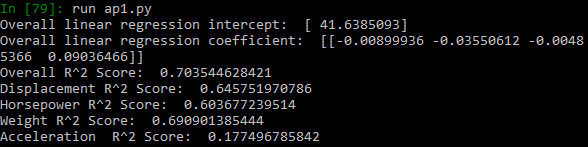
To determine the most effective predictor in forecasting a vehicle’s MPG, I regressed each predictor independently of the others. To accomplish this I created a [n x 1] vector with the respective feature I was getting the R2 score for. Then using the score method within sklearn’s LinearRegression object, I was able to produce the R2 value. This was repeated for each of the four requested features. The results, shown below, demonstrate that the weight feature is the most effective predictor of MPG with a R2 score of (.691).

When using the single-predictor model, weight had the best R2 value. Weight’s coefficient was (-.00739). This means that for every 100 units of weight the MPG would decrease by .739 MPG. So the relationship is an inverse linear function. This makes sense as we would expect the MPG to decrease in real life as the weight increased, with all else held the same.

When using the linear regression method on all four features the coefficient of determination or R2 score, is (.7035). When compared to the other R2 scores using a single predictor, as seen below, it is shown that the four features together provide a slightly more accurate model than the best single predictor.



Yes, the cylinders information can be included in the dataset into our model. The R2 value becomes .70349 including the new feature versus .70354 without it. If you use the single-predictor model the cylinder’s R2 value is .6016. These values help determine that the relationship between cylinders and MPG is virtually the same as the other features. This means the linear relationship is just as effective. Using the same python libraries I just had to add ‘Cylinders’ to the below line. The rest of the code is generic enough to accommodate the feature add. You can also follow the same template I used in the single-predictor model to find the cylinder information. 