**Resubmission Reflection 1**

The specifications I did not complete were the creation of testing and training data sets to test generalizability, and testing for multicollinearity using VIF testing. I did not achieve these specifications because I didn’t know how to complete them.

**Resubmission Reflection 2**

In order to complete these specifications, I went back to the in class exercises we previously completed and took notes. I used the classwork as guidelines for completing the specifications. All revisions for this memo are in the regression analysis paragraph following the **(NEW)** tag. New additions to the appendix also emboldened.

**Predicting the Average Mileage of Charged Cars Using Emission Statistics**

**Executive analysis**

Based on the analysis performed using engine emissions, the average mileage of a car who’s engine was equipped with a charger could be fairly accurately predicted. The cluster analysis scatter plot visualized the relationship between average mileage and emissions per mile. As average mileage of the cars increased, the emissions per mile decreased. The less efficient form of charger (Supercharger being represented in the lower bounds of mileage, and the more efficient form (Turbocharger) being represented at the higher bound of mileage and lower emissions. The linear regression performed showed that the mileage could be predicted within 3.31 miles to the gallon. The model accounted for 96.1% of variance within the model. When referring to miles to the gallon, an error of 3.31 is on the higher side. Using the actual values of C02 emissions instead of emission ratios would most likely serve to create a more accurate prediction model.

**Preliminary analysis**

Cars that are equipped with engine chargers tend to produce different average mileage and emissions compared to non-charged cars. The goal of this analysis was to find if there was a relationship between mileage of these cars and emissions created. Within the data set containing only charged cars, no missing values impacted the predictions model. On average, supercharged cars produced higher emissions, (Figure 1) and lower average mile per gallon. (Figure 2) Additionally, supercharged cars consumed the largest amount of fuel per year, compared to turbo and twin chargers. (Figure 3)

**Feature Engineering**

The original data set was reduced to only include cars which were equipped with either super, turbo, or twin charged engines. This cut the original data set to approximately 8000 individual cars. In order to better understand the relationship between fuel mileage and emissions, a number of features were created. The first being Emission ratios for both city and highway driving. This feature represented the C02 tailpipe emissions per mile for both city and highway driving. (Figure 4) An average of the values was then supplemented. Additionally, an average of both city and highway miles per gallon was created. (Figure 5) These features were the primary predictors used in the analysis. A gallons per year feature was tested, but was not used in the final analysis. The feature represented the barrels of fuel used per year, multiplied by 42 gallons, which represents the standard size of a barrel of petroleum.

**Cluster Analysis**

A number of sets of cluster analysis were performed. The most effective scatter plot representing the relationship between average fuel mileage and average emissions ratio. The elbow curve generated indicated that 4 clusters would be most effective in modeling the data before creating diminishing returns in adding more clusters. (Figure 6) The resulting scatter plot showed the relationship between mileage and emissions. (Figure 7) The clusters at the low end of average mileage displayed higher levels of C02 emissions. The second scatter plot generated color coded the data points. Green representing turbocharged cars, blue representing supercharged cars, and yellow representing twin charged cars. (Figure 8) In the plot, the super charged cars are on the lower end of mileage, and higher end of emissions. This supports the idea that superchargers are less effective in how they receive power from the crankshaft of the car. The points representing turbochargers are distributed closer to the higher end of mileage, and lower end of emissions. This supports the idea that powering turbochargers from the exhaust of the car is both more effective at lowering emissions, but more economical in the average mile per gallon the car receives.

**Regression Analysis**

The goal of the regression analysis was to test the accuracy of predicting a charged car’s mileage using emissions statistics. The dependent variable being average mileage. The independent variables being average emission ratio, car year, Gallons consumed per year and the C02 tailpipe emission (Grams Per Mile). (Figure 9) The model generated an R^2 score of .961, which meant that 96.1% of variance in the data was accounted for by the model. The coefficients of the features all decreased as the average miles mileage increased, which supports the idea that more fuel efficient cars produce fewer C02 emissions**.**

**(NEW)** In order to better test the generalizability of the model, the data was split into training and testing sets containing 80% and 20% of the data respectively. (Figure 10) Regression tests were then run for both the training and test data. (Figures 11, 12 respectively) The two regression tests performed similarly after being exposed to new data sets, which indicates good adaptability when exposed to new conditions. However, each of the regression models indicated strong levels of multicollinearity. A VIF test was run for the testing data (Figure 13). The “GallonsPerYear” and” C02TailpipGpm” variables both had large inflation factors, which indicates that it is difficult to accurately assess the contribution of these two variables to the overall models.

**Appendix**

Figure 1

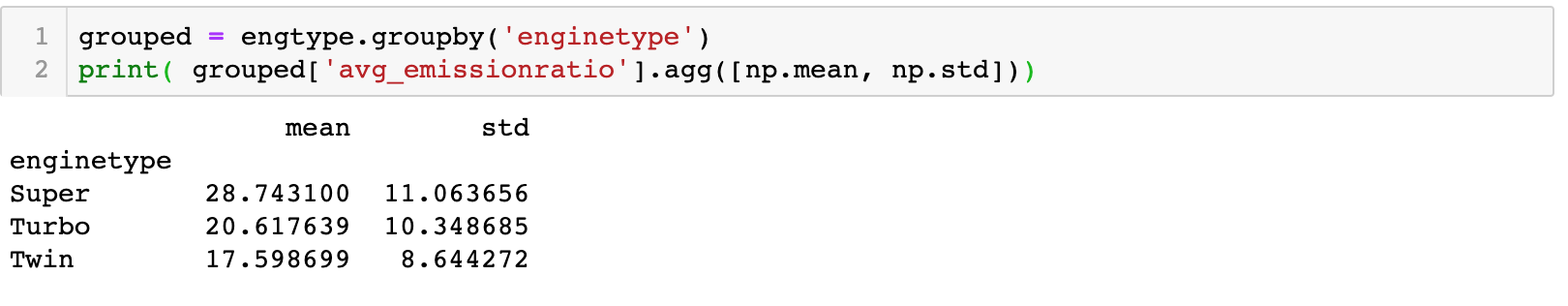


Figure 2

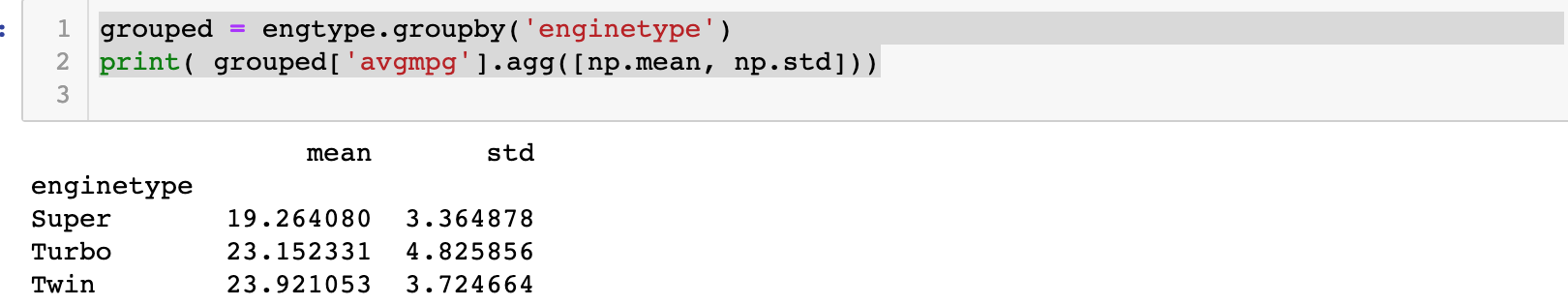


Figure 3

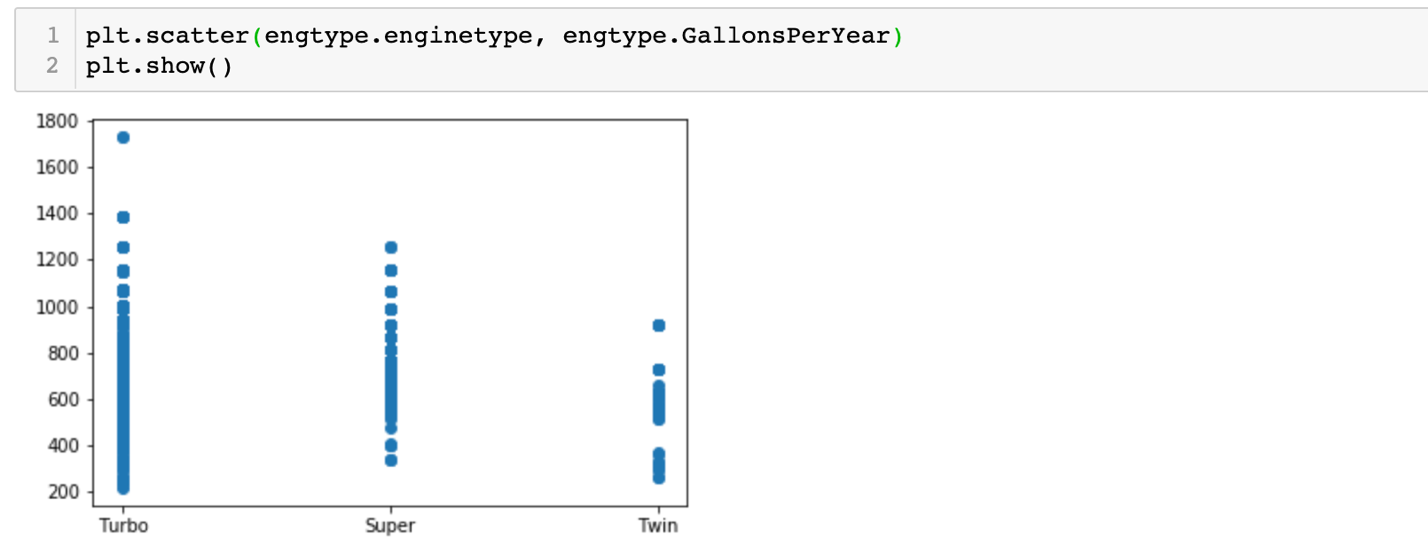


Figure 4

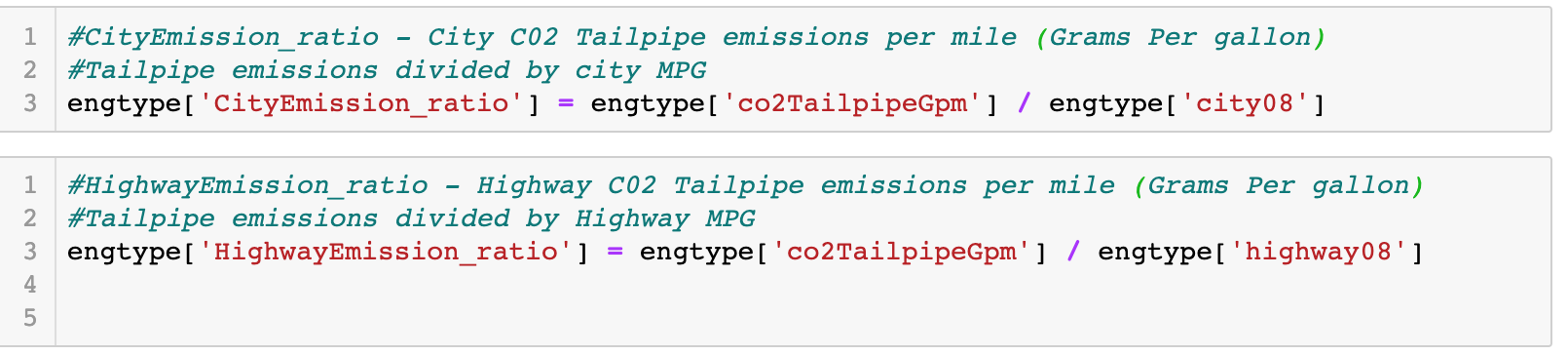


Figure 5



Figure 6

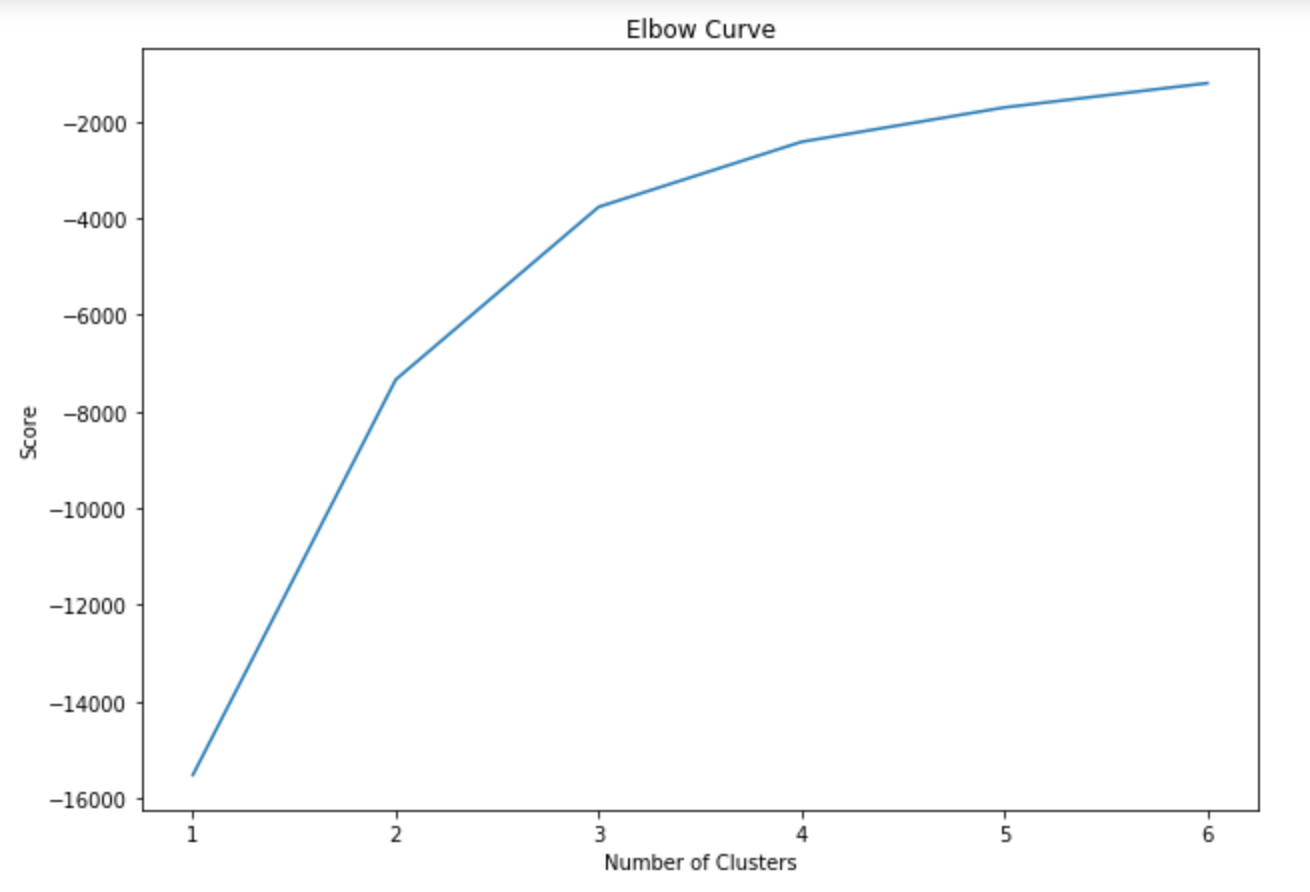


Figure 7

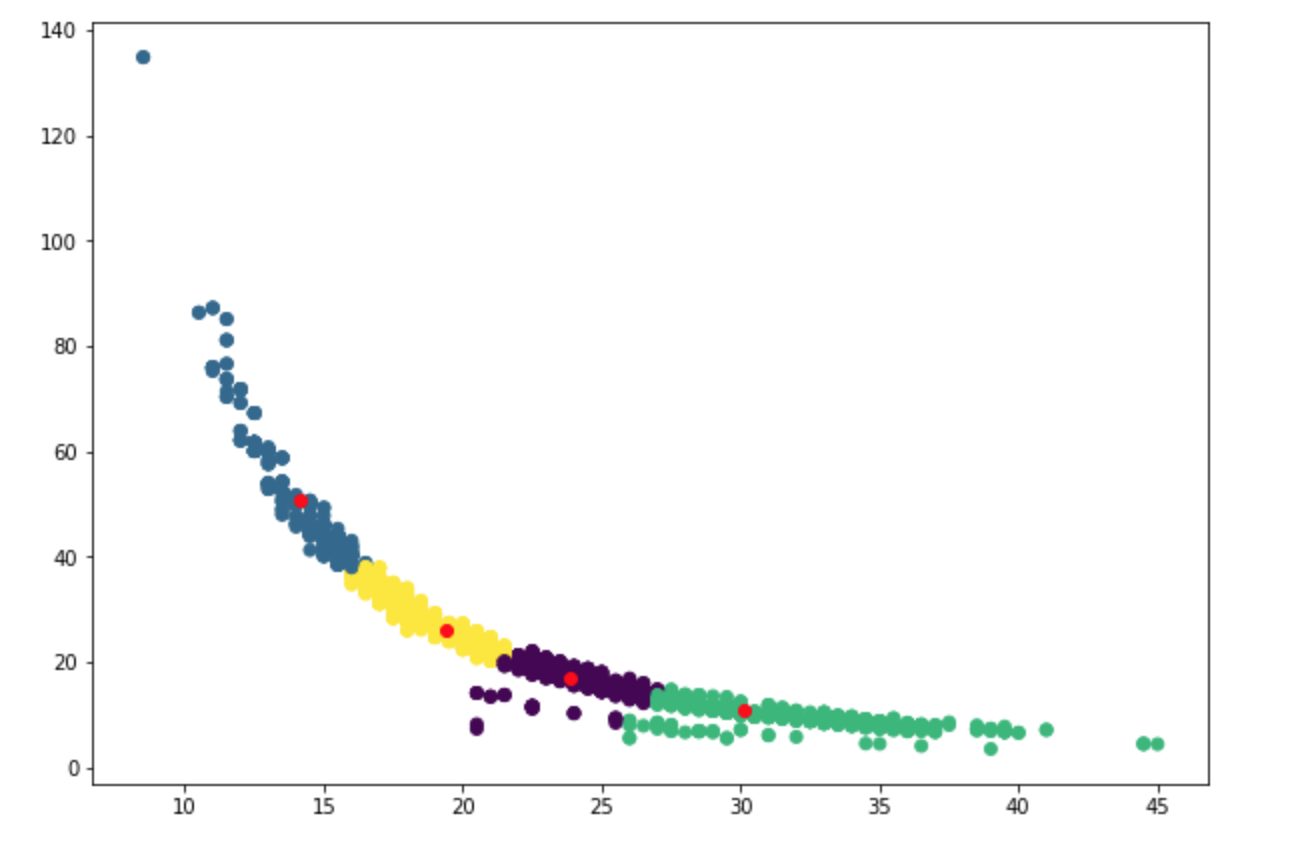


Figure 8

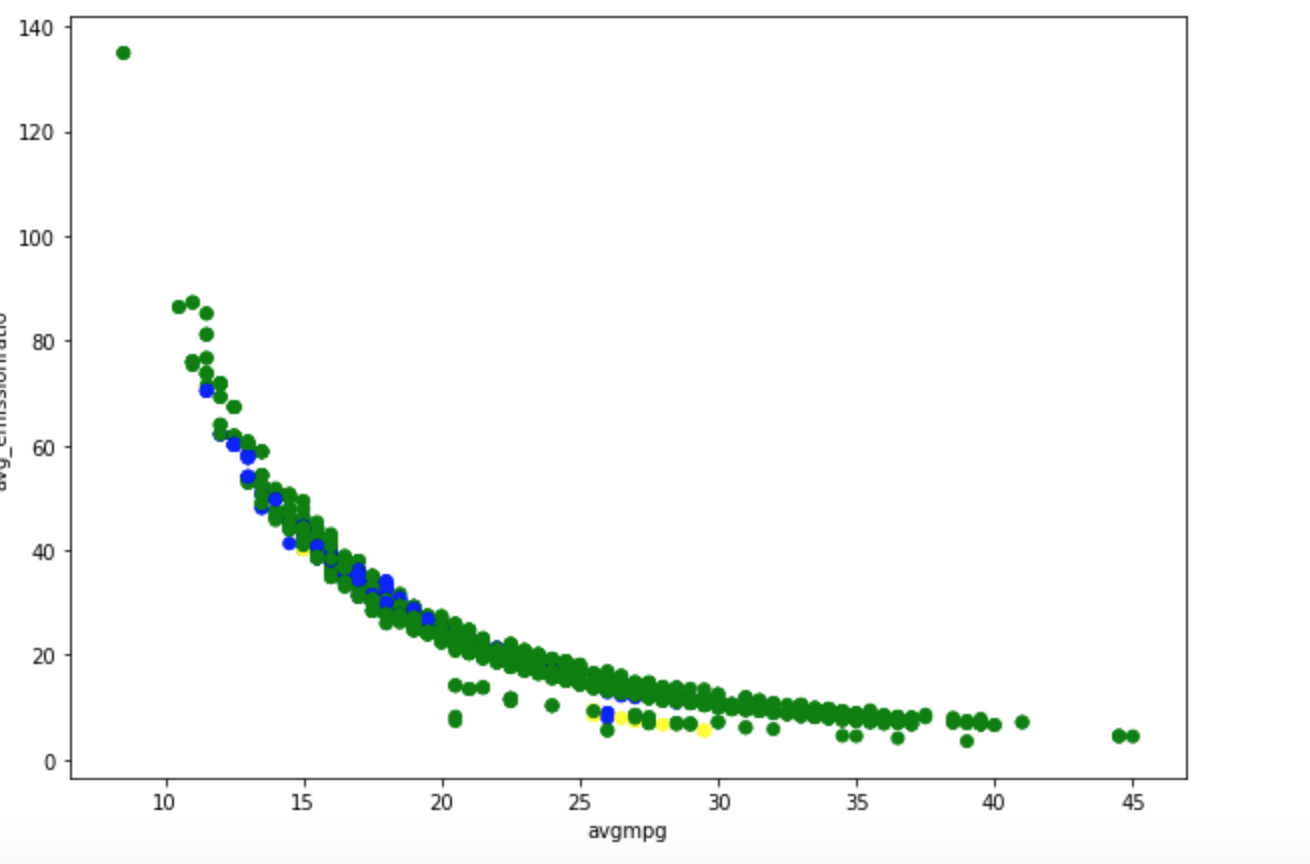
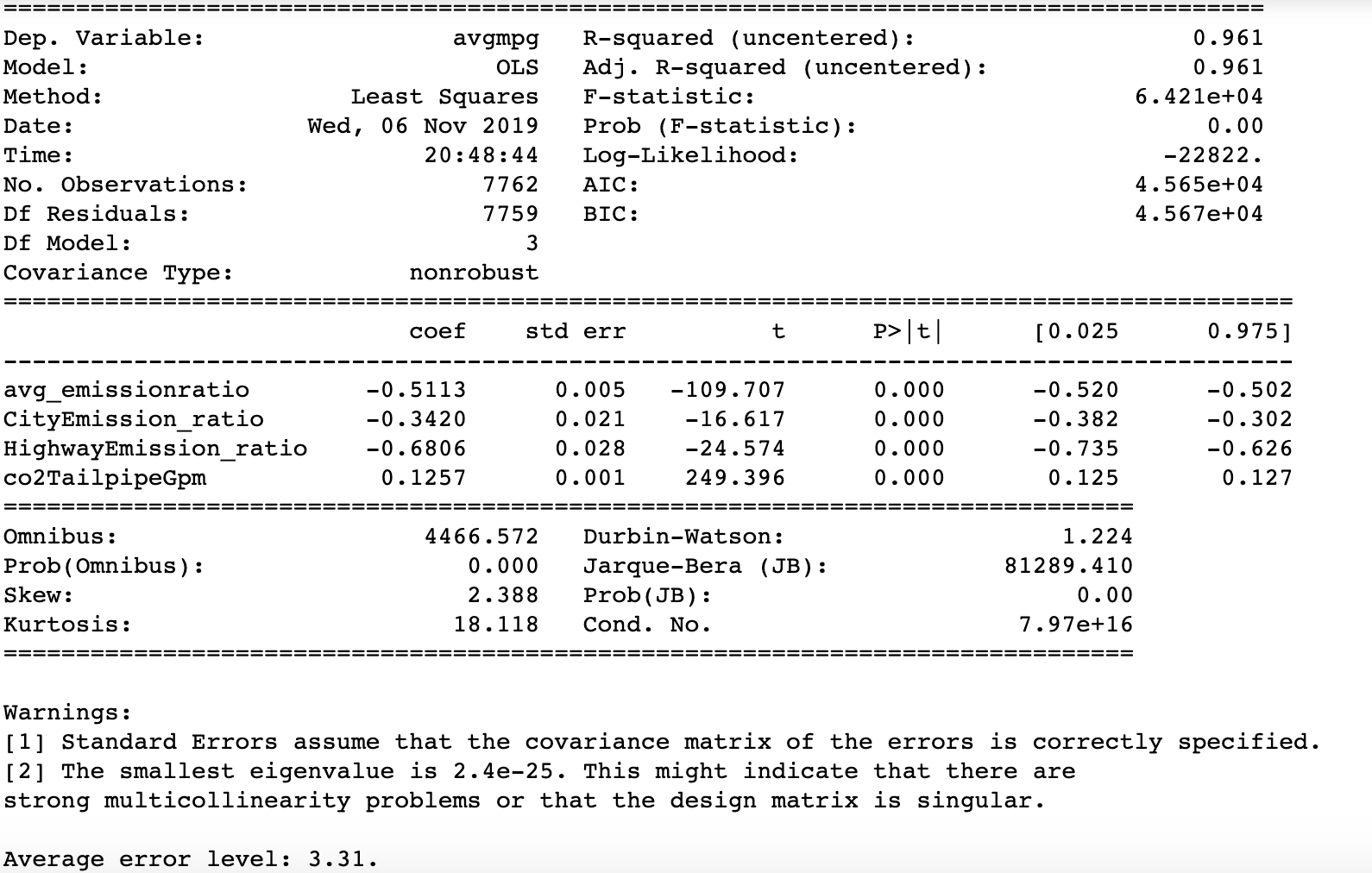
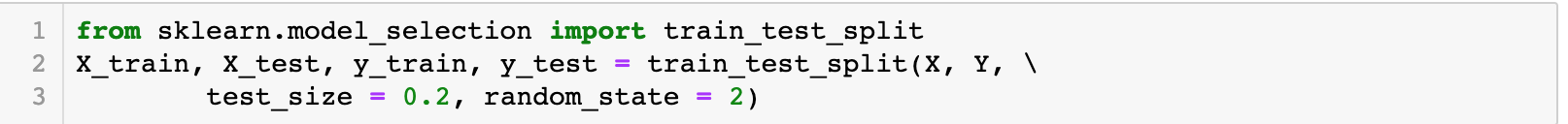


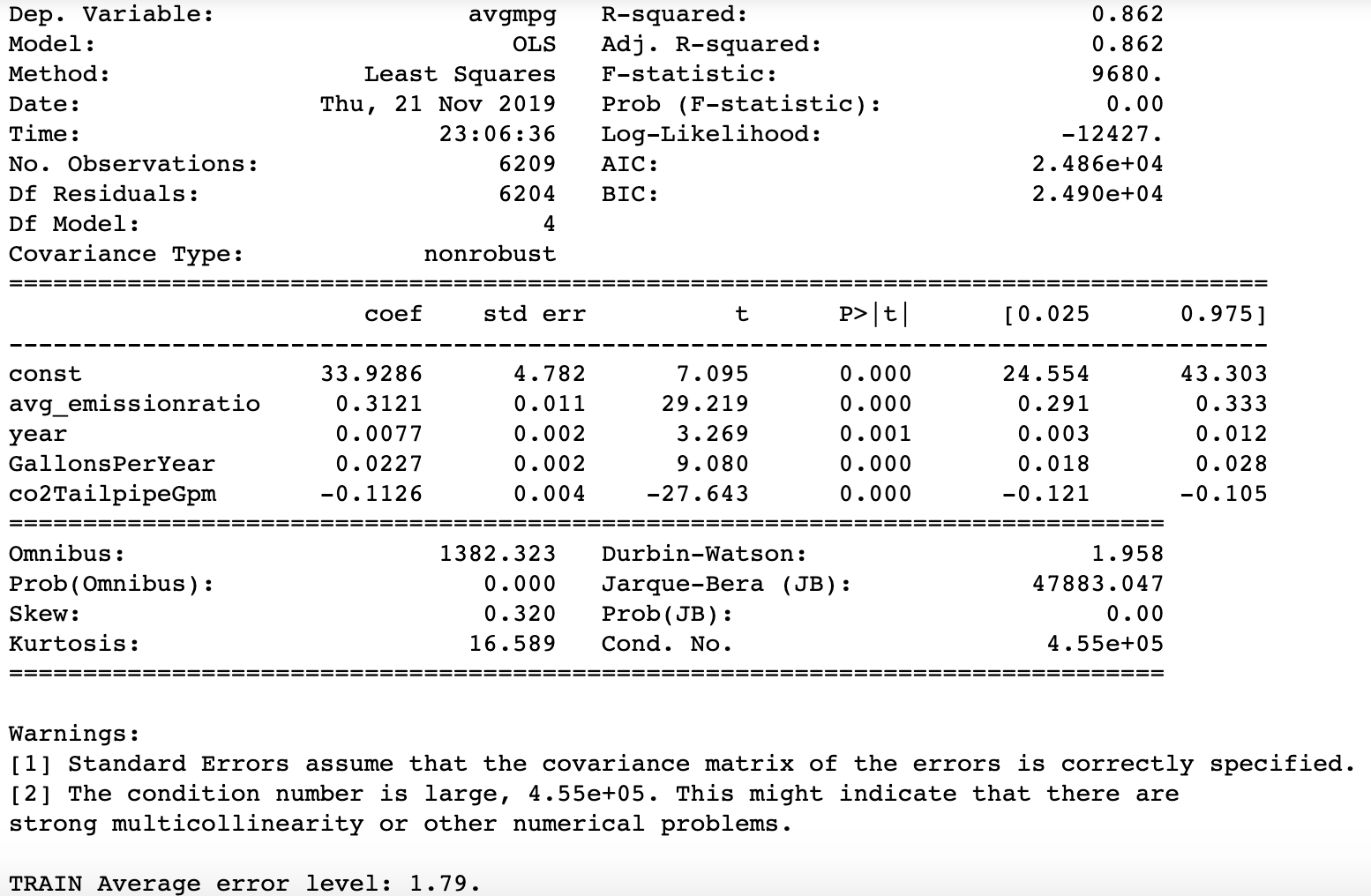
Figure 9



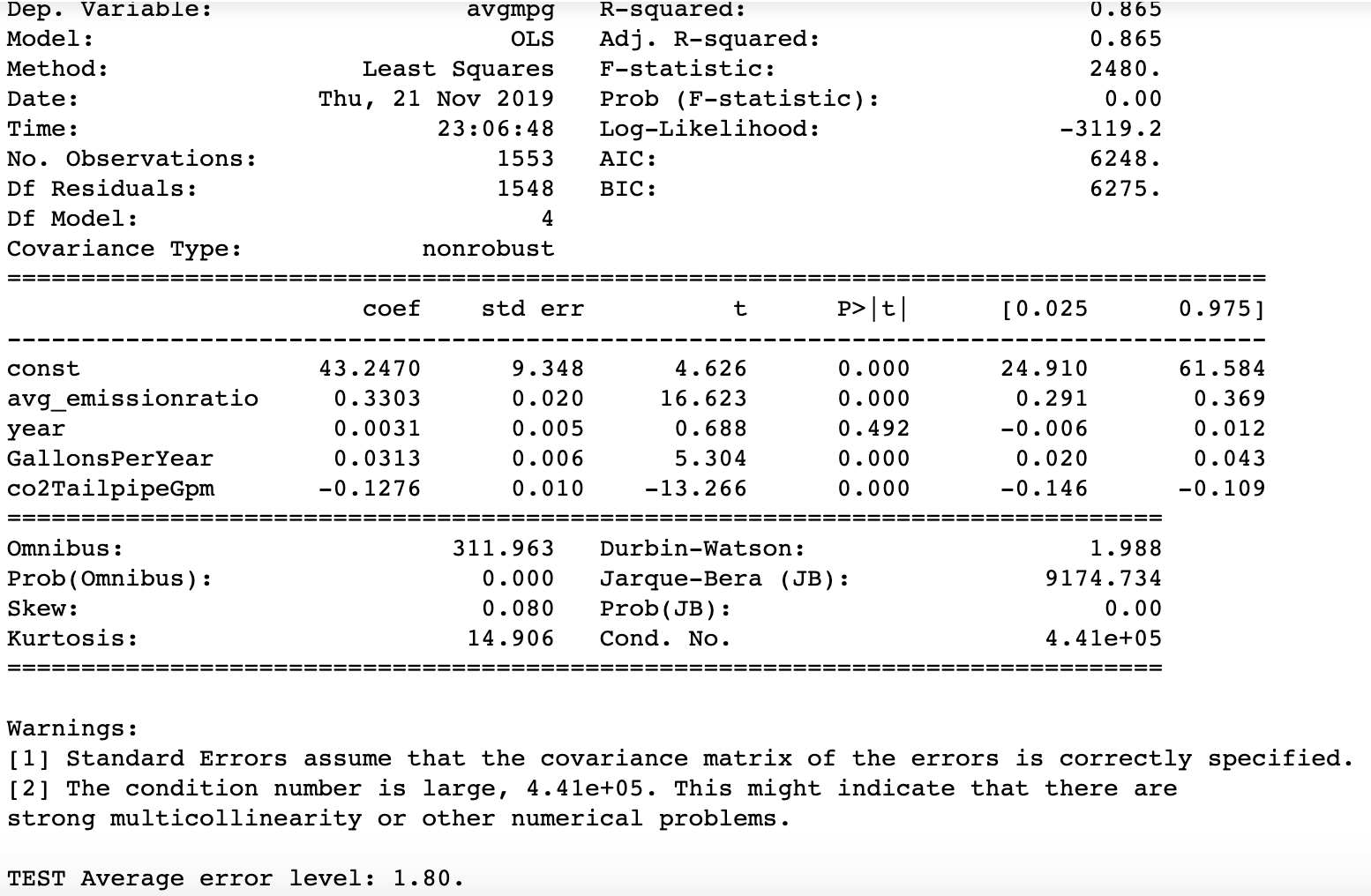
**Figure 10**

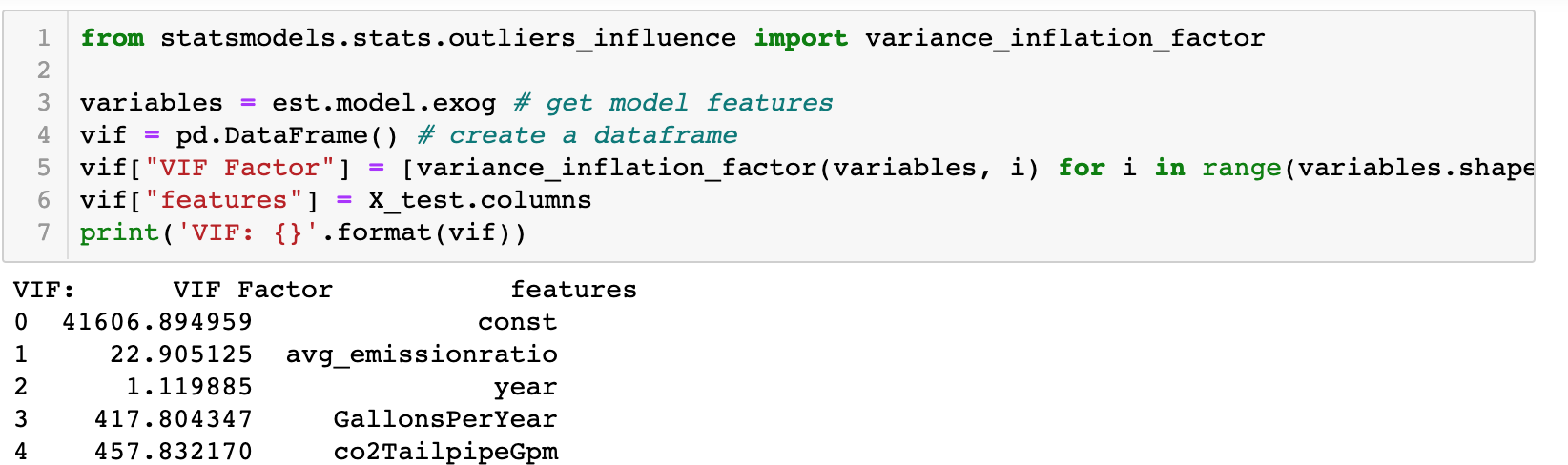


**Figure 11**



**Figure 12**

****

**Figure 13**