**Terro's Real Estate Agency (1)**

| **Criteria** |
| --- |
| **1) Generate the summary statistics for each variable in the table. (Use Data analysis tool pack). Write down your observation.** |
| **2) Plot a histogram of the Avg\_Price variable. What do you infer?** |
| **3. Compute the covariance matrix. Share your observations.** |
| **4) Create a correlation matrix of all the variables (Use Data analysis tool pack).**  a) Which are the top 3 positively correlated pairs and b) Which are the top 3 negatively correlated pairs. |
| **5) Build an initial regression model with AVG\_PRICE as ‘y’ (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot:**  a) What do you infer from the Regression Summary output in terms of variance explained, coefficient value, Intercept, and Residual plot? b) Is LSTAT variable significant for the analysis based on your model? |
| **6) Build a new Regression model including LSTAT and AVG\_ROOM together as independent variables and AVG\_PRICE as dependent variable:**  a) Write the Regression equation. If a new house in this locality has 7 rooms (on an average) and has a value of 20 for L-STAT, then what will be the value of AVG\_PRICE? How does it compare to the company quoting a value of 30000 USD for this locality? Is the company Overcharging/ Undercharging? b) Is the performance of this model better than the previous model you built in Question 5? Compare in terms of adjusted R-square and explain |
| **7) Build another Regression model with all variables where AVG\_PRICE alone be the Dependent Variable and all the other variables are independent. Interpret the output in terms of adjusted R square, coefficient and Intercept values. Explain the significance of each independent variable with respect to AVG\_PRICE.** |
| **8) Pick out only the significant variables from the previous question. Make another instance of the Regression model using only the significant variables you just picked and answer the questions below:**  a) Interpret the output of this model. b) Compare the adjusted R-square value of this model with the model in the previous question, which model performs better according to the value of adjusted R-square? c) Sort the values of the Coefficients in ascending order. What will happen to the average price if the value of NOX is more in a locality in this town? d) Write the regression equation from this model. |