BUSINESS REPORT OF

TERRO’S REAL ESTATE AGENCY

DONE BY

S.PRAVEEN RAJ

10.09.2023

**Introduction:**

Terro's Real Estate is an agency specializing in estimating the pricing of houses in a specific locality. This pricing is determined based on various factors and features of a property, allowing the agency to assess the business value of a property accurately. In this report, we will analyze a dataset of 506 houses in Boston and discuss how geographic features impact property prices. The dataset includes several attributes, and we will examine their significance in predicting house prices.

**Data Dictionary:**

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 the highway (in miles).

TAX: Full-value property-tax rate per $10,000.

PTRATIO: Pupil-teacher ratio by town.

LSTAT: Percentage of the lower status of the population.

AVG\_PRICE: Average value of houses in $1000.

**Observations:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Crime Rate: | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1) The Crime Rate Ranging from 0.04 minimum to 9.99 maximum. | | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2) It shows that most areas have only average Crime rate of 4.87. | | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The Distribution of data is Positively Skewed this show that there are more areas with lower  crime rate than higher crime rate. | | | | | | | | | | | | | | | | | | | | | | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The Kurtosis shows that the distribution of data is platykurtic which means it is less peaked  than a normal distribution. | | | | | | | | | | | | | | | | | | | | |  | |
| Age: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The Age of Houses is ranging from 2.9 years to 100 years. | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most of houses have an Average Age of 77.5 years. | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The Distribution of data is Negatively Skewed, Showing that there are more new houses. | | | | | | | | | | | | | | | | | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The Kurtosis value Shows that the distribution is platykurtic. | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| Industry: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The proportion of non-retail business acres per town varies, with a range from 0.46% to 27.74%. | | | | | | | | | | | | | | | | | | | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most towns have proportions of non-retail business acres around the median value of  approximately 9.69%. | | | | | | | | | | | | | | | | | | | | | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution of data is slightly positively skewed, showing that there are lower  proportions of non-retail business acres. | | | | | | | | | | | | | | | | | | | | | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value suggests that the distribution is platykurtic. | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| Nox: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The NOX is ranging from 0.385 to 0.871. | | | | | | | |  | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most areas have nitric oxide concentrations around the median value of approximately 0.538. | | | | | | | | | | | | | | | | | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution of data is positively skewed, Show more NOX is there. | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value shows a distribution nearest to 0. | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| Distance: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The distance has a range from 1 to 24. | | | | | | | |  | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most areas have distances around the median value of approximately 5. | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution of data for distance is positively skewed, indicating more areas closer  to Highways | | | | | | | | | | | | | | | | | | | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value shows it is a platykurtic. | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| Tax: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)Tax has a range from 187 to 711. | | | | | | | |  | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)The distribution of tax rates is positively skewed, It shows more areas with  higher tax rates. | | | | | | | | | | | | | | | | | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The kurtosis value shows it is a platykurtic | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The Average of the Tax is 408 and it shows that areas are with higher tax. | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| PTRATIO: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)Pupil-teacher ratios have range of 12.6 to 22. | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most areas have pupil-teacher ratios around the median value of approximately 19.05. | | | | | | | | | | | | | | | | | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution is negatively skewed, showing that more areas with lower pupil-teacher ratios. | | | | | | | | | | | | | | | | | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value is leptokurtic is more peaked than a normal distribution. | | | | | | | | | | | | | | | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| AVG\_ROOM: | | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The Average Room range, with a maximum range of 5.219. | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most Houses have Room counts of the median value of approximately 6.2085. | | | | | | | | | | | | | | | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution is positively skewed, it shows the higher room counts. | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value is leptokurtic is more peaked than a normal distribution | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| LSTAT: |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The percentage of the lower status of the population varies widely, with a range  from 1.73% to 37.97%. | | | | | | | | | | | | | | | | | | | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)Most areas have around the median value of approximately 11.36%. | | | | | | | | | | | | | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The distribution is positively skewed, showing that higher percentages of lower-income residents. | | | | | | | | | | | | | | | | | | | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The kurtosis value is leptokurtic is more peaked than a normal distribution. | | | | | | | | | | | | | | | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
| AVG\_PRICE: | | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 1)The Average number of houses ranges from 5 to 50. | | | | | | | | | |  | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 2)House prices around the median value of approximately 21.2. | | | | | | | | | | | |  | |  | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 3)The Distribution is positively skewed and shows areas with higher house prices. | | | | | | | | | | | | | | | |  | |  | |  | |
|  |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
|  | 4)The Kurtosis values shows that it is leptokurtic is more peaked than normal distribution. | | | | | | | | | | | | | | | | | |  | |  | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *CRIME\_RATE* | *AGE* | *INDUS* | *NOX* | *DISTANCE* | *TAX* | *PTRATIO* | *AVG\_ROOM* | *LSTAT* | *AVG\_PRICE* |
| CRIME\_RATE | 1 |  |  |  |  |  |  |  |  |  |
| AGE | 0.006859 | 1 |  |  |  |  |  |  |  |  |
| INDUS | -0.00551 | 0.644779 | 1 |  |  |  |  |  |  |  |
| NOX | 0.001851 | 0.73147 | 0.763651 | 1 |  |  |  |  |  |  |
| DISTANCE | -0.00906 | 0.456022 | 0.595129 | 0.611441 | 1 |  |  |  |  |  |
| TAX | -0.01675 | 0.506456 | 0.72076 | 0.668023 | 0.910228 | 1 |  |  |  |  |
| PTRATIO | 0.010801 | 0.261515 | 0.383248 | 0.188933 | 0.464741 | 0.460853 | 1 |  |  |  |
| AVG\_ROOM | 0.027396 | -0.24026 | -0.39168 | -0.30219 | -0.20985 | -0.29205 | -0.3555 | 1 |  |  |
| LSTAT | -0.0424 | 0.602339 | 0.6038 | 0.590879 | 0.488676 | 0.543993 | 0.374044 | -0.61381 | 1 |  |
| AVG\_PRICE | 0.043338 | -0.37695 | -0.48373 | -0.42732 | -0.38163 | -0.46854 | -0.50779 | 0.69536 | -0.73766 | 1 |

**Correlation**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TOP THREE POSITIVELY CORRELATED PAIRS** | | | | |  |  |  |  |  |  |  |  |  |  |
| 1)DISTANCE AND TAX | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:0.910228 | | |  |  |  |  |  |  |  |  |  |  |  |
|  | The pair has the Highest Correlation in the Matrix,It shows that there is a strong relationship  between Distance from the Highway and Tax. | | | | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2)NOX AND INDUSTRY | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:0.7637 | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | The pair is the second-highest Correlation in the Matrix. It shows that there is a Moderate  relationship between Nox and Industry. | | | | | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3)AGE AND NOX | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:0.73147 | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | The Pair is the third-highest Correlation in the Matrix. It shows that there is a Moderate  relationship between Age and Nox. | | | | | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **TOP THREE NEGATIVELY CORRELATED** | | | |  |  |  |  |  |  |  |  |  |  |  |
| 1)AVG PRICE AND LSTAT | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:-0.73766 | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | The Pair has the Highest negatively Correlation in the Matrix.It shows it has a strong  negatively relationship between Avg Price and Lstat. | | | | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2)LSTAT VS AVG ROOM | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:-0.61381 | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | The Pair has the second-highest negative Correlation in the matrix.It shows it has a  moderate negative relationship between Lstat and Avg Room. | | | | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3)AVG PRICE VS PTRATIO | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Correlation:-0.50779 | |  |  |  |  |  |  |  |  |  |  |  |  |
|  | The Pair has the third highest negatively Correlation in the Matrix.It show it has a  Moderate negative relationship between Avg price and Ptratio. | | | | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Other Charts

**Analysis and Insights:**

1. **Feature Significance**:

We performed a multiple linear regression analysis to understand the significance of each feature in predicting house prices. Here are the key findings:

* + NOX (Nitric Oxides Concentration): A higher concentration of nitric oxides is associated with significantly lower house prices. A one-unit increase in NOX leads to a substantial decrease in property values.
  + DISTANCE (Distance from Highway): Houses located closer to highways tend to have higher prices. A one-unit increase in distance from the highway is associated with a moderate increase in property values.
  + TAX (Property Tax Rate): Higher property tax rates are negatively correlated with house prices. A one-unit increase in property tax rate leads to a modest decrease in property values.
  + PTRATIO (Pupil-Teacher Ratio): Areas with a lower pupil-teacher ratio tend to have higher property prices. A one-unit decrease in PTRATIO results in a moderate increase in property values.
  + AVG\_ROOM (Average Number of Rooms): More rooms per house positively impact property values. A one-unit increase in the average number of rooms leads to a significant increase in property prices.
  + LSTAT (Percentage of Lower Status Population): Areas with a higher percentage of lower-status population tend to have lower property values. A one-unit increase in LSTAT results in a significant decrease in property prices.
  + AGE (Proportion of Older Houses): The proportion of older houses has a moderate impact on property prices. However, other factors are more influential in predicting prices.
  + INDUSTRY (Proportion of Non-Retail Business): While it has some influence, the proportion of non-retail business acres has a relatively weaker correlation with property prices.
  + CRIME RATE: The crime rate does not appear to be a statistically significant predictor of property prices.

1. **Model Performance:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **The Significance of each independent variable with respect to AVG\_PRICE.** | | | | | | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Crime Rate | |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of The Crime rate is 0.5346. The p-value suggests that  CRIME\_RATE is not statistically significant in predicting | | | | | | | | | | |
|  | AVG\_PRICE because it's greater than the commonly used  significance level of 0.05. | | | | | | | |  |  |  |
|  | This means that changes in CRIME\_RATE do not have a significant  impact on AVG\_PRICE. | | | | | | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Age |  |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of the Age 0.01267.he p-value is less than 0.05, indicating  that AGE is statistically significant in predicting AVG\_PRICE. | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Industry |  |  |  |  |  |  |  |  |  |  |  |
|  | The P-Value of the Industry is 0.039121.Industry is statistically  significant in predicting AVG\_PRICE. | | | | | | | | | |  |
|  | An increase in INDUSTRY is associated with a positive increase in  AVG\_PRICE. | | | | | | | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Nox |  |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of Nox is 0.008294. Nox is statistically significant and  negatively related to AVG\_PRICE. | | | | | | | | | |  |
|  | An increase in NOX is associated with a decrease in AVG\_PRICE. | | | | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Distance |  |  |  |  |  |  |  |  |  |  |  |
|  | The P value of Distance is 0.2611. DISTANCE is highly statistically  significant and positively related to AVG\_PRICE. | | | | | | | | | | |
|  | An increase in DISTANCE is associated with an increase in AVG\_PRICE. | | | | | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Tax |  |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of Tax is 0.000251.TAX is highly statistically significant  and negatively related to AVG\_PRICE. . | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Ptratio |  |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of ptratio is 6.59E-15. PTRATIO is highly statistically  significant and negatively related to AVG\_PRICE. | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Avg\_Room | |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of Avg\_Room is 3.89E-19.AVG\_ROOM is highly  statistically significant and positively related to AVG\_PRICE. | | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Lstat |  |  |  |  |  |  |  |  |  |  |  |
|  | The p-value of LSTAT is 8.91E-27. LSTAT is highly statistically  significant and negatively related to AVG\_PRICE. | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

We constructed a regression model using the significant variables (NOX, DISTANCE, TAX, PTRATIO, AVG\_ROOM, LSTAT) and achieved an adjusted R-squared value of approximately 0.6835. This indicates that about 68.35% of the variability in property prices can be explained by these significant features.

**Conclusion:**

In summary, Terro's Real Estate has valuable insights into the factors affecting property prices in Boston. The concentration of nitric oxides, distance from highways, property tax rates, pupil-teacher ratios, the average number of rooms, and the percentage of lower-status population are key determinants of house prices. Terro's can use this knowledge to provide accurate property pricing estimates and offer valuable guidance to clients in making informed decisions.

**Recommendations:**

1. Consider Geographic Factors: Terro's should pay particular attention to geographic features like NOX concentration and proximity to highways when estimating property prices.
2. Highlight Education Facilities: Properties located in areas with lower pupil-teacher ratios may command higher prices. Terro's can emphasize this in property listings.
3. Room Quality Matters: The number of rooms in a house significantly affects its value. Terro's should factor this into their pricing assessments.
4. Population Demographics: The social and economic composition of neighborhoods, as represented by LSTAT, plays a crucial role in property values.
5. Regular Updates: Keep the pricing model up-to-date with current data and market trends to provide the most accurate estimates.

By incorporating these insights into their practices, Terro's Real Estate can offer exceptional value to clients and maintain a strong position in the real estate market.