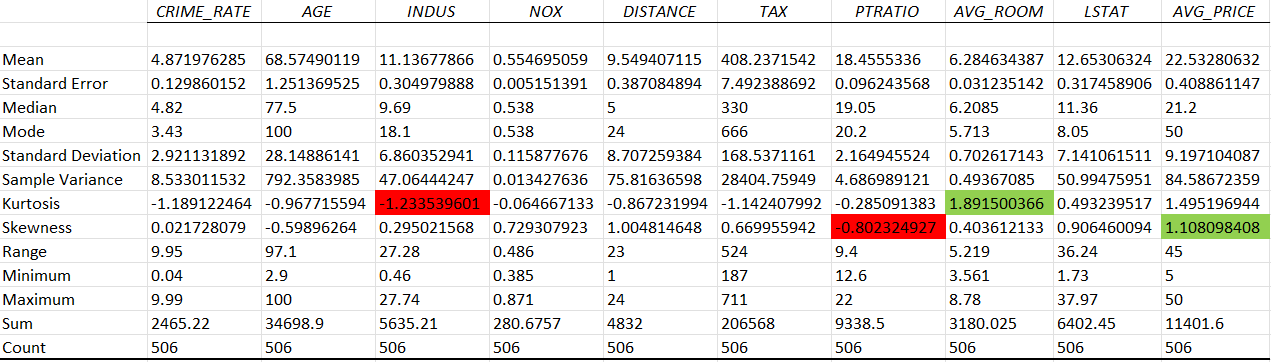
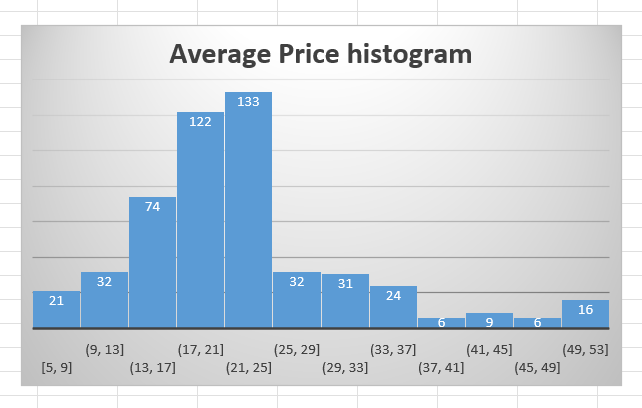
Question1. Generate the summary statistics for each variable in the table. (Use Data analysis tool pack). Write down your observation.?



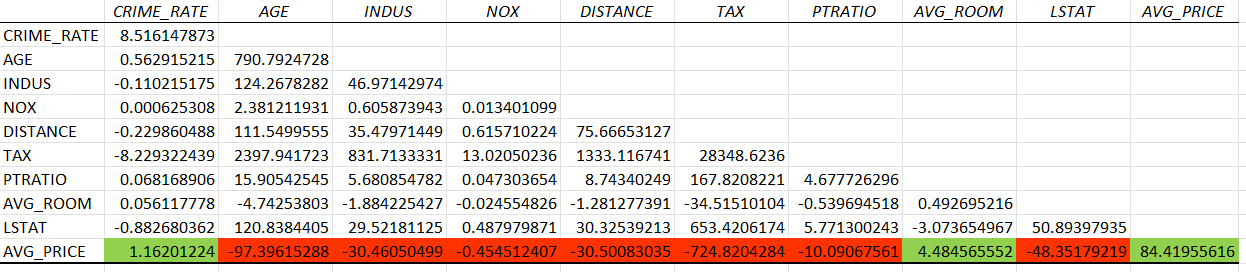
By generating the summary statistics of all the variable we can say that Average price has the most positive skewness, PTRatio has the most negetive skewness and Average Room has the highest Kurtosis, Indus has the most negetive Kurtosis.

Question2. Plot a histogram of the Avg\_Price variable. What do you infer?



133 people belong in the Range between 21-25 which is the highest number of people in the Range 6 people belong in the range between 37-41 and 45-49 which is the lowest number of people in the Range.

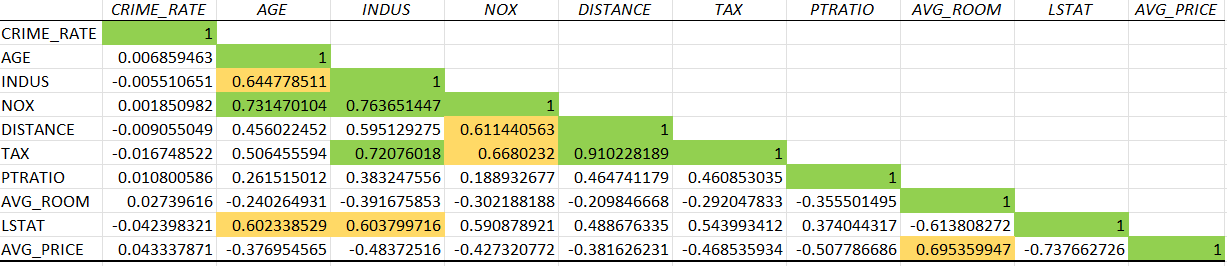
Question3. Compute the covariance matrix. Share your observations.



After generating the covariance matrix we can say that the Dependent variable Average price has negetive covariance with the most of the variable by that we can say that our regression model will not be Accurate.

Question4. 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.



a)TOP 3 +VE CORRELATED PAIRS

DISTANCE-TAX

INDUS-NOX

AGE-NOX

b)TOP 3 -VE CORRELATED PAIRS

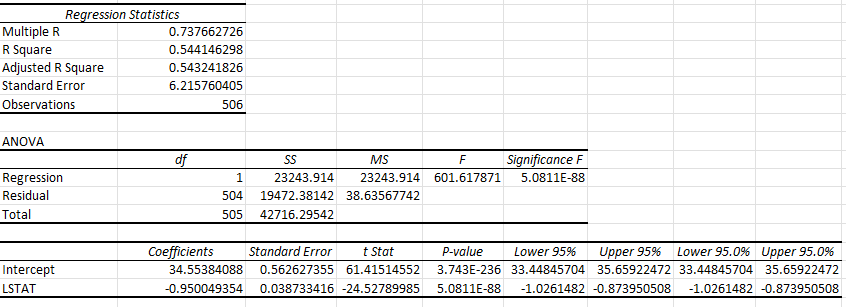
LSTAT-AVG\_PRICE

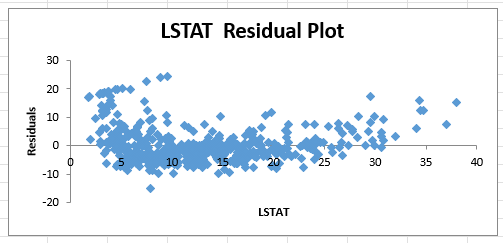
AVG\_ROOM-LSTAT

PTRATIO-AVG\_PRICE

Question5. 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?





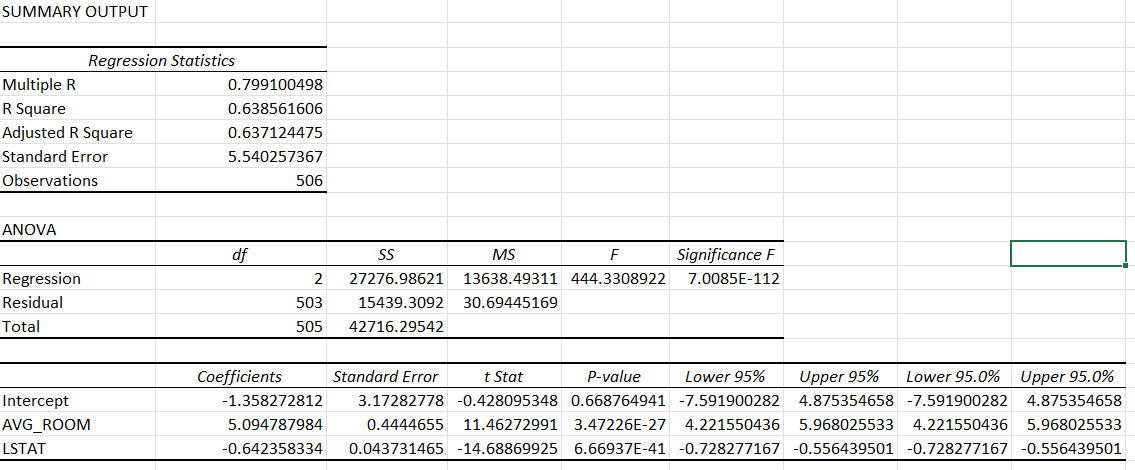
a)Based on Rsquare we can say that this model has not predicted the Average price acurately and variance is high between the Average predicted price and actual price,

The coefficient value for LSAT is -0.95.

b)LSTAT is a significant variable while pedicting the Average price but just based on LSTAT we can't predict the Average price acurately.

Question6. 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



a) Regression Equation = -1.35 + 5.09\*(Avg\_Room) - 0.64\*(LStat)

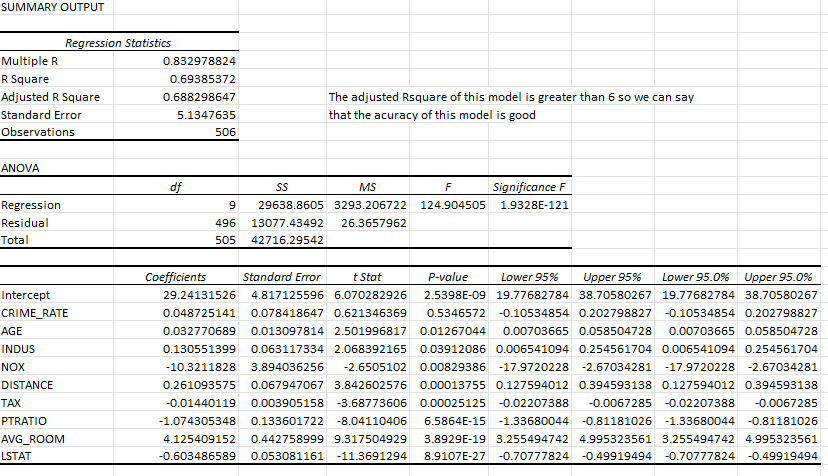
Regression Equation = -1.35 + 5.09\*(7) - 0.64\*(20)

Regression Equation = 21.48

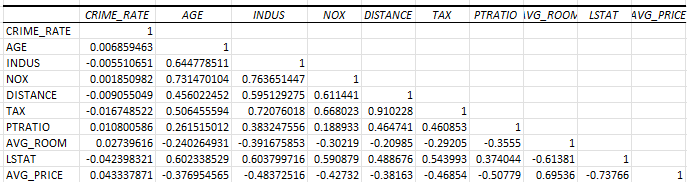
by comparing the predicted price of my model and companies price the company is Overcharging.

b)Yes this model is better than previous model compared to previous model the Rsquare is high in this model. The Rsquare of previous model is 0.54 and the Rsquare of current model is 0.63

Question7. 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.



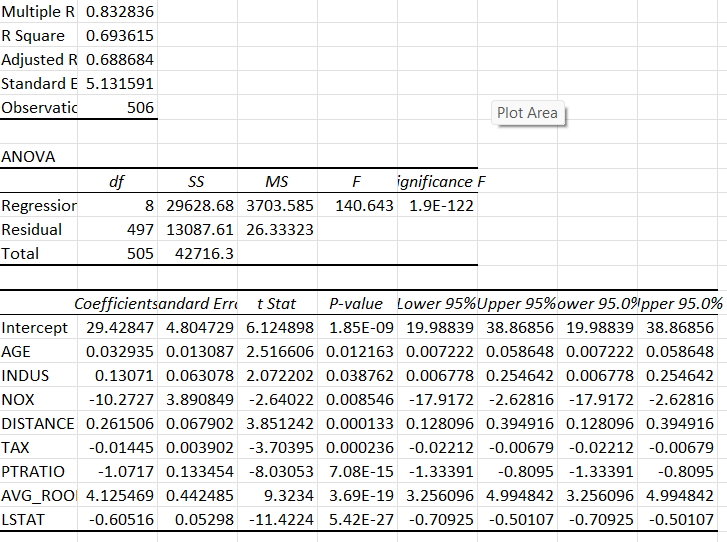
The adjusted Rsquare of this model is greater than 6 so we can say that the acuracy of this model is good.



By looking the correalation between variables and dependant variable we can say that crime rate is not related with Average price, the relation between nox and indus is high, even nox and age are highly correlated, and Pvalue of crime rate is high based on those reason we can say that this is not the best model.

Question8. 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.



a)The accuracy of this model based on Rsquare is 0.69, adjusted Rsquare is 0.68

And the standard error is 5.13

b)By comparing the Adjusted Rsquare from this model and the previous model the Adjusted Rsquare has increased by 0.04