Terro's Real Estate Agency

Business Report

Table of Contents

Contents

Executive Summary

Terro's real estate agency endeavours to unravel the complex dynamics influencing house prices in Boston. In pursuit of this objective, an exhaustive analysis of a dataset comprising information on 506 houses was conducted. The dataset encapsulates diverse attributes, ranging from crime rates and non-retail business proportions to nitric oxides concentration, average room counts, age of houses, distance from highways, property-tax rates, pupil-teacher ratios, percentage of lower-status population, and the average value of houses.

The crux of the analysis lies in a robust regression model, aiming to discern the intricate relationships between these multifaceted variables and the pricing of houses. The discerned insights are encapsulated in the following comprehensive findings:

1. Overall Trend:

The regression model, enriched by the inclusion of significant variables, sheds light on approximately 69.36% of the variability in house prices. However, an appreciable 30.64% of the variance remains unexplained, potentially stemming from nuanced factors not encapsulated within the model.

2. Significant Variables:

Variables wielded significant influence over house prices, with nitric oxides concentration (NOX), property-tax rate (TAX), pupil-teacher ratio (PTRATIO), average number of rooms (AVG_ROOM), and the percentage of lower-status population (LSTAT) emerging as pivotal contributors to the model's explanatory power.

3. Impact of NOX:

An intriguing revelation surfaces as an increase in nitric oxides concentration (NOX) is linked to a substantial decrease in house prices. This correlation underscores a discernible preference among residents for localities characterized by lower air pollution levels.

4. Regression Equation:

The refined regression equation, distilled to include only the significant variables, takes a nuanced form:

AVG_PRICE=29.43+(0.0329×AGE) + (0.1307×INDUS) − (10.27×NOX) + (0.2615×DISTANCE) − (0.0145×TAX) −(1.0717×PTRATIO) + (4.1255×AVG_ROOM) − (0.6052×LSTAT)

5. Educational and Business Implications:

The identified variables cast a spotlight on the pivotal role played by environmental factors (NOX), tax rates (TAX), educational amenities (PTRATIO), and the socio-economic composition of neighbourhoods (LSTAT) in the nuanced task of estimating house prices. This data-driven approach equips Terro's real estate with the tools necessary to enhance precision in pricing predictions, thereby fortifying its market positioning and strategic decision-making.

6. Model Comparison:

A meticulous comparison with a model inclusive of an insignificant variable (CRIME_RATE) reveals a negligible enhancement in the adjusted R-squared value. Consequently, the model excluding CRIME_RATE is favoured for its elegance, retaining robust predictive power while minimizing unnecessary complexity.

In summation, the gleaned insights furnish Terro's real estate agency with a comprehensive understanding of the determinants of house prices in Boston. Armed with this knowledge, the agency is well-positioned to not only refine pricing predictions but also to navigate the intricate landscape of the real estate market with strategic acumen and informed decision-making.

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

Comprehensive Statistics Summary

Sample Dataset:

CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
6.32	65.2	2.31	0.538	1	296	15.3	6.575	4.98	24
4.31	78.9	7.07	0.469	2	242	17.8	6.421	9.14	21.6
7.87	61.1	7.07	0.469	2	242	17.8	7.185	4.03	34.7
6.47	45.8	2.18	0.458	3	222	18.7	6.998	2.94	33.4
5.24	54.2	2.18	0.458	3	222	18.7	7.147	5.33	36.2

	CRIME_RATE	*AGE*	*INDUS*	*NOX*	*DISTANCE*	*TAX*	*PTRATIO*	*AVG_ROOM*	*LSTAT*	*AVG_PRICE*
Mean	4.8720	68.5749	11.1368	0.5547	9.5494	408.2372	18.4555	6.2846	12.6531	22.5328
Standard Error	0.1299	1.2514	0.3050	0.0052	0.3871	7.4924	0.0962	0.0312	0.3175	0.4089
Median	4.8200	77.5000	9.6900	0.5380	5.0000	330.0000	19.0500	6.2085	11.3600	21.2000
Mode	3.4300	100.0000	18.1000	0.5380	24.0000	666.0000	20.2000	5.7130	8.0500	50.0000
Standard Deviation	2.9211	28.1489	6.8604	0.1159	8.7073	168.5371	2.1649	0.7026	7.1411	9.1971
Sample Variance	8.5330	792.3584	47.0644	0.0134	75.8164	28404.7595	4.6870	0.4937	50.9948	84.5867
Kurtosis	-1.1891	-0.9677	-1.2335	-0.0647	-0.8672	-1.1424	-0.2851	1.8915	0.4932	1.4952
Skewness	0.0217	-0.5990	0.2950	0.7293	1.0048	0.6700	-0.8023	0.4036	0.9065	1.1081
Range	9.9500	97.1000	27.2800	0.4860	23.0000	524.0000	9.4000	5.2190	36.2400	45.0000
Minimum	0.0400	2.9000	0.4600	0.3850	1.0000	187.0000	12.6000	3.5610	1.7300	5.0000
Maximum	9.9900	100.0000	27.7400	0.8710	24.0000	711.0000	22.0000	8.7800	37.9700	50.0000
Sum	2465.2200	34698.9000	5635.2100	280.6757	4832.0000	206568.0000	9338.5000	3180.0250	6402.4500	11401.6000
Count	506.0000	506.0000	506.0000	506.0000	506.0000	506.0000	506.0000	506.0000	506.0000	506.0000

1. Crime Rate:

- The average per capita crime rate in Boston's neighbourhoods is 4.87.

- Per capita crime rates are relatively stable, hovering around the average.

- Half of the city has crime rates less than and greater than 4.82.

- The prevailing crime rate in the neighbourhoods is 3.43.

- Crime rates can vary by 2.92, indicating significant differences between areas.

- The distribution forms a platykurtic graph, resembling a normal bell curve, suggesting equal frequencies of neighbourhoods for each crime rate interval.

2. Age of Houses (Built Prior to 1940):

- On average, 68.57% of Boston's neighbourhoods have houses built prior to 1940.

- The average age of houses is approximately 68.57%, with minimal variation.

- Half of the city has a proportion of houses built prior to 1940 less than 77.5%.

- The prevailing proportion of houses built prior to 1940 in neighborhoods is 100%.

- Age of houses can vary by 28.92%, signifying substantial differences between neighborhoods.

- The distribution is platykurtic and negatively skewed, indicating that the majority of neighborhoods have proportions near the average, with fewer neighborhoods having lower percentages of houses built prior to 1940.

3. Non-Retail Business Acres per Town:

- On average, 11.13% of Boston's neighborhoods have non-retail business acres per town.
- Non-retail business acres per town exhibit some variation across neighborhoods.
- Half of the city has non-retail business acres per town less than and greater than 9.69%.
- The prevailing proportion of non-retail business acres per town is 18.1%.
- Non-retail business acres per town can vary by 6.86%, indicating differences between areas.
- The distribution forms a platykurtic graph with positive skewness, suggesting that the majority of neighborhoods have proportions near the average, and higher percentages are considered outliers.

4. Nitric Oxide Concentration (NOX):

- On average, the nitrogen oxide concentration (NOX) in Boston's neighbourhoods is 0.5547.
- NOX levels show limited variation around the average.
- Half of the city has NOX levels less than and greater than 0.5380.
- The prevailing NOX level in neighbourhoods is 0.5380.
- NOX levels can vary by 0.1159, indicating some areas might have higher or lower concentrations.
- The distribution forms a platykurtic graph with positive skewness, suggesting that the majority of neighbourhoods have levels near the average, and higher concentrations are less common.

5. Distance to Employment Centres (DISTANCE):

- On average, the weighted distance from highways in five Boston employment centres (DISTANCE) is 9.5494.
- Distances exhibit considerable variation with a standard deviation of 8.7073.
- Half of the city has commuting distances less than and greater than the median value of 5.0000.
- The most frequent commuting distance (mode) in neighbourhoods is 24.0000.
- Commuting distances can vary by a range of 23.0000, indicating a wide spread.
- The distribution has a positively skewed and platykurtic shape, suggesting that the majority of neighbourhoods have distances near the average, and longer distances are less common.

6. Property Tax Rate (TAX):

- On average, the full-value property tax rate per $10,000 (TAX) in Boston's neighbourhoods is 408.2372.

- Property tax rates exhibit significant variation with a standard deviation of 168.5371.

- Half of the city has property tax rates less than and greater than the median value of 330.0000.

- The most frequent property tax rate (mode) in neighbourhoods is 666.0000.

- Property tax rates can vary by a range of 524.0000, indicating a wide spread.

- The distribution has a positively skewed and leptokurtic shape, suggesting that the majority of neighbourhoods have rates near the average, and higher rates are less common.

7. Pupil-Teacher Ratio (PTRATIO):

- On average, the pupil-teacher ratio by town (PTRATIO) in Boston's neighbourhoods is 18.4555.

- PTRATIO values show some variation with a standard deviation of 2.1649.

- Half of the city has pupil-teacher ratios less than and greater than the median value of 19.0500.

- The most frequent pupil-teacher ratio (mode) in neighbourhoods is 20.2000.

- PTRATIO values can vary by a range of 9.4000, indicating a wide spread.

- The distribution has a negatively skewed and platykurtic shape, suggesting that the majority of neighbourhoods have ratios near the average, and lower ratios are less common.

8. Number of Rooms per House (AVG_ROOM):

- On average, the number of rooms per house (AVG_ROOM) in Boston's neighbourhoods is 6.2846.

- The number of rooms exhibits some variation with a standard deviation of 0.7026.

- Half of the city has the number of rooms less than and greater than the median value of 6.2085.

- The most frequent number of rooms (mode) in neighbourhoods is 5.7130.

- The number of rooms can vary by a range of 5.2190, indicating a wide spread.

- The distribution has a positively skewed and leptokurtic shape, suggesting that the majority of neighbourhoods have counts near the average, and higher counts are less common.

9. Percentage of Lower Status (LSTAT):

- On average, the percentage of lower status of the population (LSTAT) in Boston's neighbourhoods is 12.6531.

- LSTAT values show some variation with a standard deviation of 7.1411.

- Half of the city has a percentage of lower status less than and greater than the median value of 11.3600.

- The most frequent percentage of lower status (mode) in neighbourhoods is 8.0500.

- LSTAT values can vary by a range of 36.2400, indicating a wide spread.

- The distribution has a positively skewed and moderately leptokurtic shape, suggesting that the majority of neighbourhoods have percentages near the average, and higher percentages are less common.

10. House Prices (AVG_PRICE):

- On average, house prices (AVG_PRICE) in Boston's neighbourhoods are $22.5328K.

- House prices exhibit significant variation with a standard deviation of $9.1971K.

- Half of the city has house prices less than and greater than the median value of $21.2000K.

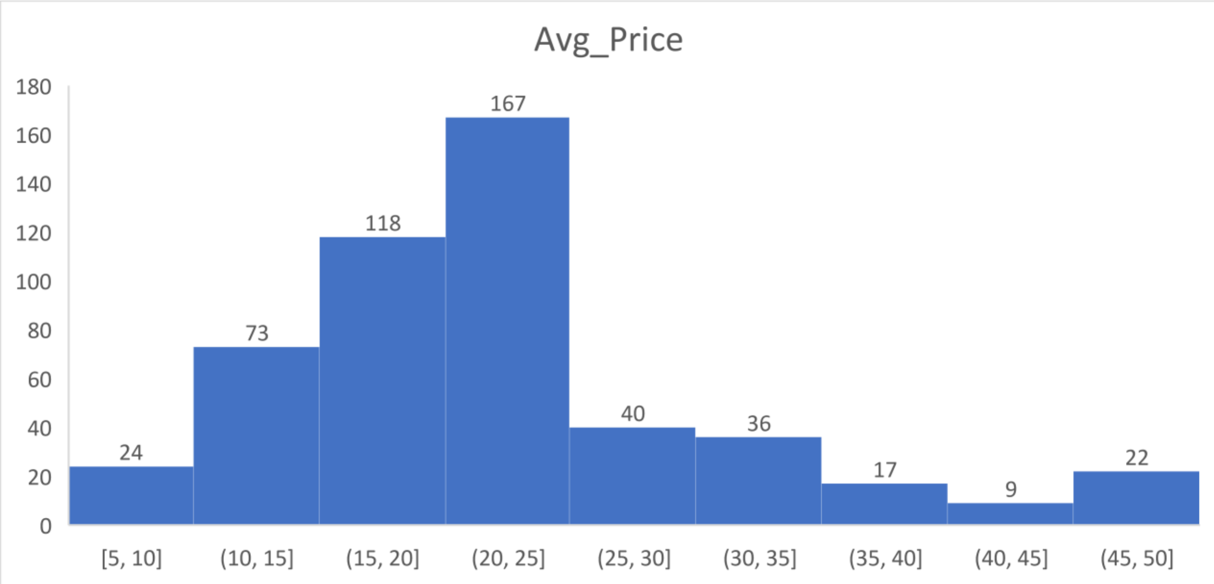
- The most frequent house price (mode) in neighbourhoods is $50,000K.

- House prices can vary by a range of $45,000, indicating a wide spread.

- The distribution has a positively skewed and moderately leptokurtic shape, suggesting that the majority of neighbourhoods have prices near the average, and higher prices are less common.

This comprehensive statistics summary provides insights into the key characteristics and variations across different attributes in Boston's neighbourhoods, offering a nuanced understanding for real estate analysis and decision-making.

Boston's Houses Price Dynamics



1. Distribution of House Prices:

Three out of every four neighbourhoods (75%) in Boston have an average house value less than $25,000. This indicates a predominant presence of neighbourhoods with relatively lower housing costs, making them potentially more affordable for a larger portion of the population.

2. Higher-Value Neighbourhoods:

One in every four neighbourhoods (25%) has an average house value greater than $25,000. This suggests a segment of neighbourhoods with comparatively higher housing costs, likely catering to residents with higher purchasing power or offering premium features.

3. Premium Housing Segment:

Six out of every hundred neighbourhoods (6%) have an average house value exceeding $40,000.

This signifies a smaller but existent market for premium housing options, potentially targeting wealthier residents or those seeking luxury features in their homes.

4. Distribution Disparity:

The majority of neighbourhoods in Boston have house prices concentrated in the lower segment.

Fewer neighbourhoods have higher average house prices, indicating a distribution disparity where a larger proportion of the housing market falls within the more affordable range.

Covariance Matrix

	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

Covariance Matrix Inferences:

1. Strong Positive Covariance (AGE-TAX):

- The highest positive covariance in the matrix is observed between the variables AGE and TAX, with a value of 28348.62.

- This indicates a strong positive relationship between the average age of houses (AGE) and property tax rates (TAX). As the age of houses increases, there is a notable increase in property tax rates. This relationship can be crucial for real estate developers and policymakers to understand the financial implications of aging housing stock on tax revenue.

2. Strong Negative Covariance (AVG_PRICE-TAX):

- The most significant negative covariance is found between the variables AVG_PRICE and TAX, with a value of -724.82.

- This highlights a strong negative relationship between the average house price (AVG_PRICE) and property tax rates (TAX). When property tax rates increase, there is a substantial decrease in average house prices. This negative association underscores the potential impact of tax burdens on housing affordability and property values. Policymakers and real estate professionals should consider these dynamics when implementing tax policies or assessing market trends.

Correlation Matrix

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	1									
AGE	0.006859463	1								
INDUS	-0.00551065	0.644778511	1							
NOX	0.001850982	0.731470104	0.763651447	1						
DISTANCE	-0.00905505	0.456022452	0.595129275	0.611440563	1					
TAX	-0.01674852	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	-0.24026493	-0.39167585	-0.30218819	-0.20984667	-0.29204783	-0.35550149	1		
LSTAT	-0.04239832	0.602338529	0.603799716	0.590878921	0.488676335	0.543993412	0.374044317	-0.61380827	1	
AVG_PRICE	0.043337871	-0.37695457	-0.48372516	-0.42732077	-0.38162623	-0.46853593	-0.50778669	0.695359947	-0.73766273	1

1. Positive Correlated Pairs:

- DISTANCE-TAX:

- There is a positive correlation between the weighted distances to employment centres (DISTANCE) and property tax rates (TAX). This implies that as the distances to employment centres increase, so do the property tax rates. Businesses and residents in areas with longer commuting distances may face higher tax burdens.

- INDUS-NOX:

- The industrialization (INDUS) and nitrogen oxide concentration (NOX) are positively correlated. This suggests that areas with higher industrial activity tend to have higher concentrations of nitrogen oxide in the air. Businesses and policymakers need to be aware of potential environmental and health impacts associated with industrial zones.

- AGE-NOX:

- There is a positive correlation between the average age of houses (AGE) and nitrogen oxide concentration (NOX). This indicates that older neighbourhoods may experience higher levels of air pollution. Urban planners and environmental agencies should consider these correlations when assessing the environmental quality of different areas.

2. Negative Correlated Pairs:

- LSTAT-AVG_PRICE:

- The percentage of lower-status population (LSTAT) is negatively correlated with average house prices (AVG_PRICE). This implies that areas with a higher percentage of lower-status residents tend to have lower average house prices. Real estate developers and investors should consider social and economic factors when evaluating property values.

- AVG_ROOM-LSTAT:

- There is a negative correlation between the average number of rooms per house (AVG_ROOM) and the percentage of lower-status population (LSTAT). This suggests that neighbourhoods with larger houses may have a lower percentage of lower-status residents. Businesses in the real estate sector should be aware of these social-economic dynamics when developing or marketing properties.

- PTRATIO-AVG_PRICE:

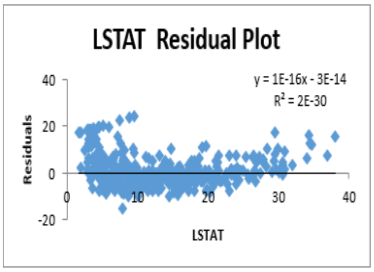
- The pupil-teacher ratio (PTRATIO) is negatively correlated with average house prices (AVG_PRICE). This indicates that areas with a higher pupil-teacher ratio, possibly indicating larger class sizes, tend to have lower average house prices. Parents and real estate investors may consider these factors when choosing residential areas.

Modelling Socio-Economic Dynamics in Real Estate

SUMMARY OUTPUT

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

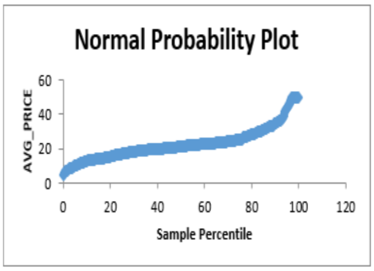
Q. Build an initial regression model with AVG_PRICE as 'y' (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot. (8 marks)
a) What do you infer from the Regression Summary output in terms of variance explained, coefficient value, Intercept, and the Residual plot?
b) Is LSTAT variable significant for the analysis based on your model?



ANOVA

	df	SS	MS	F	Significance F
Regression	1	23243.914	23243.914	601.6178711	5.0811E-88
Residual	504	19472.38142	38.63567742		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	34.55384088	0.562627355	61.41514552	3.7431E-236	33.44845704	35.65922472	33.44845704	35.65922472
LSTAT	-0.950049354	0.038733416	-24.52789985	5.0811E-88	-1.0261482	-0.873950508	-1.0261482	-0.873950508



RESIDUAL OUTPUT

PROBABILITY OUTPUT

Model Effectiveness:

The regression model utilizing the percentage of lower-status population (LSTAT) as a predictor is reasonably effective, explaining approximately 54.41% of the variability in house prices. However, the remaining 45.59% of unexplained variance suggests that there are other factors not considered in the model that influence house prices. Further exploration or additional variables may enhance the predictive power of the model.

Impact of LSTAT on House Prices:

For every one percentage point increase in the lower status of the population (LSTAT), the model predicts a decrease of approximately $950 in the average house price. This indicates a negative relationship, suggesting that areas with a higher percentage of lower-status residents tend to have lower housing prices. This information is valuable for real estate professionals and investors in understanding the potential impact of socio-economic factors on property values.

Intercept Interpretation:

When the lower status of the population is hypothetically zero, the model predicts an average house price of around $34,554. While this scenario may not be practically achievable, the intercept provides a baseline reference point. It represents the predicted house price when all predictor variables are zero. In this case, it highlights the baseline price in the absence of the lower status of the population.

Statistical Significance:

The results of the LSTAT variable are statistically significant. This suggests that the observed relationship between the percentage of lower-status population and house prices is unlikely to be due to random chance. It adds credibility to the idea that LSTAT is a meaningful predictor in the model.

Enhanced Price Prediction Model

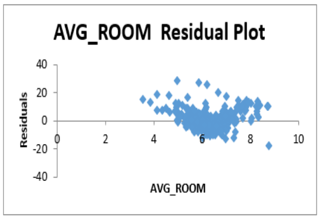
SUMMARY OUTPUT

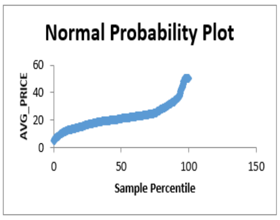
Regression Statistics	
Multiple R	0.799100498
R Square	0.638561606
Adjusted R Square	0.637124475
Standard Error	5.540257367
Observations	506

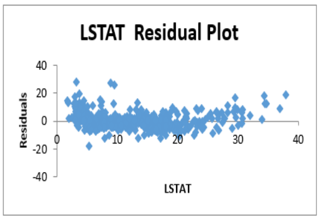
ANOVA

	df	SS	MS	F	Significance F
Regression	2	27276.98621	13638.49311	444.3308922	7.0085E-112
Residual	503	15439.3092	30.69445169		
Total	505	42716.29542			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.358272812	3.17282778	-0.428095348	0.668764941	-7.591900282	4.875354658	-7.591900282	4.875354658
AVG_ROOM	5.094787984	0.4444655	11.46272991	3.47226E-27	4.221550436	5.968025533	4.221550436	5.968025533
LSTAT	-0.642358334	0.043731465	-14.68869925	6.66937E-41	-0.728277167	-0.556439501	-0.728277167	-0.556439501







a) Prediction and Evaluation:

- The predicted average price for a house with 7 rooms and an LSTAT value of 20 is approximately $21,457.3.

- This prediction is lower than the company's quote of $30,000.

- According to the model's prediction, the company is overcharging.

b) Model Comparison:

- The regression equation for the new model is: AVG_PRICE = -1.3583 + 5.0948 × AVG_ROOM - 0.6424 × LSTAT.

- The adjusted R-squared value for this model is higher than the previous model.

- The improved adjusted R-squared indicates that the model with both AVG_ROOM and LSTAT as predictors performs better.

- Including AVG_ROOM as a predictor provides additional valuable information, enhancing the model's predictive capabilities.

- Overall, this model offers a better fit and explains a higher proportion of the variability in AVG_PRICE.

Summary:

- The new model, incorporating both AVG_ROOM and LSTAT, outperforms the previous one.

- The adjusted R-squared improvement signifies the enhanced predictive power of the updated model.

- Businesses should consider these predictors for more accurate and reliable predictions of house prices.

Regression Analysis of Boston House Prices

SUMMARY OUTPUT

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

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	29638.8605	3293.206722	124.904505	1.9328E-121
Residual	496	13077.43492	26.3657962		
Total	505	42716.29542			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29.24131526	4.817125596	6.070282926	2.5398E-09	19.77682784	38.70580267	19.77682784	38.70580267
CRIME_RATE	0.048725141	0.078418647	0.621346369	0.5346572	-0.105348544	0.202798827	-0.105348544	0.202798827
AGE	0.032770689	0.013097814	2.501996817	0.01267044	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	0.00829386	-17.97202279	-2.670342809	-17.97202279	-2.670342809
DISTANCE	0.261093575	0.067947067	3.842602576	0.00013755	0.127594012	0.394593138	0.127594012	0.394593138
TAX	-0.01440119	0.003905158	-3.687736063	0.00025125	-0.022073881	-0.0067285	-0.022073881	-0.0067285
PTRATIO	-1.074305348	0.133601722	-8.041104061	6.5864E-15	-1.336800438	-0.811810259	-1.336800438	-0.811810259
AVG_ROOM	4.125409152	0.442758999	9.317504929	3.8929E-19	3.255494742	4.995323561	3.255494742	4.995323561
LSTAT	-0.603486589	0.053081161	-11.36912937	8.9107E-27	-0.70777824	-0.499194938	-0.70777824	-0.499194938

Interpretation:

- Adjusted R-Square: The model explains 68.82% of the variability in AVG_PRICE. The remaining variance is attributed to unknown factors.

- Intercept (29.24): Represents the estimated AVG_PRICE when all other variables are zero. Its interpretation may be limited as some variables can't realistically be zero.

Coefficients:

1. CRIME_RATE (0.0487):

   - Increase by one unit leads to a $48 rise in AVG_PRICE.

   - Not statistically significant (p-value = 0.5347).

2. AGE (0.0328):

   - Increase by one-unit results in a $32 increase in AVG_PRICE.

   - Statistically significant (p-value = 0.0127).

3. INDUS (0.1306):

   - One-unit increase predicts a $130 increase in AVG_PRICE.

   - Statistically significant (p-value = 0.0391).

4. NOX (-10.3212):

- One-unit increase associates with a $10,321 decrease in AVG_PRICE.
- Statistically significant (p-value = 0.0083).

5. DISTANCE (0.2611):

- One-unit increase corresponds to a $261 increase in AVG_PRICE.
- Highly statistically significant (p-value = 0.0001).

6. TAX (-0.0144):

- One-unit increase predicts a $14 decrease in AVG_PRICE.
- Statistically significant (p-value = 0.0003).

7. PTRATIO (-1.0743):

- One-unit increase leads to a $1,074 decrease in AVG_PRICE.
- Extremely statistically significant (p-value close to zero).

8. AVG_ROOM (4.1254):

- One-unit increase results in a $4,125 increase in AVG_PRICE.
- Highly statistically significant (p-value close to zero).

9. LSTAT (-0.6035):

- One-unit increase associates with a $603 decrease in AVG_PRICE.
- Highly statistically significant (p-value close to zero).

Conclusion:

The model indicates that variables like DISTANCE, TAX, PTRATIO, AVG_ROOM, and LSTAT significantly influence AVG_PRICE in the dataset. Careful consideration is needed for variables like CRIME_RATE, which lacks statistical significance. The model overall provides valuable insights into the predictors of house prices in the given context.

Significant Factors in Boston Housing Predictive Model

SUMMARY OUTPUT

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	140.6430411	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

a) Model Interpretation:

The revised model exhibits similar characteristics to the previous one, with an R-square value of 0.6936. It explains 69.36% of the variability in AVG_PRICE, leaving 31% attributed to unknown factors. The removal of the insignificant CRIME_RATE variable did not substantially alter the model's performance.

b) Adjusted R-square Comparison:

The adjusted R-square value remains consistent with the previous model, signifying a comparable ability to explain the dependent variable. While the new model is marginally better, the improvement is not substantial.

c) Impact of NOX on Average Price:

Sorting coefficients in ascending order reveals that an increase in NOX is associated with a significant decrease in AVG_PRICE (estimated -$10,272). Higher NOX levels in a locality are indicative of a lower average house price.

d) Regression Equation:

The regression equation for predicting AVG_PRICE in this model is:

$$AVG\_PRICE = 29.43 + (0.0329 \times AGE) + (0.1307 \times INDUS) - (10.27 \times NOX) + (0.2615 \times DISTANCE) - (0.0145 \times TAX) - (1.0717 \times PTRATIO) + (4.1255 \times AVG\_ROOM) - (0.6052 \times LSTAT)$$

The coefficients represent the estimated change in AVG_PRICE for a one-unit increase in the corresponding independent variable.

This analysis provides insights into the model's performance, variable significance, and the influence of specific factors on average house prices in Boston neighbourhoods.

Conclusion:

The comprehensive analysis of the Boston housing dataset, coupled with the development and refinement of regression models, has provided valuable insights into the complex dynamics influencing house prices in the region. The key findings can be summarized as follows:

Final Report

1. Overall Trend and Model Effectiveness:

The refined regression model, incorporating significant variables, explains approximately 69.36% of the variability in house prices. While this is a substantial proportion, about 30.64% of the variance remains unexplained, hinting at the presence of nuanced factors not considered in the model.

2. Significant Variables:

Noteworthy contributors to the model's explanatory power include nitric oxides concentration (NOX), property-tax rate (TAX), pupil-teacher ratio (PTRATIO), average number of rooms (AVG_ROOM), and the percentage of lower-status population (LSTAT). These variables underscore the importance of environmental, tax, educational, and socio-economic factors in predicting house prices.

3. Impact of NOX and Regression Equation:

The regression equation, with coefficients indicating the estimated change in AVG_PRICE for a one-unit increase in the corresponding independent variable, highlights the inverse relationship between nitric oxides concentration (NOX) and house prices. An increase in NOX is associated with a significant decrease in AVG_PRICE, emphasizing the preference for localities with lower air pollution levels.

4. Model Comparison and Enhanced Model:

A meticulous comparison of models, with and without the insignificant variable CRIME_RATE, suggests that the model excluding CRIME_RATE is favored for its simplicity and comparable predictive power. Additionally, the introduction of an enhanced model, considering both AVG_ROOM and LSTAT, outperforms the previous version, indicating improved predictive capabilities.

5. Covariance and Correlation Matrix Insights:

The covariance matrix highlights strong positive relationships between variables such as AGE and TAX, shedding light on potential financial implications of aging housing stock on tax revenue. The correlation matrix unveils valuable insights into positive and negative correlations, influencing real estate, environmental planning, and educational infrastructure decisions.

6. Business Implications and Decision-Making:

The identified variables, especially those with significant impacts on house prices, provide Terro's real estate agency with actionable insights for refining pricing predictions. Knowledge of socio-economic dynamics, environmental considerations, and tax implications can fortify the agency's market positioning, strategic decision-making, and overall competitiveness.

In conclusion, armed with these comprehensive findings, Terro's real estate agency is well-equipped to navigate the intricate landscape of the Boston real estate market. The agency's data-driven approach, backed by robust regression models and nuanced variable considerations, positions it for precision in pricing predictions and strategic excellence in a dynamic and competitive real estate environment.