

**TASK**

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

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

**Introduction**

The automobile.txt dataset contains various attributes of cars, including their specifications and prices. This dataset is useful for performing exploratory data analysis (EDA) to uncover insights about car features, pricing, and market trends.

The dataset includes columns such as make, fuel-type, body-style, horsepower, price, and more. These attributes help in understanding the characteristics and pricing of different car models.

**DATA CLEANING**

Replaced missing values denoted by ‘?’ with NaN.

Dropped rows with missing values.

Converted columns price to appropriate data types.

Removing Duplicate Rows

Dropped columns 'normalized-losses' that had a lot of missing data.

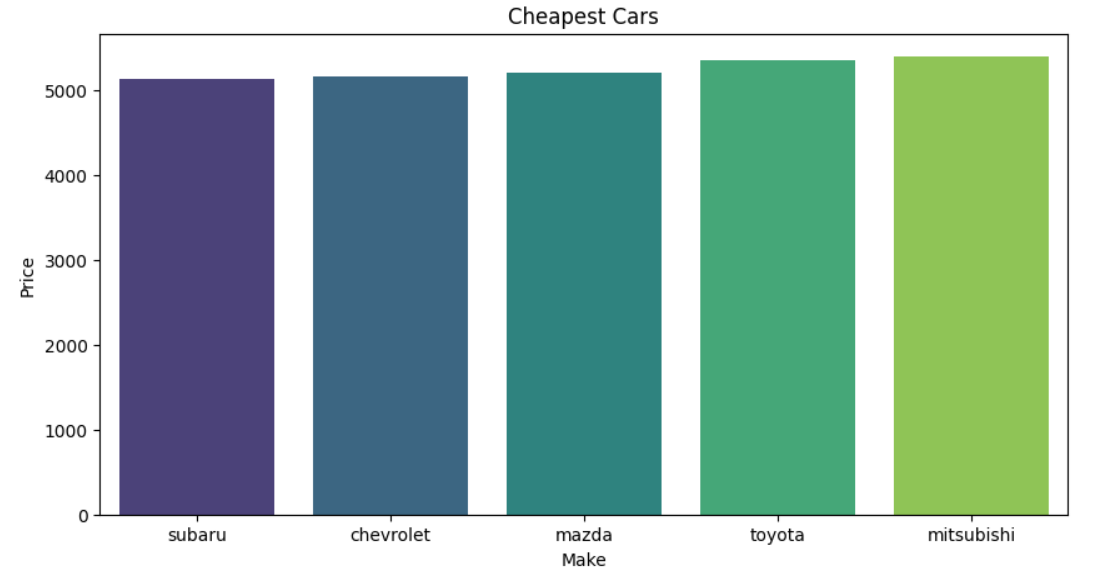
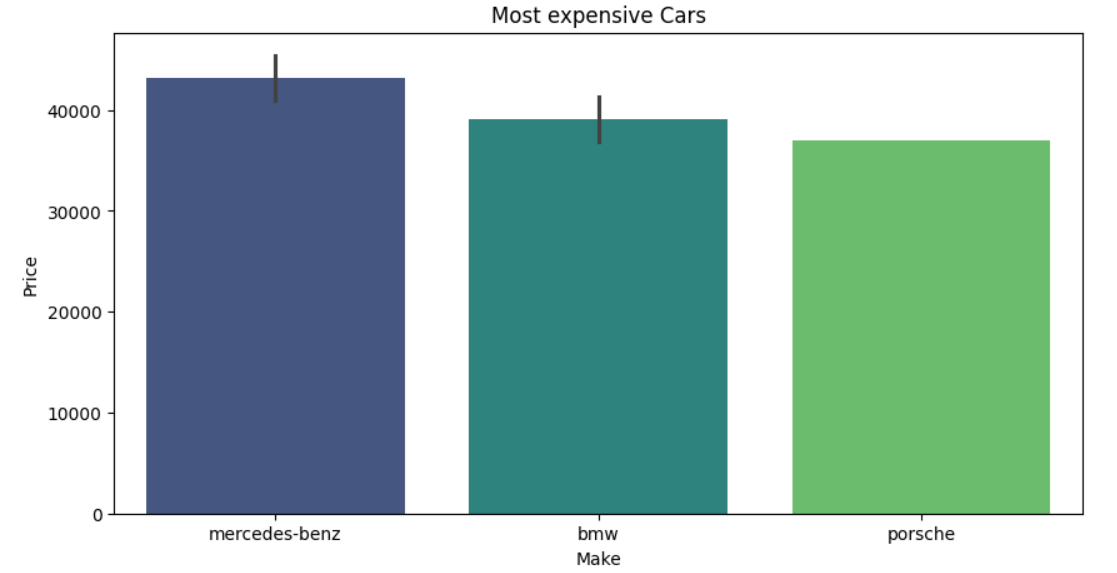
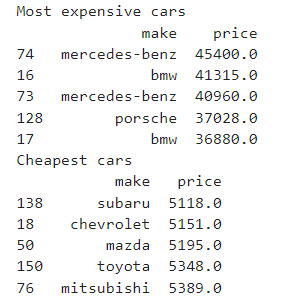
**MISSING DATA**

Yes, the dataset had missing values represented by ‘?’. These were handled by replacing them with NaN and subsequently dropping the rows with missing values to ensure data integrity.

**DATA STORIES AND VISUALISATIONS**

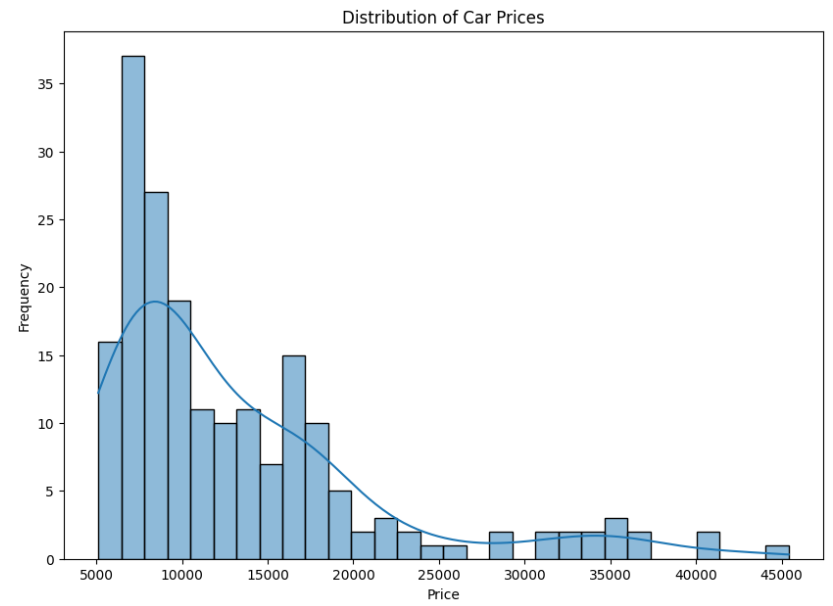
Which are the 5 most expensive cars?

How do the most expensive and cheapest cars compare?

  
  
Mrecedes-Benz, BMW and Porsche have a high price with the prices being in and around 40000. Subaru appears to have car with small price just above 5000 following Chevrolet.

Distribution of car prices

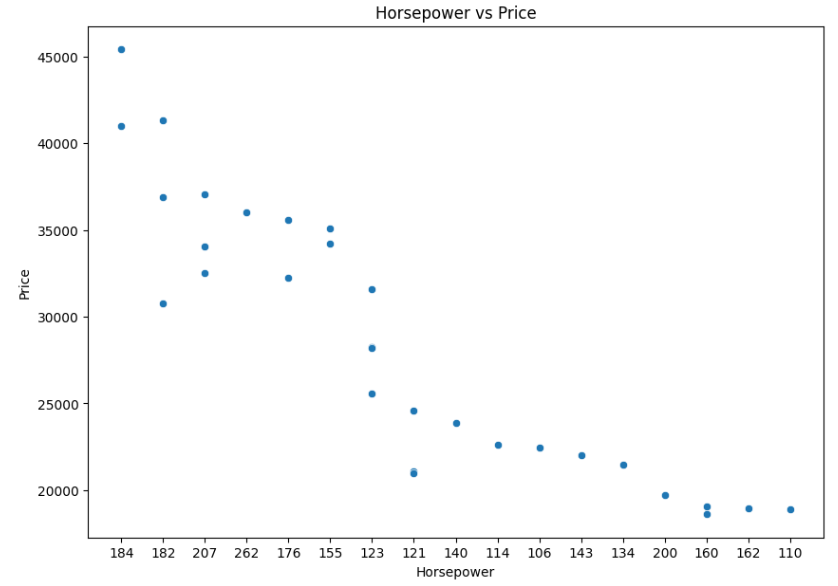
Get the distribution of car price to identify the most common price range on outliers



The distribution appears to be right-skewed, meaning there are a few cars with very high price compared to the rest.

Compare horsepower and price.

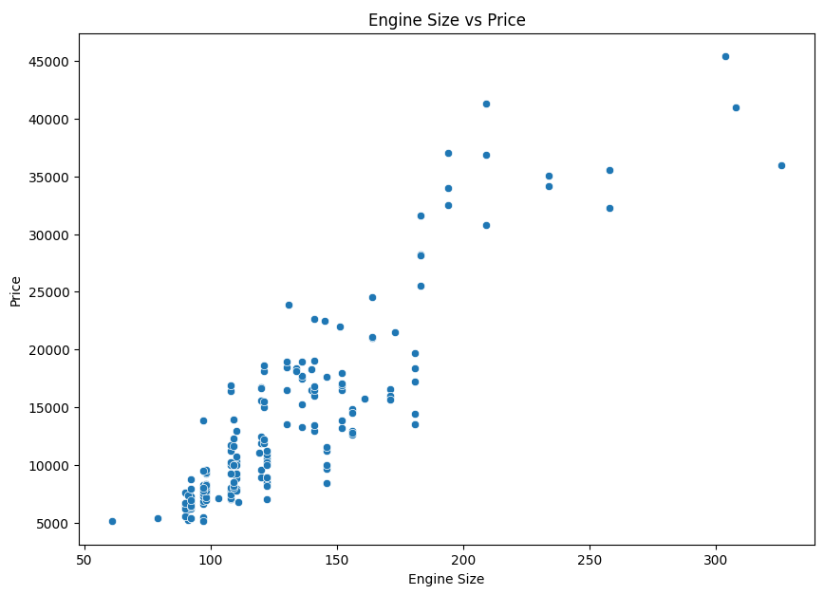
Explore the relationship between horsepower and top 30 expensive cars and see how engine power influence car pricing.



The dots represent individual data, indicating that generally, as the horsepower increases, the price also tends to increase. However high horsepower may not always mean high price as shown in the scatter plot above (example, 200 horsepower with around 20000 price and 184 horsepower with around 45000 price)

Compare Engine size and price

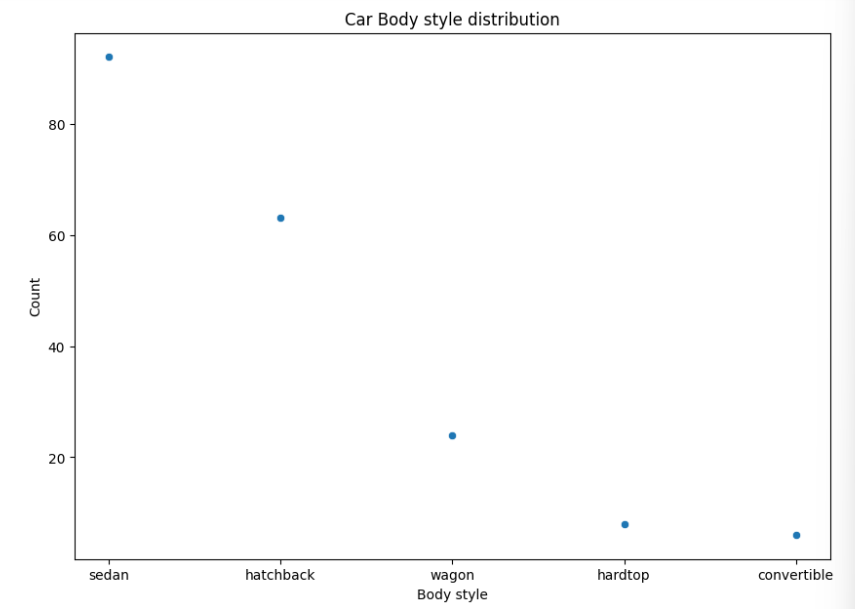
Explore the relationship between engine size and price to see how engine size influences car pricing.



This scatter plot shows that as the engine size increases the price also tends to increase.

Car body style distribution

Explore the distribution of car body style to see which body style is most preferred.



Sedan is the most common body style with counts above 80. Hatchback and wagon are moderately common with counts between 40 and 60. Hardtop and Convertible are least common with counts around or below 20.

**THIS REPORT WAS WRITTEN BY: SPHELELE KHUMALO**

