*Business report of Terro’s Real Estate Agency*

1. ***summary statistics for each of the variables.***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CRIME\_RATE |  | AGE |  | INDUS |  | NOX |  |
|  |  |  |  |  |  |  |  |
| Mean | 4.871976 | Mean | 68.574901 | Mean | 11.13678 | Mean | 0.554695 |
| Standard Error | 0.12986 | Standard Error | 1.2513695 | Standard Error | 0.30498 | Standard Error | 0.005151 |
| Median | 4.82 | Median | 77.5 | Median | 9.69 | Median | 0.538 |
| Mode | 3.43 | Mode | 100 | Mode | 18.1 | Mode | 0.538 |
| Standard Deviation | 2.921132 | Standard Deviation | 28.148861 | Standard Deviation | 6.860353 | Standard Deviation | 0.115878 |
| Sample Variance | 8.533012 | Sample Variance | 792.3584 | Sample Variance | 47.06444 | Sample Variance | 0.013428 |
| Kurtosis | -1.18912 | Kurtosis | -0.967716 | Kurtosis | -1.23354 | Kurtosis | -0.06467 |
| Skewness | 0.021728 | Skewness | -0.598963 | Skewness | 0.295022 | Skewness | 0.729308 |
| Range | 9.95 | Range | 97.1 | Range | 27.28 | Range | 0.486 |
| Minimum | 0.04 | Minimum | 2.9 | Minimum | 0.46 | Minimum | 0.385 |
| Maximum | 9.99 | Maximum | 100 | Maximum | 27.74 | Maximum | 0.871 |
| Sum | 2465.22 | Sum | 34698.9 | Sum | 5635.21 | Sum | 280.6757 |
| Count | 506 | Count | 506 | Count | 506 | Count | 506 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DISTANCE | | TAX |  | PTRATIO |  | AVG\_ROOM | |
|  |  |  |  |  |  |  |  |
| Mean | 9.549407 | Mean | 408.2372 | Mean | 18.45553 | Mean | 6.284634 |
| Standard Error | 0.387085 | Standard Error | 7.492389 | Standard Error | 0.096244 | Standard Error | 0.031235 |
| Median | 5 | Median | 330 | Median | 19.05 | Median | 6.2085 |
| Mode | 24 | Mode | 666 | Mode | 20.2 | Mode | 5.713 |
| Standard Deviation | 8.707259 | Standard Deviation | 168.5371 | Standard Deviation | 2.164946 | Standard Deviation | 0.702617 |
| Sample Variance | 75.81637 | Sample Variance | 28404.76 | Sample Variance | 4.686989 | Sample Variance | 0.493671 |
| Kurtosis | -0.86723 | Kurtosis | -1.14241 | Kurtosis | -0.28509 | Kurtosis | 1.8915 |
| Skewness | 1.004815 | Skewness | 0.669956 | Skewness | -0.80232 | Skewness | 0.403612 |
| Range | 23 | Range | 524 | Range | 9.4 | Range | 5.219 |
| Minimum | 1 | Minimum | 187 | Minimum | 12.6 | Minimum | 3.561 |
| Maximum | 24 | Maximum | 711 | Maximum | 22 | Maximum | 8.78 |
| Sum | 4832 | Sum | 206568 | Sum | 9338.5 | Sum | 3180.025 |
| Count | 506 | Count | 506 | Count | 506 | Count | 506 |
|  |  |  |  |  |  |  |  |

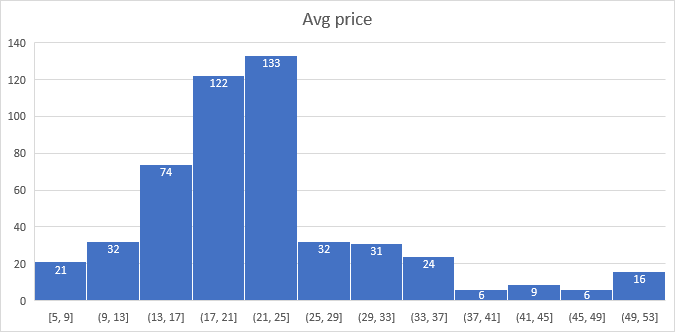
|  |  |  |  |
| --- | --- | --- | --- |
| LSTAT |  | AVG\_PRICE | |
|  |  |  |  |
| Mean | 12.65306 | Mean | 22.53280632 |
| Standard Error | 0.317459 | Standard Error | 0.408861147 |
| Median | 11.36 | Median | 21.2 |
| Mode | 8.05 | Mode | 50 |
| Standard Deviation | 7.141062 | Standard Deviation | 9.197104087 |
| Sample Variance | 50.99476 | Sample Variance | 84.58672359 |
| Kurtosis | 0.49324 | Kurtosis | 1.49196944 |
| Skewness | 0.90646 | Skewness | 1.108098408 |
| Range | 36.24 | Range | 45 |
| Minimum | 1.73 | Minimum | 5 |
| Maximum | 37.97 | Maximum | 50 |
| Sum | 6402.45 | Sum | 11401.6 |
| Count | 506 | Count | 506 |

***Detailed analysis:***

*The Mean, Median, Standard Deviation, range, minimum and maximum value of each aspect can be easily seen in the above tables.*

*Minimum and maximum age ranging between 2.9 to 100 and range is 97.1 The total count is 506. Mean(average) is of Avg price is 22.53. The average price of household is $ 22000.*

1. ***histogram of the Avg\_Price Variable.***



*The most household price ranging between the $21000 to $25000 and the range is 133. the next range $ 17000 to $ 21000 and the range 122. Lowest range is 6 and the price ranging from the $31000 to $41000 and the $45000 to $49000. The Average price range is inferred from this histogram.*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CRIME\_RATE | AGE | INDUS | NOX | DISTANCE | TAX | PTRATIO | AVG\_ROOM | LSTAT | AVG\_PRICE |
| CRIME\_RATE | 8.5161 |  |  |  |  |  |  |  |  |  |
| AGE | 0.5629 | 790.79 |  |  |  |  |  |  |  |  |
| INDUS | -0.11 | 124.27 | 46.971 |  |  |  |  |  |  |  |
| NOX | 0.0006 | 2.3812 | 0.6059 | 0.0134 |  |  |  |  |  |  |
| DISTANCE | -0.23 | 111.55 | 35.48 | 0.6157 | 75.667 |  |  |  |  |  |
| TAX | -8.229 | 2397.9 | 831.71 | 13.021 | 1333.1 | 28349 |  |  |  |  |
| PTRATIO | 0.0682 | 15.905 | 5.6809 | 0.0473 | 8.7434 | 167.82 | 4.6777 |  |  |  |
| AVG\_ROOM | 0.0561 | -4.743 | -1.884 | -0.025 | -1.281 | -34.52 | -0.54 | 0.4927 |  |  |
| LSTAT | -0.883 | 120.84 | 29.522 | 0.488 | 30.325 | 653.42 | 5.7713 | -3.074 | 50.894 |  |
| AVG\_PRICE | 1.162 | -97.4 | -30.46 | -0.455 | -30.5 | -724.8 | -10.09 | 4.4846 | -48.35 | 84.42 |

1. ***covariance matrix.***

*The covariance matrix shown above shows the symmetry and covariance for each pair of variables. Also, the covariance between the same variables is equal to their variance.*

*Therefore, the diagonal lines show the variance of each variable.*

*We can interpret the direction of multivariate data in multidimensional space.*

*About the* ***covariance matrix:*** *A covariance matrix is ​​a type of matrix used to represent covariance values ​​between pairs of elements specified in a random vector.*

1. ***Correlation matrix***

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

*The correlation matrix showing the correlation coefficients between different variables****.***

*A correlation coefficient of 1 indicates a perfect positive correlation.*

*This above correlation matrix is prefect positive correlation 1.*

***Top 3 positively correlated pairs****:*

*AVG\_PRICE and AVG\_ROOM (correlation coefficient = 0.376954565)*

*INDUS and NOX (correlation coefficient = 0.763651447)*

*AGE and CRIME\_RATE (correlation coefficient = 0.006859463)*

***Top 3 negatively correlated pairs:***

*AVG\_PRICE and LSTAT (correlation coefficient = -0.737662726)*

*AVG\_PRICE and DISTANCE (correlation coefficient = -0.381626231)*

*AVG\_ROOM and LSTAT (correlation coefficient = -0.613808272)*

1. ***initial regression model***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SUMMARY OUTPUT** | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| **Regression Statistics** | |  |  |  |  |  |  |  |
| **Multiple R** | 0.737663 |  |  |  |  |  |  |  |
| **R Square** | 0.544146 |  |  |  |  |  |  |  |
| **Adjusted R Square** | 0.543242 |  |  |  |  |  |  |  |
| **Standard Error** | 6.21576 |  |  |  |  |  |  |  |
| **Observations** | 506 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| **ANOVA** |  |  |  |  |  |  |  |  |
|  | **df** | **SS** | **MS** | **F** | **Significance F** |  |  |  |
| **Regression** | 1 | 23243.91 | 23243.91 | 601.6179 | 5.08E-88 |  |  |  |
| **Residual** | 504 | 19472.38 | 38.63568 |  |  |  |  |  |
| **Total** | 505 | 42716.3 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** | **Lower 95%** | **Upper 95%** | **Lower 95.0%** | **Upper 95.0%** |
| **Intercept** | 34.55384 | 0.562627 | 61.41515 | 3.7E-236 | 33.44846 | 35.65922 | 33.44846 | 35.65922472 |
| **LSTAT** | -0.95005 | 0.038733 | -24.5279 | 5.08E-88 | -1.02615 | -0.87395 | -1.02615 | -0.873950508 |

*Here the regression model with the scatter plot, residual plot is present.* ***R Square*** *is 0.54 means the 54% of the variation in the dependent variable is explained by the independent variable (Good fit) And the* ***Multiple R*** *show the its highly correlated.* *The intercept value of the coefficient is the constant of the regression equation.* ***P- value*** *much be less than the 0.05 % show that LSTAT variable is significant.*

***Coefficient value*** *the coefficient for LSTAT is -0.95, indicating that for each unit increase in LSTAT, the dependent variable (house price) decreases by 0.95 units. This suggests a negative linear relationship between LSTAT and house prices.*

*The* ***intercept*** *is 34.55. This means that if LSTAT is zero, the predicted value of the dependent variable is 34.55. This intercept value may or may not have a practical interpretation, depending on the context of the data.*

1. ***Regression model.***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Regression Statistics | |  |  |  |  |  |  |  |
| Multiple R | 0.799100 |  |  |  |  |  |  |  |
| R Square | 0.638562 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.637124 |  |  |  |  |  |  |  |
| Standard Error | 5.540257 |  |  |  |  |  |  |  |
| Observations | 506 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | df | SS | MS | F | Significance F |  |  |  |
| Regression | 2 | 27276.99 | 13638.49 | 444.3309 | 7E-112 |  |  |  |
| Residual | 503 | 15439.31 | 30.69445 |  |  |  |  |  |
| Total | 505 | 42716.3 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | -1.35827 | 3.172828 | -0.42810 | 0.668765 | -7.59190 | 4.875355 | -7.591900282 | 4.875354658 |
| AVG\_ROOM | 5.094788 | 0.444466 | 11.46273 | 3.47E-27 | 4.22155 | 5.968026 | 4.221550436 | 5.968025533 |
| LSTAT | -0.64236 | 0.043731 | -14.6887 | 6.67E-41 | -0.72828 | -0.55644 | -0.728277167 | -0.556439501 |

***Regression equation: y= -****1.35827+ 5.094788\*AVG\_ROOM-0.64236\*LSTAT*

*The average price of house with 7 room is the 21.46 (21,000$) and the company selling the house for 30,000 $, its clearly* ***Overcharging*.**

*The performance of this model of previous is adjusted R square is 0.543 and, on this question, adjusted R square is 0.637 that is good as compared to previous is adjusted R square.*

*Higher adjusted r square is good for regression model.*

1. ***Regression model with all variables.***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Regression Statistics | |  |  |  |  |  |  |  |
| Multiple R | 0.83297882 |  |  |  |  |  |  |  |
| R Square | 0.69385372 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.68829865 |  |  |  |  |  |  |  |
| Standard Error | 5.1347635 |  |  |  |  |  |  |  |
| Observations | 506 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | df | SS | MS | F | Significance F |  |  |  |
| Regression | 9 | 29638.8605 | 3293.20672 | 124.904505 | 1.933E-121 |  |  |  |
| Residual | 496 | 13077.4349 | 26.3657962 |  |  |  |  |  |
| Total | 505 | 42716.2954 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 29.2413153 | 4.8171256 | 6.07028293 | 2.5398E-09 | 19.7768278 | 38.7058027 | 19.7768278 | 38.7058027 |
| CRIME\_RATE | 0.04872514 | 0.07841865 | 0.62134637 | 0.5346572 | -0.10534854 | 0.20279883 | -0.10534854 | 0.20279883 |
| AGE | 0.03277069 | 0.01309781 | 2.50199682 | 0.01267044 | 0.00703665 | 0.05850473 | 0.00703665 | 0.05850473 |
| INDUS | 0.1305514 | 0.06311733 | 2.06839217 | 0.03912086 | 0.00654109 | 0.2545617 | 0.00654109 | 0.2545617 |
| NOX | -10.3211828 | 3.89403626 | -2.6505102 | 0.00829386 | -17.9720228 | -2.67034281 | -17.9720228 | -2.67034281 |
| DISTANCE | 0.26109357 | 0.06794707 | 3.84260258 | 0.00013755 | 0.12759401 | 0.39459314 | 0.12759401 | 0.39459314 |
| TAX | -0.01440119 | 0.00390516 | -3.68773606 | 0.00025125 | -0.02207388 | -0.0067285 | -0.02207388 | -0.0067285 |
| PTRATIO | -1.07430535 | 0.13360172 | -8.04110406 | 6.5864E-15 | -1.33680044 | -0.81181026 | -1.33680044 | -0.81181026 |
| AVG\_ROOM | 4.12540915 | 0.442759 | 9.31750493 | 3.8929E-19 | 3.25549474 | 4.99532356 | 3.25549474 | 4.99532356 |
| LSTAT | -0.60348659 | 0.05308116 | -11.3691294 | 8.9107E-27 | -0.70777824 | -0.49919494 | -0.70777824 | -0.49919494 |

*The adjusted R square is 0.68 indicate that 68% variability depended variable is explained by the independent variable.*

*The coefficient is the beta of the specified variable.*

*The intercept value of 29.241 indicates that if all independent variables are zero, the predicted value of the dependent variable (house price) is 29.241.*

*The significance of each variable compared to AVG\_PRICE can be measured using the p-value. Except for CRIME\_RATE, all other variables have p-values ​​less than 0.05, proving their significance.*

1. ***Full Regression model***



*Here the summary output and residual plot provided above. This multiple regression shows the relationship between the independent variable (AGE, INDUS, NOX, DISTANCE, TAX, PTRATIO, AVG\_ROOM, LSTAT) and dependent variable (Avg­\_price) except the crime rate.*

*The adjusted R square is 0.68868368 that 68% variability depended variable is explained by the independent variable.*

*The adjusted R square of preview question was 0.68829865* *and, on this question, the adjusted R square is 0.688683681872453. This adjusted R square increase by 0.000385035. higher adjusted R square is good.*

*Regression equation = 29.43 + 0.03(AGE) + 0.13(INDUS) - 10.27(NOX) + 0.26(DISTANCE) - 0.01(TAX) - 1.07(PTRATIO) + 4.13(AVG\_ROOM) - 0.61(LSTAT)*