TERRO'S_REA

REAL ESTATE AGENCY
REPORT TOWN BOSTON

BALAJI K| PRICE ESTIMATION | 11th June 2023

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

CRIME_RATE	
Mean	4.871976285
Standard Error	0.129860152
Median	4.82
Mode	3.43
Standard Deviation	2.921131892
Sample Variance	8.533011532
Kurtosis	-1.189122464
Skewness	0.021728079
Range	9.95
Minimum	0.04
Maximum	9.99
Sum	2465.22
Count	506

Crime Rate: -

The Average crime rate is about 4.87 per person.

The Median is also almost equal to the average like 4.82 per person.

The mode is about 3.43

The SD is unlikely to be 2.92 which is not good, the spread is high which makes crime rate to become little unpredictable.

The Kurtosis is negative about -1.18 called Platykurtic or Flattened curve.

The Skewness is Positively skewed which is about 0.02

and the distribution is found to be normal.

The range is about 9.95 which says maximum crime rate of 9.99 and the minimum crime rate of 0.04.

The total count is about 506 in Boston Town.

AGE	
Mean	68.57490119
Standard Error	1.251369525
Median	77.5
Mode	100
Standard Deviation	28.14886141
Sample Variance	792.3583985
Kurtosis	-0.967715594
Skewness	-0.59896264
Range	97.1
Minimum	2.9
Maximum	100
Sum	34698.9
Count	506

Age: -

The Average proportion of house build prior to 1940 was about 68.57 for about 506 Houses.

The Median is found to about 77.5

The mode is about 100 says frequent age of House in Boston

The SD is likely to be 28.14 which is acceptable case the spread is little high.

The Kurtosis is negative about -0.96 called Platykurtic or Flattened curve.

The Skewness is negatively skewed which is about -0.598

and says the median is greater than median the Boston town has high frequency of aged houses.

The range is about 97.1 which says maximum age of 100 and the minimum of 2.9

The total count is about 506 Houses in Boston Town.

INDUS	
Mean	11.13677866
Standard Error	0.304979888
Median	9.69
Mode	18.1
Standard Deviation	6.860352941
Sample Variance	47.06444247
Kurtosis	-1.233539601
Skewness	0.295021568
Range	27.28
Minimum	0.46
Maximum	27.74
Sum	5635.21
Count	506

Indus: -

The Average Indus rate is about 11.13%

The Median is 9.69%

The mode is about 18.1 %

The SD is about 6.86 which says the spread is high.

The Kurtosis is negative about -1.23 called Platykurtic or Flattened curve.

The Skewness is Positively skewed which is about 0.295

The range is about 27.28 which says maximum % of 27.74 and the minimum % ratio of 0.46, there will be an outlier in this type which makes SD to high

The total count is about 506 in Boston Town.

NOX	
Mean	0.554695059
Standard Error	0.005151391
Median	0.538
Mode	0.538
Standard Deviation	0.115877676
Sample Variance	0.013427636
Kurtosis	-0.064667133
Skewness	0.729307923
Range	0.486
Minimum	0.385
Maximum	0.871
Sum	280.6757
Count	506

NOx: -

The Average PPM is about 0.55 Which is found to be safer PPM

The Median is 0.538 PPM is almost equal to mean.

The mode is similar to median about 0.538PPM

The SD is about 0.115 PPM which is good, the spread is good.

The Kurtosis is negative about -0.06 denotes Platykurtic or Flattened curve.

The Skewness is Positively skewed which is about 0.729

The range is about 0.486 which says maximum PPM of 0.871 and the minimum PPM of 0.385, so the SD is found Good.

The total count is about 506 in Boston Town.

DISTANCE	
Mean	9.549407115
Standard Error	0.387084894
Median	5
Mode	24
Standard Deviation	8.707259384
Sample Variance	75.81636598
Kurtosis	-0.867231994
Skewness	1.004814648
Range	23
Minimum	1
Maximum	24
Sum	4832
Count	506

Distance: -

The Average miles is about 9.54 away from highway

The Median is 5 miles.

The mode is greater than that of mean and median which is about 24 miles.

The SD is about 8.70 miles the spread is not good almost near to mean.

The Kurtosis is negative about -0.86 denotes Platykurtic or Flattened curve.

The Skewness is Positively skewed which is about 1.00

The range is about 23 which says maximum miles away from highway is 24 and the minimum 1, so the SD is not good.

The total count is about 506 in Boston Town.

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

Tax: -

The Average Tax is about $408.23 per $10,000

The Median Tax is $330

The repeated tax was about $666 which is so called Mode.

The SD is about $168.53 which says the spread is ok.

The Kurtosis is negative about -1.14 denotes Platykurtic or Flattened curve.

The Skewness is Positively skewed which is about 0.66

The range is about $524 which says maximum Tax is $711and the minimum $187

The total count is about 506 in Boston Town.

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

PTRATIO:-

The Average PT ratio is about 18.45 by town.

The Median is 19.05 which is greater than mean.

The mode is greater than that of mean and median which holds ratio of 20.2

The SD is about 2.16 which is good, the spread is good

The Kurtosis is negative about -0.28 denotes Platykurtic or Flattened curve.

The Skewness is negatively skewed which is about -0.80, also says that median is greater than that of mean and median.

The range is about 9.4 which says maximum ratio of22 and the minimum 12.6, so the SD is not good.

The total count is about 506 in Boston Town.

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

AVG_ROOM: -

The Average number of rooms is about 6.28

The Median is 6.20 which is almost similar to mean.

The mode is 5.713 Rooms, considering all these 3 central tendencies the rooms having Boston town is 6 per house.

The SD is about 0.70 which is good, the spread is good

The Kurtosis is Positive about 1.89 denotes Leptokurtic or meso curve.

The Skewness is Positively skewed which is about 0.40 mean is greater than that of median and mode.

The range is about 5.219 which says maximum ratio of 8.78 and the minimum 3.561.

The total count is about 506 in Boston Town.

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

LSTAT: -

The Average LSTAT% is about 12.65 in Boston.

The Median is 11.36% which is near to mean.

The mode is 8.05%

The SD is about 7.14 which is not really good, the spread is will be high.

The Kurtosis is positive about 0.49 denotes Leptokurtic or Meso the peak is tending to sharp.

The Skewness is Positively skewed which is about 0.90 mean is greater than median.

The range is about 36.24% which says maximum of 37.97% of lower the population and the minimum 1.73%, so the spread will be high impacts the SD.

The total count is about 506 in Boston Town.

AVG_PRICE	
Mean	22.53280632
Standard Error	0.408861147
Median	21.2
Mode	50
Standard Deviation	9.197104087
Sample Variance	84.58672359
Kurtosis	1.495196944
Skewness	1.108098408
Range	45
Minimum	5
Maximum	50
Sum	11401.6
Count	506

- **AVG_PRICE: -**
- The Average Value of House is about 22.53 ($$$) in Boston.
- The Median is 21.2 which is almost near to mean.
- The mode is 50 which is having the outliers so the median will be considered in this variable.
- The SD is about 9.19 which says the spread is considerably good.
- The Kurtosis is positive about 1.49 denotes Leptokurtic or Meso the peak is tending to sharp.
- The Skewness is Positively skewed which is about 1.10
- The range is about 45 which says maximum of 50 thousand dollars and the minimum of 5 which is having the outliers.
- The total count is about 506 in Boston Town.

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

AVG_Price Histogram: -

- The average price of Boston Housing is about 22.53($$$)
- Since this a financial term and also this type of variable will be having a outlier which will tend to impact the mean of 22.53($$$), in this case I tend to choose median as AVG_Price model which will more considerable in this type of variable
- So, the median is about 21.2($$$)
- In this Histogram graph the max frequency holds 167 for bin range (20,25]
- This also shows the curve Positively Skewed where it is having the tails towards its right, in other words mean more values were around the mean and median.
- And the Kurtosis also tends to sharp which will holds a positive value called Leptokurtic which is also explained in summary statistics
- The outliers in this case consist of 22 number frequency which is having avg_price of 45 to 50 thousand $

Q3) Compute the covariance matrix. Share your observations

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	8.52									
AGE	0.56	790.79								
INDUS	-0.11	124.27	46.97							
NOX	0.00	2.38	0.61	0.01						
DISTANCE	-0.23	111.55	35.48	0.62	75.67					
TAX	-8.23	2397.94	831.71	13.02	1333.12	28348.62				
PTRATIO	0.07	15.91	5.68	0.05	8.74	167.82	4.68			
AVG_ROOM	0.06	-4.74	-1.88	-0.02	-1.28	-34.52	-0.54	0.49		
LSTAT	-0.88	120.84	29.52	0.49	30.33	653.42	5.77	-3.07	50.89	
AVG_PRICE	1.16	-97.40	-30.46	-0.45	-30.50	-724.82	-10.09	4.48	-48.35	84.42

Co-Variance Measure: - (Deep Summary)

- The term generally defines that how far the observations from the mean, in case of two variables it is called as Co-Variance
- Crime Rate & Age their values are positive which their values of both were above or below average.
- Indus & Crime rate their values tend to have negative value, so one of their variable averages is likely to have negative side
- Indus & Age both variable values were above their averages so it is Positive.
- NOx & Crime rate has no relation to each other in value tend to have Zero as average.

- NOx & Age both of their values were above or below their averages, so its covariation tends to remain Positive.
- NOx & Indus Both of their values were above or below their averages, so its covariation tends to remain Positive.
- Distance & NOx one of their variables tend to have less value than their average so it tends to have the negative value.
- Distance & age, Distance & Indus, Distance & NOx all these tend to have both respective variable value High or Low than their averages so it is also positive
- Tax & crime rate, one of the variables mean tend to have lower mean so this says it is a negative co variance.
- Tax & Age, Tax & Indus, Tax & NOx, Tax & Distance all the variables of each column gave their mean above or below their average, so it is having positive Co-variation
- PTRATIO & Crime rate, PTRATIO & Age, PTRATIO & Indus, PTRATIO & NOx, PTRATIO & Distance, PTRATIO & Tax all these tend to have their above or below so it is having positive Co-Variation number.
- AVG_ROOM & Crime Rate tend to have their variable value above or below their average so it holds the positive Co-Variation.
- AVG_ROOM & Age, AVG_ROOM & Indus, AVG_ROOM & NOx, AVG_ROOM & Distance, AVG_ROOM & Tax, AVG_ROOM & PTRATIO all the other holds a Negative Co-variance.
- LSTAT & Crime rate holds negative Co-variance which holds one of their variables is having low value than their average.

- LSTAT & Age, LSTAT & Indus, LSTAT & NOx, LSTAT & Distance, LSTAT & Tax, LSTAT & PTRATIO all the other holds a positive Co-variation.
- LSTAT & AVG_ROOM holds a negative Co-variation.
- AVG_Price & Crime Rate holds a Positive Co-variation which holds a value of above or below their averages.
- AVG_Price & Age, AVG_Price & Indus, AVG_Price & NOx, AVG_Price & Distance, AVG_Price & Tax, AVG_Price & PTRATIO, AVG_Price & LSTAT Holds a Negative Co-Variation.
- AVG_Price & AVG_Room Has Positive Co-Variation.

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

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	1.00									
AGE	0.01	1.00								
INDUS	-0.01	0.64	1.00							
NOX	0.00	0.73	0.76	1.00						
DISTANCE	-0.01	0.46	0.60	0.61	1.00					
TAX	-0.02	0.51	0.72	0.67	0.91	1.00				
PTRATIO	0.01	0.26	0.38	0.19	0.46	0.46	1.00			
AVG_ROOM	0.03	-0.24	-0.39	-0.30	-0.21	-0.29	-0.36	1.00		
LSTAT	-0.04	0.60	0.60	0.59	0.49	0.54	0.37	-0.61	1.00	
AVG_PRICE	0.04	-0.38	-0.48	-0.43	-0.38	-0.47	-0.51	0.70	-0.74	1.00

4a) Top 3 Positively Co-Related Pairs: -

1.Distance & Tax (0.91)

2.NOx & Indus (0.76)

3.NOx & Age (0.73)

4b) Top 3 Negatively Co-Related Pairs: -

1.AVG_PRICE & LSTAT (-0.74)

2.LSTAT & AVG_ROOM (-0.61)

3.AVG_PRICE & PTRATIO (-0.51)

Q5) Build an initial regression model with AVG_PRICE as 'y' (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot
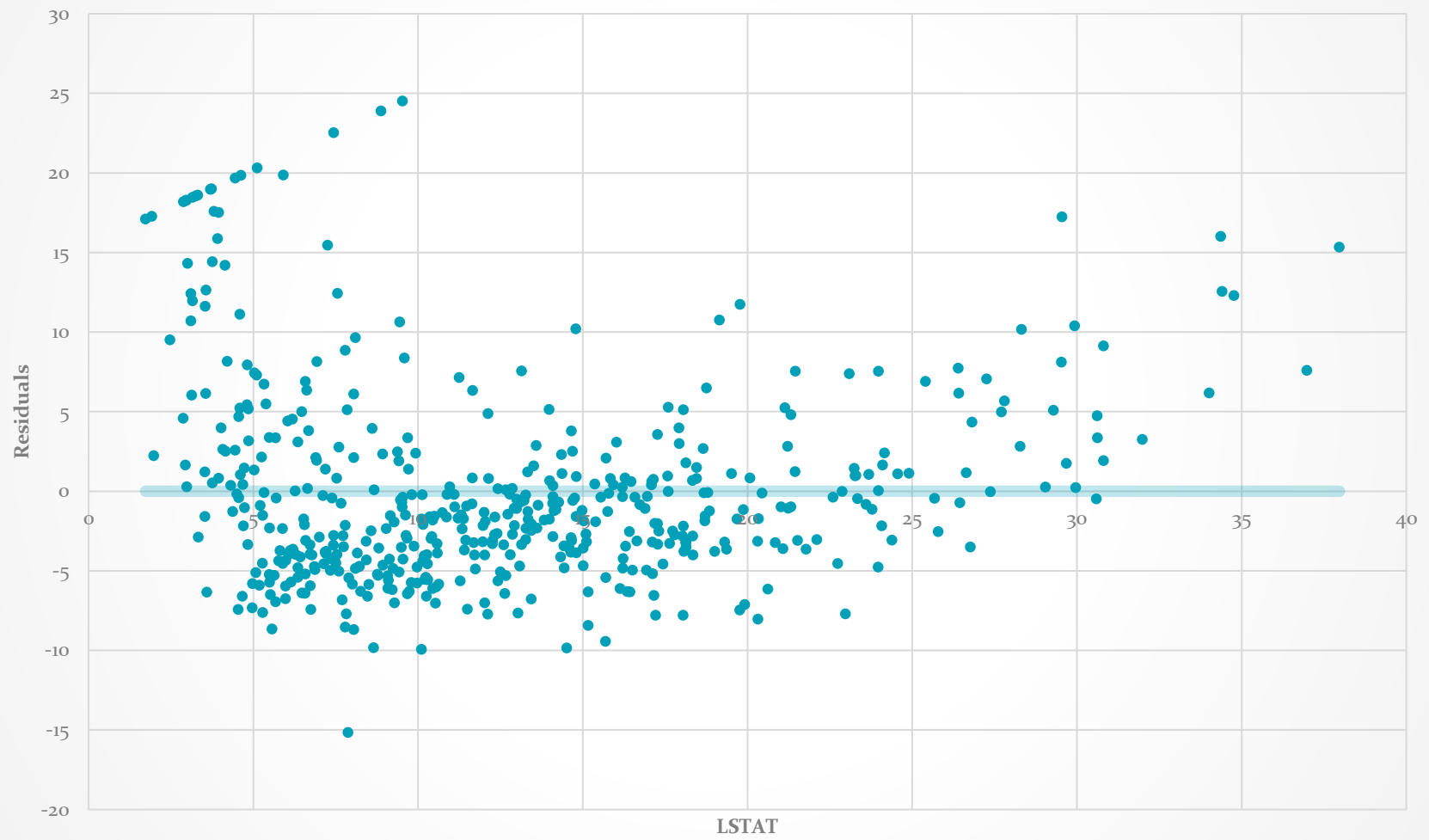
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.737662726
R Square	0.544146298
Adjusted R Square	0.543241826
Standard Error	6.215760405
Observations	506

ANOVA

	df	SS	MS	F	Significance F
Regression	1	23243.914	23243.914	601.6178711	5.0811E-88
Residual	504	19472.38142	38.63567742		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	34.55384088	0.562627355	61.41514552	3.7431E-236	33.44845704	35.65922472	33.44845704	35.65922472
LSTAT	-0.950049354	0.038733416	-24.52789985	5.0811E-88	-1.0261482	-0.873950508	-1.0261482	-0.873950508

LSTAT Residual Plot

Regression Summary Output: -

As mentioned in the question this is the regression summary output which holds Avg_price as (dependent variable) & LSTAT as (Independent variable)

5)a)

The Co-Variation of this model holds -48.35 & the Co-relation of -0.74 which is a negative Linear relation

The Adjusted **R Square Value = 0.543**

P-Value of Simple Linear Regression model = 5.0811E-88

P-Value of Intercept = 3.743E-236

Co-efficient of LSTAT = -0.95004

Co-efficient of Intercepts = 34.55

The residual plot showing that there is no patterned error in the observation.

The SLR equation would be **Predicted AVG_PRICE = 34.5538 - 0.9500 * LSTAT**

5b) **All the conditions hold good enough to progress this model, but the adjusted R-squared value is 0.543 not at all satisfied, still this regression model needs Improvements, Considering the P-value this LSTAT variable is significant model for predicting the avg_price.**

6) Build a new Regression model including LSTAT and AVG_ROOM together as independent variables and AVG_PRICE as dependent variable.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.799100498
R Square	0.638561606
Adjusted R Square	0.637124475
Standard Error	5.540257367
Observations	506

ANOVA

	df	SS	MS	F	Significance F
Regression	2	27276.98621	13638.49311	444.3308922	7.0085E-112
Residual	503	15439.3092	30.69445169		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.358272812	3.17282778	0.428095348	0.668764941	-7.591900282	4.875354658	-7.591900282	4.875354658
AVG_ROOM	5.094787984	0.4444655	11.46272991	3.47226E-27	4.221550436	5.968025533	4.221550436	5.968025533
LSTAT	-0.642358334	0.043731465	-14.68869925	6.66937E-41	-0.728277167	-0.556439501	-0.728277167	-0.556439501

6.a) **Regression Equation for the Following model which holds dependent variable as(AVG_PRICE) & Independent variables (AVG_ROOM), (LSTAT) is**

Predicted AVG_PRICE = -1.358 + 5.094 * AVG_ROOM - 0.642 * LSTAT

Average = 7

LSTAT Value = 20

Predicted AVG_PRICE = -1.358 + 5.094 * (7) - 0.642 * 20

= 21.458($$$)

Based on the given data the company quoted $30,000 for this location which would be overcharging as this holds a predicted value of $21,458

6)b) **Comparing the two models: -**

In the **first model** the adjusted R-square holds value of **0.543**

In this **MLR model** the adjusted R-square holds value of **0.637**

Comparing the two this MLR model has improved Adjusted R-square of 0.637 which is considerably good than the First SLR model.

This is because the Co-relation between AVG_PRICE & AVG_ROOM is 0.70 & AVG_PRICE & LSTAT is -0.74 (Positive & Negative) respectively.

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.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832978824
R Square	0.69385372
Adjusted R Square	0.688298647
Standard Error	5.1347635
Observations	506

ANOVA

	df	SS	MS	F	Significance F
Regression	9	29638.8605	3293.206722	124.9045049	1.9328E-121
Residual	496	13077.43492	26.3657962		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	29.24131526	4.817125596	6.070282926	2.53978E-09	19.77682784	38.70580267	19.77682784	38.70580267
CRIME_RATE	0.048725141	0.078418647	0.621346369	0.534657201	-0.105348544	0.202798827	-0.105348544	0.202798827
AGE	0.032770689	0.013097814	2.501996817	0.012670437	0.00703665	0.058504728	0.00703665	0.058504728
INDUS	0.130551399	0.063117334	2.068392165	0.03912086	0.006541094	0.254561704	0.006541094	0.254561704
NOX	-10.3211828	3.894036256	-2.650510195	0.008293859	-17.97202279	-2.670342809	-17.97202279	-2.670342809
DISTANCE	0.261093575	0.067947067	3.842602576	0.000137546	0.127594012	0.394593138	0.127594012	0.394593138
TAX	-0.01440119	0.003905158	-3.687736063	0.000251247	-0.022073881	-0.0067285	-0.022073881	-0.0067285
PTRATIO	-1.074305348	0.133601722	-8.041104061	6.58642E-15	-1.336800438	-0.811810259	-1.336800438	-0.811810259
AVG_ROOM	4.125409152	0.442758999	9.317504929	3.89287E-19	3.255494742	4.995323561	3.255494742	4.995323561
LSTAT	-0.603486589	0.053081161	-11.36912937	8.91071E-27	-0.70777824	-0.499194938	-0.70777824	-0.499194938

7)Multilinear Regression Summary: -

As Mentioned, this model holds dependent variable as AVG_PRICE & all the other given variable as independent variable.

Adjusted R-Square: -

The Adjusted R-Square Value is about 0.688 which is considerably good when compared to other two models created before & it is also greater than 0.5, still this model needs Improvements.

P-value: -

But still this model's variable crime rate holds P-value of 0.53 which is significantly high, so this model drops out because of this variable respect to AVG_PRICE, since the co-relation of these two says that it has almost no relation to each other (0.04)

All the other variable holds P-value less than 0.05.

Co-efficient: -

The Intercepts holds a good rate of 29.24 ($$$)

	Coefficients	Description
Intercept	29.24131526	Intercept Holds a pretty good value of 29.24($$$)
CRIME_RATE	0.048725141	Since this value having the P-value of 0.53 which is not acceptable should be < 0.05, so it is not considered as good predictor and also this contains almost no relation with Avg_Price (0.04)
AGE	0.032770689	For each value of age variable, the average predicted price of House increases by 0.0327
INDUS	0.130551399	For each value of Indus variable predicted price of House increases by 0.1305 which Impacts the House price to increase (as this increases Employment chances when Industries started to Implement their Plants over Boston.
NOX	-10.3211828	For each value of NOx variable, the average predicted price of House decreases by -10.32, this affects the pricing because Higher NOx in the air tends to be Poisonous & also leads to Pollution of air creates brown layer in Air causing accidents, etc
DISTANCE	0.261093575	For each value Increase of distance variable predicted price of House increases by 0.2610, so people want to stay in City side rather than staying nearby highways.
TAX	-0.01440119	For each value Increase of tax variable, the average predicted price of House decreases by -0.014, which affects the house pricing when tax Increases

PTRATIO	-1.074305348	For each value of ptratio variable the average predicted price of House decreases by -1.074, this area contains lack of Pupil or Lack of Teacher which Impacts the future student's life & also affects the house pricing
AVG_ROOM	4.125409152	For each value of avg_room variable predicted price of House increases by 4.1254, when number of rooms increases the average price of House also Increases, Considered as good factor.
LSTAT	-0.603486589	For each value Increase of Lstat variable the average predicted price of House decreases by -0.603, the lower the status of people tend to affect the house pricing, maybe they lack in paying the house rent or some other factors

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:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832835773
R Square	0.693615426
Adjusted R Square	0.688683682
Standard Error	5.131591113
Observations	506

ANOVA

	df	SS	MS	F	Significance F
Regression	8	29628.68142	3703.585178	140.6430411	1.911E-122
Residual	497	13087.61399	26.33322735		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
NOX	-10.27270508	3.890849222	-2.640221837	0.008545718	-17.9172457	-2.628164466	-17.9172457	-2.628164466
PTRATIO	-1.071702473	0.133453529	-8.030529271	7.08251E-15	-1.333905109	-0.809499836	-1.333905109	-0.809499836
LSTAT	-0.605159282	0.0529801	-11.42238841	5.41844E-27	-0.70925186	-0.501066704	-0.70925186	-0.501066704
TAX	-0.014452345	0.003901877	-3.703946406	0.000236072	-0.022118553	-0.006786137	-0.022118553	-0.006786137
AGE	0.03293496	0.013087055	2.516605952	0.012162875	0.007222187	0.058647734	0.007222187	0.058647734
INDUS	0.130710007	0.063077823	2.072202264	0.038761669	0.006777942	0.254642071	0.006777942	0.254642071
DISTANCE	0.261506423	0.067901841	3.851242024	0.000132887	0.128096375	0.394916471	0.128096375	0.394916471
AVG_ROOM	4.125468959	0.44248544	9.323400461	3.68969E-19	3.256096304	4.994841615	3.256096304	4.994841615
Intercept	29.42847349	4.804728624	6.124898157	1.84597E-09	19.98838959	38.8685574	19.98838959	38.8685574

8)Multilinear Regression summary:-(Final Model)

a) **Adjusted R-square value = 0.688,** which is considerably good

This model after Ignoring the Crime rate all the other P-value becomes significant.

b) **Comparing the adjusted R-square with previous model, it is slightly Improved in this model 0.68868>0.68829**

This model performs better than the previous model & also its intercept value also slightly improved due to ignorance of Crime rate.

P-Value: -

All the P-Value of this model holds value less than 0.05

c) **Co-efficient sorted in ascending order,**

	Coefficients
NOX	-10.27270508
PTRATIO	-1.071702473
LSTAT	-0.605159282
TAX	-0.014452345
AGE	0.03293496
INDUS	0.130710007
DISTANCE	0.261506423
AVG_ROOM	4.125468959
Intercept	29.42847349

If the value of NOX increases, the average price of House decreases. Since the NOX (Nitrous Oxide in air is considered as pollution, if the pollution increases it affects ultimately the pricing factor of House in Boston.

d) **Regression equation of the model: -**

Predicted AVG_PRICE = 29.428 + 0.0329 * Age + 0.130 * Indus - 10.272 * NOX + 0.261 * Distance - 0.0144 * Tax - 1.071 * PTRATIO + 4.125 * AVG_ROOM - 0.605 * LSTAT