

	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

Solution 1:

1. The mean crime rate is 4.87, indicating that on average, there are about 5 crimes per capita in the given population.
2. The standard deviation is 2.92, indicating that there is a significant amount of variability in the crime rate data.
3. The kurtosis value is negative (-1.19), indicating that the distribution of crime rate data is flatter than a normal distribution.

	AGE
Mean	68.5749012
Standard Error	1.25136953
Median	77.5
Mode	100
Standard Deviation	28.1488614
Sample Variance	792.358399
Kurtosis	-0.96771559
Skewness	-0.59896264
Range	97.1
Minimum	2.9
Maximum	100
Sum	34698.9
Count	506

- The mean age is approximately 68.57, and the median age is 77.5. This suggests that the distribution of ages is skewed to the left, since the median is higher than the mean. The mode age is 100, indicating that there are a relatively large number of people with this age in the sample.
- The standard deviation of age is high at 28.15, which indicates that the ages in the sample are widely dispersed around the mean.
- The skewness of age is negative (-0.60), indicating that the distribution is skewed to the left.
- The kurtosis of age is negative (-0.97), suggesting that the distribution of age is flatter than a normal distribution. This means that there are fewer outliers or extreme values.

	INDUS
Mean	11.1367787
Standard Error	0.30497989
Median	9.69
Mode	18.1
Standard Deviation	6.86035294
Sample Variance	47.0644425
Kurtosis	-1.2335396
Skewness	0.29502157
Range	27.28
Minimum	0.46
Maximum	27.74
Sum	5635.21
Count	506

- The mean value of INDUS is 11.14, and the median value is 9.69. This suggests that the distribution of INDUS is slightly skewed to the right, since the mean is higher than the median. The mode value of INDUS is 18.1, indicating that this value appears most frequently in the sample.
- The standard deviation of INDUS is 6.86, indicating that the values of INDUS in the sample are spread out from the mean.
- The skewness of INDUS is positive (0.30), indicating that the distribution is slightly skewed to the right.
- The kurtosis of INDUS is negative (-1.23), suggesting that the distribution of INDUS flatter than a normal distribution. This means that there are fewer outliers or extreme values.

	NOX
Mean	0.55469506
Standard Error	0.00515139
Median	0.538
Mode	0.538
Standard Deviation	0.11587768
Sample Variance	0.01342764
Kurtosis	-0.06466713
Skewness	0.72930792
Range	0.486
Minimum	0.385
Maximum	0.871
Sum	280.6757
Count	506

- The mean value of NOX is 0.5547, and the median value is 0.538. This suggests that the distribution of NOX is slightly skewed to the right, since the mean is higher than the median. The mode value of NOX is also 0.538, indicating that this value appears most frequently in the sample.
- The standard deviation of NOX is 0.1159, indicating that the values of NOX in the sample are somewhat spread out from the mean
- The skewness of NOX is positive (0.729), indicating that the distribution is moderately skewed to the right.
- The kurtosis of NOX is negative (-0.064), suggesting that the distribution of NOX is slightly flatter than a normal distribution. This means that there are fewer outliers or extreme values.

	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

- The mean value of DISTANCE is 9.5494, and the median value is 5. This suggests that the distribution of DISTANCE is heavily skewed to the right, since the mean is much higher than the median. The mode value of DISTANCE is 24, indicating that this value appears most frequently in the sample.
- The standard deviation of DISTANCE is 8.7073, indicating that the values of DISTANCE in the sample are widely spread out from the mean. Additionally, the range of DISTANCE is 23, which further highlights the large variability in the values.
- The skewness of DISTANCE is positive (1.0048), indicating that the distribution is heavily skewed to the right.
 - The kurtosis of DISTANCE is negative (-0.8672), suggesting that the distribution of DISTANCE is slightly flatter than a normal distribution.

	TAX
Mean	408.237154
Standard Error	7.49238869
Median	330
Mode	666
Standard Deviation	168.537116
Sample Variance	28404.7595
Kurtosis	-1.14240799
Skewness	0.66995594
Range	524
Minimum	187
Maximum	711
Sum	206568
Count	506

1. The mean of the data is \$408.24, indicating that the average tax amount paid by the sample of 506 individuals is around \$408.
2. The median value of the data is \$330, which is less than the mean, indicating that the data is skewed to the right.
3. The mode of the data is \$666, indicating that this value occurs most frequently in the data set.
4. The standard deviation of the data is \$168.54, indicating that the data has a relatively large spread around the mean.
5. The sample variance of the data is 28404.76, indicating that the data has a wide spread and is highly variable

	PTRATIO
Mean	18.4555336
Standard Error	0.09624357
Median	19.05
Mode	20.2
Standard Deviation	2.16494552
Sample Variance	4.68698912
Kurtosis	-0.28509138
Skewness	-0.80232493
Range	9.4
Minimum	12.6
Maximum	22
Sum	9338.5
Count	506

1. The mean of the data is 18.46, indicating that the average pupil-teacher ratio in the sample of 506 schools is around 18.5 students per teacher.
2. The median value of the data is 19.05, which is slightly higher than the mean, indicating that the data is skewed to the left.
3. The mode of the data is 20.2, indicating that this value occurs most frequently in the data set.
4. The standard deviation of the data is 2.16, indicating that the data has a relatively low spread around the mean.

	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

1. The mean of the data is 6.28, indicating that the average number of rooms per dwelling in the sample of 506 houses is around 6.3.
2. The median value of the data is 6.21, which is slightly lower than the mean, indicating that the data is slightly skewed to the right.
3. The mode of the data is 5.713, indicating that this value occurs most frequently in the data set.
4. The standard deviation of the data is 0.70, indicating that the data has a relatively low spread around the mean.

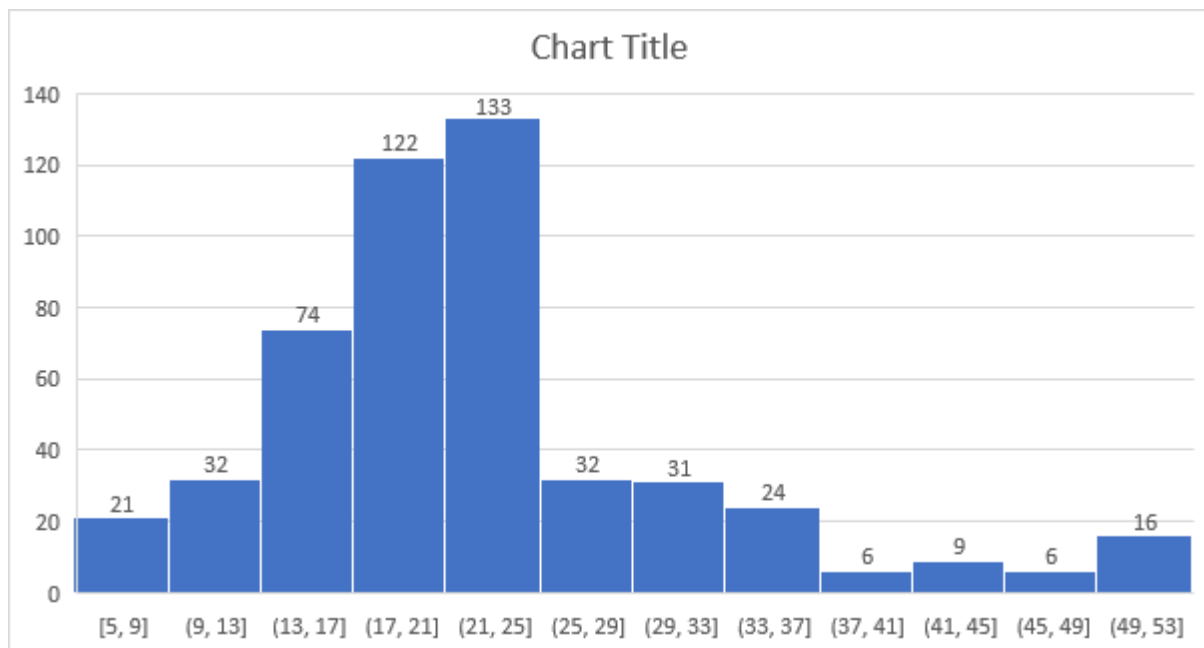
	LSTAT
Mean	12.6530632
Standard Error	0.31745891
Median	11.36
Mode	8.05
Standard Deviation	7.14106151
Sample Variance	50.9947595
Kurtosis	0.49323952
Skewness	0.90646009
Range	36.24
Minimum	1.73
Maximum	37.97
Sum	6402.45
Count	506

1. The mean of the data is 12.65, indicating that the average percentage of lower status of the population in the sample of 506 areas is around 12.7%.
2. The median value of the data is 11.36, which is lower than the mean, indicating that the data is slightly skewed to the right.
3. The mode of the data is 8.05, indicating that this value occurs most frequently in the data set.
4. The standard deviation of the data is 7.14, indicating that the data has a relatively large spread around the mean.

	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

1. The mean of the data is 22.53, indicating that the average value of owner-occupied homes in the sample of 506 areas is around \$22,530.
2. The median value of the data is 21.2, which is lower than the mean, indicating that the data is slightly skewed to the right.
3. The mode of the data is 50, indicating that this value occurs most frequently in the data set.
4. The standard deviation of the data is 9.20, indicating that the data has a relatively large spread around the mean.

Solution 2:



Inference: From the data we can infer that the price range from min of 5 to max of 53 in thousands with most of the prices around 21 to 25

Solution 3:

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	8.5									
AGE	0.6	790.8								
INDUS	-0.1	124.3	47.0							
NOX	0.0	2.4	0.6	0.0						
DISTANCE	-0.2	111.5	35.5	0.6	75.7					
TAX	-8.2	2397.9	831.7	13.0	1333.1	28348.6				
PTRATIO	0.1	15.9	5.7	0.0	8.7	167.8	4.7			
AVG_ROOM	0.1	-4.7	-1.9	0.0	-1.3	-34.5	-0.5	0.5		
LSTAT	-0.9	120.8	29.5	0.5	30.3	653.4	5.8	-3.1	50.9	
AVG_PRICE	1.2	-97.4	-30.5	-0.5	-30.5	-724.8	-10.1	4.5	-48.4	84.4

- Positive values in the off-diagonal elements indicate a positive relationship between the corresponding variables. Negative values indicate a negative relationship.
- TAX and AVG_PRICE have large values of negative covariance.

Solution 4:

	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

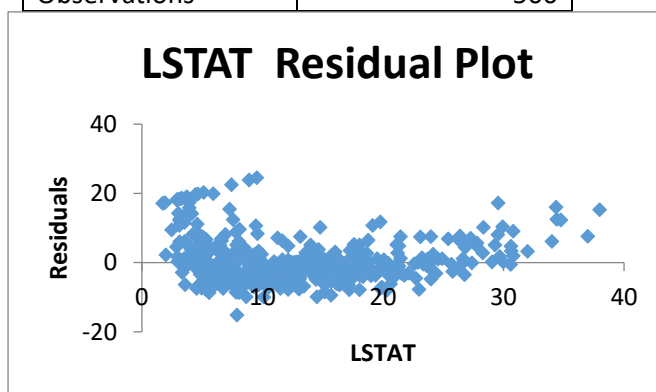
a) Age & Nox, Nox & Indus, and Tax & Distance are the Top 3 Positively correlated pairs represented in yellow.

b) Lstat & Avg_room, Avg_price & Ptratio, and Avg_price & Lstat are the Top 3 Negatively correlated pairs represented in green

Solution 5:

SUMMARY OUTPUT

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



a)

1. The multiple R-value is 0.7377, which indicates a strong positive correlation between the dependent variable and the independent variables.
2. The R-squared value is 0.5441. This indicates that the model provides a moderate fit to the data by 54.41%
3. The adjusted R-squared value is slightly lower than the R-squared value, which indicates that adding more independent variables to the model may not improve its fit significantly.
4. The coefficient values represent the change in the dependent variable for a one-unit change in the corresponding independent variable, holding all other variables constant. The coefficients can be used to determine the strength and direction of the relationship between the dependent variable and each independent variable.
5. The intercept value represents the estimated value of the dependent variable when all independent variables are set to zero. In this case, it may or may not be a meaningful value depending on the context of the data.

b) As the P-value is less than 0.05 so it is significant for the analysis based on the model.

Solution 6:

SUMMARY
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.799100498
R Square	0.638561606
Adjusted R Square	0.637124475
Standard Error	5.540257367
Observations	506

a)

AVG_ROOM	LSTAT	Predicted Avg price
7	20	21.45807639

The company is currently undercharging as compared to the company quoting 30000 USD. The company is currently undercharging as compared to the company quoting 30000 USD.

b) Adjusted r square is more in this model than the previous one the regression line fits well as compared to observed data in this model.

Solution 7:

<i>Regression Statistics</i>	
Multiple R	0.832978824
R Square	0.69385372
Adjusted R Square	0.688298647
Standard Error	5.1347635
Observations	506

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	29.24131526	4.817125596	6.070283	0.00	19.77683	38.7058	19.77683	38.7058
CRIME_RATE	0.048725141	0.078418647	0.621346	0.53	-0.10535	0.202799	-0.10535	0.202799
AGE	0.032770689	0.013097814	2.501997	0.01	0.007037	0.058505	0.007037	0.058505
INDUS	0.130551399	0.063117334	2.068392	0.04	0.006541	0.254562	0.006541	0.254562
NOX	-10.3211828	3.894036256	-2.65051	0.01	-17.972	-2.67034	-17.972	-2.67034
DISTANCE	0.261093575	0.067947067	3.842603	0.00	0.127594	0.394593	0.127594	0.394593
TAX	-0.01440119	0.003905158	-3.68774	0.00	-0.02207	-0.00673	-0.02207	-0.00673
	-							
PTRATIO	1.074305348	0.133601722	-8.0411	0.00	-1.3368	-0.81181	-1.3368	-0.81181
AVG_ROOM	4.125409152	0.442758999	9.317505	0.00	3.255495	4.995324	3.255495	4.995324
	-							
LSTAT	0.603486589	0.053081161	-11.3691	0.00	-0.70778	-0.49919	-0.70778	-0.49919

1. As the adjusted r square is close to 1 so the regression line is a good fit for the model. The coefficient of CRIME_RATE, AGE, INDUS, DISTANCE, AVG_ROOM is Positive which means they move in the same direction. Negative coefficient means they move inversely. Magnitude of coefficient determines relationship more value means stronger relationship.
2. The intercept is simply the value of the dependent variable when all the predictor variables are equal to zero.
3. All of them are significant as the P-value is less than 0.05 except CRIME_RATE where P-value is more than 0.05 which is insignificant.

Solution 8:

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.832835773
R Square	0.693615426
Adjusted R Square	0.688683682
Standard Error	5.131591113
Observations	506

a) R square value with 0.69 signifies 69% of the dependent variable is explained by the independent variable, Adjusted R square value is close to 1 so it is a good fit for the model.

b) Adjusted R square is almost the same, so both are good fit for the model.

c)

<i>Coefficients</i>	
Intercept	29.43
NOX	-10.27
PTRATIO	-1.07
LSTAT	-0.61
TAX	-0.01
AGE	0.03
INDUS	0.13
DISTANCE	0.26
AVG_ROOM	4.13

➤ The values of the coefficient have been sorted in ascending order.

d) $\text{AVG_PRICE} = \text{Coefficient of AGE} \times \text{AGE} + \text{Coefficient of INDUS} \times \text{INDUS} + \text{Coefficient of NOX} \times \text{NOX} + \text{Coefficient of DISTANCE} \times \text{DISTANCE} + \text{Coefficient of TAX} \times \text{TAX} + \text{Coefficient of PTRATIO} \times \text{PTRATIO} + \text{Coefficient of AVG_ROOM} \times \text{AVG_ROOM} + \text{Coefficient of LSTAT} \times \text{LSTAT} + \text{Intercept}$