Automobile Dataset Analysis Report

Name: Stefan Willem Vorster

Date: 22 April 2025

Introduction

This report presents an Exploratory Data Analysis (EDA) of an Automobile dataset. The primary goal is to extract meaningful insights into the relationships between various automobile features—such as fuel efficiency, engine size, manufacturer, and price—and how they interplay across different car attributes. Special attention is given to how car manufacturers and specifications affect consumer-centric metrics like fuel efficiency and cost. The following visualizations and discussions explore trends in car design, market placement, and economic efficiency across a range of vehicle types and brands.

Discussions

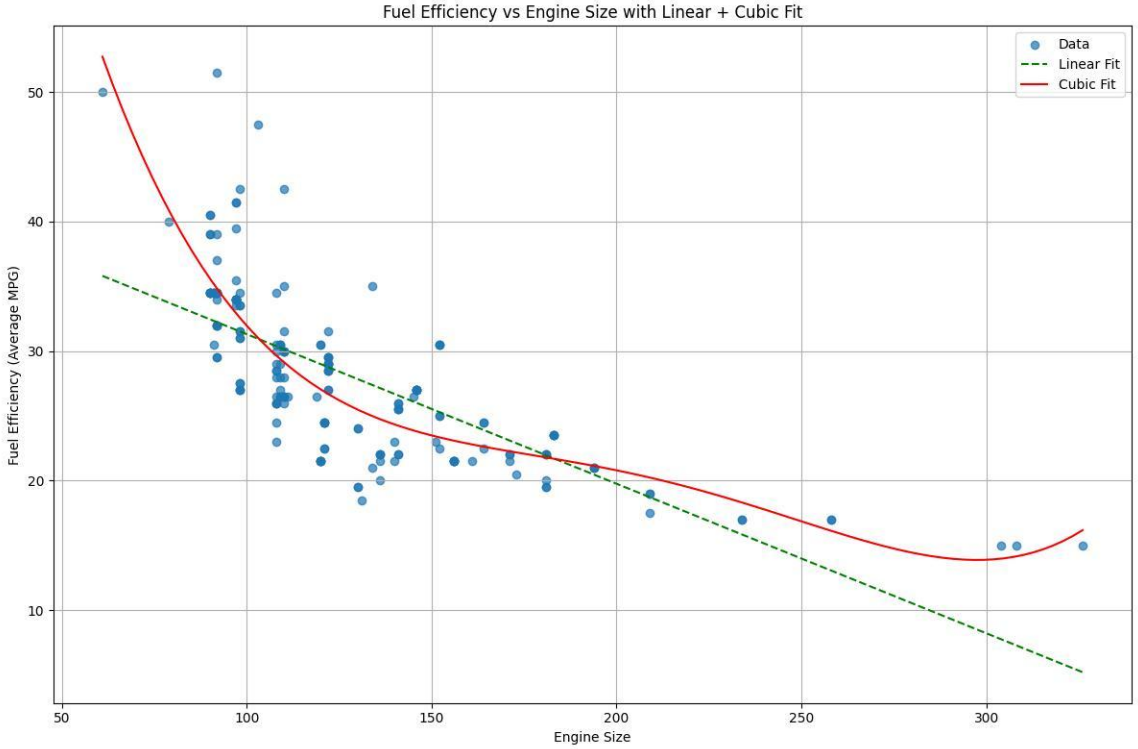

Figure 1: Relationship Between Engine Size and Fuel Efficiency

This scatter plot illustrates the inverse relationship between engine size and fuel efficiency (miles per gallon, MPG). As expected, cars with larger engines tend to have lower MPG

ratings. This suggests a clear trade-off between performance (often associated with larger engines) and fuel economy. Outliers in the plot may represent either extremely efficient small-engine cars or high-performance vehicles with unusually large engines and poor fuel economy. The negative trend confirms that engine size is a significant determinant of fuel efficiency and can inform eco-conscious consumer decisions.

This histogram illustrates the distribution of car prices across the dataset. The graph shows a right-skewed distribution, indicating that most cars are priced below $20,000, with a few high-end vehicles reaching up to $45,000. This skew suggests that the market is saturated with budget and mid-range cars, while luxury cars form a small portion of the dataset. Identifying outliers here is essential for accurate modeling and prediction.

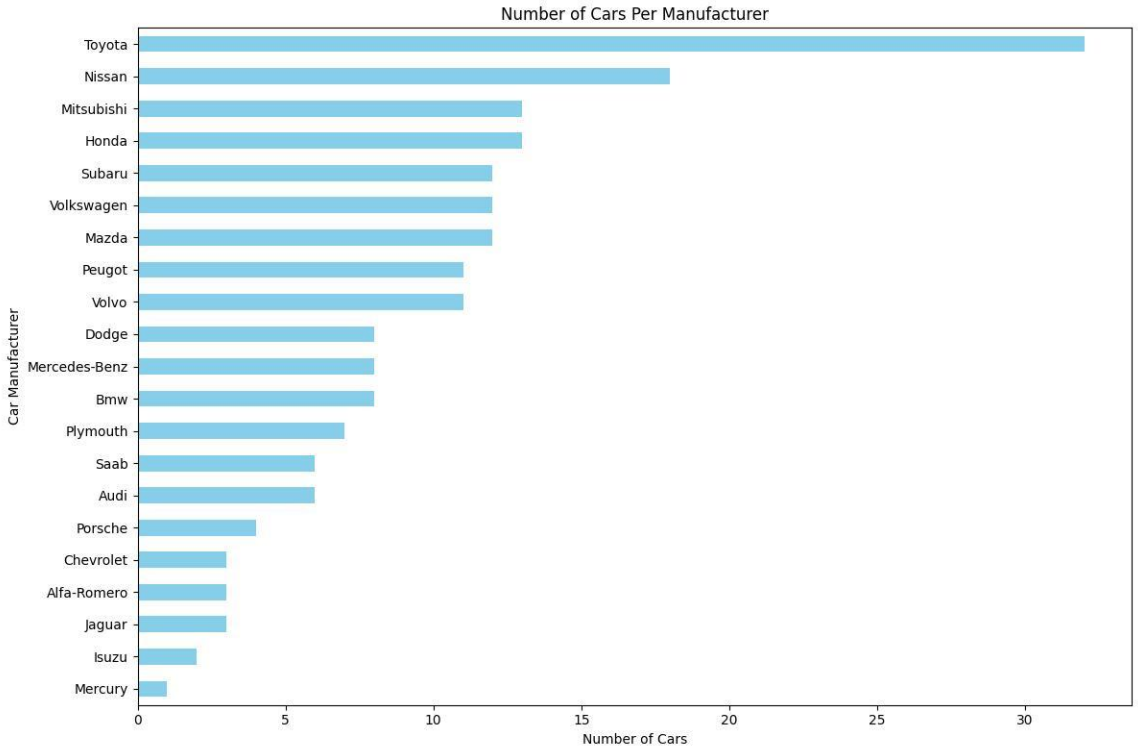

Figure 2: Number of Cars per Manufacturer

This bar chart shows the number of cars represented in the dataset per manufacturer. Brands such as Toyota and Nissan have the highest frequency, indicating their broad market presence and diverse vehicle offerings. In contrast, luxury or niche manufacturers like Porsche or Jaguar appear less frequently, likely reflecting their focus on high-end or limited-edition vehicles. This distribution helps contextualize later findings—such as average MPG or pricing per brand—based on representation.

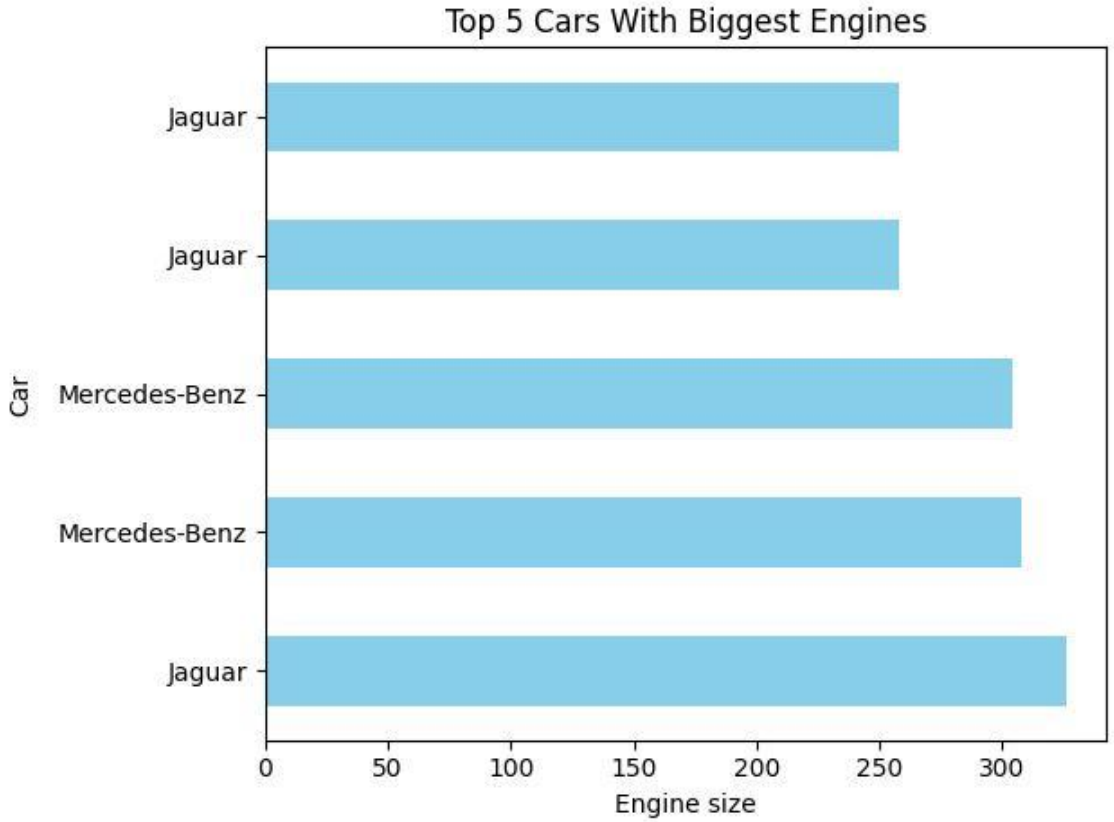

Figure 3: Top 5 Cars with the Largest Engine Sizes

This bar chart displays the top five vehicles in terms of engine displacement. These vehicles are often positioned as performance-oriented or luxury models. Notably, brands like Porsche and Jaguar appear here, which aligns with their focus on delivering high-performance engines. These large engines suggest high horsepower and torque outputs, though they likely come at the expense of fuel efficiency and affordability. Such cars appeal to a niche market prioritizing power and prestige over economy.

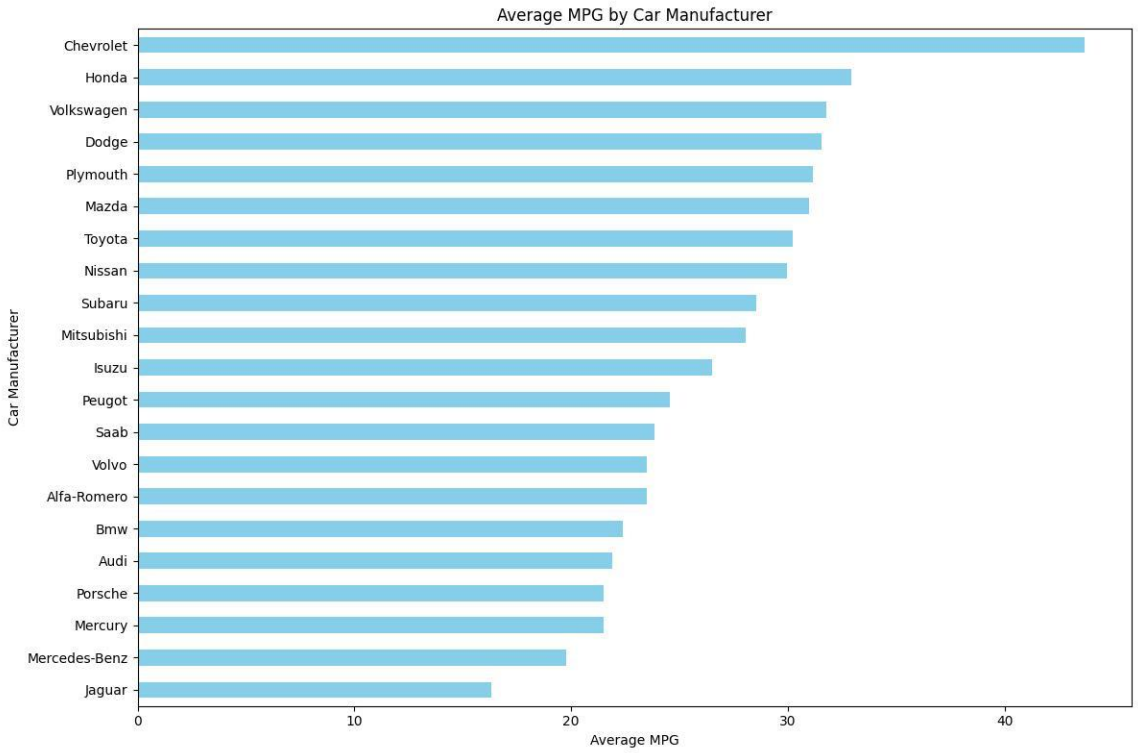

Figure 4: Average MPG by Manufacturer

In this bar chart, average miles per gallon is computed per manufacturer. Brands like Honda and Subaru tend to dominate in fuel efficiency, emphasizing their strategy of producing economical and commuter-friendly vehicles. On the other hand, manufacturers such as BMW or Porsche fall on the lower end of the MPG spectrum, which is consistent with their production of sportier or more luxurious vehicles. This analysis provides an important perspective for environmentally conscious consumers when choosing a brand.

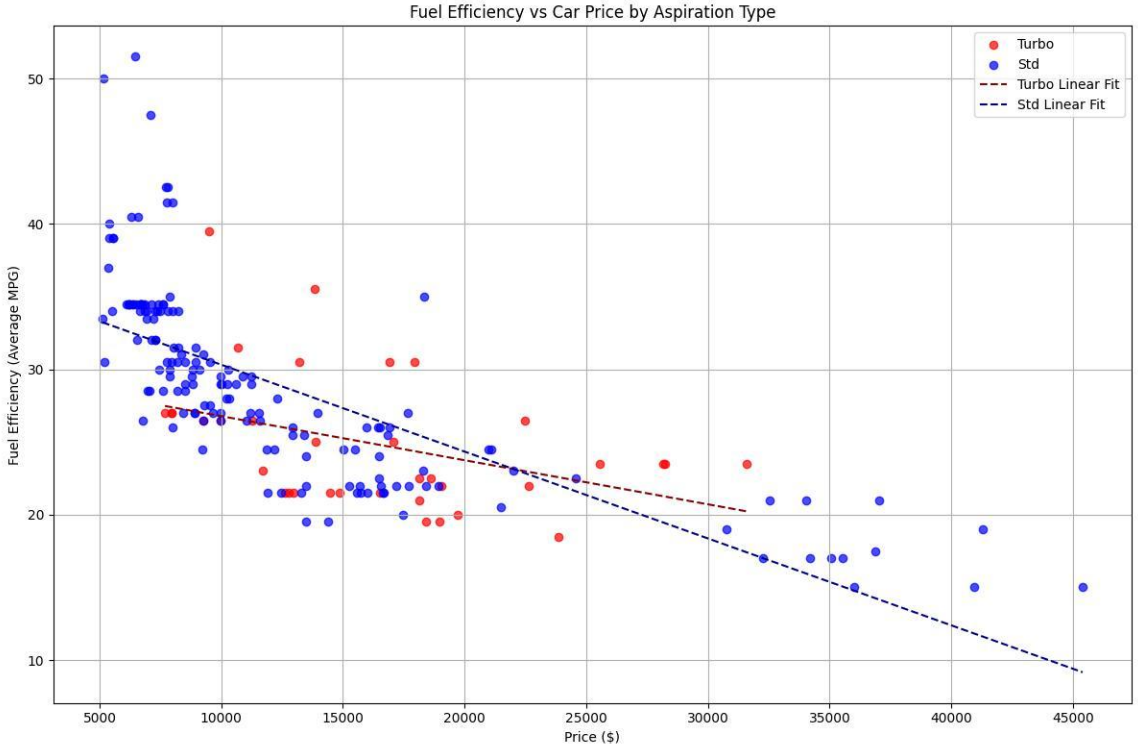

Figure 5: Fuel Efficiency vs Price by Aspiration Type

This scatter plot compares car price against fuel efficiency, with points differentiated by aspiration type (standard vs. turbocharged). The chart reveals that turbocharged cars, while often more expensive, do not always provide better fuel efficiency. In fact, many fall below the fuel economy of standard aspiration vehicles. This suggests that the turbo feature may be marketed for performance enhancement rather than economic efficiency. The trade-off highlighted here is between performance and long-term running costs, especially in fuel consumption.

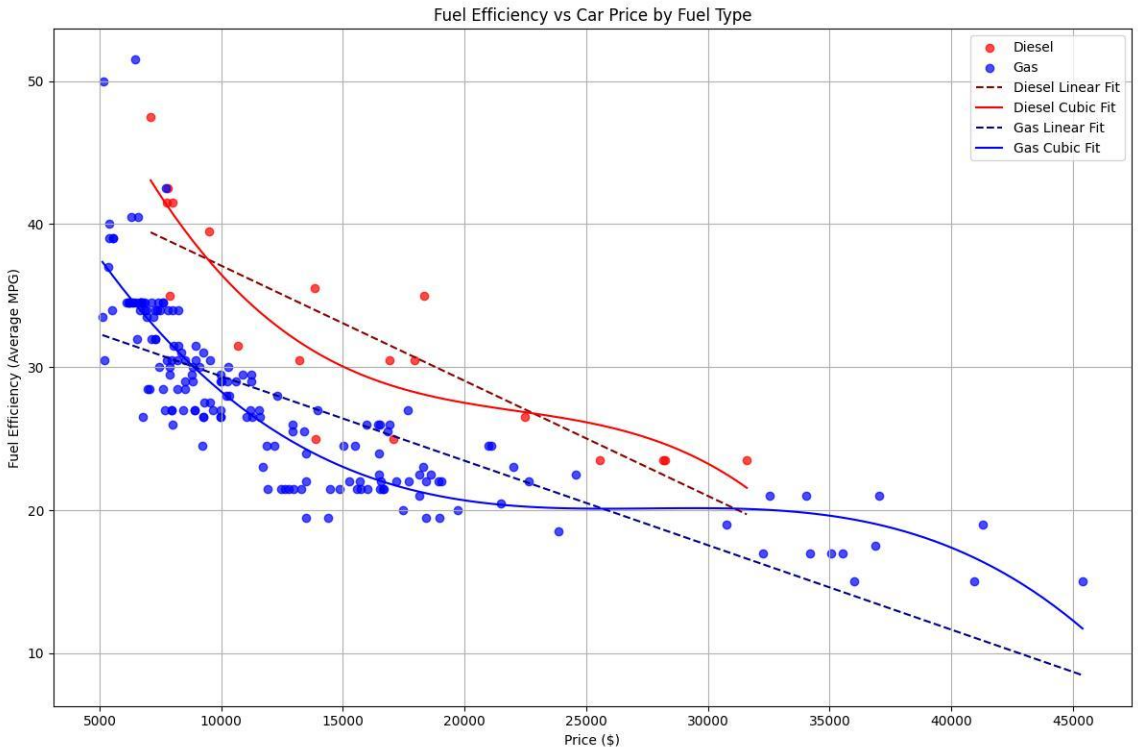

Figure 6: Fuel Efficiency vs Price by Fuel Type

This scatter plot analyses the relationship between fuel efficiency and car price based on fuel type (gas vs. diesel). Diesel vehicles appear to cluster at slightly higher prices but offer comparable or better fuel efficiency than gasoline cars. This supports the notion that diesel engines, while initially more expensive, can deliver better mileage and may be more cost-effective in the long run. However, the variance in pricing also suggests that other features (e.g., luxury finishes, brand value) significantly influence overall cost beyond just the engine's fuel type.

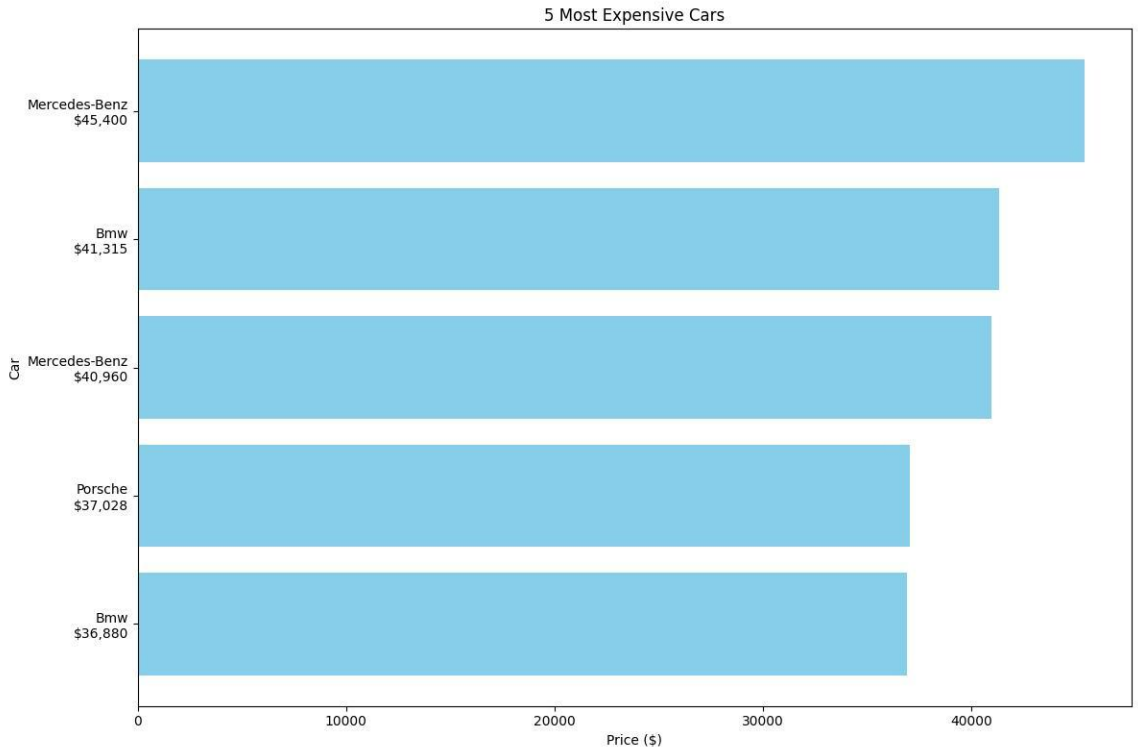

Figure 7: Top 5 Most Expensive Cars

This bar chart identifies the top five most expensive vehicles in the dataset. These vehicles, belonging to high-end manufacturers like Porsche or BMW, likely feature luxury interiors, powerful engines, and advanced safety or entertainment systems. The significant jump in price compared to average cars underlines the impact of brand prestige and performance specifications. This chart also offers a glimpse into the upper echelon of automobile pricing within the dataset.

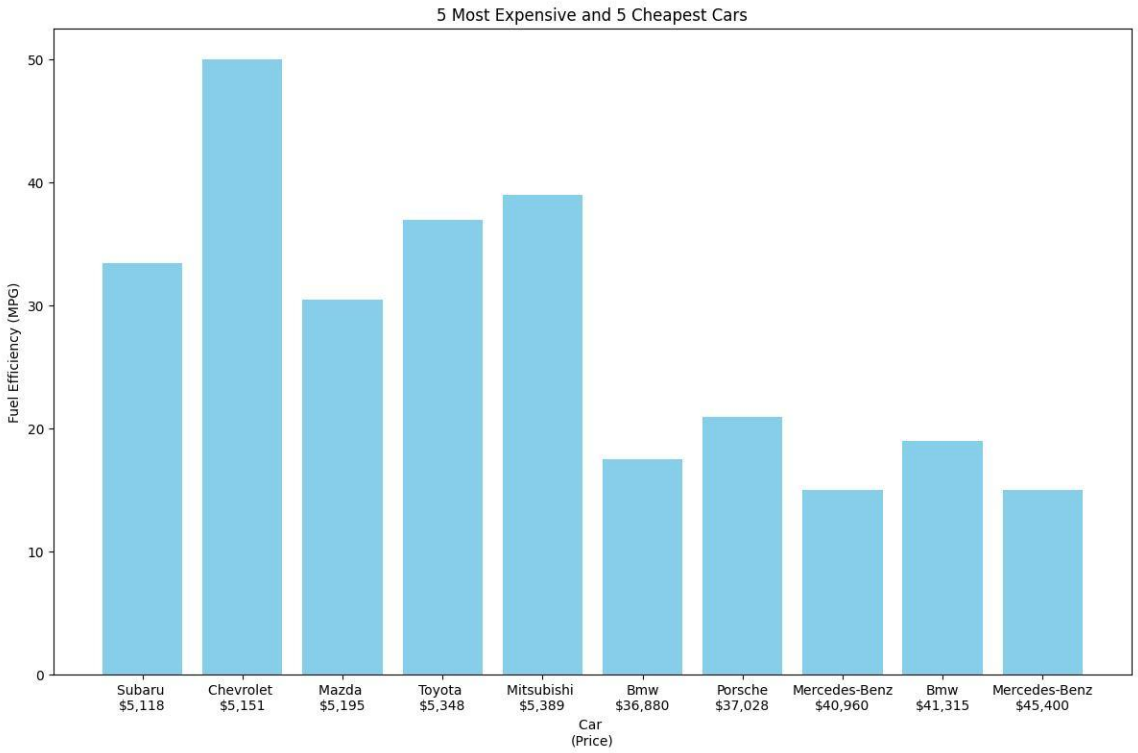

Figure 8: Fuel Efficiency of the Top 5 Most and Least Expensive Cars

In this comparative bar chart, the top five most expensive and top five cheapest cars are juxtaposed based on their fuel efficiency. Interestingly, while the most expensive cars tend to have lower MPG values—indicative of performance-oriented engineering—the cheapest cars often have higher fuel efficiency. This reinforces a common trade-off in the automobile industry: affordability and fuel economy typically go hand-in-hand, while high cost often reflects power and luxury over efficiency. This visualization is particularly useful for contrasting consumer needs and manufacturer goals.

Conclusion

This exploratory analysis of the Automobile dataset reveals strong relationships between engine size, fuel efficiency, price, and vehicle features like aspiration type and fuel type. Manufacturers strategically position their vehicles across the price and efficiency spectrum to cater to varying market segments—from budget-conscious consumers seeking fuel-efficient cars to enthusiasts desiring high-performance luxury vehicles. Key insights include:

- A negative correlation between engine size and MPG.

- Turbocharged and diesel engines generally cost more but vary in fuel savings.
- Manufacturers differ significantly in the average fuel economy of their vehicles.

Overall, the visualizations provide a well-rounded understanding of how technical specifications and brand choices affect consumer-relevant variables such as fuel economy and cost. These findings can aid both buyers in making informed decisions and analysts in building predictive models or performing further segmentation.