Business report

Real estate data analysis- exploratory data analysis and linear regression

Terro’s real estate agency

By

**LEELA ABINAYA V**

**1)Generate the summary statistics for each variable in the table .(use data analysis tool pack).write down your observation**

Ans) we use the descriptive statistics function inside the data analysis tool pack which can be found in the data ribbon for this task . the observation that we have arrived at from the table are that

\*.among the given data TAX have a highest standard error (7.49) and NOX (nitric oxides concertation)

Has the lowest standard error of (0.00515),thus sample means are widely spread around the population mean for NOX and sample means are closely distributed around the population mean for TAX

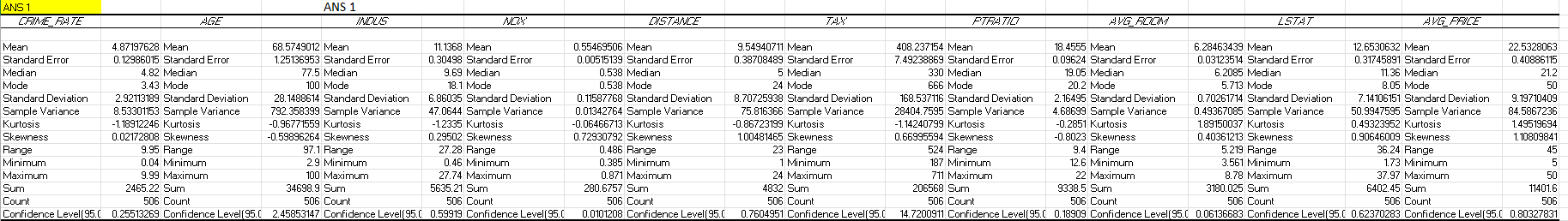
\* TAX holds the highest sample variance( 280404.579 )while NOX have a lowest value ( 0.0134 ), thus the data are more spread out the mean for Tax ,and data are clustered around the mean for NOX.

\*.TAX has the highest minimum value (187,) while crime rate have a lowest value (0.04) thus the data spreads around the tax , and the data are clustered

around for the crime rate .

\*.TAX have a highest mode value ( 666 ),NOX have a lowest mode value (0.538).,thus the data spreads around the tax, and the data are clustered around for the NOX

\*TAX has the highest median value (330),NOX have a lowest median value (0.538),thus the data spreads around the tax, and the data are clustered around the NOX.

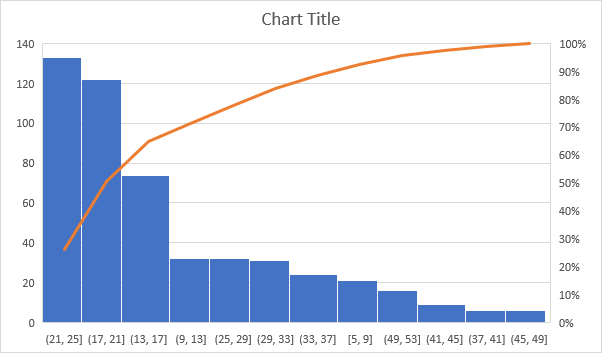


**2) plot the histogram of the average price variable .what do you infer?**

Ans :

To find the histogram for the average price variable .we have to copy the data which is given from the original data .after that select the insert from the menu bar .

Selecting the histogram for this chart



BY observing the data in the histogram we can say that the data is more spread towards the leftside of the histogram ,ie having a long left tail , thus we can say that the data is positively skewed

**3)**

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

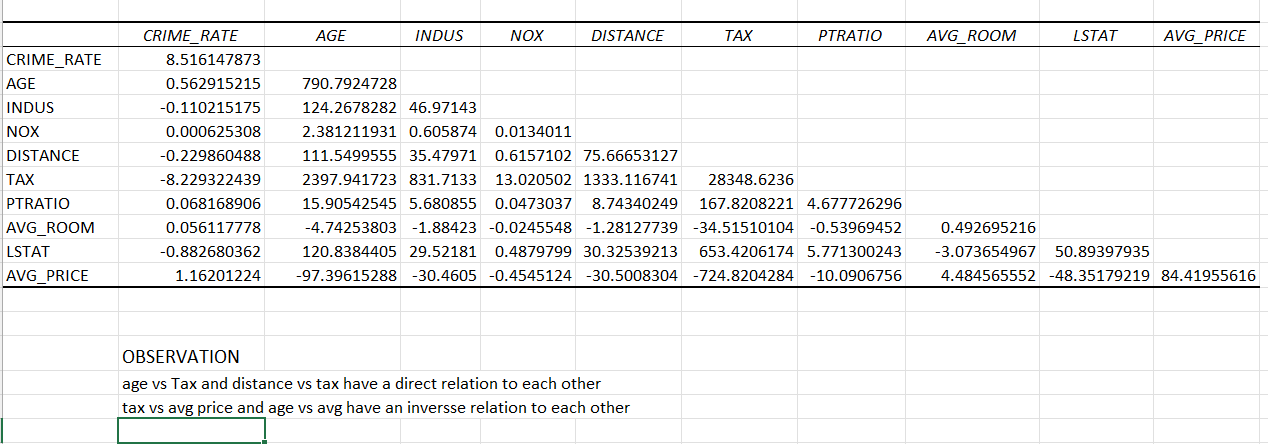
**Ans in this given data , select data from the menu bar and with in it click data analysis.**

**In that select the covariance and insert the respective x range and y range , and select the label.**

Ans :for the covariance matrix . select data which is in menu bar and within it click data analysis

Option . in that we have to select the respective x and y range and also enable the label option to prior the headlines which is given from the original data . after we select all the features we get the

Covariance table .



Observation:

From the above covariance table, it is clear that Age vs Tax and Distance vs Tax has Direct relation with each other. And also it is inferred that Tax vs Avg.Price and Age vs Avg.Price lies in Inverse relation with each other respectively.

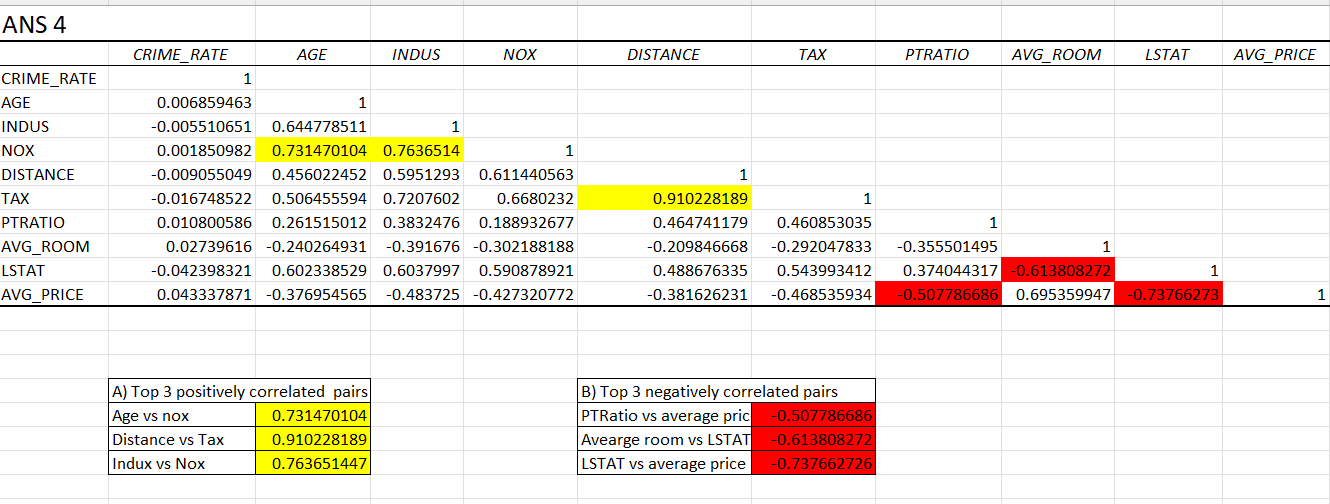
**4) Create a correlation matrix of all the variables (Use Data analysis tool pack). (5 marks)**

**a) Which are the top 3 positively correlated pairs**

**b) Which are the top 3 negatively correlated pair**

ANS :

From the given data we want to find a correlation matrix of all the variable .to find that we want to select the data from the menu bar option and within that we select data analysis . we will find further option, in that we have to select the correlation option .after that we have to select the respective range of x and y to the table extend . enable the label option to mark the prior headlines to correlate the table by the given data .



We want to select the as a slope direct which is related to the table EG : crime rate vs crime rate . after that we want to select the conditional formatting which is in the home page of excel . we want t select the top to bottom and that click more options and select the range form the table . after that we have to click over whether the option are given to us .in that table we can select the top to bottom . and also select the top 3 positively correlation and the top 3 negative correlation

**5) Build an initial regression model with AVG\_PRICE as ‘y’ (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot. (8 marks)**

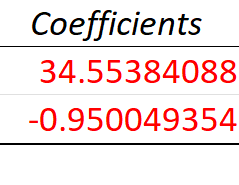
**a) What do you infer from the Regression Summary output in terms of variance explained, coefficient value, Intercept, and the Residual plot?**

**b) Is LSTAT variable significant for the analysis based on your model?**

Ans )

Regression is the methodology by which an output can be derived from the set of population and samples .

For finding the regression model we have to select the data from the menu bar and click the menu bar and option as data analysis. From the search and click the regression process,while after that select the respective range x and y . enable to label for headline prior and select all the negative and residual plot independent variable of the summary output .we will get the summary output with coefficiant , significant , p value etc



|  |  |
| --- | --- |
| A) intercept and average price is positively related and LSTAT and average price p id negatively related |  |
| B) LSTAT variable and average price are positively corelated thus it significant for our analysis |  |

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

**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?**

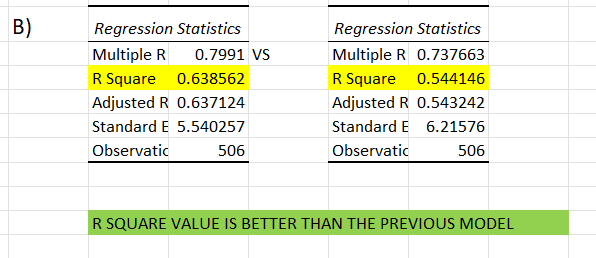
**How does it compare to the company quoting a value of 30000 USD for this locality?**

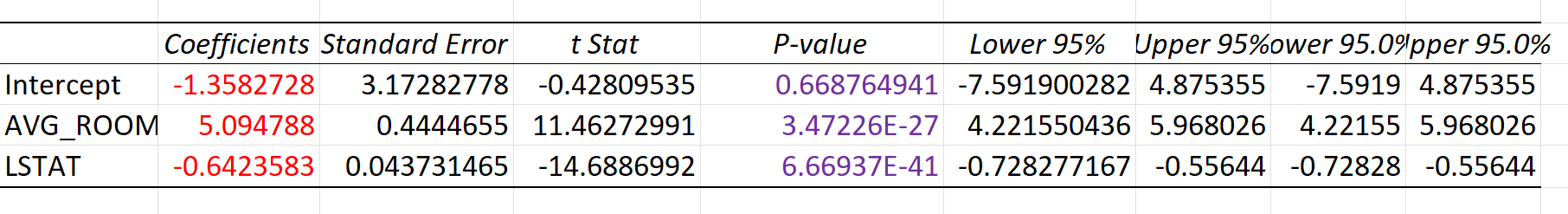
**Is the company Overcharging/undercharging?**

Ans )

|  |  |  |
| --- | --- | --- |
|  |  |  |

The company quoting a value of 30000usd but the calculated value of avg price is lesser than the company quotation  
hence we conclude that bthe company is overcharging

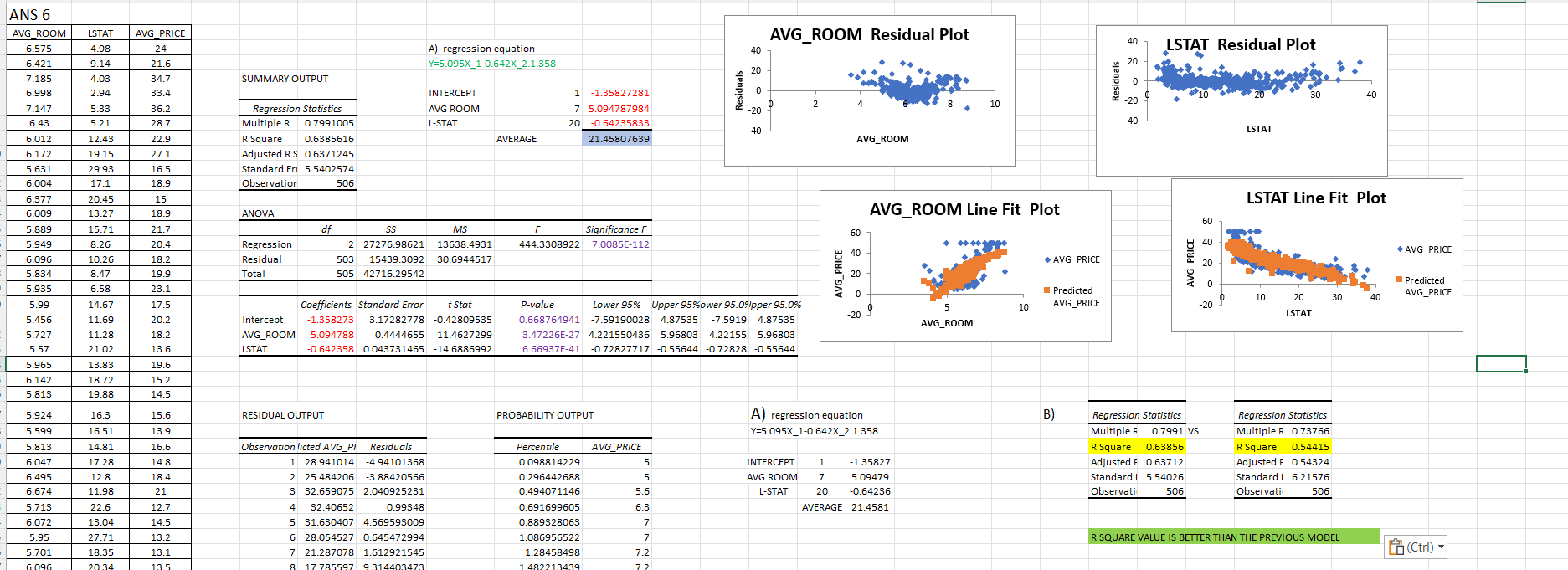




This regression statistics is form the same question

For comparing the value of r square

0.63856161 is greater than the 0.5441463 s0 we got a better value from the pervious question



**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**

Ans)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1)Comparing the r square values this model is better than others |  |  |  |  |  |  |
| 2)comparing the coefficient and intercept crime rate ,age ,indus, distance ,average room have direct relation with  Average price and the others have inverse relation    **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: (8 marks) 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.**         |  |  | | --- | --- | | C) If NOX is more ,the average price will decrease | | | if NOX is less ,the average price will increase |  |   d)  D)Y=0.0329X\_1+0.1370X\_2-10.272X\_3 0.261X\_4-0.0144X\_5-1.0717X\_6+4.414X\_7-0.605X\_8+29.428 | | | | | | |