Note- Detailed Tables are Created in attached Excel File.

Q1-Ans

RIME_RATE	
Mean	4.871976
Standard E	0.12986
Median	4.82
Mode	3.43
Standard D	2.921132
Sample Va	8.533012
Kurtosis	-1.18912
Skewness	0.021728
Range	9.95
Minimum	0.04
Maximum	9.99
Sum	2465.22
Count	506

AGE	
Mean	68.5749
Standard E	1.25137
Median	77.5
Mode	100
Standard D	28.14886
Sample Va	792.3584
Kurtosis	-0.96772
Skewness	-0.59896
Range	97.1
Minimum	2.9
Maximum	100
Sum	34698.9
Count	506

INDUS	
Mean	11.13678
Standard E	0.30498
Median	9.69
Mode	18.1
Standard D	6.860353
Sample Va	47.06444
Kurtosis	-1.23354
Skewness	0.295022
Range	27.28
Minimum	0.46
Maximum	27.74
Sum	5635.21
Count	506

NOX	
Mean	0.554695
Standard E	0.005151
Median	0.538
Mode	0.538
Standard D	0.115878
Sample Va	0.013428
Kurtosis	-0.06467
Skewness	0.729308
Range	0.486
Minimum	0.385
Maximum	0.871
Sum	280.6757
Count	506

DISTANCE	
Mean	9.549407
Standard E	0.387085
Median	5
Mode	24
Standard D	8.707259
Sample Va	75.81637
Kurtosis	-0.86723
Skewness	1.004815
Range	23
Minimum	1
Maximum	24
Sum	4832
Count	506

TAX	
Mean	408.2372
Standard E	7.492389
Median	330
Mode	666
Standard D	168.5371
Sample Va	28404.76
Kurtosis	-1.14241
Skewness	0.669956
Range	524
Minimum	187
Maximum	711
Sum	206568
Count	506

PTRATIO	
Mean	18.45553
Standard E	0.096244
Median	19.05
Mode	20.2
Standard D	2.164946
Sample Va	4.686989
Kurtosis	-0.28509
Skewness	-0.80232
Range	9.4
Minimum	12.6
Maximum	22
Sum	9338.5
Count	506

AVG_ROOM	
Mean	6.284634
Standard E	0.031235
Median	6.2085
Mode	5.713
Standard D	0.702617
Sample Va	0.493671
Kurtosis	1.8915
Skewness	0.403612
Range	5.219
Minimum	3.561
Maximum	8.78
Sum	3180.025
Count	506

LSTAT	
Mean	12.65306
Standard E	0.317459
Median	11.36
Mode	8.05
Standard D	7.141062
Sample Va	50.99476
Kurtosis	0.49324
Skewness	0.90646
Range	36.24
Minimum	1.73
Maximum	37.97
Sum	6402.45
Count	506

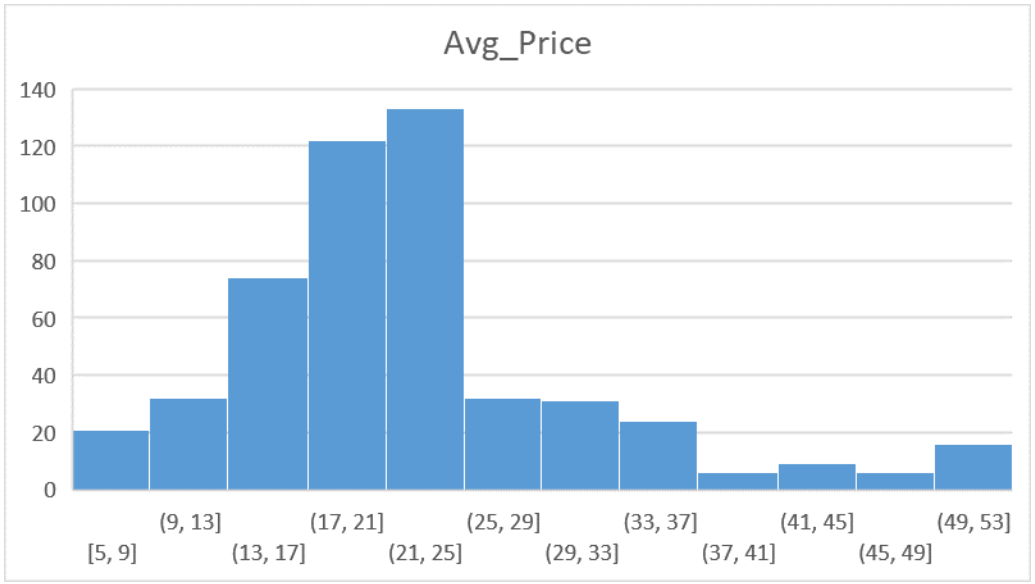
AVG_PRICE	
Mean	22.53281
Standard E	0.408861
Median	21.2
Mode	50
Standard D	9.197104
Sample Va	84.58672
Kurtosis	1.495197
Skewness	1.108098
Range	45
Minimum	5
Maximum	50
Sum	11401.6
Count	506

From the Summary Statistics it Observed That.

1. Avg_Room, LSTAT and Avg_Price having postive Kurtosis. Which means they are having Sharp peakedness.

2. Age and PTratio having negative Skewness. Which means they are Left Skew, there Peak lies on right side.

Q2-Ans



The Histogram of Avg_Price

1.The Peakedness lying to the left side.

2. Observations represents 'trailing off' to the right.

3. It Conclude that its a Positve Skew.

Q3-Ans

Covariance	*CRIME_RATE*	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
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	50.893979	
AVG_PRICE	1.16201224	-97.3962	-30.4605	-0.45451	-30.5008	-724.82	-10.0907	4.484565552	-48.351792	84.4195562

The data is not Standardized into scale of -1 to +1.

With the help of Covariance table we can Observe that some values are Positive and some are Negative.

Positive values indicates that the relatipship between them is Positive they are either increaing or decreasing mostly together.
Negative values indicates that the relationship between them is Negative they mostly moving in opposite direction.

Q4-Ans

Correlation	*CRIME_RATE*	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
CRIME_RATE	1									
AGE	0.006859463	1								
INDUS	-0.005510651	0.644779	1							
NOX	0.001850982	0.73147	0.763651	1						
DISTANCE	-0.009055049	0.456022	0.595129	0.611441	1					
TAX	-0.016748522	0.506456	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	-0.24026	-0.39168	-0.30219	-0.20985	-0.29205	-0.3555	1		
LSTAT	-0.042398321	0.602339	0.6038	0.590879	0.488676	0.543993	0.374044	-0.61380827	1	
AVG_PRICE	0.043337871	-0.37695	-0.48373	-0.42732	-0.38163	-0.46854	-0.50779	0.695359947	-0.73766	1

a.	Top 3 positively correlated pairs	b.	Top 3 negatively correlated pairs
	They are almost Moving together.		They are moving most of the time Opposite.
1	Distance and Tax	1	LSTAT and Avg_Price
2	Indus and NOX	2	Avg_Room and LSTAT
3	Age and NOX	3	PTRatio and Avg_Price

Q5-Ans: RESIDUAL PLOT are created in attached Excel (Q5-Sheet)

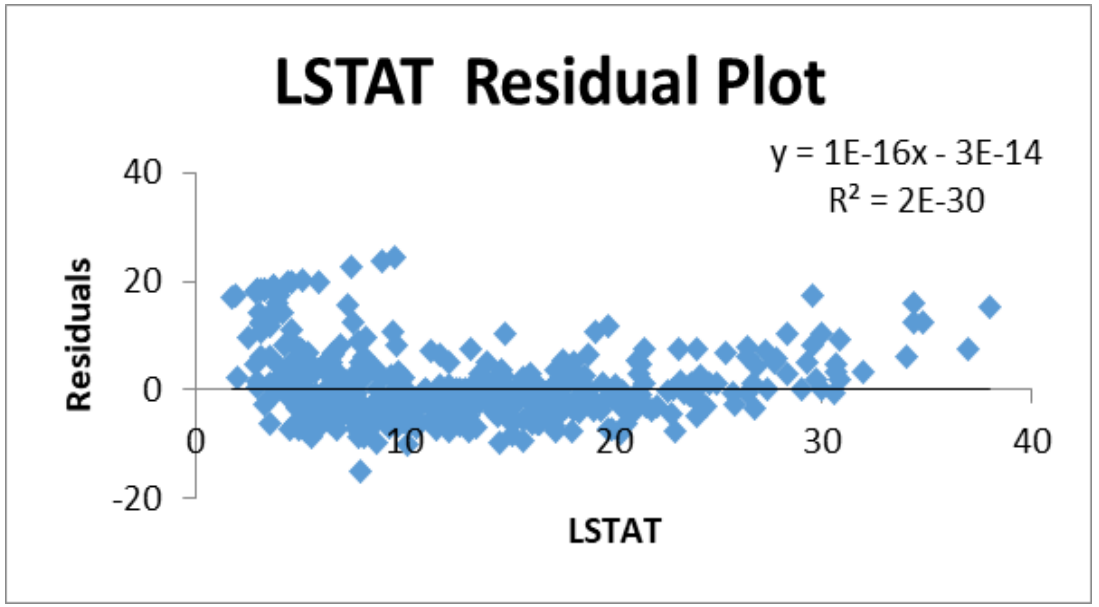
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.737663
R Square	0.544146
Adjusted R Square	0.543242
Standard Error	6.21576
Observations	506

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	23243.91	23243.91	601.617871	5.08E-88
Residual	504	19472.38	38.63568		
Total	505	42716.3			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	34.55384	0.562627	61.41515	3.743E-236	33.44846	35.65922	33.44846	35.65922
LSTAT	-0.95005	0.038733	-24.5279	5.08E-88	-1.02615	-0.87395	-1.02615	-0.87395



a. The Multile R value is almost near to +1. Its indicates that both are strong positively Correlated.

The R Square value is good enough which indicates that 54% of the dependent variables (y-value) are explained by the independent variables (x-value).

b. LSTAT variable significant for the analysis as it having P-value less than 0.05.

Q6-Ans

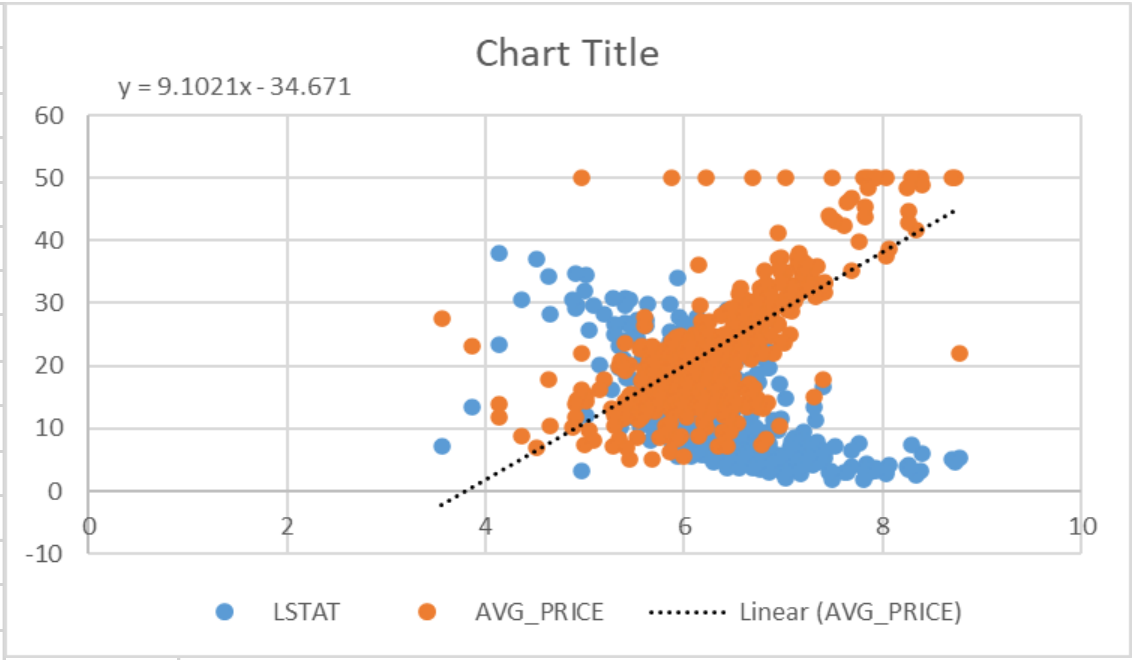
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7991
R Square	0.638562
Adjusted R	0.637124
Standard E	5.540257
Observati	506

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressio	2	27276.99	13638.49	444.3309	7E-112
Residual	503	15439.31	30.69445		
Total	505	42716.3			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-1.35827	3.172828	-0.4281	0.668765	-7.5919	4.875355	-7.5919	4.875355
AVG_ROO	5.094788	0.444466	11.46273	3.47E-27	4.22155	5.968026	4.22155	5.968026
LSTAT	-0.64236	0.043731	-14.6887	6.67E-41	-0.72828	-0.55644	-0.72828	-0.55644



a. Regression equation

Y = 5.0948 X_1(Average room)-0.642 X_2(LSAT) -1.358

AVG_ROO	7
L-STAT	20
Avg_Price	21.45808

b. As compare to previous observation the R Square value slightly increased which good that the 63.8% of the dependent variables (y-value) are explained by the independent variables (x-value).

Q7-Ans

SUMMARY OUTPUT								
Regression Statistics								
Multiple F	0.832979							
R Square	0.693854							
Adjusted I	0.688299							
Standard E	5.134764							
Observati	506							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regressio	9	29638.86	3293.207	124.9045	1.9E-121			
Residual	496	13077.43	26.3658					
Total	505	42716.3						
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	29.24132	4.817126	6.070283	2.54E-09	19.77683	38.7058	19.77683	38.7058
CRIME_RA	0.048725	0.078419	0.621346	0.534657	-0.10535	0.202799	-0.10535	0.202799
AGE	0.032771	0.013098	2.501997	0.01267	0.007037	0.058505	0.007037	0.058505
INDUS	0.130551	0.063117	2.068392	0.039121	0.006541	0.254562	0.006541	0.254562
NOX	-10.3212	3.894036	-2.65051	0.008294	-17.972	-2.67034	-17.972	-2.67034
DISTANCE	0.261094	0.067947	3.842603	0.000138	0.127594	0.394593	0.127594	0.394593
TAX	-0.0144	0.003905	-3.68774	0.000251	-0.02207	-0.00673	-0.02207	-0.00673
PTRATIO	-1.07431	0.133602	-8.0411	6.59E-15	-1.3368	-0.81181	-1.3368	-0.81181
AVG_ROO	4.125409	0.442759	9.317505	3.89E-19	3.255495	4.995324	3.255495	4.995324
LSTAT	-0.60349	0.053081	-11.3691	8.91E-27	-0.70778	-0.49919	-0.70778	-0.49919

The Adj R Sqaure value is approximate 0.7 which indicates that the 70% of the y-values are explained by x-values. Which is best.

As only CRIME_RATE is having P-value greater than 0.05 which indicates that it is Insignificant, all other are significant as their P-value is less than 0.05.

Q8-Ans: The detailed Summary Statistics present in attached Excel (Q8-Sheet)

a. While observing the accuracy (81.52%) of this model it was a good than other models.

b.

Q7-R Sqr val	Q8-R Sqr val
0.69385372	0.693615

The R Sqaure value in this(Q8) regression slightly decreased.

Q7-Adj R Sqr	Q8-Adj R Sqr val
0.68829865	0.688684

This Regression Model (Q8) performs better according to the value of adjusted R-square.

c. The coefficient(-10.272) and correlation value(-0.427) of NOX is negative so we can summarise that as the value of NOX increases the value of AVG price decreases, thus they are inversely related.

d. $Y = 0.0329X\_1(Age) + 0.13X\_2(Indus) - 10.272X\_3(NOX) + 0.261X\_4(Distance) - 0.0144X\_5(Tax) - 1.0717X\_6(PTRatio) + 4.125\ X\_7(Avg\ Room) - 0.605\ X\_8(LSTAT) + 29.428$