Linear Regression Assignment 1
Big Data Science

For all problems, submit your python code. Comment out your responses in the code and submit 1 file.

For problems 1 and 2, the height_weight files contain 1 column of heights (in inches) and 1 column of weights (in pounds). We are interested in the effect of height on weight.

1. Using the height_weight1.csv file, fit a linear model to the height and weight data that includes an intercept. Fit another model that does not include the intercept. In terms of our 3 residual assumptions, how do these two models compare? Can we use R^2 or another metric to determine which fits the data better?

2. Using the heigh_weight2.csv data, fit another linear model. This time, do not include an intercept. Does this model meet the assumptions of residuals? If not, explain why not. Additionally, if one of the assumptions is not met, how do you think that impacts our prediction? Does it impact our point prediction? What about our prediction interval?

For Problem 3, the cars.csv file has age of the car (in years), the brand/make of a car (Toyota, Ford, or BMW), the type (1 for SUV, 2 for sedan, and 3 for convertible), number of miles, and price the car recently sold for.

3. Our goal is to determine a model to predict the price of a used car. Fit 3 different models to the data and pick the "best" one. Why do you consider this model the best? Using this "best model", predict what a 7 year old, BMW convertible with 67,000 miles would cost.