A Study On the Relationship Between Car Specifications and Fuel Economy Using Multivariate Linear Regression

Yujie Wu, 1003968904

22/12/2020

Abstract

The fuel economy of an automobile is not only a very important indicator for normal families to choose a vehicle, but it also determines the extent of damage to the natural environment that driving this vehicle can potentially cause. In this report, we will be using multivariate linear regression to explore the relationship between fuel economy (in terms of MPG) and the specs of a car. Specifically, we will be analyzing how engine displacement, air aspiration method, drive mode and transmission type will affect fuel economy. Since the dataset that we are going to perform analysis are based on 2021 models, the conclusion is especially aiming for automobiles produced in the 2021 model year. Our findings displayed a negative relationship between engine displacement and the MPG rates. As for other specifications, results vary and will be discussed in detail in later parts.

Keywords

Fuel Economy, City/Highway/Combined MPG (miles per gallon), Engine Displacement, Air Aspiration Method, Transmission, Driving Mode, Multivariate Linear Regression

Introduction

As the world economy has developed step by step in these few decades, people's life qualities are growing gradually and they get more spare money to distribute. Meanwhile, the cost of purchasing a car compare to the past has also dropped with the establishment of new automobile manufacturers. Therefore, the majority of people or families choose to buy one or more cars to improve their quality of life. A study from "Statista 2020" has shown that the number of cars sold in 2018 one year (78.9 million units) exceeded the sum in the previous 8 years (2010-2018: 69.7 million units) (Statista 2020). With the dramatic increases in automobile annual sales, its impact on the natural environment can not be ignored. A car with better fuel economy will cause less damage to the environment. Moreover, fuel economy has become a household issue, since refueling is the biggest cost of owning a car if the depreciation is ignored (Wade, 2020).

MPG, also known as miles per gallon, it denotes how far (in terms of miles) a car can go by only consuming one gallon of gasoline/diesel. This is an important measurement of fuel economy. With higher MPG rates, cars go farther under the same consumption than the cars with lower MPG rates, which eventually corresponds to spend less on refueling and cause less damage to the environment. The MPG rate can be affected by many aspects of a car. From common sense, cars with more cylinders will have less MPG rates. Engine displacement is also inversely related to the fuel economy in general. However, it is not absolute and there are definitely some outliers. In order to figure out a more accurate relationship with fuel economy, this report will take various aspects into consideration.

One dataset will be used to investigate how car specifications will affect fuel economy. In the Methodology section, there will be an introduction to the dataset that will be used. After that, part of the variables from that dataset will be selected to obtain a multivariate linear regression model. There will also be a section to explicitly explain the meaning of each variable. Moreover, a Result section is also included in this report, which will include statistical analysis of the relationship between fuel economy and car specs from the model. Finally, a short summary and a comprehensive conclusion of this report will be included in the Discussion section.

Methodology (Data and Model)

As mentioned above, this section will be a detailed introduction to the dataset that will be using, as well as the multivariate linear regression model.

Data

The *"2021 FE Guide for DOE"* dataset:

This dataset was provided by EPA (The United States Environmental Protection Agency). The data in this dataset is obtained from the experiments performed by the Environmental Protection Agency's National Vehicle and Fuel Emissions Laboratory. Each entry in this dataset is a car model manufactured in the 2021 model year and the dataset contains two types of information in general for each entry. The first type is the general information of the car model, such as model year, manufacture name, division, carline, transmission, number of cylinders and air aspiration method… These are mostly categorical data with few exceptions. The rest of variables are fuel economy stats, such as city FE, highway FE and combined FE, which are all numerical data. The goal of this report is to use the general information of a given car model to predict the fuel economy figures of that car. Therefore, this dataset is selected as it is extremely comprehensive and it provides every specification that can potentially influence the fuel economy of a car. The source of the data is also reliable since the data collection process was oversight by EPA, which is the official US. government source for fuel economy information. One weakness of this dataset is that this dataset only contains information for cars manufactured in the 2021 model year.

The *"2021 FE Guide for DOE"* dataset can be downloaded from: https://www.fueleconomy.gov/feg/download.shtml

Table 1: Baseline Characteristics Table

```
##                                                                     Stratified by comb FE
##                                                                      10 to 19
##   n                                                                   254
##   Eng Displ (mean (SD))                                              4.84 (1.12)
##   Air Aspiration Method Desc (%)
##      Naturally Aspirated                                              125 (49.2)
##      Other                                                              1 ( 0.4)
##      Supercharged                                                      17 ( 6.7)
##      Turbocharged                                                     111 (43.7)
##      Turbocharged+Supercharged                                          0 ( 0.0)
##   Trans Desc (%)
##      Automated Manual                                                  12 ( 4.7)
##      Automated Manual- Selectable (e.g. Automated Manual with paddles)  25 ( 9.8)
##      Automatic                                                        113 (44.5)
##      Continuously Variable                                              0 ( 0.0)
##      Manual                                                            14 ( 5.5)
```

```
##     Selectable Continuously Variable (e.g. CVT with paddles)           0 ( 0.0)
##     Semi-Automatic                                                    90 (35.4)
##   Drive Desc (%)
##     2-Wheel Drive, Front                                               1 ( 0.4)
##     2-Wheel Drive, Rear                                               87 (34.3)
##     4-Wheel Drive                                                     80 (31.5)
##     All Wheel Drive                                                   77 (30.3)
##     Part-time 4-Wheel Drive                                            9 ( 3.5)
##                                                                    Stratified by comb FE
##                                                                     20 to 29
##   n                                                                  570
##   Eng Displ (mean (SD))                                             2.59 (0.68)
##   Air Aspiration Method Desc (%)
##     Naturally Aspirated                                              158 (27.7)
##     Other                                                              7 ( 1.2)
##     Supercharged                                                       6 ( 1.1)
##     Turbocharged                                                     393 (68.9)
##     Turbocharged+Supercharged                                          6 ( 1.1)
##   Trans Desc (%)
##     Automated Manual                                                  15 ( 2.6)
##     Automated Manual- Selectable (e.g. Automated Manual with paddles)   49 ( 8.6)
##     Automatic                                                        144 (25.3)
##     Continuously Variable                                              6 ( 1.1)
##     Manual                                                            38 ( 6.7)
##     Selectable Continuously Variable (e.g. CVT with paddles)          22 ( 3.9)
##     Semi-Automatic                                                   296 (51.9)
##   Drive Desc (%)
##     2-Wheel Drive, Front                                             135 (23.7)
##     2-Wheel Drive, Rear                                              133 (23.3)
##     4-Wheel Drive                                                     75 (13.2)
##     All Wheel Drive                                                  214 (37.5)
##     Part-time 4-Wheel Drive                                           13 ( 2.3)
##                                                                    Stratified by comb FE
##                                                                     30 to 39
##   n                                                                  108
##   Eng Displ (mean (SD))                                             1.89 (0.43)
##   Air Aspiration Method Desc (%)
##     Naturally Aspirated                                               76 (70.4)
##     Other                                                              0 ( 0.0)
##     Supercharged                                                       0 ( 0.0)
##     Turbocharged                                                      32 (29.6)
##     Turbocharged+Supercharged                                          0 ( 0.0)
##   Trans Desc (%)
##     Automated Manual                                                   2 ( 1.9)
##     Automated Manual- Selectable (e.g. Automated Manual with paddles)    8 ( 7.4)
##     Automatic                                                          0 ( 0.0)
##     Continuously Variable                                             27 (25.0)
##     Manual                                                            18 (16.7)
##     Selectable Continuously Variable (e.g. CVT with paddles)          33 (30.6)
##     Semi-Automatic                                                    20 (18.5)
##   Drive Desc (%)
##     2-Wheel Drive, Front                                              91 (84.3)
##     2-Wheel Drive, Rear                                                1 ( 0.9)
##     4-Wheel Drive                                                      0 ( 0.0)
```

```
##      All Wheel Drive                                                     16 (14.8)
##      Part-time 4-Wheel Drive                                              0 ( 0.0)
##                                                                       Stratified by comb FE
##                                                                        40 to 49
##   n                                                                      11
##   Eng Displ (mean (SD))                                                2.16 (0.35)
##   Air Aspiration Method Desc (%)
##      Naturally Aspirated                                                 11 (100.0)
##      Other                                                                0 (  0.0)
##      Supercharged                                                         0 (  0.0)
##      Turbocharged                                                         0 (  0.0)
##      Turbocharged+Supercharged                                            0 (  0.0)
##   Trans Desc (%)
##      Automated Manual                                                     0 (  0.0)
##      Automated Manual- Selectable (e.g. Automated Manual with paddles)    1 (  9.1)
##      Automatic                                                            0 (  0.0)
##      Continuously Variable                                                4 ( 36.4)
##      Manual                                                               0 (  0.0)
##      Selectable Continuously Variable (e.g. CVT with paddles)             6 ( 54.5)
##      Semi-Automatic                                                       0 (  0.0)
##   Drive Desc (%)
##      2-Wheel Drive, Front                                                 9 ( 81.8)
##      2-Wheel Drive, Rear                                                  0 (  0.0)
##      4-Wheel Drive                                                        0 (  0.0)
##      All Wheel Drive                                                      1 (  9.1)
##      Part-time 4-Wheel Drive                                              1 (  9.1)
##                                                                       Stratified by comb FE
##                                                                        50 to 59
##   n                                                                       8
##   Eng Displ (mean (SD))                                                1.82 (0.32)
##   Air Aspiration Method Desc (%)
##      Naturally Aspirated                                                  8 (100.0)
##      Other                                                                0 (  0.0)
##      Supercharged                                                         0 (  0.0)
##      Turbocharged                                                         0 (  0.0)
##      Turbocharged+Supercharged                                            0 (  0.0)
##   Trans Desc (%)
##      Automated Manual                                                     0 (  0.0)
##      Automated Manual- Selectable (e.g. Automated Manual with paddles)    3 ( 37.5)
##      Automatic                                                            0 (  0.0)
##      Continuously Variable                                                4 ( 50.0)
##      Manual                                                               0 (  0.0)
##      Selectable Continuously Variable (e.g. CVT with paddles)             1 ( 12.5)
##      Semi-Automatic                                                       0 (  0.0)
##   Drive Desc (%)
##      2-Wheel Drive, Front                                                 8 (100.0)
##      2-Wheel Drive, Rear                                                  0 (  0.0)
##      4-Wheel Drive                                                        0 (  0.0)
##      All Wheel Drive                                                      0 (  0.0)
##      Part-time 4-Wheel Drive                                              0 (  0.0)
##                                                                       Stratified by comb FE
##                                                                        p
##   n
##   Eng Displ (mean (SD))                                                <0.001
```

```
##   Air Aspiration Method Desc (%)                                       <0.001
##      Naturally Aspirated
##      Other
##      Supercharged
##      Turbocharged
##      Turbocharged+Supercharged
##   Trans Desc (%)                                                       <0.001
##      Automated Manual
##      Automated Manual- Selectable (e.g. Automated Manual with paddles)
##      Automatic
##      Continuously Variable
##      Manual
##      Selectable Continuously Variable (e.g. CVT with paddles)
##      Semi-Automatic
##   Drive Desc (%)                                                       <0.001
##      2-Wheel Drive, Front
##      2-Wheel Drive, Rear
##      4-Wheel Drive
##      All Wheel Drive
##      Part-time 4-Wheel Drive
##                                                                       Stratified by comb FE
##                                                                        test
##   n
##   Eng Displ (mean (SD))
##   Air Aspiration Method Desc (%)
##      Naturally Aspirated
##      Other
##      Supercharged
##      Turbocharged
##      Turbocharged+Supercharged
##   Trans Desc (%)
##      Automated Manual
##      Automated Manual- Selectable (e.g. Automated Manual with paddles)
##      Automatic
##      Continuously Variable
##      Manual
##      Selectable Continuously Variable (e.g. CVT with paddles)
##      Semi-Automatic
##   Drive Desc (%)
##      2-Wheel Drive, Front
##      2-Wheel Drive, Rear
##      4-Wheel Drive
##      All Wheel Drive
##      Part-time 4-Wheel Drive
```

From table 1, we can easily observe that the comb FE for the majority of the cars is between 20 to 29. it is clear that the comb FE is decreasing as the Eng Displ increases, this strongly suggests that there exists some sort of linear relationship between the comb FE and Eng Displ. Moreover, cars with larger comb FE (e.g. 40 to 49 and 50 to 50) are all naturally aspirated and supercharged and turbocharged cars' comb FE are relatively lower. However, there are also 49.2% of cars that have low comb FE but is naturally aspirated. As for the transmission method, cars with continuously variable and Selectable Continuously Variable (e.g. CVT with paddles) transmission types generally have higher comb FE, the majority of them fall into the interval of 30 to 39. Finally, 91 cars in the dataset are front-wheel drive and have comb FE within 30 to 39, which almost occupies 85% of the population of this interval.

Model

In this report, I will be using a multivariate linear regression model to model how will the specifications of a car affects its fuel economy. The main reason for selecting a multivariate linear regression model is that we assume all the car specifications, which are the independent variables, have linear relationships with the dependent variable, which is a numerical variable and is the fuel economy of a car model. As previously stated, the goal is to predict the fuel economy figure of a car model given its specifications. Therefore, a linear regression model best fits in this scenario. In order to make the model more accurate, I tried to take as many related specifications into account as possible. The selected specifications are Eng Displ, Air Aspiration Method Desc, Trans Desc, Drive Desc and, without losing generality, combined fuel economy (Comb FE (Guide) - Conventional Fuel) will be used as a measurement of the fuel economy.
Specifically:

- "Eng Displ" is a numerical variable with a value from 1.0 to 8.0. This represents the sum of volume that the pistons swiped in all cylinders. Cars with higher engine displacement can burst out more power, and consequently, they will consume more fuel. Thus this variable is selected. Notice that number of cylinders is positively related to engine displacement, thus the number of cylinders is not selected as a predictor.
- "Air Aspiration Method Desc" is a categorical variable with five possible values: "Turbocharged", "Naturally Aspirated", "Supercharged", "Supercharged+Turbocharged" and "Other". This variable indicates how does air enter the engine. Generally, naturally aspirated engines are equipped on those baseline cars while performance cars usually have turbocharged or supercharged engines.
- "Trans Desc" is a categorical variable with seven possible values: "Automated Manual- Selectable (e.g. Automated Manual with paddles)", "Automated Manual", "Automatic", "Semi-Automatic", "Manual", "Selectable Continuously Variable (e.g. CVT with paddles)" and "Continuously Variable". This represents the transmission method of a car model. This variable is selected since the transmission type can affect how the car works, for example, manual cars can put the gear into neutral while driving, which may save some gas.
- "Drive Desc" is a categorical variable with five possible values: "All Wheel Drive", "2-Wheel Drive, Rear", "Part-time 4-Wheel Drive", "4-Wheel Drive" and "2-Wheel Drive, Front". This variable represents which wheels the engine will transmit its power. This variable also can potentially affect the fuel economy since part of the engine power can be consumed or converted to other energy types during transportation. With more power lost, the engine needs to work harder to get to the same level.

Below is the linear regression model we are using:

$$
\begin{gathered}
MPG_{combined} = \beta_0 + \beta_1 X_{ED} \\
+\beta_2 X_{AAMO} + \beta_3 X_{AAMS} + \beta_4 X_{AAMT} + \beta_5 X_{AAMTS} \\
+\beta_6 X_{TAMS} + \beta_7 X_{TA} + \beta_8 X_{TCV} + \beta_9 X_{TM} + \beta_{10} X_{TSCV} + \beta_{11} X_{TSA} \\
+\beta_{12} X_{2WDR} + \beta_{13} X_{4WD} + \beta_{14} X_{AWD} + \beta_{15} X_{PT4WD} \\
+\epsilon
\end{gathered}
$$

- $MPG_{combined}$ is the combined MPG of a car, which corresponds to the variable "combined fuel economy (Comb FE (Guide) - Conventional Fuel)".

- X_{ED} represents the engine displacement of a car, which corresponds to the variable "Eng Displ".

- $< \beta_2 X_{AAMO}, \beta_3 X_{AAMS}, \beta_4 X_{AAMT}, \beta_5 X_{AAMTS} >$ represent the air aspiration method of a car, which corresponds to the variable "Air Aspiration Method Desc". Only one of them should be evaluated to 1 for each car model.

- $< \beta_6 X_{TAMS}, \beta_7 X_{TA}, \beta_8 X_{TCV}, \beta_9 X_{TM}, \beta_{10} X_{TSCV}, \beta_{11} X_{TSA} >$ represent the transmission method of a car, which corresponds to the variable “Trans Desc”. Only one of them should be evaluated to 1 for each car model.

- $< \beta_{12} X_{2WDR}, \beta_{13} X_{4WD}, \beta_{14} X_{AWD}, \beta_{15} X_{PT4WD} >$ represent the driving mode of a car, which corresponds to the variable “Drive Desc”. Only one of them should be evaluated to 1 for each car model.

When all X values are zero, this model shows the MPG of a car that is naturally aspirated, automated manual transmission, front 2 Wheel Drive, with engine displacement equals to 0. This is the same as β_0. Similarly,

- β_1 represents the slope of engine displacement on MPG rates. If β_1 is negative, it means the higher the engine displacement, the lower the MPG rate will be and vice verse.

- β_2 to β_5 represent the slope of different air aspiration method for this model. If β_i is negative, it means that the MPG rate is inversely proportional to the corresponding air aspiration method and vice verse.

- β_6 to β_{11} represent the slope of different transmission type for this model. If β_i is negative, it means that the MPG rate is inversely proportional to the corresponding transmission type and vice verse.

- β_{12} to β_{15} represent the slope of driving mode for this model. If β_i is negative, it means that the MPG rate is inversely proportional to the corresponding driving mode and vice verse.

- ϵ is the error term for this model.

Table 2: The Model

term	estimate	std.error	statistic	p.value
(Intercept)	35.0281433	0.7594900	46.1206146	0.0000000
Eng Displ	-2.8985760	0.1075873	-26.9416203	0.0000000
Air Aspiration Method DescOther	-1.3255975	1.2209163	-1.0857399	0.2778739
Air Aspiration Method DescSupercharged	-1.0095326	0.7355105	-1.3725604	0.1702182
Air Aspiration Method DescTurbocharged	-2.5494551	0.2718144	-9.3793970	0.0000000
Air Aspiration Method DescTurbocharged+Supercharged	-2.5274161	1.3958342	-1.8106851	0.0705105
Trans DescAutomated Manual- Selectable (e.g. Automated Manual with paddles)	1.6520560	0.7328935	2.2541556	0.0244172
Trans DescAutomatic	-0.2448826	0.6613406	-0.3702820	0.7112562
Trans DescContinuously Variable	6.8981250	0.8529897	8.0869973	0.0000000

term	estimate	std.error	statistic	p.value
`Trans Desc`Manual	-0.0334434	0.7492109	-0.0446382	0.9644052
`Trans Desc`Selectable Continuously Variable (e.g. CVT with paddles)	5.5469092	0.7977834	6.9529009	0.0000000
`Trans Desc`Semi-Automatic	0.2807820	0.6571478	0.4272737	0.6692784
`Drive Desc`2-Wheel Drive, Rear	-1.8857188	0.3768111	-5.0044145	0.0000007
`Drive Desc`4-Wheel Drive	-2.5119353	0.4307103	-5.8320768	0.0000000
`Drive Desc`All Wheel Drive	-2.9843572	0.3209119	-9.2996148	0.0000000
`Drive Desc`Part-time 4-Wheel Drive	-3.2847348	0.7426479	-4.4230042	0.0000109

The LM function in R is used to simulate the model of the relationship between the MPG rate and the specifications of a car and got the above summary table. Based on this summary table, we could get the estimated model as followed:

$$\widehat{MPG_{combined}} = 35.0281433 - 2.8985760X_{ED}$$
$$-1.3255975X_{AAMO} - 1.0095326X_{AAMS} - 2.5494551X_{AAMT} - 2.5274161X_{AAMTS}$$
$$+1.6520560X_{TAMS} - 0.2448826X_{TA} + 6.8981250X_{TCV} - 0.0334434X_{TM} + 5.546909210X_{TSCV} + 0.2807820X_{TSA}$$
$$-1.8857188X_{2WDR} - 2.5119353X_{4WD} - 2.9843572X_{AWD} - 3.2847348X_{PT4WD}$$

Results

Table 2 summarizes the estimation of the relationship between car specifications and fuel economy given by our multivariate linear regression model. Based on this table, the slope of engine displacement is -2.8985760, which means we estimate that the MPG rate and engine displacement are inversely related. Similarly, the slopes of other, supercharged, turbocharged and turbocharged+supercharged air aspiration methods are - 1.3255975, - 1.0095326, - 2.5494551 and - 2.5274161, which are all negative. This means the MPG rate will also decrease if the air aspiration method is switched from naturally aspirated to any of these air aspiration methods. Furthermore, the model suggests that the automated manual- selectable (e.g. Automated Manual with paddles), selectable continuously variable (e.g. CVT with paddles), semi-automatic and continuously variable transmission methods are proportional to the MPG rate since their slopes are positive numbers (1.6520560, 5.5469092, 0.2807820 and 6.8981250). Whereas the slope for manual and automatic transmissions are -0.0334434 and -0.2448826 respectively, which means they are inversely proportional to the MPG rate. Finally, all of the driving modes on the table have negative slopes with the minimum being part-time 4-wheel drive, which is -3.2847348 and maximum being 2-wheel drive, rear, which is -1.8857188. The rests are 4-wheel drive: -2.5119353 and all wheel drive: -2.9843572.

This table also provides a column with p-values for each coefficient. There are six of them (except β_0) that have a p-value equals to zero, they are other and turbocharged from air aspiration method, continuously variable and selectable continuously variable (e.g. CVT with paddles) form transmission type, 4-wheel drive and all wheel drive from driving modes. The p-values for 2-wheel drive, rear, part-time 4-wheel drive and automated manual- selectable (e.g. Automated Manual with paddles) are also lower than the common alpha level of 0.05. This means the null hypothesis can be rejected and these predictors are likely to be meaningful additions to the model (Minitab Blog editor, 2013). Whereas, the p-values for other, supercharged and supercharged+turbocharged air aspiration method, automatic, manual and semi-automatic transmission methods are all larger than 0.05, which indicates these predictors are unlikely to be meaningful additions to the model, they are not statistically significant.

Discussion

In this study, we explored the relationship between the MPG rates and car specifications using a multivariate linear regression model on the dataset named *"2021 FE Guide for DOE"*. The prediction variables that we used include are Eng Displ, Air Aspiration Method Desc, Trans Desc, Drive Desc. The result of the model can be found in table 2 and is summarized in above section. In this section, we will make a conclusion and discuss the weakness and next steps of this study.

In the above section, we stated that the engine displacement is inversely proportional to the MPG rate. Therefore, with a one unit increase of engine displacement, the MPG rate will decrease by almost 2.9. In other words, the higher the power of a car is, the faster it will drain its energy. It is always a good thing to consider the trade-off between performance and range before purchasing a car. Moreover, the model also suggests the naturally aspirated cars are the best energy-saving machines compare to other air aspiration methods cars. Naturally aspirated engines "breathe" without any help from others while there is a turbo in front of the turbocharged engine which will intervene at a certain speed. There is also a turbo-liked component in the supercharged car which alters the "breathing" process of the engine. (Wingate 2017) The power delivery of naturally aspirated cars is more smooth-going and ferocious for others. This should also be a good reference for car selection, people who put more value on fuel economy than power delivery method (performance) should pay more attention to naturally aspirated cars and vice versa.

There is no direct relationship between performance and driving mode and transmission types. However, we can still conclude that the automated manual- selectable (e.g. Automated Manual with paddles), selectable continuously variable (e.g. CVT with paddles), semi-automatic and continuously variable transmission methods have better fuel efficiency than the automated manual transmission, while the fuel efficiency for manual and automatic transmission cars is lower than the automatic manual cars, which contradicts to our initial assumption. One of the reasons why the transmission type is chosen is that we assume the manual transmission can save some gas by putting gear into neutral while driving, but the result is exactly the opposite. Lastly, the front 2 wheel drive is the most basic and most energy-saving driving mode among all options while the part time 4 wheel drive is the most fuel consuming driving mode. With one gallon of gas, the ranges for part time 4 wheel drive vehicles are almost 3.3 miles less than front wheel drive vehicles. (If other specifications stay the same)

Weakness

One weakness of this model is, as previously stated, the dataset chosen only contains information for cars manufactured in the 2021 model year. Therefore, this model is overfitting to newer cars and may produce inaccurate estimations for cars manufactured in the previous years. Moreover, cars are manufactured for different purposes, different types of vehicles may have very different fuel economies to meet the demand. Multicollinearity may exist since this model does not take the general usage of the cars into considerations. Mostly, this model only considers the combined MPG as the measurement of fuel economy due to the time and space constraints of this study/report. However, the city fuel efficiency and the highway fuel efficiency of a car may vary. Cars that have outstanding highway fuel efficiency might not have the same performance in the city.

Next Steps

One possible future work is to divide the current model into two, which means we can analyze the relationship between car specifications and the city and highway MPG individually, instead of analyzing the combined MPG. On the other hand, we can also consider the functionality of cars in the future to solve the multicollinearity issue. The website which provides the *"2021 FE Guide for DOE"* dataset also includes fuel economy guide datasets for vehicles manufactured in the past few decades. With these datasets, we can examine the performance of the model we got on previously manufactured cars and determine whether the model is overfitting for cars manufactured in the 2021 model year or not.

Github Repo

https://github.com/wuyujie1/STA304-Final-Project

References

Download Fuel Economy Data. (n.d.). Retrieved December 06, 2020, from https://www.fueleconomy.gov/feg/download.shtml

Editor, M. (n.d.). How to Interpret Regression Analysis Results: P-values and Coefficients. Retrieved December 21, 2020, from https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients

Gasoline Vehicles:Learn More About the Label. (n.d.). Retrieved December 06, 2020, from https://www.fueleconomy.gov/feg/label/learn-more-gasoline-label.shtml

Tyler Wade. (2020, April 24). What is the total cost of ownership for a car? Retrieved December 06, 2020, from https://www.ratehub.ca/blog/what-is-the-total-cost-of-owning-a-car/

Wagner, I. (2020, November 04). Global car sales 2010-2020. Retrieved December 06, 2020, from https://www.statista.com/statistics/200002/international-car-sales-since-1990/

Wingate, A. (2017, April 09). Turbocharger/ supercharger vs naturally aspirated engine. Retrieved December 21, 2020, from https://drivetribe.com/p/turbocharger-supercharger-vs-naturally-J0Yivq7KSvG3yDDH9Xvm7g?iid=EM2uGkYgRRSdkGK0qG-hrQ