ASSIGNMENT: TERRO'S REAL ESTATE AGENCY

Date:14/01/2023

Author: Ishaan Kohli

Problem Statement:

"Finding out the most relevant features for pricing of a house" Terro's real-estate is an agency that estimates the pricing of houses in a certain locality. The pricing is concluded based on different features / factors of a property. This also helps them in identifying the business value of a property. To do this activity the company employs an "Auditor", who studies various geographic features of a property like pollution level (NOX), crime rate, education facilities (pupil to teacher ratio), connectivity (distance from highway), etc. This helps in determining the price of a property

The agency has provided a dataset of 506 houses in Boston. Following are the details of the dataset:

Data Dictionary:

Attribute	Description
CRIME RATE	per capita crime rate by town
INDUSTRY	proportion of non-retail business acres per town (in percentage terms)
NOX	nitric oxides concentration (parts per 10 million)
AVG_ROOM	average number of rooms per house
AGE	proportion of houses built prior to 1940 (in percentage terms)
DISTANCE	distance from highway (in miles)
TAX	full-value property-tax rate per $10,000
PTRATIO	pupil-teacher ratio by town
LSTAT	% lower status of the population
AVG_PRICE	Average value of houses in $1000's

Objective (Task)

Your job, as an auditor, is to analyze the magnitude of each variable to which it can affect the price of a house in a particular locality.

Assumptions – Following assumptions were taken in to consideration for the Analysis

1. The relation Independent Variable & Target Variable should be in a Linear Relation.
2. It assumes that there is no relationship between Independent Variable's.
3. It assumes that the points on the points on the Best Fit Line have a Normal Distribution.
4. The Errors are Independent.
5. There should be homoscedasticity of Errors.

QUESTION 1: GENERATE THE SUMMARY STATISTICS FOR EACH VARIABLE IN THE TABLE. (USE DATA ANALYSIS TOOL PACK). WRITE DOWN YOUR OBSERVATION.

OBSERVATIONS

Table 1: Crime Rate

Column1	Column2
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
Largest(1)	9.99
Smallest(1)	0.04
Confidence Level(95.0%)	0.255132688

- The Average Crime rate of Boston is 4.8719% per Capita Crime rate.
- The Graph of Crime Rate is Right Skewed.
- The graph is Platykurtic in Nature.

Table 2: Age

Column1	Column2
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
Largest(1)	100
Smallest(1)	2.9

- The Average Age of House in Boston is 68 years(Approx).
- The oldest house in Boston is 100 Years old
- The most recent built house is 3 years old in Boston.
- Majority of the House in Boston are around 100 Years old.
- The Graph is Platykurtic in nature.
- The Graph of Age is Left Skewed.

Confidence Level(95.0%)	2.458531467

Table 3: Indus

Column1	Column2
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
Largest(1)	27.74
Smallest(1)	0.46
Confidence Level(95.0%)	0.599185642

- There is around 11.13% of Wholesale Business in Boston Where our houses are constructed.
- The Graph is Right Skewed.
- The graph is Platykurtic in Nature.

Table 4: NOX

Column1	Column2
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
Largest(1)	0.871
Smallest(1)	0.385
Confidence Level(95.0%)	0.010120797

- There is around 55.4696% Nitric Oxides present in parts (per 10 million).
- The Maximum NOX present in an area is 87.1% And the minimum is 38.5%.
- The Graph is Right Skewed.
- The graph is Platykurtic in Nature

Table 5: Distance

Column1	Column2
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
Largest(1)	24
Smallest(1)	1
Confidence Level(95.0%)	0.760495101

- The houses in Boston are having a distance of 9.54 miles on an average from the Highway.
- The maximum distance of a house from the Highway is 24 miles.
- There are also houses which are only 1 mile away from the Highway.
- The graph is Platykurtic in Nature
- The Graph of Distance is Right Skewed.

Table 6: Tax

Column1	Column2
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
Largest(1)	711
Smallest(1)	187
Confidence Level(95.0%)	14.72009106

- The Average tax of a House in Boston per $10,000 is $408.237.
- Majority of the Houses are having Tax of $666 in Boston.
- The Minimal Tax payable for a house In Boston is $167 per $10,000.
- The Maximum Tax payable for a house In Boston is $711 per $10,000.
- The Graph is Right Skewed.
- The graph is Platykurtic in Nature

Table 7: PTRATIO

Column1	Column2
PTRATIO	

- The Average Pupil-Teacher ratio by town is around 18.45.
- The Graph of PTRATIO is Left Skewed.

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
Largest(1)	22
Smallest(1)	12.6
Confidence Level(95.0%)	0.189087104

- The graph is Platykurtic in Nature.

Table 8: AVG_ROOM

Column1	*Column2*
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
Largest(1)	8.78
Smallest(1)	3.561
Confidence Level(95.0%)	0.061366829

- The Average Rooms in a House in Boston is 6.
- Majority of the houses in Boston are having 5 rooms.
- There are a maximum of 8 rooms in a house.
- The minimum no .of rooms in a house in Boston are 3.
- The graph is Leptokurtic in Nature
- The Graph is Right Skewed.

Table 9: LSTAT

Column1	*Column2*
LSTAT	
Mean	12.65306324
Standard Error	0.317458906
Median	11.36
Mode	8.05
Standard Deviation	7.141061511

- There is an average around 12.65% of houses In the LSTAT Variable.
- The Graph of LSTAT is Left Skewed.
- The graph is Leptokurtic in Nature

Sample Variance	50.99475951
Kurtosis	0.493239517
Skewness	0.906460094
Range	36.24
Minimum	1.73
Maximum	37.97
Sum	6402.45
Count	506
Largest(1)	37.97
Smallest(1)	1.73
Confidence Level(95.0%)	0.623702827

Table 10: AVG_PRICE

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
Largest(1)	50
Smallest(1)	5
Confidence Level(95.0%)	0.80327831

- The Average Price of a house in Boston is $22,532.80.
- The Minimum cost of a house in Boston is $5,000.
- The Most Expensive House of Boston costs $50,000.
- Majority of the House In Boston costs around $50,000.
- The graph is Leptokurtic in Nature
- The Graph is Right Skewed.

QUESTION 2: PLOT A HISTOGRAM OF THE AVG_PRICE VARIABLE. WHAT DO YOU INFER?

INFERENCE

Considering Graph 1,

- **A house in Boston minimum costs $ 5,000 and the highest price goes to around $53000.**

Considering the Plot of Avg. Price in Boston, it is visible that the majority of house price ranges from $13,000 to $25,000. To be precise around 135 houses are under the range $21,000-$25000; 122 houses under $17,000-$21,000 & 74 Houses having prices between $13,000-$17,000.

- With This it can be inferred that the population of Boston would prefer to live in the houses which costs **between $17K and $25K.**

Graph 1

QUESTION 3: COMPUTE THE COVARIANCE MATRIX. SHARE YOUR OBSERVATIONS.

Table 11: Covariance Matrix

	CRIME_RATE	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
CRIME_RATE	8.516147873									
AGE	0.562915215	790.7924728								
INDUS	-0.110215175	124.2678282	46.97142974							
NOX	0.000625308	2.381211931	0.605873943	0.013401099						
DISTANCE	-0.229860488	111.5499555	35.47971449	0.615710224	75.66653127					
TAX	-8.229322439	2397.941723	831.7133331	13.02050236	1333.116741	28348.6236				
PTRATIO	0.068168906	15.90542545	5.680854782	0.047303654	8.74340249	167.8208221	4.677726296			
AVG_ROOM	0.056117778	-4.74253803	-1.884225427	-0.024554826	-1.281277391	-34.51510104	-0.539694518	0.492695216		
LSTAT	-0.882680362	120.8384405	29.52181125	0.487979871	30.32539213	653.4206174	5.771300243	-3.073654967	50.89397935	
AVG_PRICE	1.16201224	-97.39615288	-30.46050499	-0.454512407	-30.50083035	-724.8204284	-10.09067561	4.484565552	-48.35179219	84.41955616

INFERENCE

- The **Diagonal Elements** (Highlighted) Indicate the **Variance** in the Data sets respectively.
  Tax Variable shows **the highest variance &** Nox shows the **lowest variance.**
- **Variable's having a positive covariance means when value of X increases (or decrease) Y also increases (or decrease).**
- **Any Variable having negative Covariance implies that when value of X Increases Y decreases and vice-versa.**
- **Referring to the above matrix:**
  Tax & Age has a very high Positive covariance, followed by Tax & distance, Tax & PT ratio.
- **It Can be inferred as:**

The Tax increases as the house age increases, similarly tax Increases when Distance & Pt ratio increases.
Similarly, Average Price of a house increase when the Avg_Room increases.
Average Price Variable is having an inverse relationship with Age, Tax and LSTAT.

- **It implies as the value in the above Variables increases the Average Price of Houses Decrease.**

QUESTION 4: 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

Table 12: Correlation Matrix

	CRIME_RATE	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
CRIME_RATE	1									
AGE	0.006859463	1								
INDUS	-0.005510651	0.644778511	1							
NOX	0.001850982	0.731470104	0.763651447	1						
DISTANCE	-0.009055049	0.456022452	0.595129275	0.611440563	1					
TAX	-0.016748522	0.506455594	0.720760183	0.668023201	0.910228189	1				
PTRATIO	0.010800586	0.261515012	0.383247556	0.188932677	0.464741179	0.460853035	1			
AVG_ROOM	-0.027396162	0.240264931	0.391675853	0.302188188	0.209846668	0.292047833	0.355501495	1		
LSTAT	0.042398321	0.602338529	0.603799716	0.590878921	0.488676335	0.543993412	0.374044317	-0.613808272	1	
AVG_PRICE	-0.043337871	0.376954565	-0.483725162	0.427320772	0.381626231	0.468535934	0.507786686	0.695359947	-0.737662726	1

SOLUTION

The Top Three Positively Correlated Pairs are:	The Top Three Negatively Correlated Pairs are:
• **TAX** And **DISTANCE** • **NOX** And **INDUS** • **NOX** And **AGE**	• **LSTAT** And **AVG_PRICE** • **AVG_ROOM** And **LSTAT** • **AVG_PRICE** And **PTRATIO**

QUESTION 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 THE RESIDUAL PLOT?

B) IS LSTAT VARIABLE SIGNIFICANT FOR THE ANALYSIS BASED ON YOUR MODEL?

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7376627
R Square	0.5441463
Adjusted R Square	0.5432418
Standard Error	6.2157604
Observations	506

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	23243.914	23243.914	601.61787	5.081E-88
Residual	504	19472.381	38.635677		
Total	505	42716.295			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	34.553841	0.5626274	61.415146	3.74E-236	33.448457	35.659225	33.448457	35.659225
LSTAT	-0.9500494	0.0387334	-24.5279	5.081E-88	-1.0261482	-0.8739505	-1.0261482	-0.8739505

SOLUTION 5A.

R-squared is a statistical measure that represents the goodness of fit of a regression model. Ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted.

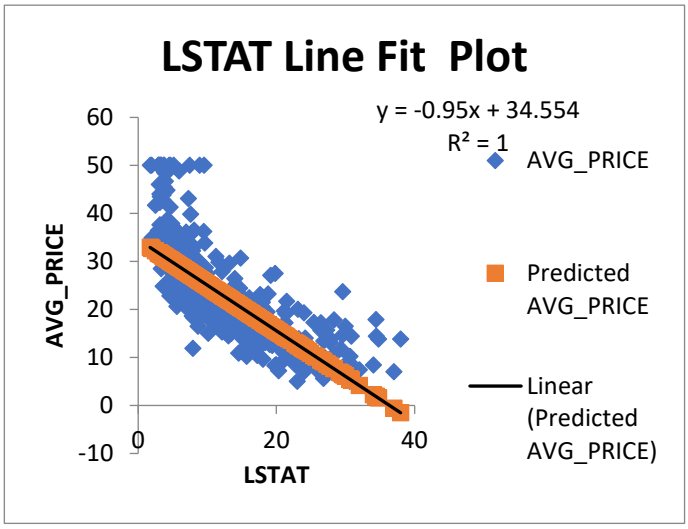
- As taken Avg Price as Dependent Variable and LSTAT as Independent Variable, the Following Regression Equation is formed:

Avg_Price =34.553841+(0.95*LSTAT)

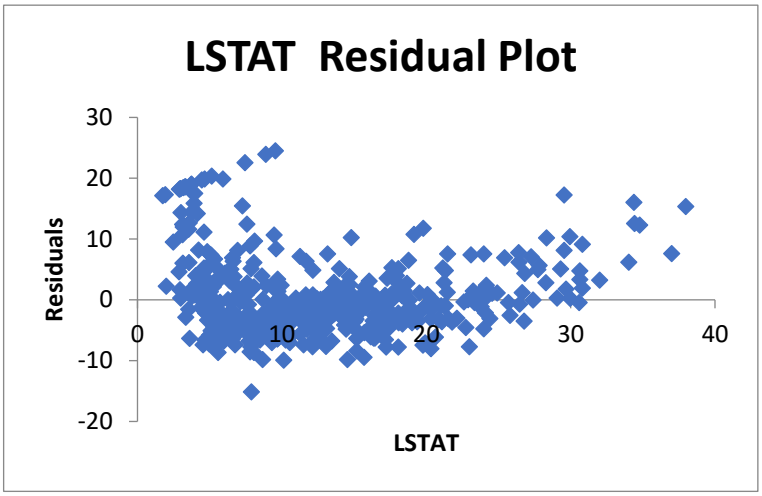
- Considering R^2, it is visible that the model predicts the Variability by **54.32% (Approx**). The Coefficient value of LSTAT is a negative value. This Proves that there's **an Inverse Relation between Avg_Price & LSTAT**. For every unit value increase in LSTAT variable, there is a '-0.9500493' decrease in Avg_Price value.

- The residual plot shows no signs of uniformity and is completely random, hence, we can infer that it has **no correlation with any of the variables taken above**. Considering the rules of writing a regression equation, we can conclude that the equation can be written as the residuals are random and the variables have a linear relation.

Graph 2



Graph 3



SOLUTION 5B.

- The Significance of a variable depends on it's P-Value. If the P-Value is less than 0.05 then the Variable is significant for our analysis, else insignificant.
- Considering the above Model, the P-value is 5.088E-88 ($5.0811*10^{-88}$), which is clearly way less than 0.05; **resulting making the variable significant**.
- Among the significant values, the variables **Age** and **Industry** have **less significance** when compared to the other variables.

QUESTION 6: 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.

SUMMARY OUTPUT

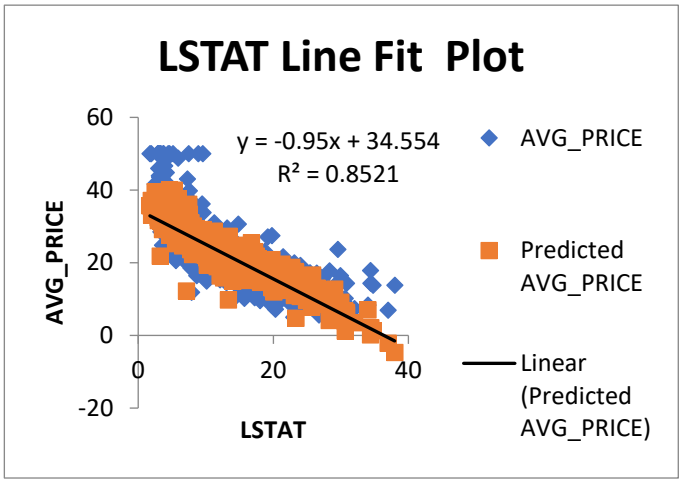
Regression Statistics	
Multiple R	0.799100498
R Square	0.638561606
Adjusted R Square	0.637124475
Standard Error	5.540257367
Observations	506

ANOVA

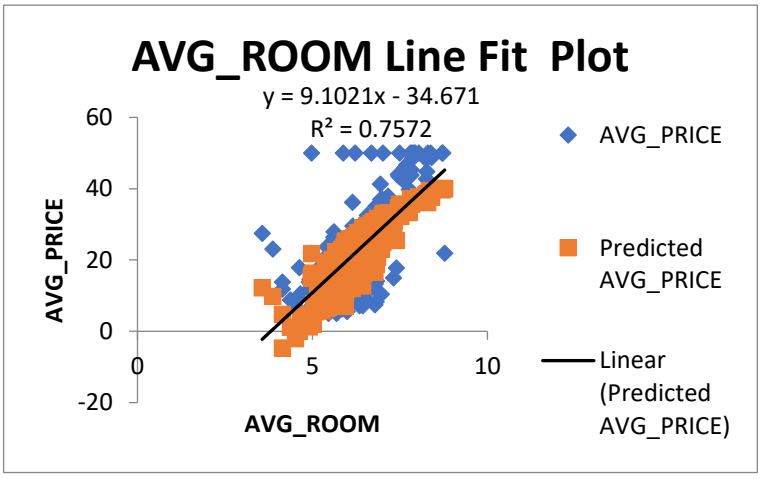
	Df	*SS*	*MS*	*F*	*Significance F*
Regression	2	27276.9862	13638.4931	444.330892	7.0085E-112
Residual	503	15439.3092	30.6944517		
Total	505	42716.2954			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.358272812	3.172828	-0.4281	0.668765	-7.5919	4.875355	-7.5919	4.875355
AVG_ROOM	5.094787984	0.444466	11.46273	3.47E-27	4.2215504	5.968026	4.22155	5.968026
LSTAT	-0.642358334	0.043731	-14.6887	6.67E-41	-0.728277	-0.55644	-0.72828	-0.55644

Graph 4



Graph 5



SOLUTION 6A.

Regression Equation/ Best Fit Line Equation: $\mathbf{Y=M_1X_1+M_2X_2+C}$

The Values to find the equation are:

$\mathbf{M_{1=}}$ **5.094788** $\mathbf{X_{1=7}}$

$\mathbf{M_{2=}}$ **-0.64236** $\mathbf{X_{2=}}$ **20** **C=-1.35827**

Substituting the Values in the provided Equation:

Best Fit Line= (5.094788*7) +(-0.64236*20) + (-1.35827)

= 21.48505

=**$21,485.05**

The Average price for the house from the provided variables data results $21,458(Approx). If a company is charging $30,000 for the house with the same features; then it is a case of **Overcharging**.

SOLUTION 6B.

- The **adjusted R-squared** is a modified version of R-squared that adjusts for the number of predictors in a regression model.
- **The Adjusted R square value should lie between 0 to 1.**

The adjusted R square for the previous Question is **0.543241826** & for this Question is **0.637124475**. The More the value closer to 1, the more precise our model will be. On comparing the values, I can conclude that my **model is more accurate in this Question with the available Independent Variable's rather than the previous Question**.

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 WITH RESPECT TO AVG_PRICE.

SOLUTION 7. SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832978824
R Square	0.69385372
Adjusted R Square	0.6882986469
Standard Error	5.1347635
Observations	506

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	29638.8667	3293.2074	124.9045049	1.93E-121
Residual	496	13077.435	26.365796		
Total	505	42716.295			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29.24131526	4.8171256	6.07028	0.00000000253977646360	19.776828	38.705803	19.776828	38.705803
CRIME_RATE	0.048725141	0.0784186	0.62134	0.534657201	-0.105348	0.202798	-0.105348	0.202798
AGE	0.032770689	0.0130978	2.50199	0.012670437	0.007036	0.058504	0.007036	0.058504
INDUS	0.130551399	0.0631173	2.06839	0.03912086	0.006541	0.254561	0.006541	0.254561
NOX	-10.3211828	3.8940363	-2.65051	0.008293859	-17.97202	-2.670328	-17.97202	-2.670328
DISTANCE	0.261093575	0.0679471	3.84260	0.000137546	0.127594	0.394593	0.127594	0.394593
TAX	-0.01440119	0.0039052	-3.68773	0.000251247	-0.022073	-0.006729	-0.022073	-0.006729
PTRATIO	-1.074305348	0.1336017	-8.04110	0.00000000000000658641598235524	-1.336800	-0.811810	-1.336800	-0.811810
AVG_ROOM	4.125409152	0.4427599	9.31750	0.00000000000000000389286981580	3.255494	4.995323	3.255494	4.995323
LSTAT	-0.603486589	0.0530812	-11.3691	0.00000000000000000000000008911	-0.707778	-0.499194	-0.707778	-0.499194

INTERPRETATION

- The **adjusted R-squared** is a modified version of R-squared that adjusts for the number of predictors in a regression model.

- **The Adjusted R square value should lie between 0 to 1.**

  The Adjusted R Square of the model is 0.68829. The More the value closer to 1, the more precise our model will be.

- The coefficients of the independent variables define their relation with the dependent variable.
- A negative coefficient means that the independent variable value decreases with an increase in the dependent variable's value and a positive coefficient means that there will be an increase.
- From the Summary Output above, we can see that Crime Rate, Age, Industry, Distance, Avg_Room have a positive relation whereas the other variables have a negative relation with Avg_Price.
- The intercept value here, is 29.24131415, which is closer to the average of the Avg_Price variable (22.53) compared to the previous model.
- **As Mentioned above, Significance of a Variable is measured om the Basis of it's P-Value. If the Variable has a P-Value less than 0.05, they are known to have significance in Our Linear Regression Model.**
- Referring to the P-Values of variables in the Summary Output of the current model it can be seen that all variables **Except Crime Rate have a P-Value less than 0.05,** resulting making them significant for our Analysis

QUESTION 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.

SOLUTION 8A. SUMMARY OUTPUT

In Linear Regression any variable whose p-value is less than 0.05, is considered to be a significant variable for our model. While Performing Linear Regression in the previous question, it was noticed that the variable "Crime Rate" is having a P-value greater than 0.05. So, while creating this model the mentioned variable is not used for regression. **Below is the Output of our new Linear Regression model without "Crime Rate" included as a variable:**

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	140.643041	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%*
Intercept	29.42847349	4.804728624	6.124898157	1.84597E-09	19.98838959	38.8685574	19.98838959	38.8685574
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
NOX	-10.27270508	3.890849222	-2.640221837	0.008545718	-17.9172457	-2.628164466	-17.9172457	-2.628164466
DISTANCE	0.261506423	0.067901841	3.851242024	0.000132887	0.128096375	0.394916471	0.128096375	0.394916471
TAX	-0.014452345	0.003901877	-3.703946406	0.000236072	-0.022118553	-0.006786137	-0.022118553	-0.006786137
PTRATIO	-1.071702473	0.133453529	-8.030529271	7.08251E-15	-1.333905109	-0.809499836	-1.333905109	-0.809499836
AVG_ROOM	4.125468959	0.44248544	9.323400461	3.68969E-19	3.256096304	4.994841615	3.256096304	4.994841615
LSTAT	-0.605159282	0.0529801	-11.42238841	5.41844E-27	-0.70925186	-0.501066704	-0.70925186	-0.501066704

SOLUTION 8B.

- The **adjusted R-squared** is a modified version of R-squared that adjusts for the number of predictors in a regression model.
- **The Adjusted R square value should lie between 0 to 1.**

The adjusted R square for the previous Question is **0.688299** & for this Question is **0.688683682**. The More the value closer to 1, the more precise our model will be. On observing the values very carefully, it is visible that there is a very minor difference in the Adjusted R square value. There's a very minute difference of 0.0003850351(Approx). With Keeping this difference in mind, the Adjusted R square value of this Question exceeds the previous one.

It can be concluded that my **model is more accurate in this Question with the available Independent Variable's rather than the previous Question**.

SOLUTION 8C. COEFFICIENTS IN ASCENDING ORDER

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
	-10.27270508	3.890849222	-2.64022183	0.008545718	-17.9172457	-2.62816446	-17.9172457	-2.62816446
NOX								
	1.07170247	0.133453529	8.03052927	0.00000000000000708250990647	1.33390510	0.80949983	1.33390510	0.80949983
PTRATIO				932	9	6	9	6
	-0.60515928	0.0529801	-11.4223884	0.00000000000000000000000000	-0.70925186	-0.50106670	-0.70925186	-0.50106670
LSTAT				542		4		4
	-0.01445234	0.003901877	-3.70394640	0.000236072	-0.02211855	-0.00678613	-0.02211855	-0.00678613
TAX					3	7	3	7
	0.03293496	0.013087055	2.51660595	0.012162875	0.00722218	0.05864773	0.00722218	0.05864773
AGE			2		7	4	7	4
	0.13071000	0.063077823	2.07220226	0.038761669	0.00677794	0.25464207	0.00677794	0.25464207
INDUS	7		4		2	1	2	1
	0.26150642	0.067901841	3.85124202	0.000132887	0.12809637	0.39491647	0.12809637	0.39491647
DISTANCE	3		4		5	1	5	1
AVG_ROO	4.12546895	0.44248544	9.32340046	0.00000000000000000036896907	3.25609630	4.99484161	3.25609630	4.99484161
M	9		1	851	4	5	4	5
	29.4284734	4.804728624	6.12489815	0.0000000018459738	19.9883895	38.8685574	19.9883895	38.8685574
Intercept	9		7		9		9	

- As Per the Data coefficient of NOX is '-10.27270508'.
- Since the coefficient of NOX is negative, It can be inferred as the value of **NOX increases**, the **Avg_Price decreases**.

SOLUTION 8D.

Regression Equation/ Best Fit Line Equation: $\mathbf{Y=M_1X_1+M_2X_2+C.}$

From the above Model, the Regression Equation can be created. The Equation is:

Y=29.42847349+4.125468959(AVG_ROOM)+0.261506423(DISTANCE)+0.130710007(INDUS)+0.3293496(AGE)+0.014452345(TAX)+0.605159282(LSTAT)+1.07170247(PTRATIO)+10.27270508(NOX).

INFERENCES

- The **price of house** is least dependent on **Crime Rate** of the Locality.
- The **NOX** variable is having the most inverse relation with price of the house. Any increase in value of NOX creates a significant drop in the price of the house.
- Houses having **high number of rooms** are sold at a higher price.
- The **PTRatio** is Having an inverse relation with the price of house. An increase in PTRatio will lead to a drop in price of house.