Terro's Real Estate Agency Final Business Report

Assignment-2

BY

Yuvaraj Pandian

GLCA DA SEPT-23 BATCH

Q1. Generate the summary statistics for each variable in the table. (Use Data analysis tool pack). Write down your observation.

CRIME_RATE	
Mean	4.87197628
Standard Error	0.12986015
Median	4.82
Mode	3.43
Standard Deviation	2.92113189
Sample Variance	8.53301153
Kurtosis	-1.18912246
Skewness	0.02172808
Range	9.95
Minimum	0.04
Maximum	9.99
Sum	2465.22
Count	506

1. Crime Rate

In Crime Rate,

Mean Value -4.87

Median Value-4.82

By seeing this observation, Mean and Median have the minimum difference with each other.

StandradDevaition

Value:2.9211

By seeing this observation, the spread has a moderate level in this data.

Skewness Value-0.0217

By Seeing this observation,more data to the right,so the median and mean are trailing towards left.

Kurtosis Value--1.1891
By seeing this observation, it has flat peak.

AGE	
Mean	68.57490119
Standard Error	1.251369525
Median	77.5
Mode	100
Standard Deviation	28.14886141
Sample Variance	792.3583985
Kurtosis	-0.96771559
Skewness	-0.59896264
Range	97.1
Minimum	2.9
Maximum	100
Sum	34698.9
Count	506

2.Age
In Age,
Mean Value-68.5
Median Value-77.5
By Seeing the observation, Mean and Median have medium difference with each other
StandardDeviation- 28.14886
By Seeing the observation, the spread is far away from the mean
Skewness Value- -0.598996
By seeing the observation, it is negatively skewed more data on the right side of the data.
Maximum value from the data is 100
Minimum Value from the data is 2.9
Kurtosis Value--0.967
By seeing this observation, it is in flat peak.

INDUS	
Mean	11.13677866
Standard Error	0.304979888
Median	9.69
Mode	18.1
Standard Deviation	6.860352941
Sample Variance	47.06444247
Kurtosis	-1.2335396
Skewness	0.295021568
Range	27.28
Minimum	0.46
Maximum	27.74
Sum	5635.21
Count	506

3.Indus

In Indus, Mean Value-11.13677

Median Value-9.69

By seeing the observation, Mean and Median value have small difference with each other.

Standard Deviation - 6.8603

By seeing the observation, the spread is moderate.

SkewnessValue- 0.295021

By seeing the observation, it is positively skewed more data on the left side of the mean.

Maximum Value from the data is 27.74

Minimum Value from the data is 0.46

Kurtosis Value- - 1.2335396

By seeing this observation, it is in flat peak.

NOX	
Mean	0.554695059
Standard Error	0.005151391
Median	0.538
Mode	0.538
Standard Deviation	0.115877676
Sample Variance	0.013427636
Kurtosis	-0.06466713
Skewness	0.729307923
Range	0.486
Minimum	0.385
Maximum	0.871
Sum	280.6757
Count	506

4.Nox

In Nox, Mean Value- 0.5546

Median Value- 0.538

By seeing the observation, the mean and median values has a very small difference with each other.

Standard Deviation-0.1158

By seeing the observation, the spread is very small and close to mean.

Skewness Value- 0.72930

By seeing the observation, it is positively skewed, more data on the right side of the mean.

Maximum Value from the data is 0.871

Minimum Value from the data is 0.385

Kurtosis Value- -0.06466

By seeing this observation, it is flat peak.

DISTANCE	
Mean	9.54940711
Standard Error	0.38708489
Median	5
Mode	24
Standard Deviation	8.70725938
Sample Variance	75.816366
Kurtosis	-0.86723199
Skewness	1.00481465
Range	23
Minimum	1
Maximum	24
Sum	4832
Count	506

5.Distance

In

Distance Mean Value- 9.5494

Median

Value- 5

By seeing the observation, the mean and median values have more difference with each other.

Standard Deviation- 8.70725

By seeing the observation, the spread is far away from the mean.

Skewness Value- 1.004814

By seeing the observation, it is a positively skewed more data on left side of the mean.

Maximum Value from the data is 24

Minimum Value from the data is 1

Kurtosis Value- -0.8672

By seeing this observation, it is flat pleak.

TAX	
Mean	408.2371542
Standard Error	7.492388692
Median	330
Mode	666
Standard Deviation	168.5371161
Sample Variance	28404.75949
Kurtosis	-1.142407992
Skewness	0.669955942
Range	524
Minimum	187
Maximum	711
Sum	206568
Count	506

6.Tax

In Tax, Mean Value- 408.2371

Median

Value- 330

By seeing the observation, the mean and median values have more difference with each other.

Standard Deviation- 168.5371

By seeing the observation, the spread is widely extensive.

Skewness Value- 0.66995

By seeing the observation, it is positively skewed more data on the left side of the mean.

Maximum Value from the data is 711

Minimum Value from the data is 187

Kurtosis Value- -1.142407

By seeing this observation, it is flat peak.

PTRATIO	
Mean	18.4555336
Standard Error	0.096243568
Median	19.05
Mode	20.2
Standard Deviation	2.164945524
Sample Variance	4.686989121
Kurtosis	-0.285091383
Skewness	-0.802324927
Range	9.4
Minimum	12.6
Maximum	22
Sum	9338.5
Count	506

7.PTRATIO

In PTRatio Mean Value- 18.455

Median Value- 19.05

By seeing the observation, the mean and median values have small difference with each other.

Standard Deviation-

2.16495

By seeing the observation, the spread is smaller.

Skewness Value- -0.802324

By seeing the observation, it is negatively skewed and more data on right side of mean.

Maximum Value from the data is 22

Minimum Value from the data is 12.6

Kurtosis Value- -0.2850

By seeing this observation,it is a flat peak.

AVG_ROOM	
Mean	6.284634387
Standard Error	0.031235142
Median	6.2085
Mode	5.713
Standard Deviation	0.702617143
Sample Variance	0.49367085
Kurtosis	1.891500366
Skewness	0.403612133
Range	5.219
Minimum	3.561
Maximum	8.78
Sum	3180.025
Count	506

8.Average Room

In Average Room, Mean Value- 6.2846

Median Value- 6.2085

By seeing the observation, the mean and median values have very small difference with each other.

Standard Deviation- 0.702617143

By seeing the observation, the spread is small.

Skewness Value- 0.4036

By seeing the observation, it is positively skewed and more data on left side of mean

Maximum Value from the data is 8.78

Minimum Value from the data is 3.561

Kurtosis Value- 1.891500

By seeing this observation, it has a high peak.

LSTAT	
Mean	12.65306324
Standard Error	0.317458906
Median	11.36
Mode	8.05
Standard Deviation	7.141061511
Sample Variance	50.99475951
Kurtosis	0.493239517
Skewness	0.906460094
Range	36.24
Minimum	1.73
Maximum	37.97
Sum	6402.45
Count	506

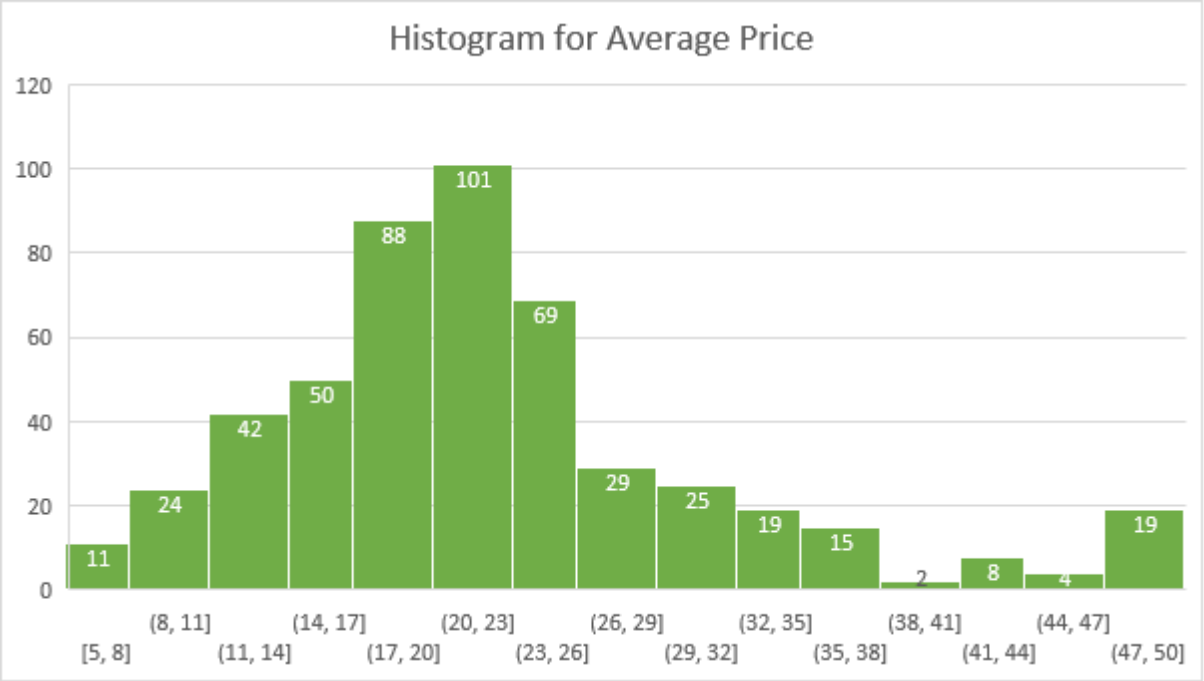
9.LSTAT
In
LSTAT Mean Value- 12.65306
Median Value- 11.36
By seeing the observation, the mean and median values have a small difference with each other.
Standard Deviation-
7.14106
By seeing the observation, the spread is moderate.
Skewness Value-
0.90646
By seeing the observation,it is positively skewed and more data on left side of mean.
Maximum Value from
the data is 37.97
Kurtosis Value- 0.4932
By seeing this observation, it has a high peak.

AVG_PRICE	
Mean	22.5328063
Standard Error	0.40886115
Median	21.2
Mode	50
Standard Deviation	9.19710409
Sample Variance	84.5867236
Kurtosis	1.49519694
Skewness	1.10809841
Range	45
Minimum	5
Maximum	50
Sum	11401.6
Count	506

10.Average Price
In Average Price
Mean Value- 22.5328
Median Value- 21.2
By seeing the observation, the mean and median values has a small difference with each other.
Standard Deviation-
9.19710
By seeing the observation, the spread is moderate.
Skewness Value-
1.10809
By seeing the observation, it is positively skewed and more data on left side of the mean.
Maximum Value from
the data is 50
Minimum Value from the
data is 5
Kurtosis Value- 1.4951
By seeing this observation,it has a high peak.

Q2. Plot a histogram of the Avg_Price variable. What do you infer?

Average Price Table in Excel File.



By seeing the above Histogram, the average is nearly 101 and kurtosis is high peak and skewness is positively skewed more data on the left, tail is in right side.

Q3. Compute the covariance matrix. Share your observations.

	CRIME_RATE	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*
CRIME_RATE	8.516147873								
AGE	0.562915215	790.7925							
INDUS	-0.11021518	124.2678	46.97143						
NOX	0.000625308	2.381212	0.605874	0.013401					
DISTANCE	-0.22986049	111.55	35.47971	0.61571	75.66653				
TAX	-8.22932244	2397.942	831.7133	13.0205	1333.117	28348.62			
PTRATIO	0.068168906	15.90543	5.680855	0.047304	8.743402	167.8208	4.677726		
AVG_ROOM	0.056117778	-4.74254	-1.88423	-0.02455	-1.28128	-34.5151	-0.53969	0.492695216	
LSTAT	-0.88268036	120.8384	29.52181	0.48798	30.32539	653.4206	5.7713	-3.07365497	50.893
AVG_PRICE	1.16201224	-97.3962	-30.4605	-0.45451	-30.5008	-724.82	-10.0907	4.484565552	-48.35

This covariance matrix provides insights into the relationships between pairs of variables.

A positive covariance between CRIME_RATE and AGE implies that as CRIME_RATE increases, AGE tends to increase as well, while a negative covariance between AGE and AVG_PRICE indicates that as AGE increases, AVG_PRICE tends to decrease.

Q4. Create a correlation matrix of all the variables (Use Data analysis tool pack).

a) Which are the top 3 positively correlated pairs
b) b) Which are the top 3 negatively correlated pairs.

	CRIME_RATE	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AV*
CRIME_RATE	1									
AGE	0.006859463	1								
INDUS	-0.00551065	0.6447785	1							
NOX	0.001850982	0.7314701	0.763651	1						
DISTANCE	-0.00905505	0.4560225	0.595129	0.611441	1					
TAX	-0.01674852	0.5064556	0.72076	0.668023	0.910228	1				
PTRATIO	0.010800586	0.261515	0.383248	0.188933	0.464741	0.460853	1			
AVG_ROOM	0.02739616	-0.2402649	-0.39168	-0.30219	-0.20985	-0.29205	-0.3555	1		
LSTAT	-0.04239832	0.6023385	0.6038	0.590879	0.488676	0.543993	0.374044	-0.6138083	1	
AVG_PRICE	0.043337871	-0.3769546	-0.48373	-0.42732	-0.38163	-0.46854	-0.50779	0.6953599	-0.73766	

Top 3 Positively Correlated

0.91022819
0.76365145
0.7314701

Top 3 Negatively Correlated

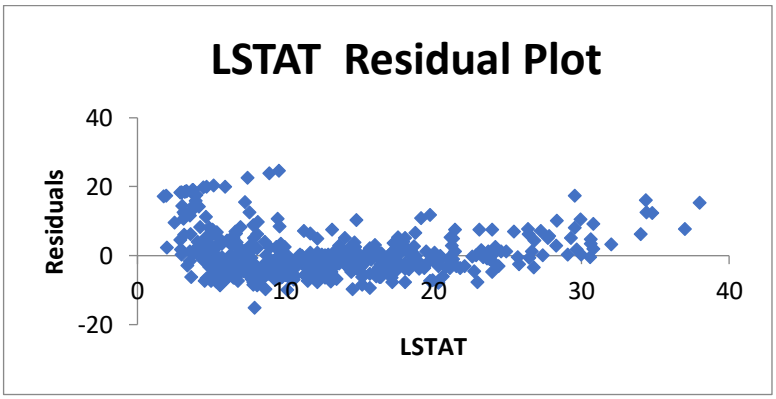
-0.73766
-0.61381
-0.50779

Q5. 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?

From the above ,

P-value
3.7E-236
5.08E-88

Adjusted R Square	0.543241826



a)

By observing the Regression, P- Value is less than 0.05

But Adjusted R square value is 0.54321
So, We cannot use this model for Regression
Resudial Plot is Random.

b) No, LSTAT Varaiable Cannot significant for the analysis based on my model.

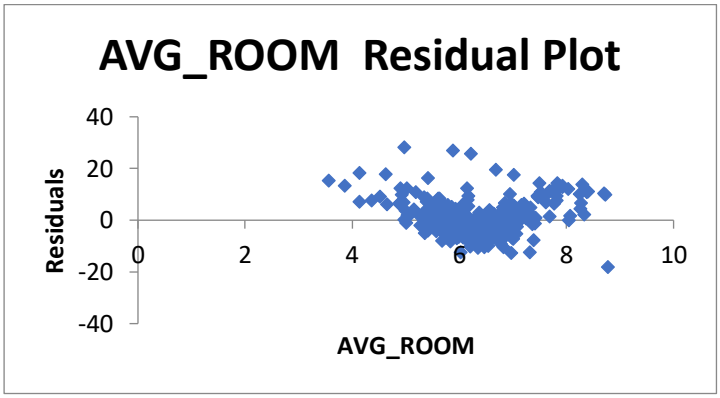
Q6. Build a new Regression model including LSTAT and AVG_ROOM together as Independent variables and AVG_PRICE as dependent variable

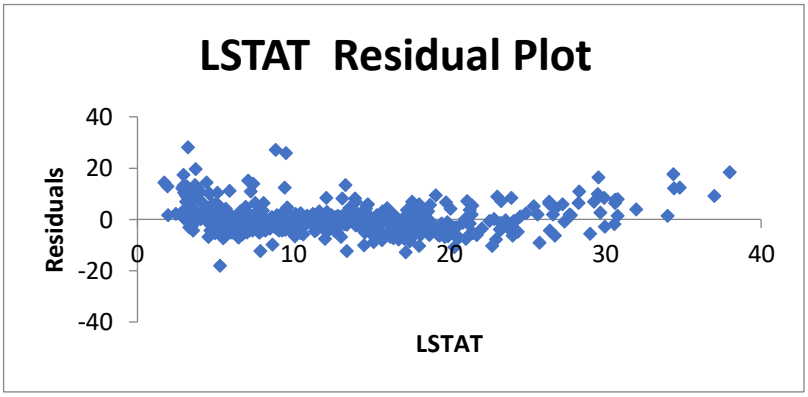
a) Write the Regression equation. If a new house in this locality has 7 rooms (on an average) and has a value of 20 for L-STAT, then what will be the value of AVG_PRICE? How does it compare to the company quoting a value of 30000 USD for this locality? Is the company Overcharging/ Undercharging?

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

P-value
0.668765
3.47E-27
6.67E-41

Adjusted R Square	0.637124475





Q6
1

The Company is Overcharging

Regression Equation=R10+R11*7+R12*20

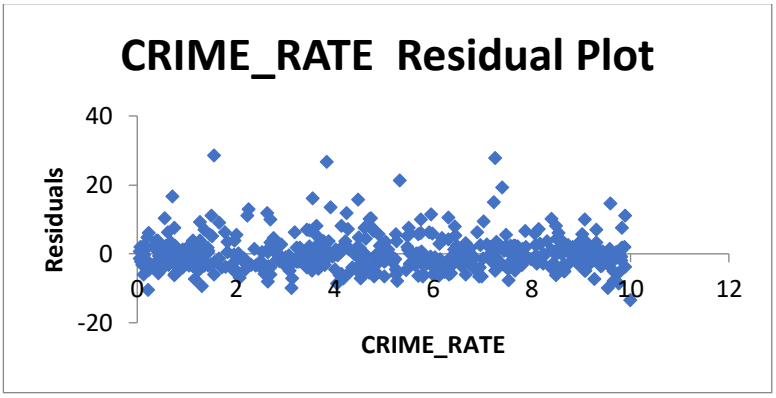
Q6
2

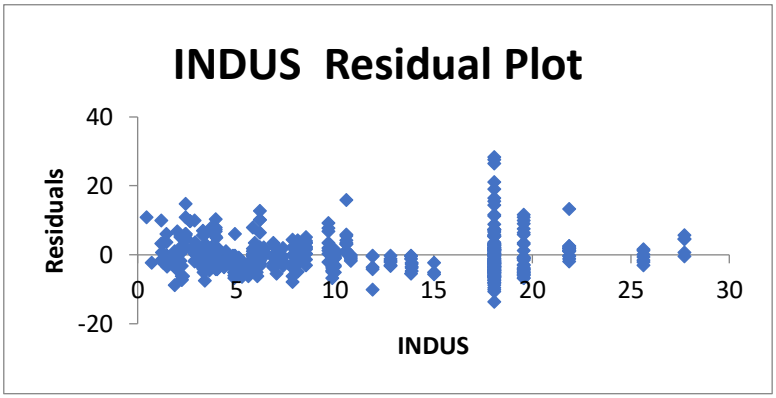
Compared to Adjusted R square in previous Question it is in 0.5 so we cannot use that model for Ana
and we can use this regression for Analysis,

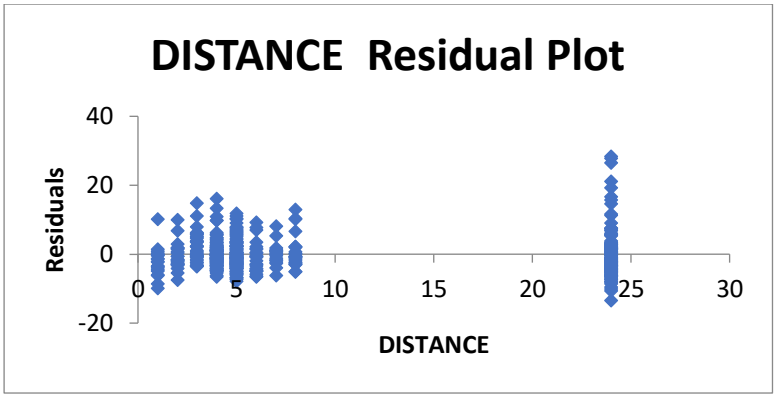
Q7.Build another Regression model with all variables where AVG_PRICE alone be the Dependent Variable and all the other variables are independent. Interpret the output in terms of adjusted R^2 square, coefficient and Intercept values. Explain the significance of each independent variable with respect to AVG_PRICE.

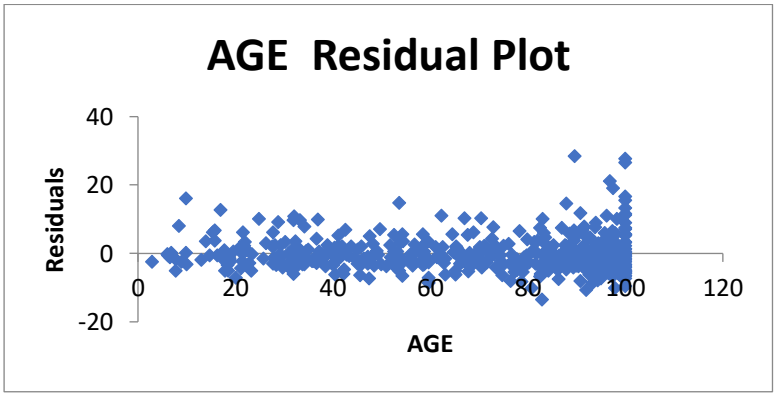
P-value
2.54E-09
0.534657
0.01267
0.039121
0.008294
0.000138
0.000251
6.59E-15
3.89E-19
8.91E-27

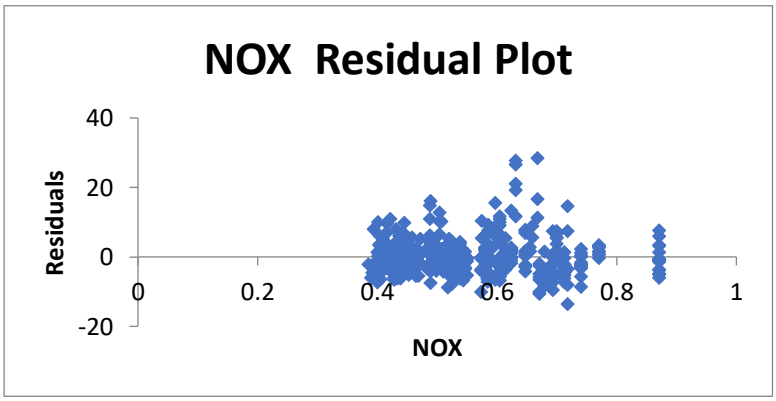
Adjusted R Square	0.688298647

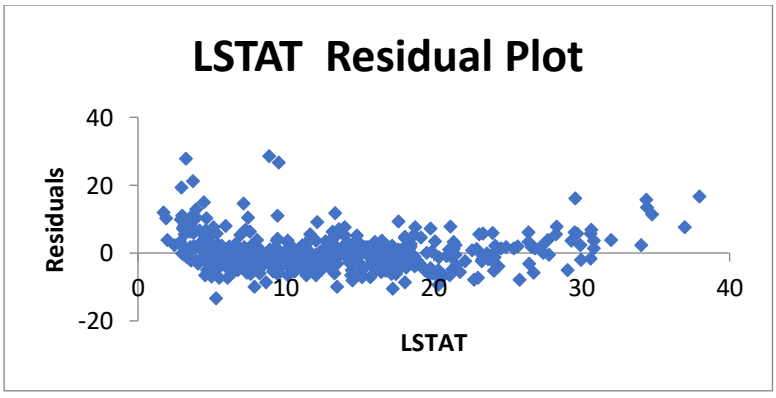


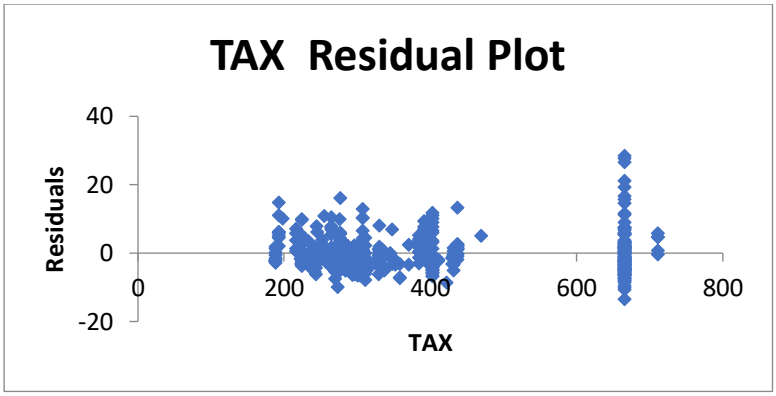


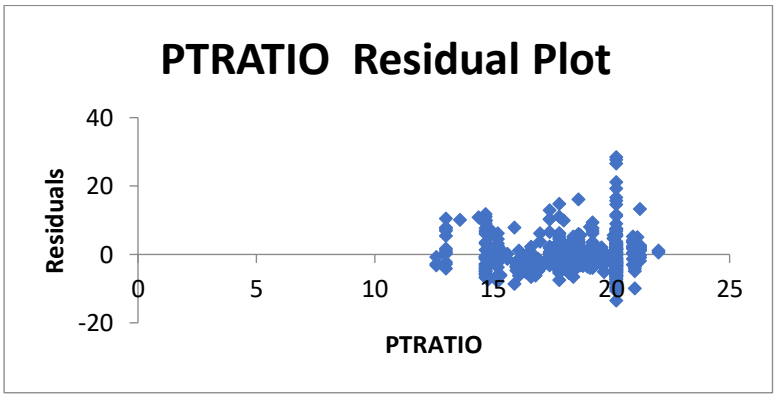

DISTANCE Residual Plot
Residuals
40
20
0
-20
0
5
10
15
20
25
30
DISTANCE

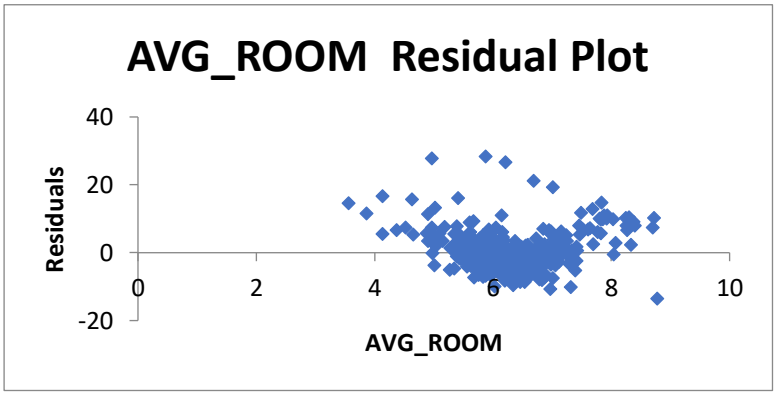

AGE Residual Plot
Residuals
40
20
0
-20
0
20
40
60
80
100
120
AGE


NOX Residual Plot
Residuals
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
NOX


LSTAT Residual Plot
Residuals
40
20
0
-20
0
10
20
30
40
LSTAT







Q7 By Observing the Regression Output, P- values are less than 0.05 except Crime_Rate.

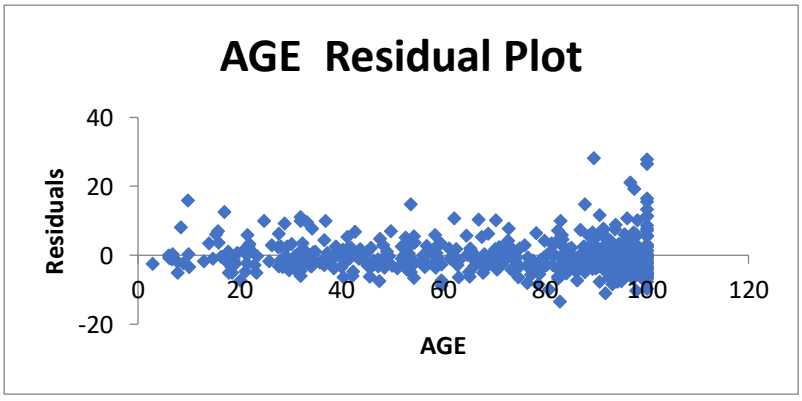
Adjusted R Square Value is 0.6 so we can use this model for analyzing and by eliminating Analyzing.

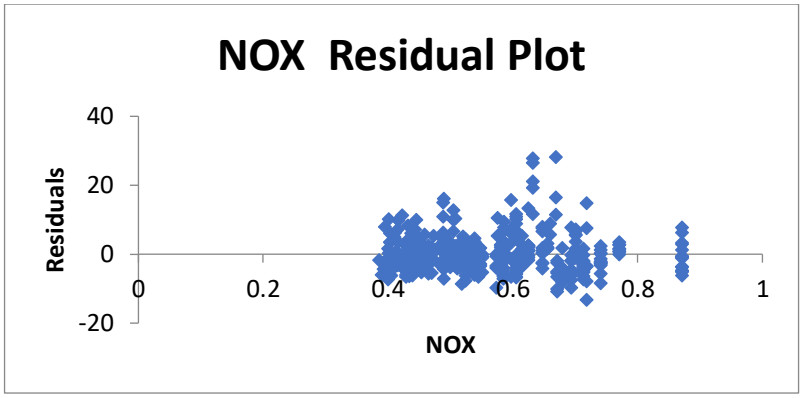
All Residuals Plots are Random plots.

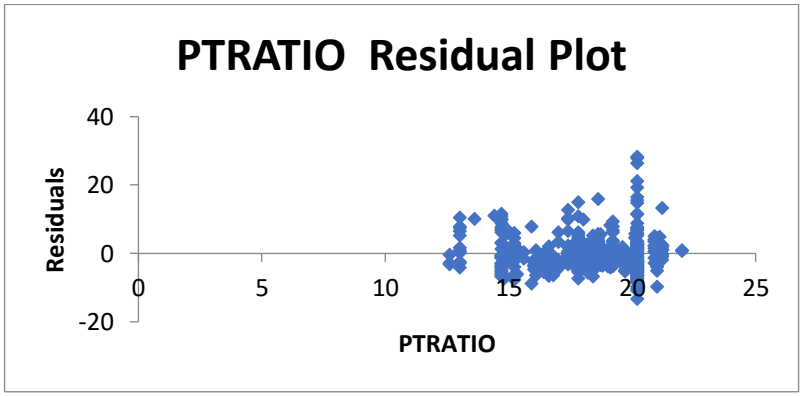
Q8. 8) Pick out only the significant variables from the previous question. Make another instance of the Regression model using only the significant variables you just picked and answer the questions below:
a) Interpret the output of this model.
b) Compare the adjusted R-square value of this model with 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.

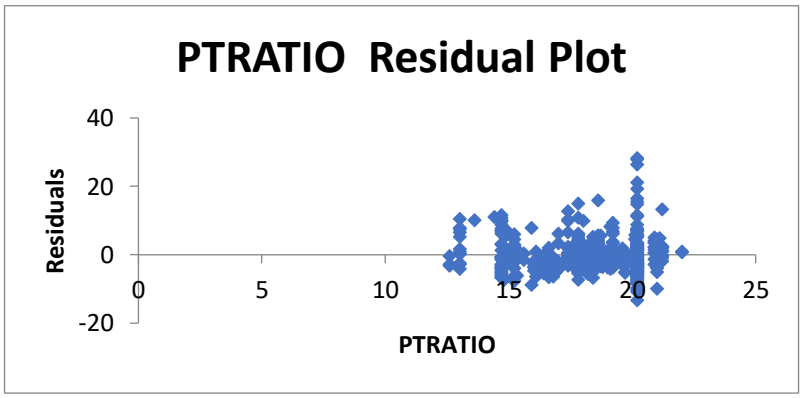
P-value
1.846E-09
0.01216288
0.03876167
0.00854572
0.00013289
0.00023607
7.0825E-15
3.6897E-19
5.4184E-27

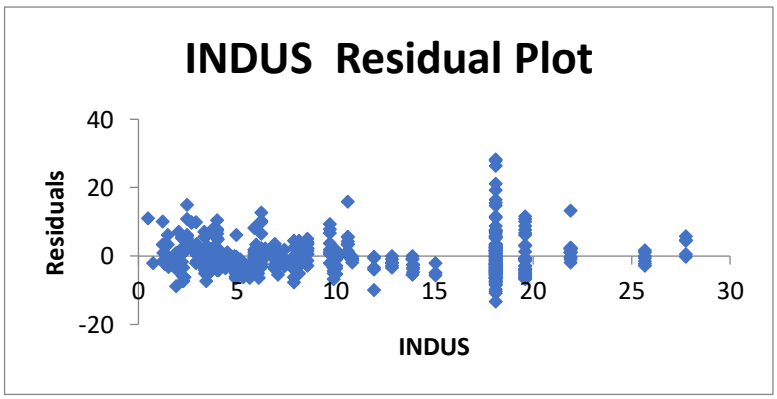
Adjusted R Square	0.688683682

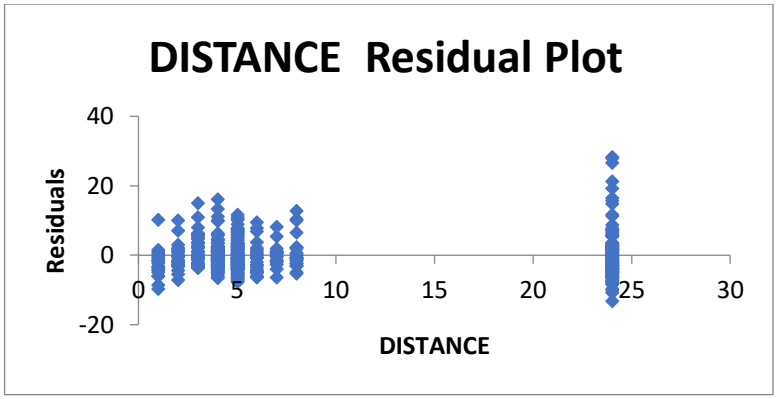


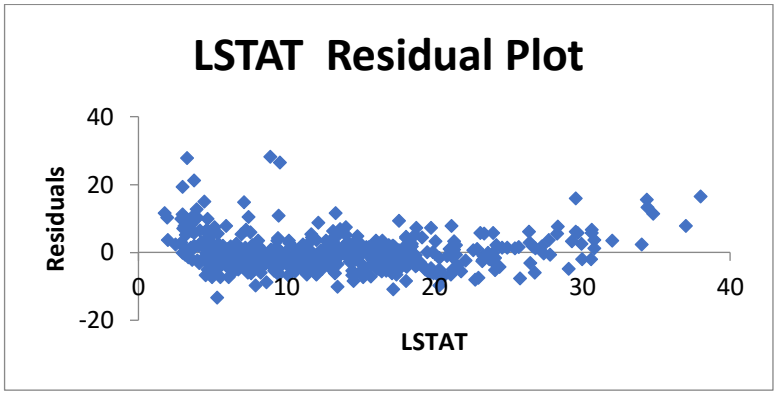

NOX Residual Plot
Residuals
40
20
0
-20
0
0.2
0.4
0.6
0.8
1
NOX

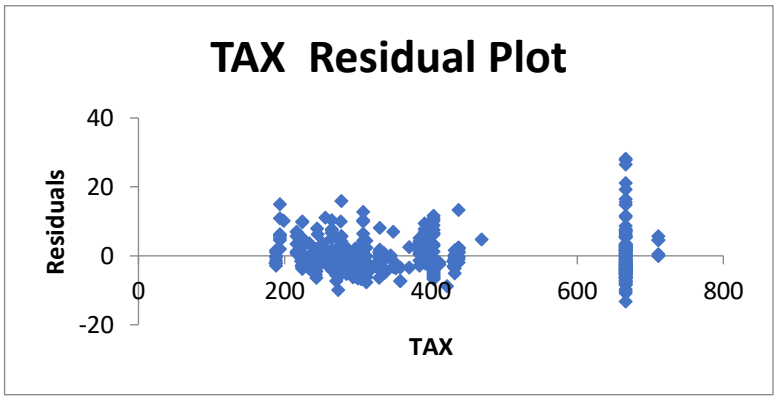

PTRATIO Residual Plot
Residuals
40
20
0
-20
0
5
10
15
20
25
PTRATIO


PTRATIO Residual Plot
Residuals
40
20
0
-20
0
5
10
15
20
25
PTRATIO


INDUS Residual Plot
Residuals
40
20
0
-20
0
5
10
15
20
25
30
INDUS







Q8 (1) By Observing the regression summary, all P- Value should be less than 0.05

Adjusted R Square value is 0.6,
All Residuals Plots should be Random

So, We can use this model for Analysis.

Q8 (2) By observing the Adjusted R Square Value from Previous Question , both the values will be same.

From Q8 Model is best for Analysis.

Q8 (3)

NOX	-10.27271
PTRATIO	-1.071702
LSTAT	-0.605159
TAX	-0.014452
AGE	0.032935
INDUS	0.13071
DISTANCE	0.2615064
AVG_ROOM	4.125469
Intercept	29.428473
	Coefficients

By Observing the values from the above table ,If the value of NOX is higher in a locality in this town, the average price of houses in that locality is expected to decrease. As NOX increases, AVG_PRICE is expected to decrease.

Q8 (4) AVG_PRICE = 29.42847349 - 10.27270508 * NOX - 1.071702473 * PTRATIO - 0.605159282 DISTANCE + 0.03293496 * AGE + 4.125468959 * AVG_ROOM