

**TASK**

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

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

**Introduction**

Summary of the data set

In this exploratory data analysis (EDA) report, we will analyse the Automobile Data Set to gain insights into the information it contains. The data set presumably contains information about automobiles, and we will perform data cleaning and exploratory analysis to better understand its characteristics.

DATA CLEANING

Data cleaning is a crucial step in the EDA process, as it ensures that the data is in a usable and consistent format. In this section, we'll summarize the methods and visualizations used during data cleaning.

Summary of the Methods and Visualizations Done During Data Cleaning

Handling Missing Data: Identify columns with missing values and decide on an appropriate strategy to handle them, such as imputation or removal.

Data Types: Check and correct data types if necessary to ensure consistency.

Duplicate Records: Identify and remove duplicate rows if they exist.

Outliers: Detect and address outliers in numerical columns if relevant.

**MISSING DATA**

We observed that the data set has missing values so we performed the query in as follows e.g df['price'] = pd.to\_numeric(df['price'], errors='coerce')

**DATA STORIES AND VISUALISATIONS**

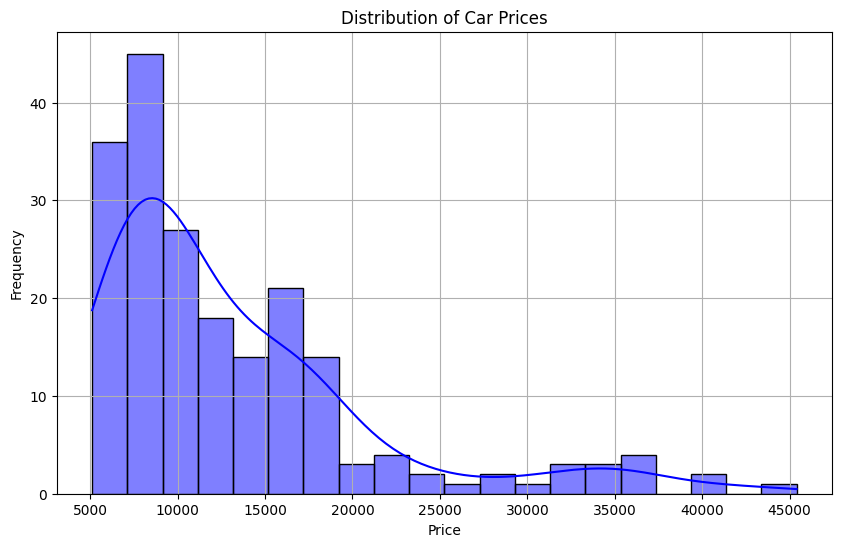
The below histogram shows the distribution of car prices as it’s indicates that the higher the price of cars the frequency and the lower the price of cars the higher the frequency.

Price Distribution: The histogram illustrates the distribution of car prices in the dataset. It appears that car prices vary widely, ranging from lower-priced cars to higher-priced ones.

Frequency: The height of each bar in the histogram represents the frequency or count of cars falling within a particular price range. In this case, we can see that there is a higher frequency of cars in the lower and mid-price ranges compared to the higher price ranges.

Skewness: The distribution of car prices is right-skewed, meaning that there are more cars with lower prices than with higher prices. This is a common characteristic of many datasets related to consumer goods, where lower-priced items tend to be more prevalent.

Market Segment: The distribution of car prices can provide insights into the market segment targeted by the dataset. In this case, it suggests that the dataset may include a range of cars, from budget-friendly options to luxury vehicles. The higher frequency of lower-priced cars could indicate that the dataset includes a significant number of affordable or mid-range cars.



**THIS REPORT WAS WRITTEN BY : BONGANI ZULU**

