

**TASK**

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

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

**Introduction**

This dataset consists of numerous automobiles with their details such as price, make, body style, etc. The aim of this analysis is to find the patterns within this automobile dataset to know the factors that influence price, fuel efficiency, and other key attributes.

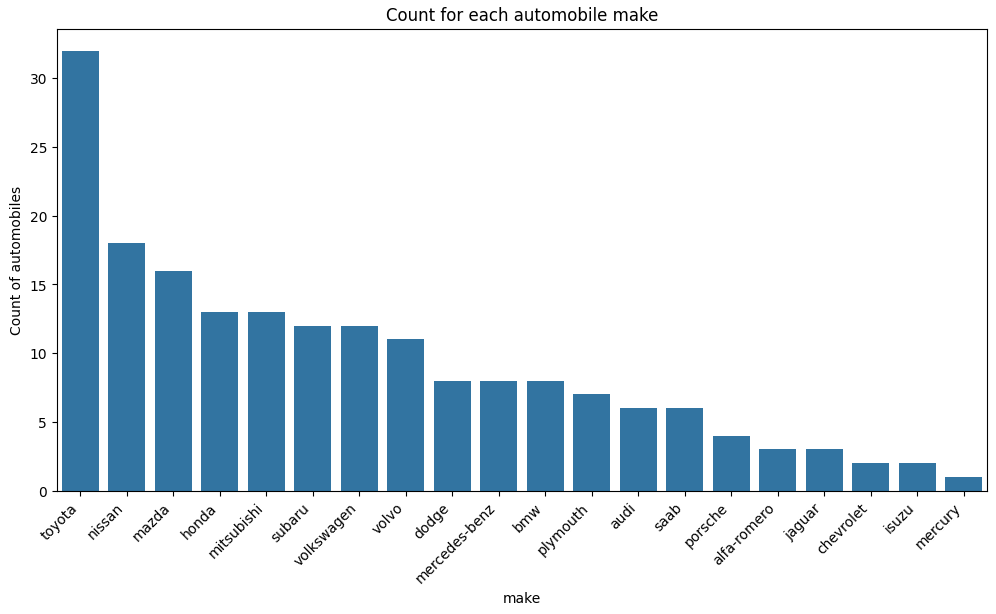
**DATA CLEANING**

Dropped the following columns: symboling, normalized-losses, engine-location, bore, stroke, compression-ratio, and peak-rpm, based on considerations of redundancy or limited direct relevance to the core analysis. I removed any duplicate rows.

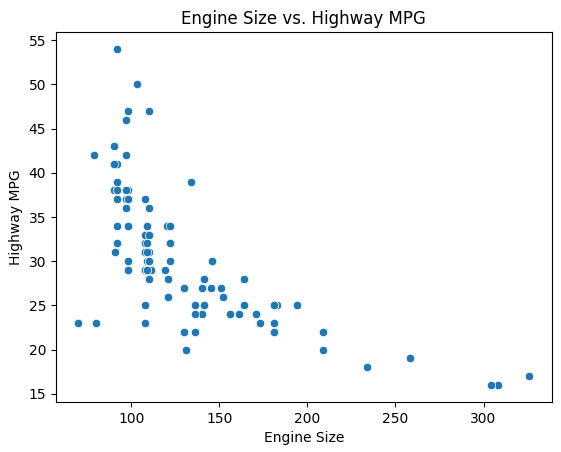
**MISSING DATA**

Missing data was represented with a question mark or a “l’. I replaced them with np.nan values and used the dropna() method.

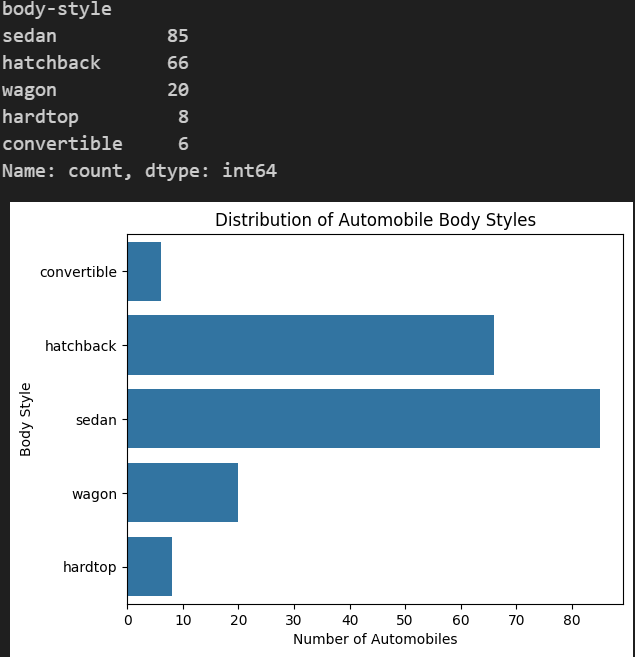
**DATA STORIES AND VISUALISATIONS**



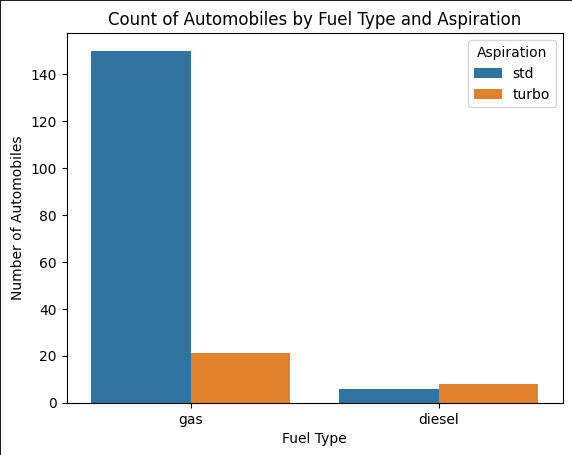
We can see that Toyota is the automobile company with the most makes with over 30 of them. Toyota significantly outnumbers other makes in the dataset, with almost 15 more entries than the next most frequent make, Nissan. The high representation of Toyota could suggest a larger market share for this brand.



The Engine Size and Highway MPG chart typically shows a negative relationship, suggesting that larger engines tend to be associated with lower highway fuel efficiency.



The hatchback and sedan auto body styles are widely the most common styles with being over 80 percent of all vehicles



The most common fuel type of automobiles is gas with a std aspiration.

**A chart with a blue rectangular object

AI-generated content may be incorrect.**

Most cars in the dataset fall within the price ranges of roughly 7,000 to 16,500, with the median price around 10,000, but there are numerous more expensive outliers, reaching up to around 45,000.

**THIS REPORT WAS WRITTEN BY : Christiaan Cronje**

