**Assignment 2**

The ***CarPrice.csv*** dataset contains the details of 205 cars. Your team is asked to build a multiple regression model to understand the factors of the car price.

The dataset contains the following variables:

* Categorical data – fueltype (either gas or diesel), doornumber (two or four) and enginelocation (front or rear)
* Numerical data- wheelbase, carlength, carwidth, carheight, curbweight, enginesize, horsepower, peakrpm, citympg, highwaympg and price

Since there are too many predictors on price, so you team will apply different variable selection methods to reduce number of predictors, and eventually build a final model to understand the pricing.

The details of the work flow is given in the jupyter notebook, make sure you have run all the previous code on the script, including given code, before you move to the next question.

Your major tasks includes the following:

1. **[Coding]** Preparing indicator variables – convert the categorical variables into indicator variables.
2. **[Coding]** Apply log transformation on price
3. **[Coding]** Best subset selection
   1. Execute best subset selection
   2. Fit the best subset model according to BIC
4. **[Coding]** Forward selection
   1. Execute the forward selection
   2. Fit the forward selection model according to BIC
5. **[Coding]** Compare the MSEs of different models
6. **[Coding]** Build a final model
7. Understanding the model/pricing