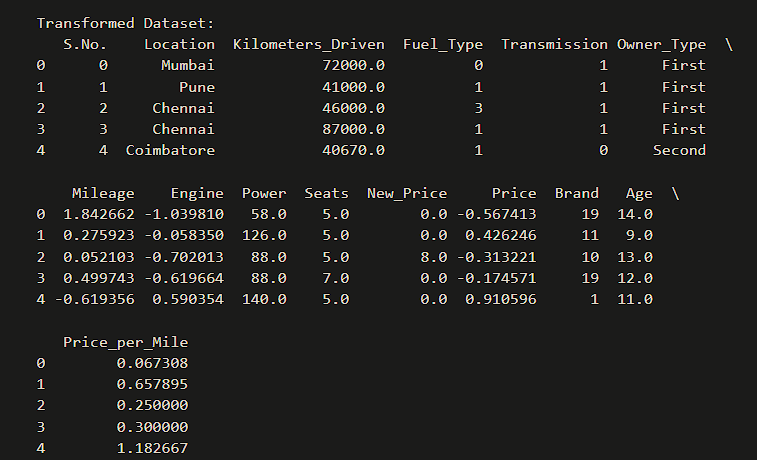


****

****

****

**9. Data Transformation**

****

****

## Conclusion:

* EDA provides a comprehensive overview of the cars dataset
* Identification and handling of missing values, outliers, and anomalies ensure data integrity and improve analysis accuracy.
* Descriptive statistics, including mean, median, and standard deviation, offer a summary of numerical attributes, aiding in understanding central tendencies and data dispersion.
* Visualization techniques, such as histograms and kernel density plots, reveal the distributions of key features, providing insights into the data's underlying patterns.
* Techniques like correlation, mutual information, or model-based feature importance assessments help prioritize variables based on their impact on the target variable.