Business Report for Terro's Real-estate

Q 1)

A Descriptive Statistics is Created in Excel and from the Output we Take Specific Summary Stats to Infer the Data

CRIME_RATE	
Mean	4.871976
Standard Deviation	2.921132
Kurtosis	-1.18912
Skewness	0.021728

From the Descriptive Stat's we can see that the Kurtosis is Negative which infers the Graph is Flat And Skewness is Positive and so the Graph leans towards left.

And Standard Deviation and Mean we can tell the Spread in Standard Deviation is Medium.

AGE	
Mean	68.5749
Standard Deviation	28.14886
Kurtosis	-0.96772
Skewness	-0.59896

From the Stat's The kurtosis is Negative and we can infer the Graph is Flat , and Skewness is also Negative and so the Graph lean Right Side. And Spread is Medium

INDUS	
Mean	11.13678
Standard Deviation	6.860353
Kurtosis	-1.23354
Skewness	0.295022

From Stats -> Kurtosis is Negative and Graph is Flat.
Skewness is Positive and Graph leans Left
Spread is Medium

NOX	
Mean	0.554695
Standard Deviation	0.115878
Kurtosis	-0.06467
Skewness	0.729308

From Stats -> Kurtosis is Negative and Graph is Flat
Skewness is Positive and Graph leans Left
Spread of Data is Less

DISTANCE	
Mean	9.549407
Standard Deviation	8.707259
Kurtosis	-0.86723
Skewness	1.004815

From Stats -> Kurtosis is Negative and Graph is Flat,
Skewness is Positive and so Graph leans Left,
Spread of Data is Wide

TAX	
Mean	408.2371542
Standard Deviation	168.5371161
Kurtosis	-1.142407992
Skewness	0.669955942

From Stats -> Kurtosis is Negative and Graph is Flat,
Skewness is Positive and so Graph leans Left,
Spread of Data is Medium

PTRATIO	
Mean	18.45553
Standard Deviation	2.164946
Kurtosis	-0.28509
Skewness	-0.80232

From Stats -> Kurtosis is Negative and Graph is Flat,
Skewness is Negative and so Graph leans Right,
Spread of Data is Less

AVG_ROOM	
Mean	6.284634
Standard Deviation	0.702617
Kurtosis	1.8915
Skewness	0.403612

From Stats -> Kurtosis is Positive and Graph gets a Sharp Peak
Skewness is Positive and so Graph leans Left
Spread of Data is Less

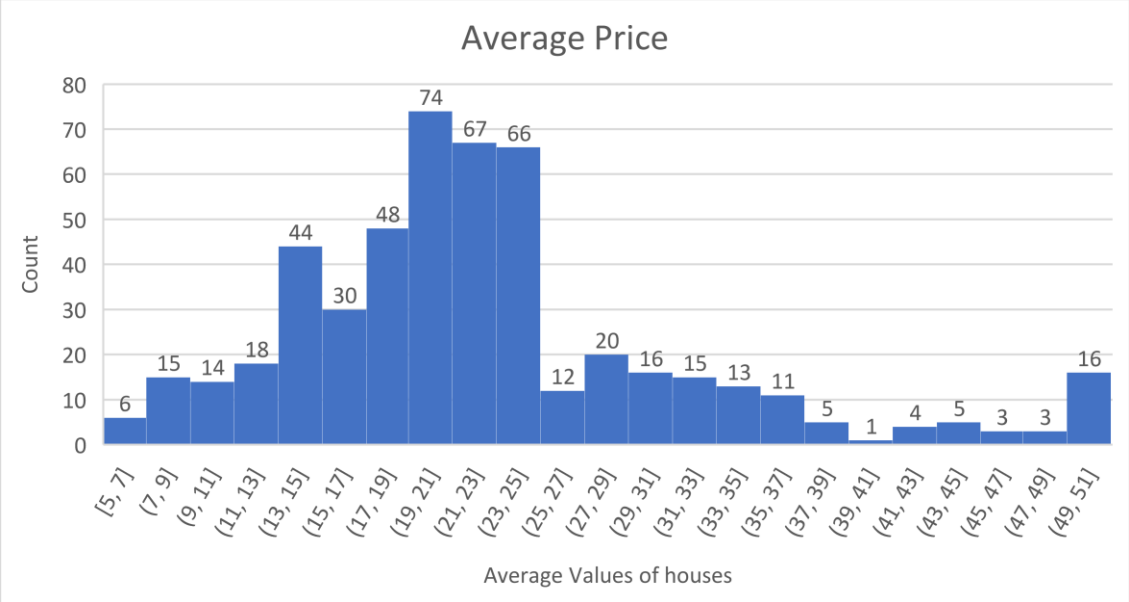
LSTAT	
Mean	12.65306324
Standard Deviation	7.141061511
Kurtosis	0.493239517
Skewness	0.906460094

From Stats -> Kurtosis is Positive and Graph is Sharp and gets a Peak
Skewness is Positive and Graph leans Left
Spread of data is Medium

AVG_PRICE	
Mean	22.53281
Standard Deviation	9.197104
Kurtosis	1.495197
Skewness	1.108098

From Stats -> Kurtosis is Positive and Graph is Sharp and gets a Peak
Skewness is Positive and Graph leans Left
Spread of data is Medium

Q 2)



From the Histogram we can see that the Average price of house are leaning left and We can say that skewness is positive

And also, we can the Graph is peaked higher and is not flat so the Kurtosis is Positive
From the Descriptive Analysis the median is 21.2, mean is 22.53, and also defines the central tendency

Q 3)

COVARIANCE	*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

From the Covariance Table we Can't Tell the Strength of a relation between two Variable
But we can say that the relation is Positive or negative
In the Above Table the Green highlighted are in Positive relation where if one grows other grows
And the Red Highlighted are Negative Relation Where if one grows other Decline

Correlation	*CRIME_RATE*	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
CRIME_RATE	1.000									
AGE	0.007	1.000								
INDUS	-0.006	0.645	1.000							
NOX	0.002	0.731	0.764	1.000						
DISTANCE	-0.009	0.456	0.595	0.611	1.000					
TAX	-0.017	0.506	0.721	0.668	0.910	1.000				
PTRATIO	0.011	0.262	0.383	0.189	0.465	0.461	1.000			
AVG_ROOM	0.027	-0.240	-0.392	-0.302	-0.210	-0.292	-0.356	1.000		
LSTAT	-0.042	0.602	0.604	0.591	0.489	0.544	0.374	-0.614	1.000	
AVG_PRICE	0.043	-0.377	-0.484	-0.427	-0.382	-0.469	-0.508	0.695	-0.738	1.000

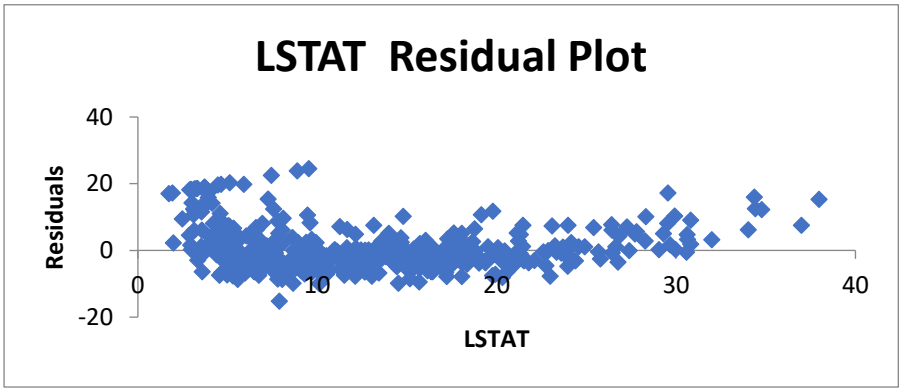
Q 4)

From Correlation Table

Top 3 Positive Relation		Top 3 Negative Relation	
1)	Distance and Tax	1)	Average Price and LSTAT
2)	Industries and NOX	2)	Average Room and LSTAT
3)	Nox and Age	3)	Average Price and PTRatio

Q 5)

From the Regression model Created with AVG_Price as Dependent Variable and LSTAT as Independent Variable



a) We can see that the Coefficient of LSTAT is negative hence the Relation is negative and P-Value is <0.05 which Make Hypothesis true and Checking out Adjusted R square Value its 0.5432 Which tells that LSTAT is reasonable for 54% and we need some more values to increase the R square Value and Seeing the Residual Plot it doesn't show any pattern

b) LSTAT Variable provides only 54% of the Prediction and also from correlation LSTAT has Higher correlation with Average Price So it's necessary for Prediction and is not enough to do a Good Prediction

Q 6)

A Regression model with Average Price as Dependent Variable and LSTAT and Average_Room as Independent Variable is Created in Excel and Adjusted R Square Value of 0.637124 is achieved.

a) Given
7 Rooms and LSTAT 20 and to find the Prediction of the Average Price.
From the Table

	Coefficients
Intercept	-1.358272812
AVG_ROOM	5.094787984
LSTAT	-0.642358334

We give the Equation for the Average Price Prediction as

$Y_{Prediction} = ß_0 + ß_1X_1 + ß_2X_2 + ß_3X_3 + \dots + ß_nX_n$

(Where n- No of Independent Variables, ß -Coefficient of corresponding Variable)

$Y_{Average\ price}$ = -1.358272 + 5.094787*(AVG_Room) + (-0.642358)*(LSTAT)
=-1.358272 + 5.094787*(7) -0.642358*(20)
$Y_{Average\ price}$ =21.458076

Average Price is given as per $1000

Now the Y predicted for 7 Average Rooms and 20 LSTAT is 21.45807639*$1000

$Y_{Predicted}$ = $ 21,458.07

The Company has Quoted a Value of $ 30,000 which by Comparison the Company has Overcharged.

b) Comparing to the Previous model this model is better than the previous one By Means of R Square Value the Previous model had only 0.543241 and in this model, it has 0.63712 Which is relatively higher than the previous performance.

Q 7)

A Regression table has been Created with Average Price as Dependent Variable and all other Variable as Independent Variable. And from the regression we get the out R^2 as 0.6882986 which is not closer to 1 and still needs to be improved. And by comparison to the previous Regressions the Adjusted R^2 Values have been improved significantly.

From the Excel we can see that P value of Crime rate is higher than 0.05 which implies that Crime Rate is not a suitable variable to predict the Average Price. And in next Regression we have to remove it to get a better accuracy.

And From the Coefficients we can the say the type of relation associated with the dependent variable

	Coefficients
Intercept	29.24131
CRIME_RATE	0.048725
AGE	0.032771
INDUS	0.130551
NOX	-10.3212
DISTANCE	0.261094
TAX	-0.0144
PTRATIO	-1.07431
AVG_ROOM	4.125409
LSTAT	-0.60349

From this table we can the see the coefficients of the Independent variable. We can see the relation type of the Independent variable.

The Cell highlighted with Red are Negative Relation where if the Dependent Variable and Independent variable are Inversely related which means if one rises and other declines and vice versa

The Cell highlighted with Green are Positive relation where if the Dependent variable and Independent variable are Directly related which means if one rises and other rises and if one falls other too falls.

From the data we can't predict the strength of the relationship and that's where correlation comes in which tells the strength of the relationship and also the type of relationship. A correlation table is created to show it clearly

Correlation	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT
AVG_PRICE	0.043	-0.377	-0.484	-0.427	-0.382	-0.469	-0.508	0.695	-0.738

From the Data we the Strength and relation is inferred and is given below in the table.

Variable	*Average Price Strength and Relation*
Crime Rate	Have Very Weak Positive Relation ~ Zero Relation
Age	Have Moderate Negative Relation
Industry	Have Moderate Negative Relation
NOX	Have Moderate Negative Relation
Distance	Have Moderate Negative Relation
Tax	Have Moderate Negative Relation
PTRatio	Have Strong Negative Relation
Average Room	Have Strong Positive Relation
LSTAT	Have Very Strong Negative Relation

Q 8)

From the Previous question and from the generated Regression we see that Crime Rate P value is greater than 0.05 and in this instance it is removed.

a) From the Model we can see that every independent variable has P-Value less than 0.05 and from the coefficients values we can see the type of relation when the Coefficients are greater than 0 the relationship is positive and when less than 0 the relationship is negative

b) Comparing the Previous Model the R^2 Value is 0.68829865 and in this model we get the R^2 value as 0.688683682 which is better than previous instance.

c) From the question we sort the coefficients in ascending order

Variable	*Coefficients*
NOX	-10.2727
PTRATIO	-1.0717
LSTAT	-0.60516
TAX	-0.01445
AGE	0.032935
INDUS	0.13071
DISTANCE	0.261506
AVG_ROOM	4.125469
Intercept	29.42847

When the NOX value increases the Average price decreases.

d) From the model we write the Regression as

$Y_{Predicted} = ß_0 + ß_1X_{1i} + ß_2X_{2i} +...+ ßpXpi$

$Y_{Predicted}$ = 29.42847+(-10.2727)*NOX + (-1.0717)*PTRatio + (-0.60516)*LSTAT + (-0.01445)*TAX +(0.032935)*AGE +(0.13071)*INDUS + (0.261506)*Distance + (4.1254)*AVG_Room

=29.42847 - 10.2727*NOX - 1.0717*PTRatio -0.60516*LSTAT -0.01445*TAX + 0.032935*AGE + 0.13071*INDUS + 0.261506*Distance + 4.1254*AVG_Room