Exploratory Data Analysis: Automobile Dataset

INTRODUCTION

This analysis explores an automobile dataset containing vehicle characteristics such as price, make, body style, engine size, and fuel efficiency. The goal is to identify patterns that influence pricing and performance, and to extract insights that could support business or market-level decisions.

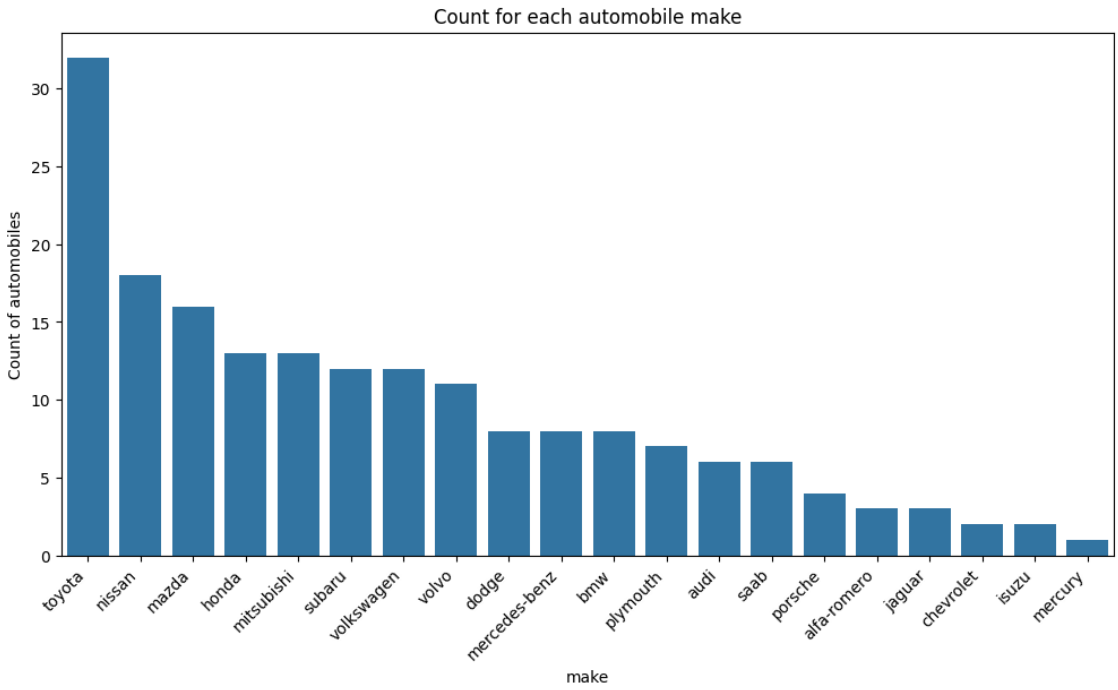
DATA CLEANING

Columns with limited analytical value or high redundancy (symboling, normalized-losses, engine-location, bore, stroke, compression-ratio, and peak-rpm) were removed to keep the analysis focused and interpretable. Duplicate rows were also removed to ensure data integrity.

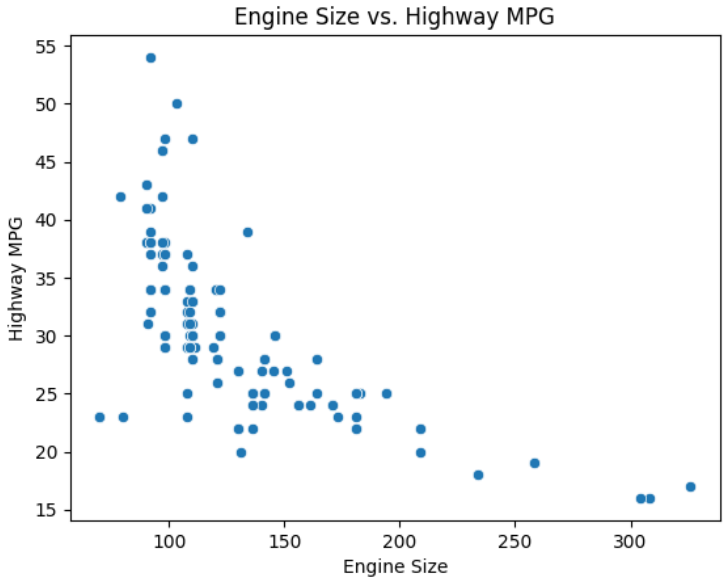
MISSING DATA

Missing values were identified in the dataset and standardised by replacing placeholder symbols (e.g. “?”) with null values. Records containing missing data were removed to ensure consistency and reliability in the analysis.

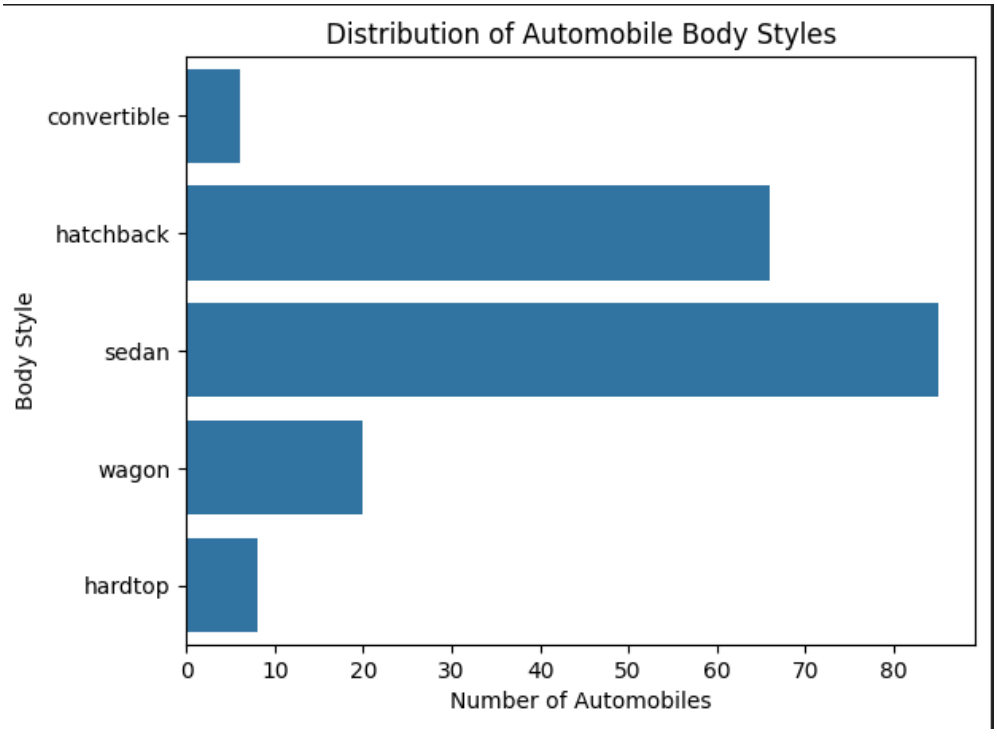
KEY INSIGHTS & VISUAL ANALYSIS



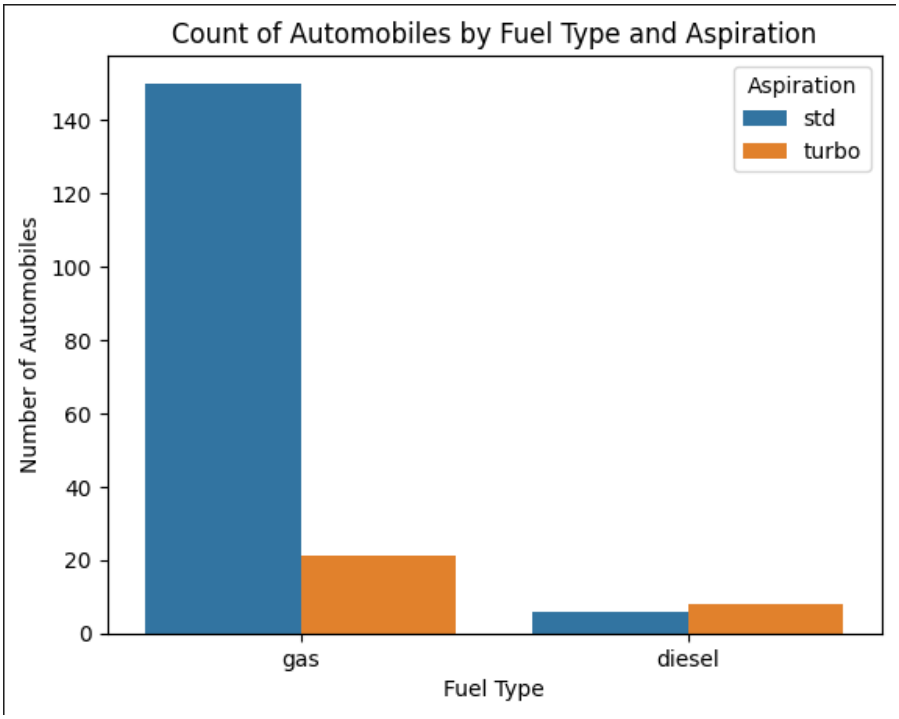
Toyota is the most frequently represented manufacturer in the dataset, with over 30 entries. This strong presence may indicate a larger market footprint or wider model variety compared to other manufacturers.



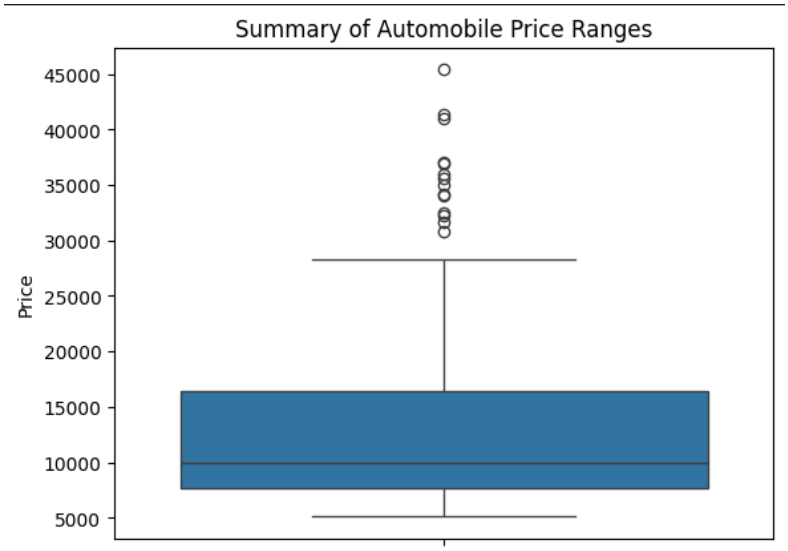
A clear negative relationship is observed between engine size and highway MPG, indicating that larger engines are generally associated with lower fuel efficiency.



Hatchback and sedan body styles dominate the dataset, accounting for over 80% of all vehicles.



The majority of vehicles in the dataset use gasoline with standard aspiration.



Most vehicles are priced between approximately 7,000 and 16,500, with a median around 10,000. A smaller number of high-priced outliers extend up to roughly 45,000.

Prepared by: Christiaan Cronje | Data Analytics Portfolio