**DS 510 W – Introduction to Data Science**

**Final Project**

In this project you will investigate the impact of a number of automobile engine factors on the vehicle’s mpg. The dataset auto-mpg.csv contains information for 398 different automobile models. Information regarding the number of cylinders, displacement, horsepower, weight, acceleration, model year, origin, and car name as well as mpg are contained in the file.

Perform some initial analysis and create visualizations using Tableau Public (reference will be available in week 9).

Create some visual plots and charts describing the data and information it is trying to give out.

Using the first 300 samples in the auto-mpg.csv, run a simple linear regression and multiple linear regression to determine the relationship between mpg and appropriate independent variable/(s). Report all the appropriate information regarding your regression.

1. Multiple R-squared
2. Adjusted R-squared
3. Complete Linear Regression equation

Maintain a log of above values for all models.

For the remaining 98 samples in the dataset, use your best linear model(s) to predict each automobile’s mpg and report how your predictions compare to the car’s actual reported mpg.

1. Residual Plot
2. Histogram.

As a part of submission, share the code and report explaining the research. You can submit your code by compiling the report on RStudio. Directions to save complete code on word / PDF file is as below.

**RStudio -> File -> Knit Document / Compile Report -> Save as Word / PDF.**

*\*Also, if at all you decide to make some addition, deletion or changes in the dataset, explain with reasons.*