Report on Housing Price Factors

By

Indresh R

Table of Content

Introduction

Terro's Real Estate, a prominent agency specializing in property valuation, seeks to determine the prices of houses in a specific locality. Their pricing methodology relies on various features and factors of a property, providing valuable insights into the property's worth. To achieve this, Terro's employs an "Auditor" tasked with studying geographic features such as crime rates, education facilities, pollution levels, and more. These features play a pivotal role in estimating property values. For this analysis, Terro's Real Estate has provided a dataset containing information on 506 houses in Boston, including details on crime rates, industrial proportions, nitric oxide concentrations, and other variables. The primary objective of this report is to conduct an exploratory data analysis (EDA) to understand the magnitude and influence of each variable on house prices in the Boston locality.

1) Generate the summary statistics for each variable in the table. Write down your observation.

- Select the "Data" tab.
- Click on "Data Analysis".
- Choose "Descriptive Statistics."
- In the "Input Range" field, select the data columns you want to analyse.
- Check the "Summary statistics" box.
- Click "OK" and view the summary statistics in a new sheet.

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
Mean	4.871976285	68.574901	11.136779	0.5546951	9.5494071	408.237154	18.4555336	6.284634387	12.65306324	22.53280632
Standard Error	0.129860152	1.2513695	0.3049799	0.0051514	0.3870849	7.49238869	0.096243568	0.031235142	0.317458906	0.408861147
Median	4.82	77.5	9.69	0.538	5	330	19.05	6.2085	11.36	21.2
Mode	3.43	100	18.1	0.538	24	666	20.2	5.713	8.05	50
Standard Deviati	2.921131892	28.148861	6.8603529	0.1158777	8.7072594	168.537116	2.164945524	0.702617143	7.141061511	9.197104087
Sample Variance	8.533011532	792.3584	47.064442	0.0134276	75.816366	28404.7595	4.686989121	0.49367085	50.99475951	84.58672359
Kurtosis	-1.189122464	-0.967716	-1.23354	-0.064667	-0.867232	-1.142408	-0.285091383	1.891500366	0.493239517	1.495196944
Skewness	0.021728079	-0.598963	0.2950216	0.7293079	1.0048146	0.66995594	-0.802324927	0.403612133	0.906460094	1.108098408
Range	9.95	97.1	27.28	0.486	23	524	9.4	5.219	36.24	45
Minimum	0.04	2.9	0.46	0.385	1	187	12.6	3.561	1.73	5
Maximum	9.99	100	27.74	0.871	24	711	22	8.78	37.97	50
Sum	2465.22	34698.9	5635.21	280.6757	4832	206568	9338.5	3180.025	6402.45	11401.6
Count	506	506	506	506	506	506	506	506	506	506
CV	0.59958	0.41048	0.61601	0.20890	0.91181	0.41284	0.11731	0.11180	0.56437	0.40817

OBSERVATION

i. **CRIME_RATE**: The average crime rate is approximately 4.872 per capita, with a relatively low standard deviation. The distribution appears to be right-skewed, as indicated by the positive skewness.

ii. **AGE**: The average age of houses is approximately 68.575 years, with a significant standard deviation, indicating a wide range of ages. The distribution is slightly left-skewed, and it has a negative kurtosis value.

iii. **INDUS**: The average proportion of non-retail business acres is around 11.137%, with a moderate standard deviation. The distribution is slightly right-skewed.

iv. **NOX**: The average nitric oxide concentration is 0.555 parts per 10 million, with a relatively low standard deviation. The distribution appears to be right-skewed, and it has a negative kurtosis value.

v. **DISTANCE**: The average distance from the highway is approximately 9.549 miles, with a relatively high standard deviation, indicating a wide range of distances. The distribution is right-skewed, and it has a positive skewness and a negative kurtosis value.

vi. **TAX**: The average property tax rate is 408.237 per $10,000, with a significant standard deviation. The distribution appears to be right-skewed.

vii. **PTRATIO**: The average pupil-teacher ratio is approximately 18.456, with a low standard deviation. The distribution is slightly left-skewed.

viii. **AVG_ROOM**: The average number of rooms per house is approximately 6.285, with a relatively low standard deviation. The distribution is slightly right-skewed.

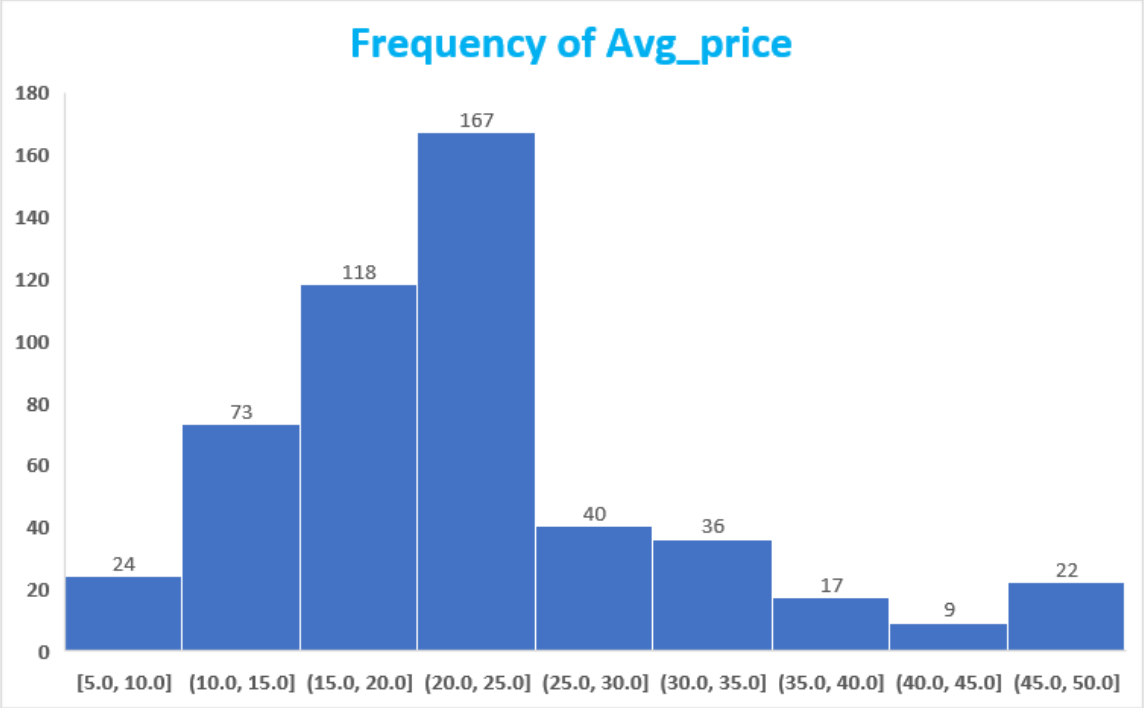
ix. **LSTAT**: The average percentage of the lower status of the population is 12.653%, with a moderate standard deviation. The distribution appears to be right-skewed, and it has a positive skewness.

x. **AVG_PRICE**: The average house price is approximately 22.533 (in $1000s), with a standard deviation of 9.197. The distribution appears to be right-skewed, and it has a positive skewness.

PTRATIO and Average room are less deviated from mean. Distance has relatively high deviation from mean.
These observations provide an initial understanding of the dataset and its variables. For further analysis, you may want to explore relationships between these variables, especially their correlations and their impact on house prices (AVG_PRICE). Additionally, you can use this information to determine which variables might be more influential in predicting house prices and which may require further investigation or preprocessing.

2) Plot a histogram of the AVG_PRICE variable. What do you infer?

- Select the column containing the "AVG_PRICE" data.
- Go to the "Insert" tab.
- Click on "Histogram."
- Customize the histogram if needed.



INFERENCE

Based on the histogram graph, it is evident that the distribution of AVG_PRICE is right-skewed on the higher end of the price range. This skewness indicates that the AVG_PRICE variable does not follow a normal distribution and majority of the house AVG_PRICE is lies below the Average. The histogram shows that there is a wide spread of house prices, with a right tail. The majority of the data points are concentrated on the left side of the histogram. a right-skewed histogram indicates that there are relatively few data points with extremely high values, pulling the mean in that direction, while the bulk of the data has lower values. This skewness suggests that there may be outliers or extreme values on the right side of the distribution that are causing the skew.

3) Compute the covariance matrix. Share your observations.

- Select the "Data" tab.
- Click on "Data Analysis".
- Choose "Covariance."
- In the "Input Range" field, select all the variable for compare.
- Click "OK" and view the covariance matrix in a new sheet.
- Use conditional formatting to differentiate positive and negative covariance.

	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

OBSERVATION

From the above analysis, it becomes evident that several positive covariance relationships exist within the dataset. These relationships imply a consistent pattern where, as one variable tends to deviate from its mean in a positive direction, the other variable also exhibits a similar tendency. Specifically, positive covariance is observed between Crime_rate and Age, Crime_rate and Nox, Crime_rate and Ptratio, Crime_rate and Avg_room, Crime_rate and Avg_price, Age and Industry, Age and Nox, Age and Distance, Age and Tax, Age and Ptratio, Age and Lstat, Industry and Nox, Industry and Distance, Industry and Tax, Industry and Ptratio, Industry and Lstat, Nox and Distance, Nox and Tax, Nox and Ptratio, Nox and Lstat, Distance and Tax, Distance and Ptratio, Distance and Lstat, Tax and Ptratio, Tax and Lstat, Ptratio vs Lstat, and Avg_room and Avg_price. These negative covariance relationships suggest a particular pattern as one variable tends to deviate from its mean in a positive direction, the other variable tends to deviate from its mean in a negative direction, and vice versa. Specifically, these negative covariance relationships are observed between Crime_rate and Distance, Crime_rate and Tax, Crime_rate and Lstat, Age and Avg_ROOM, Age and Avg_PRICE, NOX and Industry, NOX and Tax, NOX and Ptratio, NOX and Lstat, Distance and Ptratio, Distance and Lstat, and Tax and

Lstat. We can see that y variable has more negative covariance relationship with x variable. (>0 = Positive covariance, <0 = Negative covariance)

4) Create a correlation matrix of all the variables.
a) Which are the top 3 positively correlated pairs and
b) Which are the top 3 negatively correlated pairs.

- Select the "Data" tab.
- Click on "Data Analysis".
- Choose "Correlation."
- In the "Input Range" field, select all the variable for compare.
- Click "OK" and view the correlation matrix in a new sheet.
- Use conditional formatting to differentiate positive and negative covariance.

	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	0.72076018	0.6680232	0.910228189	1				
PTRATIO	0.010800586	0.261515012	0.383247556	0.188932677	0.464741179	0.460853035	1			
AVG_ROOM	0.02739616	-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	-0.48372516	-0.427320772	-0.381626231	-0.468535934	-0.507786686	0.695359947	-0.737662726	1

a) Which are the top 3 positively correlated pairs

- Distance vs Tax
- NOX vs Industry
- Age vs NOX

b) Which are the top 3 negatively correlated pairs.

- Lstat vs Avg_Price
- Avg_Room vs Lstat
- PTratio vs Avg_price

>0 = Positive correlation
<0 = Negative correlation

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

- Use Excel's Data Analysis Tool Pak to perform a Regression analysis.
- Interpret the regression summary output and the residual plot.

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

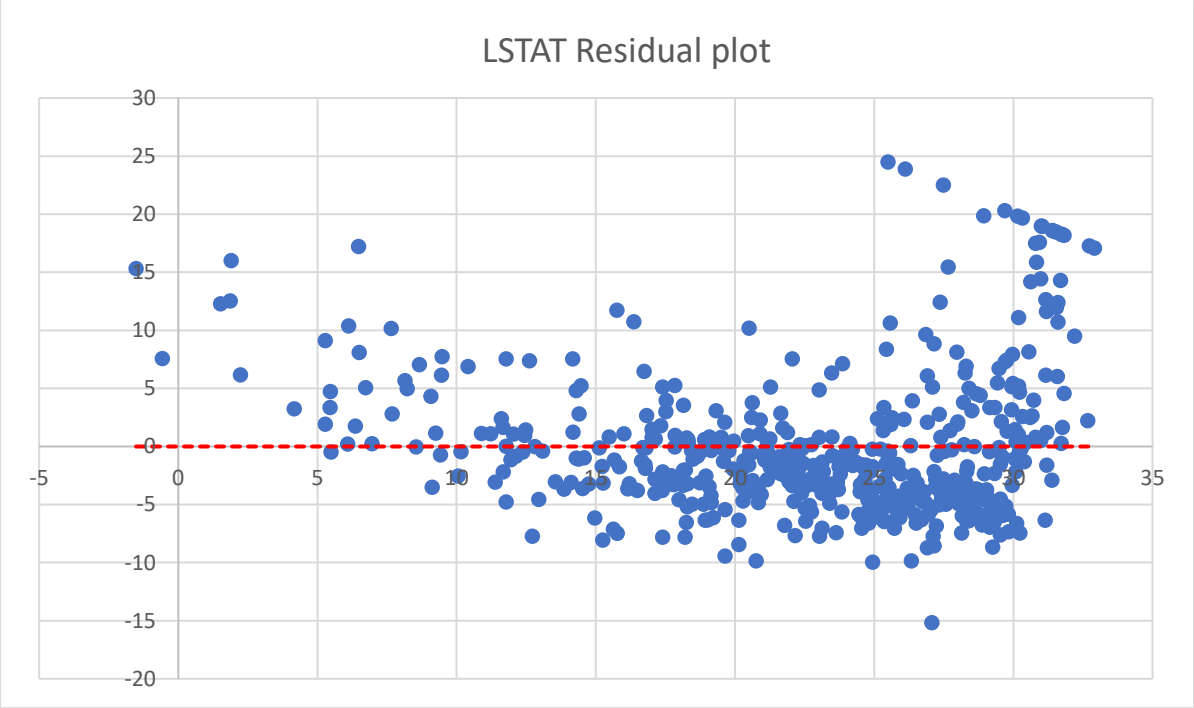
R square <60% condition not met.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	34.55384088	0.562627355	61.41514552	3.743E-236	33.44845704	35.6592247	33.44845704	35.65922472
LSTAT	-0.950049354	0.038733416	-24.52789985	5.0811E-88	-1.0261482	-0.87395051	-1.0261482	-0.873950508

P-value for LSTAT is less than 0.05 that means it is significant for prediction.

Avg_residuals	-2.737E-14
Actual Avg of Y	22.53
Avg_MSE	38.4829672
RMSE	6.20346413
Max possible Error	0.2753
skewness	1.46

- The maximum possible error is 27.5% which is greater than 10%. Condition not met.
- Mean of residuals = 0 (condition met)
- It is not normally distributed (condition not met).
- The variance of the residuals is constant (Condition met).



a) What do you infer from the Regression Summary output in terms of variance explained, coefficient value, Intercept, and Residual plot?

We can clearly observe that the R-squared value is less than 60%, which ideally should be greater than 60% for a good fit. The difference between R-squared and adjusted R-squared is less than 1%, indicating that there is no irrelevant data in the model. LSTAT is significant for prediction because its p-value is less than 0.05. Y = -0.950049354X + 34.55384088 is the equation of regression, from this equation, we can infer that if LSTAT increases by 1 unit, the price of the house will decrease by -0.950049354. The intercept value represents the average price when LSTAT is 0. The maximum possible error for this model is 27.5%, which is greater than 10%, indicating that it has prediction errors.

<u>Regarding residual properties:</u>

- The mean of residuals is 0.
- The residuals do not follow a normal distribution.
- In the residual plot, we can observe that the values exhibit a slight downward trend, which is negligible; therefore, it can be considered constant.

From this analysis, we can conclude that not all the assumptions of the model are met.

b) Is LSTAT variable significant for the analysis based on your model?

The p-value of LSTAT is 5.0811E-88, which is significantly less than 0.05. This suggests that the LSTAT variable is indeed significant for the analysis. However, it's important to note that despite the significance of LSTAT, the model still has an R-squared value of less than 60% and a maximum possible error greater than 10%. These findings indicate that the model has a high level of prediction errors.

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

- Use Excel's Data Analysis Tool Pak to perform a Regression analysis.
- Interpret the regression summary output and the residual plot.

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

R square is > 60% and difference between R square and adjusted R square is less than 1%. Condition is met.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.358272812	3.17282778	-0.428095348	0.66876494	-7.59190028	4.87535466	-7.59190028	4.875354658
AVG_ROOM	5.094787984	0.4444655	11.46272991	3.4723E-27	4.221550436	5.96802553	4.22155044	5.968025533
LSTAT	-0.642358334	0.043731465	-14.68869925	6.6694E-41	-0.72827717	-0.5564395	-0.72827717	-0.556439501

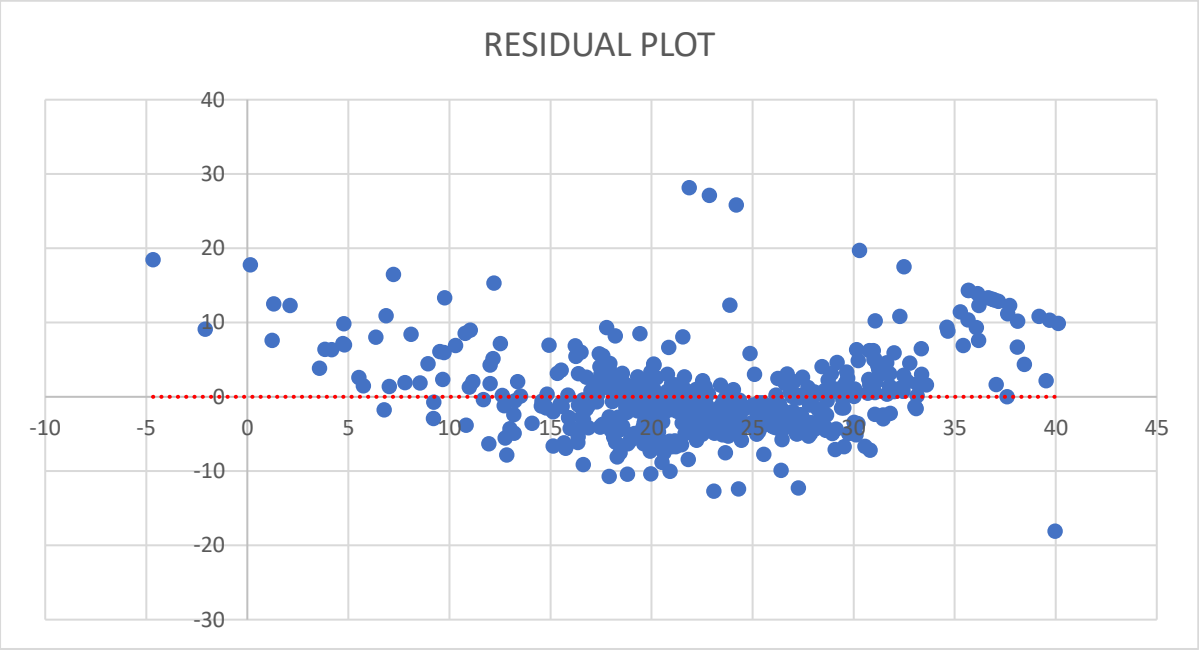
P-value for both Average room and LSTAT are less than 0.05. This both are significant variable for prediction.

Avg_Residuals	1.4474E-14
Avg_actual y	22.53
MSE	30.5124688
RMSE	5.52380926
Max possible error	0.2451
skewness	1.35

Maximum possible error is greater than 10%. Condition not met.

Assumption

- Mean of residuals = 0, met
- It doesn't follow normal distribution, not met
- Variance of residuals are constant, met



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?

Given:
L-STAT = 20, AVG_ROOM=7
Y=mx1+nx2+c
Y=5.094787984X1-0.642358334X2-1.358272812
=(E33*7) + (E34*20) + E32
= (5.094787984*7) - (0.642358334*20) - 1.358272812
=21.458076396
Y=21.458076396 * 1000
=21458.07639 USD
Company quoting value = 30000 USD
Avg_price for this locality < Company quoting value
21458 USD < 30000 USD
The company's quoted value (30,000 USD) is higher than your calculated average price (21,458.08 USD). Therefore, based on this analysis, it appears that the company is overcharging for houses in this locality.

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.

The R-squared and adjusted R-squared values have improved compared to the previous response, and they are now greater than 60%. Additionally, the difference between the R-squared and adjusted R-squared is less than 1%. Furthermore, the maximum possible error has also decreased from 27.5% to 24.5%. This indicates that the prediction errors have been minimized compared to the previous model.

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.

- Use Excel's Data Analysis Tool Pak to perform a Regression analysis.
- Interpret the regression summary output and the residual plot.

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

R square is > 60% and difference between R square and adjusted R square is less than 1%. Condition met

	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

All variables are significant expect the Crime_rate highlighted in yellow colour. we have to remove crime_rate from the model because it doesn't explain the Y variable.
After removing the irrelevant variable we have to do this process again.

OBSERVATION

- The R-squared value is greater than 60%, indicating that the model is effective in explaining the relationship between the predictors and the outcome.
- The difference between R-squared and adjusted R-squared is less than 1 percentage point, suggesting that there is no irrelevant data.
- The regression equation is as follows: Y = 0.048725141 X1 + 0.032770689 X2 + 0.130551399 X3 - 10.3211828 X4 + 0.261093575 X5 - 0.01440119 X6 - 1.074305348 X7 + 4.125409152 X8 - 0.603486589 X9 + 29.24131526.
- All the independent variables have p-values less than 0.05, except for the crime rate variable, which has a p-value greater than 0.05, making it insignificant. To obtain a good regression model, we need to remove the crime rate variable and repeat the process.

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.

- Use Excel's Data Analysis Tool pak to perform a Regression analysis.
- Interpret the regression summary output and the residual plot.

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

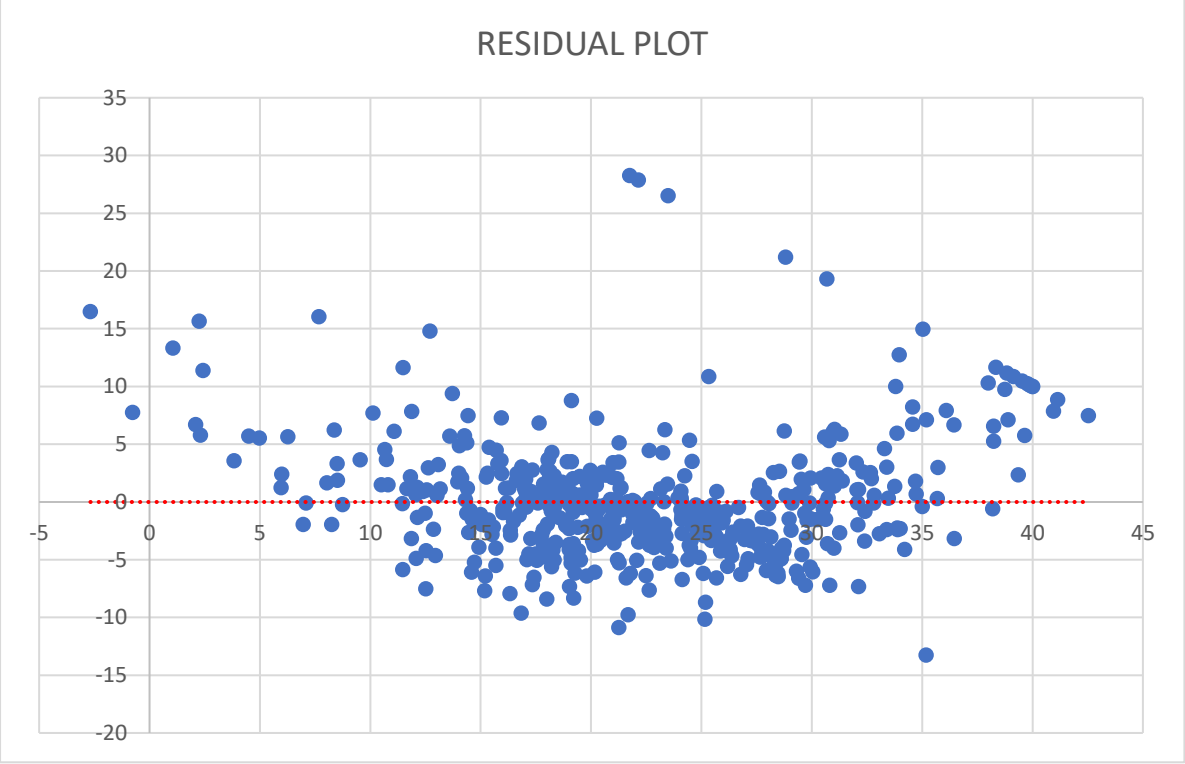
R square is > 60% and difference between R square and adjusted R square is less than 1%. Condition met.

	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

All the variables are significant. P-value <0.05

Avg_Residuals	-1.03948E-14
Avg_actual y	22.5328
MSE	25.8648
RMSE	5.0857
Max possible error	0.2257
skewness	1.6439

Maximum possible error is greater than 10%. Condition not met.
Mean of residuals = 0, met
It doesn't follow normal distribution, not met
Variance of residuals are constant, met



From above we can conclude that all the assumptions for the regression model is not met that means this data is not suitable for regression or this data has predictive outliers we have to remove the predictive outliers and we have to repeat the process again.

a) Interpret the output of this model.

The R-squared value is greater than 60%, indicating that the model is effective in explaining the relationship between the predictors and the outcome. The difference between R-squared and adjusted R-squared is less than 1 percentage point, suggesting that there is no irrelevant data.

The equation of the model is: Y = 0.03293496 * X1 + 0.130710007 * X2 - 10.27270508 * X3 + 0.261506423 * X4 - 0.014452345 * X5 - 1.071702473 * X6 + 4.125468959 * X7 - 0.605159282 * X8 + 29.42847349. The sign of each coefficient (positive or negative) indicates the direction of the relationship. A positive coefficient means that as the corresponding independent variable increases, the dependent variable is expected to increase as well. Conversely, a negative coefficient suggests that as the independent variable increases, the dependent variable is expected to decrease. The coefficient associated with the constant term (29.42847349 in this case) represents the estimated value of the dependent variable when all the independent variables are set to zero. It serves as the baseline value. All independent variables p values are less than 0.05 so all are significant.

However, the maximum possible error is greater than 10%, indicating that the model has a high prediction error.

Regarding the residuals:

- The residual mean is 0, but the residuals are not normally distributed.
- The variance of the residuals is not constant.
- Therefore, it can be concluded that this model is not suitable for Regression.

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?

When comparing the adjusted R-squared value of this model with the model in the previous question, it's evident that there is a slight difference. In the previous

model, the difference between R-squared and adjusted R-squared was 0.005555074, whereas in this model, the difference is 0.00493174, indicating a minor variation. Furthermore, the difference between the adjusted R-squared value of the previous model and this model is 0.000385035. From this, we can conclude that this model performs slightly better than the previous one in terms of adjusted R-squared.

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?

The coefficients in ascending order
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

	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

As the coefficient for NOX is -10.27270508, it means that for each one-unit increase in the NOX concentration (parts per 10 million), the average price of houses (AVG_PRICE) is expected to decrease by approximately 10.27 units, while holding all other variables constant.
In simpler terms, when the NOX concentration increases, the average house price tends to decrease significantly in this model. This suggests that higher levels of nitric oxides concentration are associated with lower property prices in this locality, according to the model.

d) Write the regression equation from this model.

Y = 0.03293496 * X1 + 0.130710007 * X2 - 10.27270508 * X3 + 0.261506423 * X4 - 0.014452345 * X5 - 1.071702473 * X6 + 4.125468959 * X7 - 0.605159282 * X8 + 29.42847349.

In conclusion

It is evident that not all the conditions for regression have been met, as the model exhibits a maximum possible error exceeding 10%. This indicates the presence of prediction errors, rendering the model unsuitable for regression analysis. Therefore, alternative modelling approaches or further data refinement may be necessary to improve the accuracy of predictions in this context.

_____END_____