Sean McLean
ALY 6020
Module 2 Project
Building The Car Of The Future

Introduction

The automotive industry is undergoing continuous transformation, driven by technological innovations and evolving consumer preferences. A major challenge for car manufacturers today is improving fuel efficiency while also maintaining performance and comfort. To tackle the issue of energy efficiency, the goal of this assignment is to analyze a comprehensive car dataset to pinpoint the attributes that significantly influence gas mileage (MPG). By identifying these key factors, we aim to provide actionable insights to guide the development of a more fuel-efficient car, helping automakers regain a competitive edge in the market.

Analysis

To tackle the challenge of designing an energy-efficient car, we began by obtaining a comprehensive dataset containing various automobile attributes. Our first step was extensive data cleansing to ensure the data was of high quality and reliable. This process involved handling missing values by either removing rows with insufficient data or adding values using the median. Additionally, we converted categorical variables into numerical formats and normalized the data to maintain consistency across different scales (Anaconda, 2024).

Once the data was clean, we conducted exploratory data analysis (EDA) to understand the distribution and relationships between the variables. We used visualization techniques such as scatter plots and residual plots to identify potential predictors of fuel efficiency. Armed with these insights, we developed a linear regression model to enhance the relationship between miles per gallon (MPG) and various car attributes. We focused on model year, acceleration, and displacement, as these showed significant correlations with MPG. Feature selection techniques helped refine the model, ensuring it was both accurate and interpretable. We evaluated the

model's performance using metrics like R-squared and Mean Squared Error (MSE), confirming that newer models, vehicles with moderate acceleration, and optimized displacement are crucial for achieving higher fuel efficiency (Anaconda, 2024).

<u>Findings</u>

After training the regression model, the results show that the only two variables with positive coefficient values are 'model year' and 'displacement,' suggesting now that most of the variables do not have a strong impact on miles per gallon of vehicles. The mean square error and coefficient of determination were accurate and contained a few errors. The variable with the highest negative value was 'weight' that was just under zero. The worst score among the variables was 'US Made' which indicates that where the vehicle comes from does not have an impact on how good the miles per gallon are on the car.

Using the feature selection process with five variables, then four variables, and then the three strongest variables by the correlation coefficient values, there were noticeable differences between the three results. The five variables selected after initializing the recursive feature elimination function (RFE) were 'cylinders,' 'horsepower,' 'acceleration,' 'model year,' and 'US made.' The mean square errors of these variables were just over 14 and the R squared value was 0.725, indicating that this model in its current state is accurate overall. The only variable with a positive coefficient was the model year, which suggests this variable from the regression model execution and analysis has a positive effect on the dependent variable.

Removing a variable from the feature selection process and running the same code again, the four variables selected were 'cylinders,' 'horsepower,' 'acceleration,' 'model year,' and 'US made.' The results were different in that the coefficients and the mean squared error were higher,

and the coefficient of determination was less accurate than the previous model that featured more variables. The one outcome that stood out was that the model year variable was not only still positive but was larger than the previous model. The variable is significant in that it can impact higher gas mileage and help automakers build more fuel-efficient vehicles in the future.

After the feature selection model was adjusted to three variables, the decision was made to use the three highest variables in terms of value which were 'model year,' 'weight,' and 'acceleration.' The model execution showed less mean squared error than the previous model and had the exact same coefficient of determination value. One potential reason for this is that both models contain the variables 'model year' and 'acceleration' which could affect the accuracy of the model overall. Once again, the model year variable had a positive coefficient as well as the acceleration variable, and the weight variable was just slightly below zero just like in the initial model with every attribute. The variables were plotted to show the relationship to the dependent variable with the car year attribute having the strongest positive linear relationship to the miles per gallon variable (See Appendix A).

Visualizing the model's performance and the relationships between variables is crucial for understanding and communicating the results. A scatter plot with regression line helps visualize the relationship between the actual MPG values and the predicted MPG values. The scatterplot points were close to the regression line, signifying high model accuracy and that the model's predictions closely match the actual values and capture the relationship between independent and dependent variables effectively. This also indicates the errors are low as the residuals which are the differences between predicted and actual values are minimal (See Appendix B).

A residual plot was also created to show the distribution of errors and checking for patterns that indicate model issues. With the residual plot points spread out, it suggests that the residuals are not consistently close to zero and may indicate issues like an incomplete model fit. The spread can also reveal patterns or variations in residuals, indicating that the model might not fully capture the relationship between variables or that it may need an adjustment (Anaconda, 2024) (See Appendix C).

Recommendations

Based on the analysis, if the car model year and acceleration attributes significantly impact gas mileage (MPG), several recommendations can be made to the car manufacturer. It is essential to focus on newer technologies and designs, and leveraging the latest advancements like engine design can enhance fuel economy. Newer car models should incorporate these efficient technologies, and ensuring frequent updates to car models to integrate the latest fuel-efficient technologies can significantly improve MPG over time.

To enhance fuel efficiency for newer car models, focus on integrating the latest technologies including advancements in areas like engine design and aerodynamics which would increase MPG. Ensure that each model year benefits from incremental improvements such as lighter materials and optimized engine performance. From an advertising standpoint, promote the fuel efficiency of newer models in marketing campaigns to help consumer interest and drive sales. Additionally, continuously analyze MPG data across different model years to identify trends and guide future design decisions, ensuring that each new model incorporates the best practices for achieving higher fuel efficiency. This research could also be applied to weights of

vehicles, looking at how car designs in the future can reduce vehicle weight without compromising safety or performance.

Lastly, designing vehicles with balanced acceleration capabilities and promoting driving modes that optimize fuel economy that can maintain fuel efficiency. Continuously collecting and analyzing data on vehicle performance and fuel economy to identify areas for improvement. This would be coupled with feedback systems that can provide real-time information to drivers, refining designs while encouraging fuel-efficient practices. Marketing fuel-efficient models by highlighting their benefits in campaigns and offering incentives for customers who choose more fuel-efficient models can also promote fuel-efficient choices.
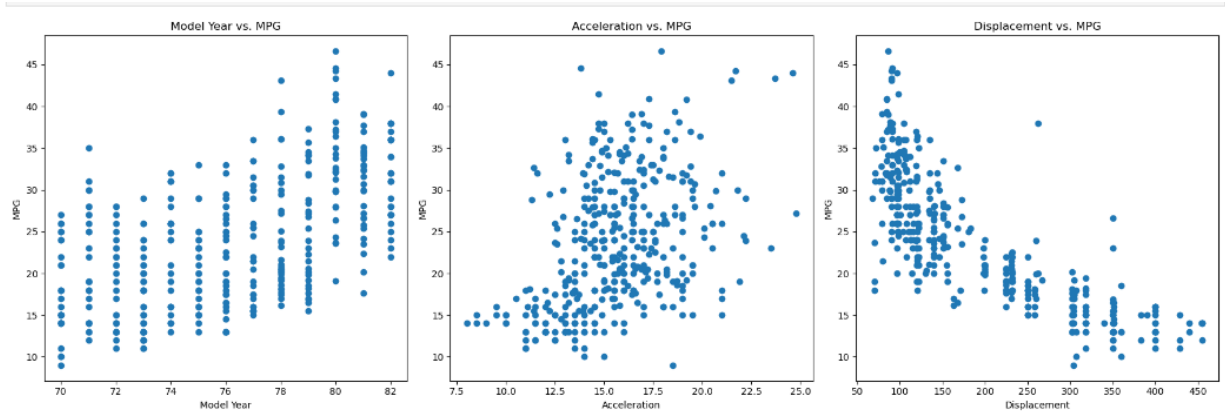
Conclusion

The analysis of the car dataset has provided valuable insights into the factors that influence fuel efficiency. By applying rigorous data cleansing, modeling, and feature selection, we identified that attributes such as model year, acceleration, and car weight have a significant impact on miles per gallon (MPG). Our linear regression model indicates that newer models, vehicles with moderate acceleration, and optimized car weights tend to achieve better fuel efficiency. These findings offer a clear direction for automakers to concentrate their design and engineering efforts on the most impactful attributes. By leveraging these insights, automakers can create a more energy-efficient vehicle, potentially reversing the decline in sales and strengthening their position in the competitive automotive market.
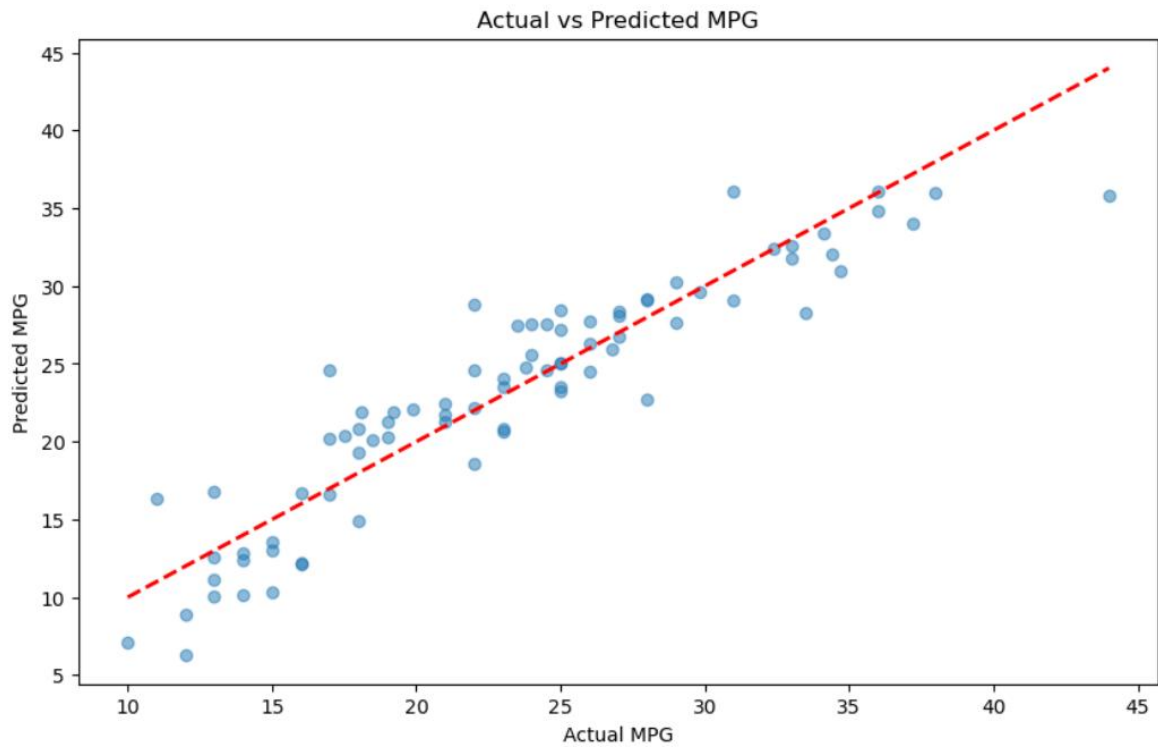
References

Anaconda Distribution. (2024). Anaconda (Version 4.12.3). Retrieved July 21, 2024, from

https://www.anaconda.com

Appendix

Appendix A

Appendix B

Actual vs Predicted MPG

Appendix C

Residuals vs Actual MPG