Sarah Milstein

Part 1:

I started this project by running the linear regression in the R code.
I first ran a str on the data so I would be able to see what we were using. I wanted to make sure all the data was in the right format so I changed the horsepower factor, which was previously done in char, to int so I would be able to include it into my linear regression model and split the data into two sets, one containing rows 1 through 300 for training and the other containing rows 301 through 398 for the testing.

Then once I ran the code on ., all the data, the model that I saw had a near perfect R score but when I excluded the car names the R score went down. I soon realized that since each car model only appeared once it did not make sense to include the name since it just overfit the model.

To find a more perfect model I ran a forward, backward and AIC test which all gave me pretty similar results. I decided to create my own model by only using the factors that had significant p values at the .05 level. Once I did that I saw that I could use minimum variables but still achieve the same R score results. My final model used the weight, model year and origin values to predict the mpg of the car.

I then calculated the residuals of the model by predicting the mpg of the remaining 98 data points and subtracting the predicted values by the actual mpg values. I ran a summary of the residuals, the mean and the median were both around 3 but the max was 16 so I figured that meant that an outlier was pulling the mean up. I then ran a summary of the actual mpg values and my predicted mpg values and saw that there was in fact at least one outlier throwing my model off, the max of the actual values was 46. Once I graphed the residuals I was able to see that many more of the points were positive values, and few were negative instead of being evenly distributed. The histogram brought out the same point but since 3 is a small percentage of the average mpg value of 23 and the R squared value is high, I think my model is still a good fit.

When I went onto Tableau to create a graph, I immediately put one together using the same factors I used in my R model, weight, origin and model year. I was able to clearly see the significance of the weight and origin on the mpg but not the model year. I then created more of the same graph using acceleration, horsepower and displacement and I was able to clearly see the effect of those variables too. I tried to add them onto my R model, hoping that they would raise my R values and decrease the residuals, but they only made a minute difference so I took them back out. I also ran a QQ plot to see how

the model's residuals rated there. The points of the residuals fit the line but start shifting off by the 1 point. It looks like there are many more larger outliers that are throwing the model off, my model is doing a better job of predicting the smaller values.

Overall I do not think my model is perfect but it is a pretty good predictor. I learnt a lot about creating models and graphs from this assignment.

https://public.tableau.com/app/profile/z.milstein/viz/Cars_17021741900080/CARS

Part 2:

I had a lot of fun with this assignment. I really enjoy working with Tableau so I decided to analyze the data on the analytics platform. I also thought that Tableau would be the best option since there were not many numeric values.

It took me a while to figure out how to format the numeric fields I did want to use. For example the CSAT scores were done as numbers but I wanted to display them as traditional percentages since I thought it would be easier to understand so I created a new field for them.

I also wanted to display the questions and answers I was analyzing on the actual dashboard without writing full sentences on it. After researching different ways to do this I decided to create icons by downloading different pngs and adding them into a new folder called custom shapes. I then made a blank calculated field and selected the custom shape I wanted to use for the icon for that blank field. With the shape now showing, I was able to write the details I wanted displayed in the tooltip so when the mouse is hovered over the icon in the dashboard, the details would be shown.

I came up with the questions just by playing around with the graphs and trying to see if there were any patterns or if there were any specific causes and effects coming out.

One of my favorite things about dashboards though is the filtering tool so by clicking on different fields, you can analyze many other different questions you might wonder about such as which reason does each channel address the most.

I really enjoyed this assignment and hope to continue learning and using Tableau more.

https://public.tableau.com/app/profile/z.milstein/viz/Calls_17022574477190/CallCenter