**Assignment on Regression Analysis**

**1. On ‘marketing’ dataset**

1. Import ‘marketing’ dataset in python
2. Perform data cleaning (if required)
3. Develop an appropriate regression model by dividing the dataset into training and testing (75 % : 25%) using (a) sklearn library (b) statsmodel library
4. What is the value of R-square and adjusted R-square. How do you interpret them?
5. What are values of coefficients. Interpret the coefficients
6. Comment on accuracy of the model using important metrics such as R-square, MSE, RMSE etc.
7. Check the important assumptions of the model

**2. On ‘diamonds’ dataset**

1. Read ‘diamonds’ dataset in python
2. Name the variables of the data.
3. How many rows are there in the data?
4. Draw a suitable diagram for the variable ‘cut’ category – bar, pie
5. Draw a suitable diagram for the variable ‘carat’ histogram, barplot
6. Determine all the descriptive statistics of the relevant continuous variables.
7. Divide the dataset into 70:30 ratio.
8. Suppose you require to predict price of the diamond given other features. Can you make use of regression for this?
9. Develop a suitable regression model to predict the price of diamond?
10. Is there a case of multicollinearity in the data? How will you check?
11. Is the normality assumptions satisfied?
12. Check for homoscedasticity in the regression model.
13. What is the value of R square and adjusted R square?
14. Predict the price for diamond given the following:

Carat = 0.25, cut = good, table = 55, x = 4, y = 4.2, z = 2.5