Terro's Real Estate Agency

Business Report

1. **Generate the summary statistics for each variable in the table. (Use Data analysis tool pack). Write down your observation.**

I Find the summary statistics for each variables using data analysis tool pack. From that analysis I got,

1) MEAN

2) STANDARD ERROR

3) MEDIAN

4) MODE

5) STANDARD DEVIATION

6) SAMPLE VARIANCE

7) KURTOSIS

8) SKEWNESS

9) RANGE

10) MINIMUM

11) MAXIMUM

12) SUM

13) COUNT

I Got these above listed measures, from the data analysis. And the values I got are clearly mentioned in excel on 1st ans sheet.

**2) Plot a histogram of the Avg\_Price variable. What do you infer?**

In statistics, a **histogram** is a graphical representation of the distribution of data. The histogram is represented by a set of rectangles, adjacent to each other, where each bar represent a kind of data.

I find the histogram of the Avg Price variable. And clearly I plot the histogram

From the histogram plot I came to a conclusion that 21 to 25 has highest average price.

**3. Compute the covariance matrix. Share your observations.**

**Covariance** is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable.

I find the covariance for each variables. And that values are clearly mentioned in excel.

**4. Create a correlation matrix of all the variables (Use Data analysis tool pack).**

**Correlation** refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a “[scatter plot](https://byjus.com/maths/scatter-plot/)”.

1. Top 3 positively correlated pairs.
2. 0.910228188533182
3. 0.763651446920914
4. 0.731470103785958
5. Top 3 negatively correlated pairs.
6. -0.737662726174014
7. -0.613808271866396
8. -0.507786685537561

**5) Build an initial regression model with AVG\_PRICE as ‘y’ (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot.**

Coefficient value : 34.5538408793831

**6) Build a new Regression model including LSTAT and AVG\_ROOM together as Independent variables and AVG\_PRICE as dependent variable**

**i clearly mentioned all the results and outputs in the excel sheet.**

**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.**

**i clearly mentioned all the results and outputs in the excel sheet.**

**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:**

**I clearly explained all the results and output in the excel sheet. That attached with this.**