Designing an Energy-Efficient Car: Analyzing Factors Contributing to Fuel Efficiency

Author(s): Pinkesh Tandel and Dennis Darko

Institution: Northeastern University

College: College of Professional Studies

Location: Vancouver, BC, Canada

Course: ALY6020: Predictive Analytics

Date: September 29, 2024

Under the guidance of Dr. Nina Rajabi Nasab

Abstract

This study analyzes vehicle attributes to help a car manufacturer struggling with sales due to fuel-inefficient vehicles. By exploring various attributes such as horsepower, weight, and model year, we aim to predict miles per gallon (MPG) using a linear regression model. Data cleansing techniques were applied, and features were selected based on their contribution to higher fuel efficiency. The model achieved an R^2 score of 0.84 and an RMSE of 2.92, indicating a good predictive fit. The findings suggest that reducing weight, optimizing horsepower, and focusing on newer technologies can significantly improve fuel efficiency.

Introduction

Fuel efficiency is a key concern for consumers and automobile manufacturers alike. A major car manufacturer, known for producing large vehicles, is seeking to design more energy-efficient cars to increase sales and meet customer demand. The goal of this study is to analyze which attributes contribute to higher MPG, allowing the manufacturer to make informed design decisions.

Using data from historical vehicle performance, this study applies linear regression modeling to predict MPG and identify significant attributes such as horsepower, weight, and model year. The model's performance will be evaluated to provide actionable insights on designing fuel-efficient automobiles.

Data Cleansing Techniques

Several data cleansing techniques were applied to ensure high-quality data for modeling:

1. **Handling Missing Data**: The dataset contained missing values in the Horsepower column. Missing values were imputed using the median to avoid skewing the distribution.

2. **Outlier Detection and Capping**: Boxplots were used to detect outliers in key attributes like MPG, Horsepower, and Weight. Extreme values were capped at the 1st and 99th percentiles to minimize their impact on the model.

3. **Multicollinearity Check**: Multicollinearity was assessed using the Variance Inflation Factor (VIF). Features like Cylinders and Displacement were removed due to high VIF scores, which would have negatively impacted the model's accuracy.

Building the Linear Regression Model

A linear regression model was built to predict MPG, with the following independent variables: Horsepower, Weight, Acceleration, Model Year, and US Made. These variables were selected after eliminating multicollinear features. The model's accuracy was evaluated using Root Mean Square Error (RMSE) and R^2 scores.

Key findings from the model include:
- **Horsepower**: Higher horsepower is negatively correlated with MPG, as more powerful

engines consume more fuel.
- **Weight**: Heavier vehicles have lower MPG, highlighting the importance of weight reduction in fuel-efficient car designs.
- **Acceleration**: Vehicles with better acceleration tend to have slightly higher MPG.
- **Model Year**: Newer models exhibit higher MPG, likely due to technological advancements.

Model Optimization and Feature Selection

To optimize the model, multicollinear variables were removed based on VIF analysis. The final model retained Horsepower, Weight, Acceleration, Model Year, and US Made as the most significant predictors of MPG.

By focusing on these attributes, the car manufacturer can design vehicles that prioritize fuel efficiency. Reducing vehicle weight, optimizing engine horsepower, and leveraging advancements in newer models are key strategies for improving MPG.

Results and Visualizations

The linear regression model demonstrated a good predictive fit, achieving an R^2 score of 0.84 and an RMSE of 2.92. These metrics indicate that the model explains 84% of the variability in MPG, and the average prediction error is 2.92 MPG.

The following visualizations illustrate the key findings:

1. **Boxplots**: Outliers before and after capping.
2. **Correlation Matrix**: Multicollinearity between attributes.
3. **Actual vs. Predicted MPG**: A scatter plot showing how well the model predicts MPG values.

Conclusion

This study demonstrates that reducing vehicle weight, optimizing horsepower, and using newer technologies are critical factors for improving fuel efficiency. The linear regression model provides actionable insights that can guide the car manufacturer in designing more energy-efficient automobiles. By focusing on these key attributes, the manufacturer can meet consumer demand for fuel-efficient cars while maintaining competitive market positioning.

Appendix

Figure 1: Boxplots showing outliers before and after capping.

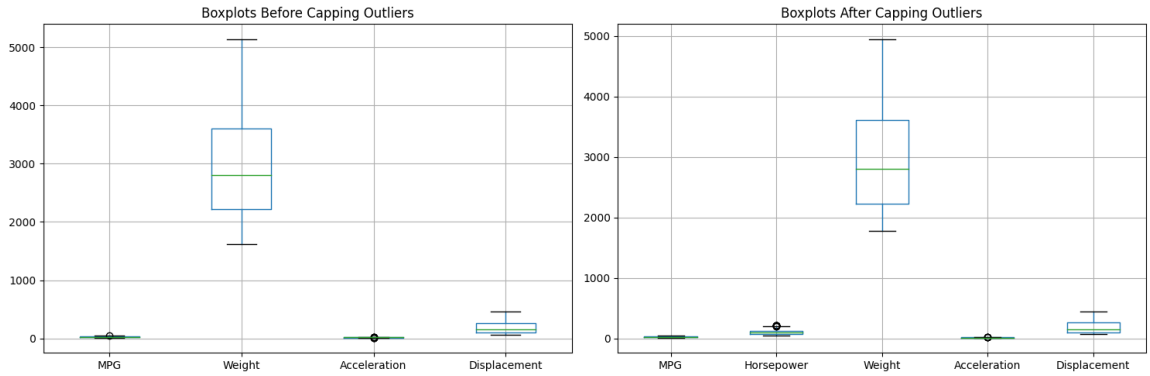

Figure 2: Correlation matrix showing multicollinearity between variables.

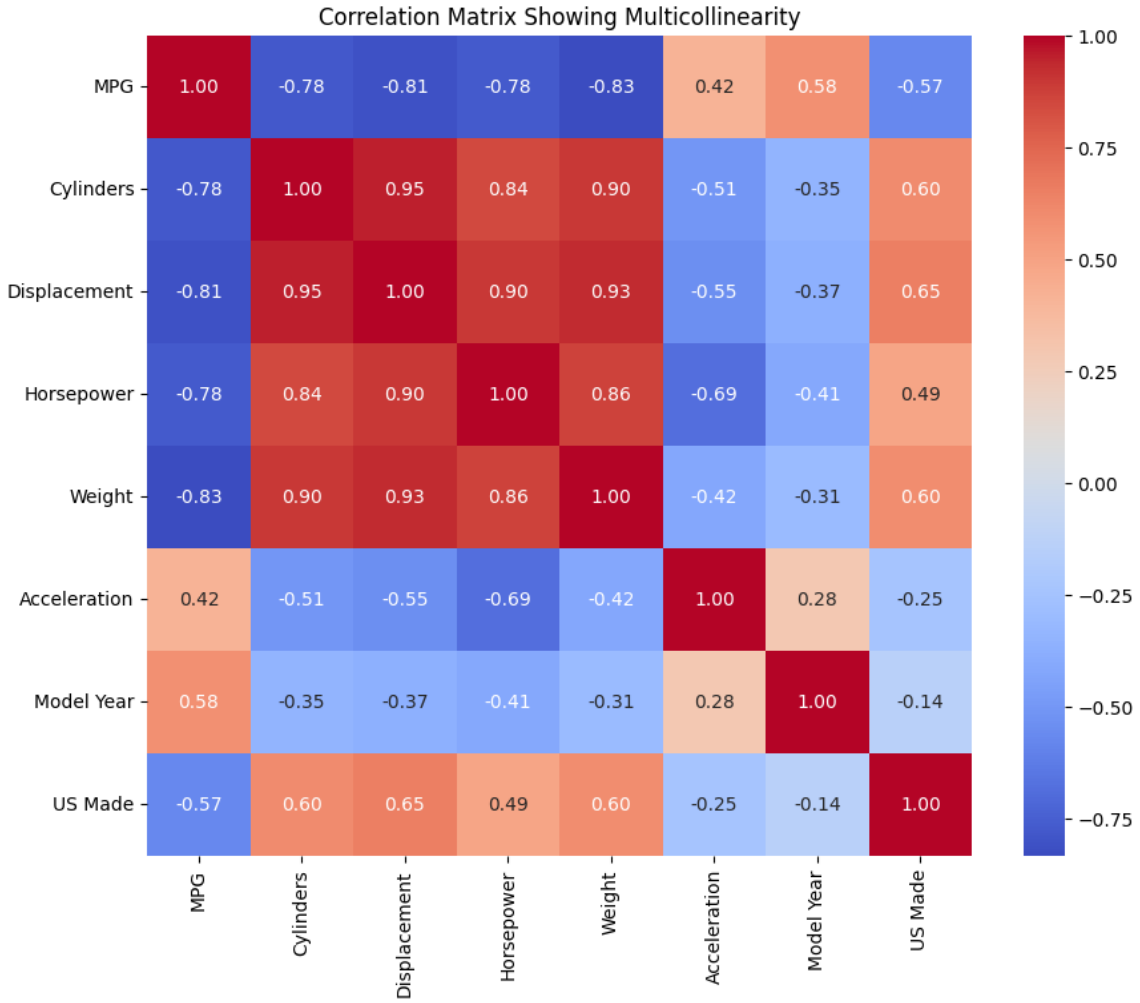

Figure 3: Actual vs. Predicted MPG plot from the linear regression model.