

Fuel Efficiency

Description and Preprocessing of Data

The Auto_MPG.csv is made up of 398 rows and 9 columns. The columns are as follows:

- **mpg** - fuel efficiency measured in miles per gallon (mpg)
- **cylinders** - number of cylinders in the engine
- **displacement** - engine displacement (in cubic inches)
- **horsepower** - engine horsepower
- **weight** - vehicle weight (in pounds)
- **acceleration** - time to accelerate from 0 to 60 mph (in seconds)
- **model year** - model year
- **origin** - origin of car (1: American, 2: European, 3: Japanese)
- **car name** - car name

Weight, Model Year and Displacement were removed due to multicollinearity issues with acceleration and horsepower. I will explore other models which replace acceleration and horsepower for the other variables and see if I can come with a model that has more predictive power.

Project Description

The aim for this project was to predict the fuel efficiency of a car using a Linear Regression model which was coded using Python. The model created finished up with an R squared value of approximately 80% and a mean difference of 10% between predicted values and actual values when the model was tested using data it had not seen before.