Exploratory Data Analysis on the Automobiles Dataset

Introduction

This exploratory data analysis was conducted in Python using Jupyter Notebook, which allowed for interactive documentation and streamlined data handling. Core libraries included Pandas for structured data manipulation, NumPy for numerical operations and managing missing values, and Matplotlib and Seaborn for building compelling visualizations and statistical plots. These open-source tools enabled efficient cleansing, transformation, and insight generation from the automobile dataset, supporting clear comparisons across price, fuel efficiency, engine specifications, and manufacturer patterns.

Data Cleansing

Removing the following columns ['normalized-losses', 'symboling'] from the data set as they will not be used in the analysis. Removing all duplicate rows within the dataset to ensure a more accurate analysis of the data. Changing all the numeric data values to appropriate data types.

Missing Rows

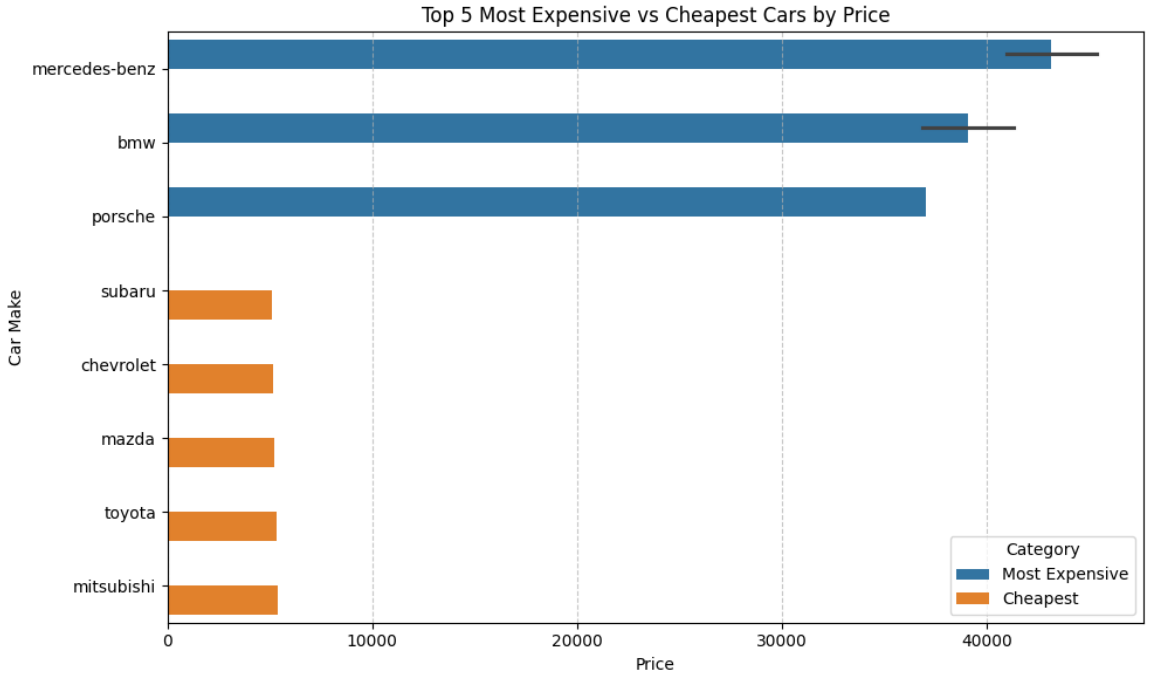
To ensure clean and reliable analysis, rows with missing values were removed using automobiles_df.dropna(inplace=True). This command drops any row in the dataset that contains at least one missing (NaN) value. While this helps avoid errors during analysis, it may also reduce the dataset size, so it's important to check how much data is lost before applying it.

This report was written by: Tlotlo Selogilwe

Data stories and visualizations

💰 Top 5 Most Expensive vs Cheapest Car

This graph compares the top 5 most expensive cars and the top 5 cheapest cars in the dataset by their price. It helps visualize the price gap between luxury and budget vehicles.



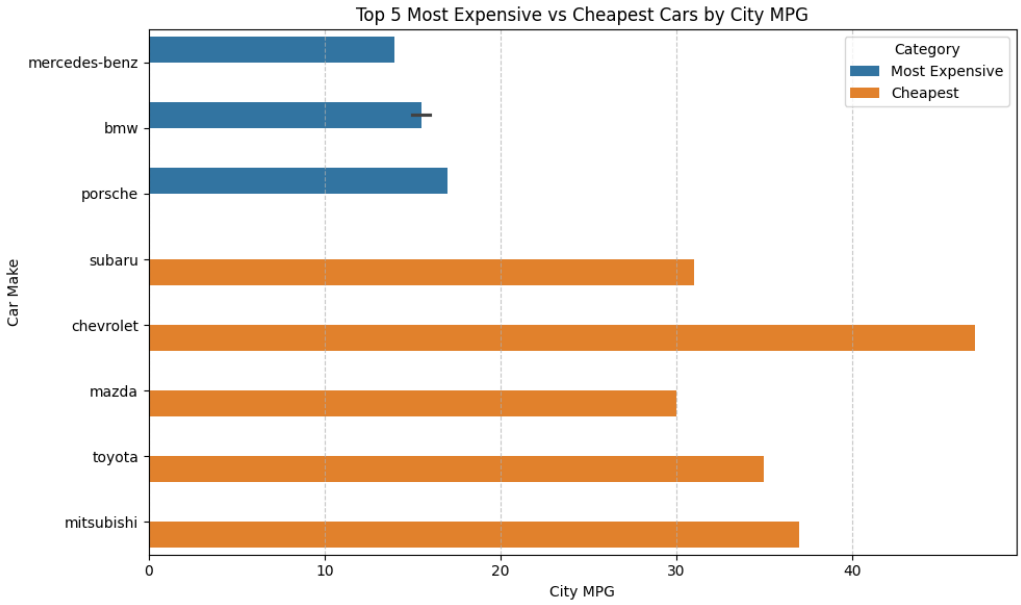
Insight:

The chart shows that the most expensive cars (Mercedes-Benz, BMW, Porsche) are much pricier than the cheapest cars (Subaru, Chevrolet, Mazda, Toyota, Mitsubishi). This highlights a clear divide between luxury and budget brands in terms of price.

This report was written by: Tlotlo Selogilwe

⛽ City MPG Comparison: Most Expensive vs Cheapest Cars

This graph compares the city MPG (miles per gallon) of the top 5 expensive and cheapest cars. It highlights the difference in fuel efficiency between luxury and economy models.



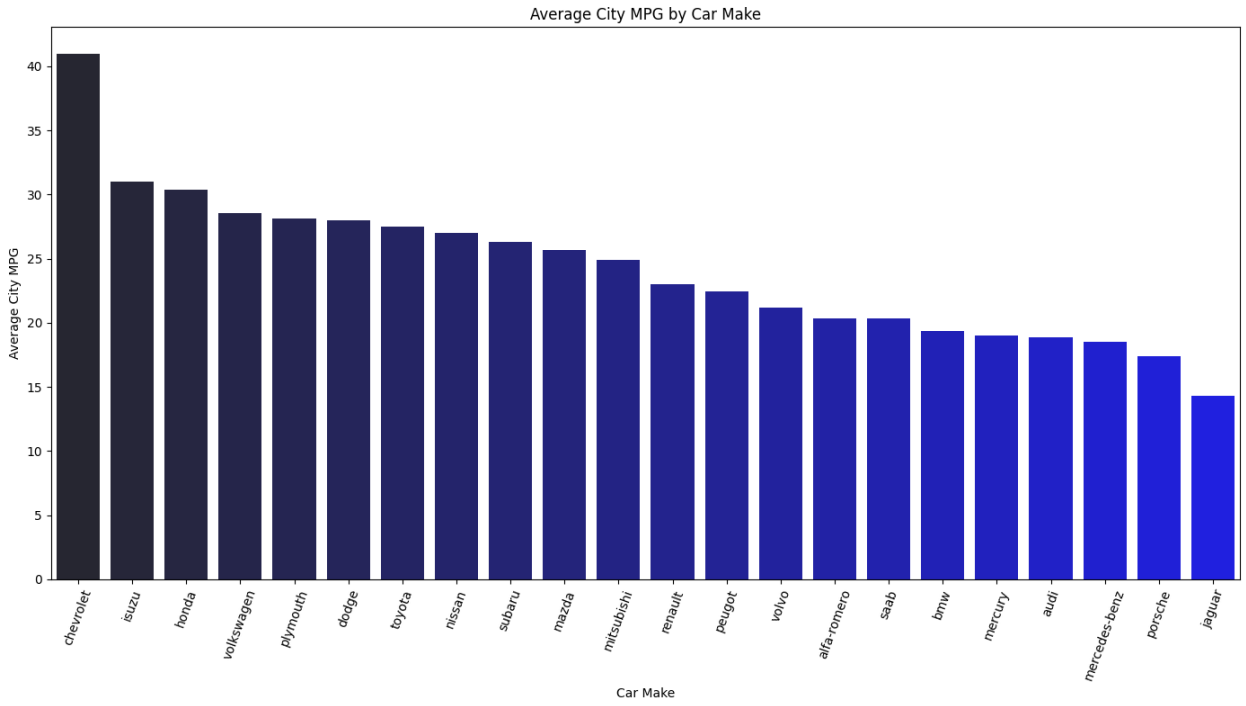
Insight:

The graph shows that Mercedes-Benz, BMW, and Porsche are the most expensive cars in the dataset, while Subaru, Chevrolet, Mazda, Toyota, and Mitsubishi are the cheapest. There is a significant price gap between luxury and budget brands. This highlights how luxury vehicles command much higher prices compared to economy models

This report was written by: Tlotlo Selogilwe

⛽ Average City MPG by Car Make

This graph shows the average city miles per gallon (MPG) for each car manufacturer, allowing you to compare fuel efficiency across brands.



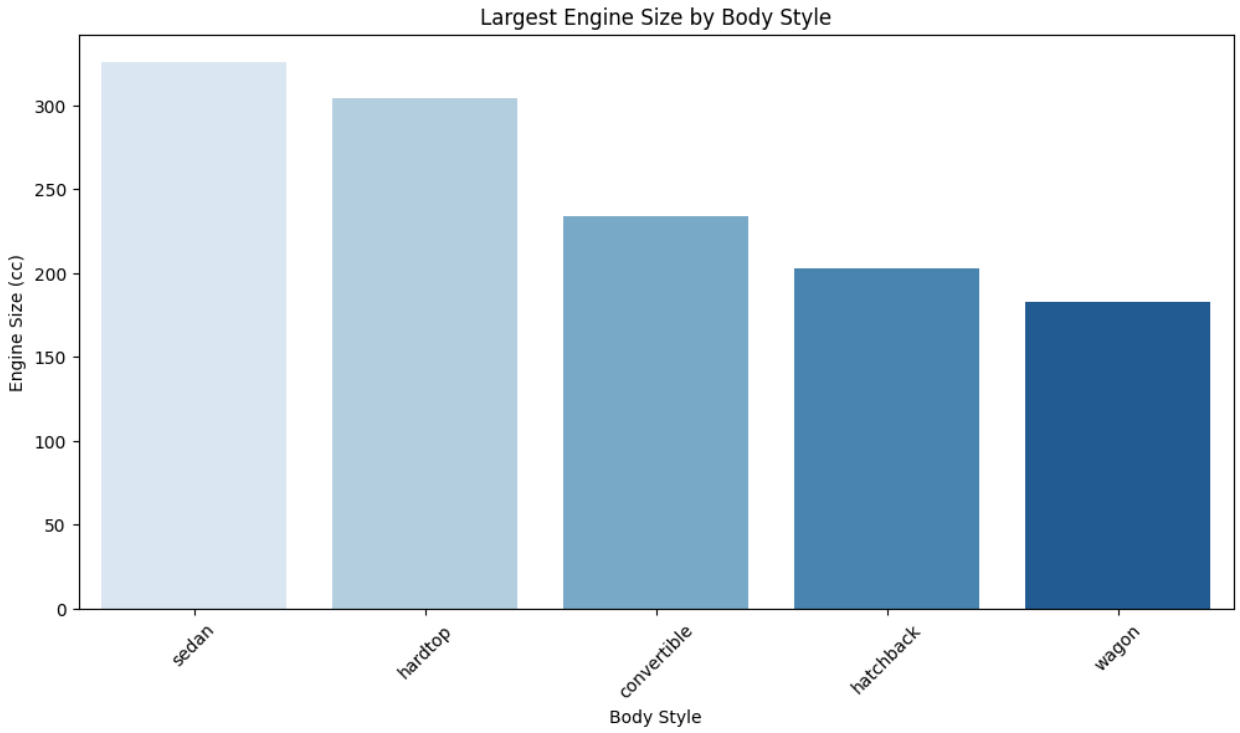
Insight:

Chevrolet, Isuzu, and Honda have the highest average city MPG, making them the most fuel-efficient manufacturers in the dataset. Luxury brands like Jaguar, Porsche, and Mercedes-Benz have the lowest average city MPG, indicating lower fuel efficiency. Generally, economy car brands outperform luxury brands in terms of fuel economy.

This report was written by: Tlotlo Selogilwe

🔧 Largest Engine Size by Body Style

This graph displays the maximum engine size for each body style, helping you compare which types of cars tend to have the largest engines.



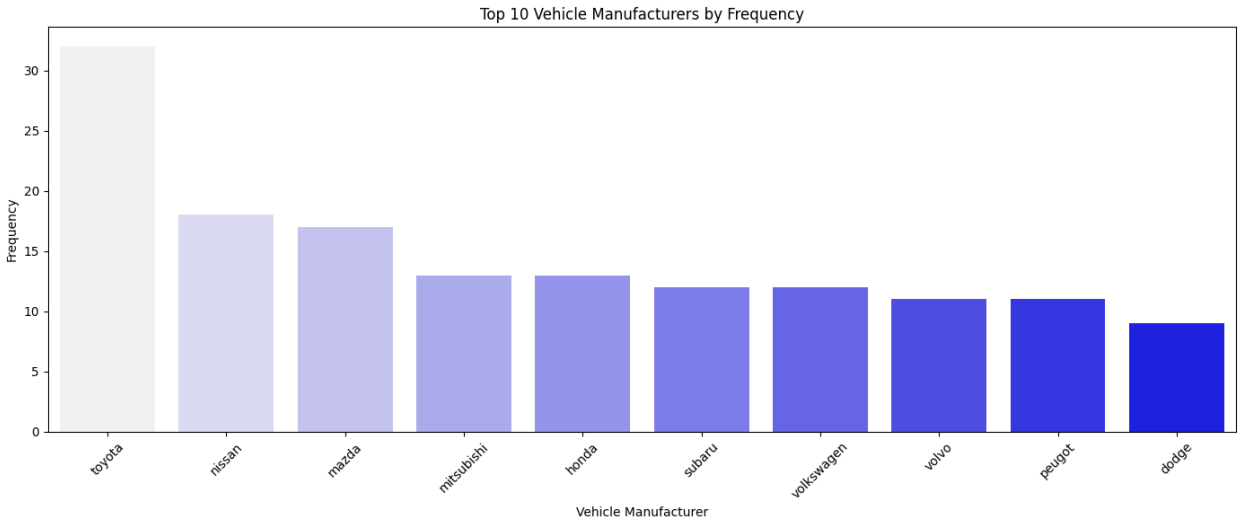
Insight:

Sedans and convertibles tend to house the largest engines, suggesting a tilt toward comfort and performance in those categories. Hatchbacks and wagons show smaller engine sizes, aligning with their roles as economical or utility-oriented models.

This report was written by: Tlotlo Selogilwe

🏭 Top 10 Most Frequent Vehicle Manufacturer

This graph shows which car manufacturers have the most models represented in the dataset, highlighting the most common brands.



Insight:

Toyota is the most frequent vehicle manufacturer in the dataset, followed by Nissan and Mazda. These brands have the widest variety of models represented. This suggests Toyota has the largest market presence or diversity in this dataset compared to other manufacturers.

This report was written by: Tlotlo Selogilwe

Conclusion

This exploratory data analysis provided valuable insights into the automobile dataset. The data cleaning process ensured the dataset was accurate and ready for analysis. Key findings include:

- There is a significant price gap between luxury and budget vehicles, with brands like Mercedes-Benz, BMW, and Porsche being the most expensive, and brands like Subaru, Chevrolet, Mazda, Toyota, and Mitsubishi being the most affordable.

- Cheaper cars tend to have better fuel efficiency (higher city MPG) compared to luxury vehicles, which often prioritize performance and features over economy.

- Sedan and hardtop body styles generally have the largest engine sizes, while wagons and hatchbacks have smaller engines.

- Toyota, Nissan, and Mazda are the most common manufacturers in the dataset, indicating a wide variety of models from these brands.

Overall, the analysis highlights clear trends in price, fuel efficiency, engine size, and manufacturer representation, providing a solid foundation for further investigation or predictive modeling.

This report was written by: Tlotlo Selogilwe