Terro’s real estate agency

Sunday, February 19, 2023

GLCA DA online

By: P Vishnu Teja

**Problem Statement:**

“Finding out the most relevant features for pricing of a house” Terro’s real-estate is an agency that estimates the pricing of houses in a certain locality. The pricing is concluded based on different features / factors of a property. This also helps them in identifying the business value of a property. To do this activity the company employs an “Auditor”, who studies various geographic features of a property like pollution level (NOX), crime rate, education facilities (pupil to teacher ratio), connectivity (distance from highway), etc. This helps in determining the price of a property.

The agency has provided a dataset of 506 houses in Boston.

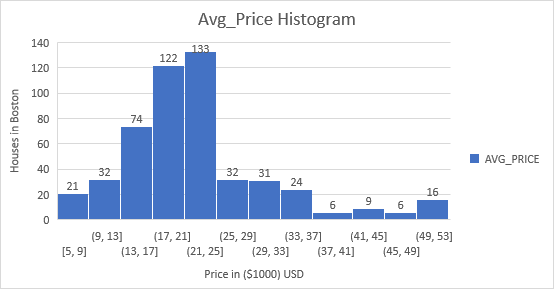
**Data Dictionary:**

|  |  |
| --- | --- |
| **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 |

1.Summary statistics:

* The variable CRIME\_RATE has a mean of 4.87 and a standard deviation of 2.92, with a minimum value of 0.04 and a maximum value of 9.99. It has a slightly positive skewness (0.02) and a slightly negative kurtosis (-1.19).
* The variable AGE has a mean of 68.57 and a standard deviation of 28.15, with a minimum value of 2.9 and a maximum value of 100. It has a slightly negative skewness (-0.60) and a slightly negative kurtosis (-0.97).
* The variable INDUS has a mean of 11.14 and a standard deviation of 6.86, with a minimum value of 0.46 and a maximum value of 27.74. It has a slightly positive skewness (0.30) and a slightly negative kurtosis (-1.23).
* The variable NOX has a mean of 0.55 and a standard deviation of 0.12, with a minimum value of 0.385 and a maximum value of 0.871. It has a positively skewed distribution with a skewness of (0.73) and a slightly negative kurtosis (-0.06).
* The variable DISTANCE has a mean of 9.55 and a standard deviation of 8.71, with a minimum value of 1 and a maximum value of 24. It has a positively skewed distribution with a skewness of (1.00) and a slightly negative kurtosis (-0.86).
* The variable TAX has a mean of 408.24 and a standard deviation of 168.54, with a minimum value of 187 and a maximum value of 711. It has a slightly positive skewness (0.67) and a slightly negative kurtosis (-1.14).
* The variable PTRATIO has a mean of 18.46 and a standard deviation of 2.16, with a minimum value of 12.6 and a maximum value of 22. It has a negatively skewed distribution with a skewness of -0.80 and a slightly negative kurtosis (-0.28).
* The variable AVG\_ROOM has a mean of 6.28 and a standard deviation of 0.70, with a minimum value of 3.56 and a maximum value of 8.78. It has a slightly positive skewness (0.40) and a kurtosis of (1.89).
* The variable LSTAT has a mean of 12.65 and a standard deviation of 9.20, with a minimum value of 1.73 and a maximum value of 37.97. It has a positively skewed distribution with a skewness of 0.91 and a kurtosis of (0.49).
* The variable AVG\_PRICE (target variable) has a mean of 22.53 and a standard deviation of 9.20, with a minimum value of 5 and a maximum value of 50. It has a positively skewed distribution with a skewness of (1.11) and a kurtosis of (1.49).
* Overall, most of the variables have a skewed distribution with some outliers. The standard deviation of some variables is relatively high, indicating that the data points are widely spread out from the mean. Additionally, some variables have a wide range, indicating a large variation in the data.

**2.Histogram of the Avg\_Price variable:**



From this histogram, we can infer that the majority of the housing prices in Boston are between 10 and 30 ($1000 USD), with a smaller number of houses having prices above 30. This could indicate that the Boston housing market is generally affordable, with some houses priced higher due to factors such as location, size, and amenities.

* The majority of the houses in Boston fall within the price range of $17,000 to $29,000, with a peak frequency of 133 houses falling in the interval of $21,000 to $25,000.
* The prices of houses in Boston range from a minimum of $5,000 to a maximum of $53,000.
* There are relatively fewer houses in the higher price ranges, with only 6 houses falling within the interval of $37,000 to $41,000, and 16 houses falling within the interval of $49,000 to $53,000.
* The distribution of house prices appears to be skewed to the right, with a longer tail on the higher end of the price range, indicating that there are fewer high-priced houses in Boston compared to the lower-priced ones.

**3.Covariance matrix observations:**

* The covariance matrix shows how the variables are related to each other.
* The diagonal elements represent the variance of each variable.
* The off-diagonal elements represent the covariance between the corresponding variables.
* A positive covariance indicates that the two variables move in the same direction, while a negative covariance indicates they move in opposite directions.
* For example, we can observe that the variables 'AGE' and 'TAX' have a very high covariance of 2397.942, which suggests that they are positively related.
* Similarly, we can observe that the variables 'DISTANCE' and 'AVG\_PRICE' have a negative covariance of -30.501, which suggests that they are

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRIME\_RATE | AGE | INDUS | NOX | DISTANCE | TAX | PTRATIO | AVG\_ROOM | LSTAT | AVG\_PRICE |
| CRIME\_RATE | 8.51614787 |  |  |  |  |  |  |  |  |  |
| AGE | 0.56291522 | 790.792473 |  |  |  |  |  |  |  |  |
| INDUS | -0.1102152 | 124.267828 | 46.9714297 |  |  |  |  |  |  |  |
| NOX | 0.00062531 | 2.38121193 | 0.60587394 | 0.0134011 |  |  |  |  |  |  |
| DISTANCE | -0.2298605 | 111.549955 | 35.4797145 | 0.61571022 | 75.6665313 |  |  |  |  |  |
| TAX | -8.2293224 | 2397.94172 | 831.713333 | 13.0205024 | 1333.11674 | 28348.6236 |  |  |  |  |
| PTRATIO | 0.06816891 | 15.9054254 | 5.68085478 | 0.04730365 | 8.74340249 | 167.820822 | 4.6777263 |  |  |  |
| AVG\_ROOM | 0.05611778 | -4.742538 | -1.8842254 | -0.0245548 | -1.2812774 | -34.515101 | -0.5396945 | 0.49269522 |  |  |
| LSTAT | -0.8826804 | 120.838441 | 29.5218113 | 0.48797987 | 30.3253921 | 653.420617 | 5.77130024 | -3.073655 | 50.8939794 |  |
| AVG\_PRICE | 1.16201224 | -97.396153 | -30.460505 | -0.4545124 | -30.50083 | -724.82043 | -10.090676 | 4.48456555 | -48.351792 | 84.4195562 |

**4.Correlation matrix**:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

|  |  |
| --- | --- |
| Positively Correlated | |
| Top1 | 0.91022818 |
| Top2 | 0.76365144 |
| Top3 | 0.73147010 |

|  |  |
| --- | --- |
| Negatively Correlated | |
| Top1 | -0.737662726 |
| Top2 | -0.613808272 |
| Top3 | -0.507786686 |

**4.a.** The top 3 positively correlated pairs are:

* Tax and Distance (0.910)
* Indus and NOx (0.763)
* Age and NOx (0.7314)

**4.b.** The top 3 negatively correlated pairs are:

* Avg\_Price and Lstat (-0.737)
* Lstat and Avg\_Room (-0.613)
* Avg\_Price and PTRATIO (-0.507)

**5.Initial regression model (Avg Price & Lstat):**

- Avg price as dependent variable

- lstat as independent variable

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

**5.a.** From the regression summary output, we can infer that:

The coefficient value for lstat is (-0.951), which means that as lstat increases by 1 unit, the avg\_price decreases by 0.95 units, all other variables held constant.

The intercept value is 34.554, which represents the predicted avg\_price when lstat is zero.

The r-squared value is 0.544, which means that approximately 54.4% of the variance in avg\_price is explained by the lstat variable.

The p-value for lstat is less than 0.05 i.e. (5.0811e-88), which indicates that the lstat variable is statistically significant in predicting the avg\_price.

From the residual plot, we can infer that:

The residual plot is a graphical representation of the differences between the predicted and actual values of the dependent variable. A good residual plot should show no pattern or trend, indicating that the model has captured the relationships between the variables.

In this case, the residual plot appears to have a slight u-shape, which may indicate that the model is not capturing all of the relevant relationships. However, the overall scatter is relatively even, which is a good sign.

**5.b.** Yes, lstat variable is significant for the analysis based on the model, as indicated by the p-value being less than 0.05. This suggests that the relationship between lstat and avg\_price is statistically significant.

But the model may not capture all of the relevant relationships between the variables. Further analysis, such as adding additional variables to the model or using more complex regression techniques, may be necessary to improve the model's performance.

**6.Regression Model (Avg\_Price, Lstat & Avg\_Room):**

- Avg\_price as dependent variable

- Lstat & Avg\_room as independent variables

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

**6.a.** Regression equation if a new house in the locality (Boston) has 7 rooms (on an average) and has a value of 20 for L-Stat, then the value of

**Avg\_Price=B0+[B1\*Avg\_Room] + [B2\*Lstat]**

Avg\_Price**=**-1.358272812 + [5.094787984\*7] + [ -0.642358334\*20]

Avg\_Price= 21.45807639

This predicted value of 21.45807639 is much lower than the company quoting a value of 30000 USD for this locality, indicating that the company is overcharging.

**6.b.** The performance of this model better than the previous model. we can compare the adjusted r-square values of the two models. The adjusted r-square value for the new model is 0.637, while the adjusted r-square value for the previous model (including only lstat as independent variable) was 0.547. This indicates that the new model with both avg\_room and lstat as independent variables is a better fit for the data than the previous model with only lstat as an independent variable.

**7.Regression model (AVG\_PRICE & All other variables):**

- AVG\_PRICE as dependent variable

- All other variables as independent

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

* Based on the regression analysis results, the adjusted r-squared value is 0.688, which means that 68.8% of the variability in the avg\_price can be explained by the independent variables included in the model.
* The intercept of the model is 29.24, which means that when all independent variables are 0, the predicted value of avg\_price is 29.24.
* The coefficient estimates for the independent variables indicate the magnitude and direction of the relationship between each independent variable and avg\_price.
* The independent variables age, Indus, distance, tax, NOx, ptratio, avg\_room, and lstat have statistically significant coefficients (p-values < 0.05), which means they are significantly related to avg\_price.
* Specifically, the coefficient for age is 0.0328, which means that for each additional year of age, the predicted avg\_price increases by 0.0328 units, holding all other variables constant.
* The coefficient for Indus is 0.131, which means that for each additional unit increase in the proportion of non-retail business acres per town, the predicted avg\_price increases by 0.131 units, holding all other variables constant.
* The coefficient for distance is 0.261, which means that for each additional unit increase in the distance in Boston area, the predicted avg\_price increases by 0.261 units, holding all other variables constant.
* The coefficient for tax is -0.0144, which means that for each additional dollar increase in the full-value property-tax rate per $10,000, the predicted avg\_price decreases by 0.0144 units, holding all other variables constant.
* The coefficient for ptratio is -1.074, which means that for each additional unit increase in the pupil-teacher ratio by town, the predicted avg\_price decreases by 1.074 units, holding all other variables constant.
* The coefficient for avg\_room is 4.13, which means that for each additional room per dwelling, the predicted avg\_price increases by 4.13 units, holding all other variables constant.
* The coefficient for lstat is -0.603, which means that for each additional percentage point increase in the proportion of lower status of the population, the predicted avg\_price decreases by 0.603 units, holding all other variables constant.
* The coefficient for NOx is -10.3212, which means that for each additional unit increase in the nitric oxide concentration the predicted avg\_price decreases by 10.3212 units, holding all other variables constant.
* On the other hand, the coefficient for Crime Rate is not statistically significant (p-value > 0.05), indicating that the variable is not significantly related to avg\_price in this model.

**8.Regression model (AVG\_PRICE & Significant**

**Variables):**

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |

**8.a.** The output of this model provides information on the relationship between the significant independent variables and the dependent variable, as well as the overall fit of the model. Here are some interpretations of the key statistics:

* Multiple R: The multiple correlation coefficient measures the strength and direction of the linear relationship between the dependent variable and the independent variables. In this model, the multiple R is 0.8328, indicating a strong positive relationship between the predictors and the target variable.
* R Square: The coefficient of determination (R-squared) measures the proportion of variance in the dependent variable that can be explained by the independent variables.
* In this model, R-squared is 0.6936, which means that about 69.4% of the variance in the dependent variable (median home value) is explained by the independent variables in the model.
* Adjusted R Square: Adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in the model. In this case, the adjusted R-squared is 0.6887, which is slightly lower than the R-squared value due to the inclusion of additional independent variables.
* Standard Error: The standard error estimates the standard deviation of the errors in the model, which represents the degree of variability in the residuals that cannot be explained by the model. In this model, the standard error is 5.1316, which means that the model's predicted median home value can vary from the actual median home value by an average of about $5,131.
* Coefficients: The coefficients represent the estimated regression coefficients for each independent variable in the model. They provide information on the direction and magnitude of the relationship between each independent variable and the dependent variable while holding all other variables constant. The intercept represents the expected median home value when all independent variables are equal to zero.
* Looking at the coefficients, we can see that several variables have a statistically significant relationship with median home value. The variables with a positive coefficient (AGE, INDUS, DISTANCE, AVG\_ROOM) are positively associated with median home value, meaning that as these variables increase, median home value tends to increase as well. The variables with a negative coefficient (NOX, TAX, PTRATIO, LSTAT) are negatively associated with median home value, meaning that as these variables increase, median home value tends to decrease.
* For example, one unit increase in LSTAT (percentage of lower status of the population) results in a 0.6051 decrease in the median home value, holding all other variables constant. Similarly, one unit increase in the average number of rooms per dwelling (AVG\_ROOM) results in a 4.1255 increase in the median home value, holding all other variables constant.
* Overall, this model can be used to predict the median home value (AVG\_PRICE) based on the significant independent variables, although the standard error indicates that there is still some unexplained variability in the model's predictions.

**8.b.** we can see that the adjusted R-square of the above model (7.AVG\_PRICE & All other variables) is 0.688298647, while the adjusted R-square of the second model (8.) is 0.688683682.

Comparing the two adjusted R-square values, we can see that the second model (8.) has a slightly higher adjusted R-square value than the above model with all variables. This suggests that the second model may perform slightly better in terms of explaining the variation in the response variable while accounting for the number of predictor variables in the model.

However, it's worth noting that the difference in adjusted R-square values between the two models is relatively small, so it's possible that other factors may need to be considered when choosing between the two models. Additionally, it's important to keep in mind that adjusted R-square is just one of several metrics that can be used to evaluate the performance of a regression model. Other metrics, such as root mean squared error or mean absolute error, may also be relevant depending on the specific context and goals of the analysis

**8.c.** The coefficients are sorted 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 |

Based on the coefficients of the model, we can see that a one-unit increase in NOX is associated with a decrease in the average housing price by $10.27 (holding all other predictors constant). However, we need to be cautious when interpreting this relationship as correlation does not imply causation. It's also worth noting that the impact of NOX on the average price may be mediated by other factors, and the relationship may be different in different localities in the town.

**8.d.** The regression equation for this model can be written as:

Y = 29.428 + 0.033age + 0.131indus - 10.273nox + 0.262distance - 0.014tax - 1.072ptratio + 4.125avg\_room - 0.605lstat

Where y represents the predicted average housing price, age is the proportion of owner-occupied units built before 1940, Indus is the proportion of non-retail business acres per town, NOx is the nitric oxides concentration, distance is the distances from Boston area to highway, tax is the full-value property-tax rate per $10,000, ptratio is the pupil-teacher ratio by town, avg\_room is the average number of rooms per dwelling(house), and lstat is the percentage of lower status of the population.

APPENDIX

Sample prediction of value of the house in Boston according to final regression equation:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AGE | INDUS | NOX | DISTANCE | TAX | PTRATIO | AVG\_ROOM | LSTAT | AVG\_PRICE | Predicted value of the house |
| 65.2 | 2.31 | 0.538 | 1 | 296 | 15.3 | 6.575 | 4.98 | 24 | 30.04888734 |
| 78.9 | 7.07 | 0.469 | 2 | 242 | 17.8 | 6.421 | 9.14 | 21.6 | 27.04098462 |
| 61.1 | 7.07 | 0.469 | 2 | 242 | 17.8 | 7.185 | 4.03 | 34.7 | 32.69896454 |
| 45.8 | 2.18 | 0.458 | 3 | 222 | 18.7 | 6.998 | 2.94 | 33.4 | 31.14306949 |
| 54.2 | 2.18 | 0.458 | 3 | 222 | 18.7 | 7.147 | 5.33 | 36.2 | 30.58808735 |
| 58.7 | 2.18 | 0.458 | 3 | 222 | 18.7 | 6.43 | 5.21 | 28.7 | 27.85095254 |
| 66.6 | 7.87 | 0.524 | 5 | 311 | 15.2 | 6.012 | 12.43 | 22.9 | 25.07089688 |
| 96.1 | 7.87 | 0.524 | 5 | 311 | 15.2 | 6.172 | 19.15 | 27.1 | 22.63588287 |
| 100 | 7.87 | 0.524 | 5 | 311 | 15.2 | 5.631 | 29.93 | 16.5 | 14.00883345 |

Y = 29.428 + 0.033age + 0.131indus - 10.273nox + 0.262distance - 0.014tax - 1.072ptratio + 4.125avg\_room - 0.605lstat

Y = 29.428 + 0.033\*65.2+ 0.13\*2.31 - 10.273\*0.538 + 0.262\*1 - 0.014296 - 1.072\*15.3 + 4.125\*6.575 - 0.605\*4.98

Y = 30.04888734