

|  |
| --- |
| **Terro’s Real Estate****Agency** |
| Real Estate Data Analysis – Exploratory Data Analysis, Linear Regression. |
| Business Report  Submitted by: Vaishnavi G  Mail Id: vaishug127@gmail.com |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Generate the Summary Statistics for each Variable in the Table. (Use Data*** ***Analysis Tool Pack). Write down your observation.***  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Observation:**  From the summary Statistics Tablegiven below, |  |  |  |  |  | | * Mean value is largest in Tax column and smallest in Nox column. | | |  |  |  | | * Nox having the largest, smallest Median value respectively. | | |  |  |  | | * When compare to all Standard Deviation value in Tax is largest among them and Nox having the   lowest one. | | | | | | | * Skewness having highest value in Average price column and smallest value in Age column. | | | | |  | | * Range value is largest in Tax column and in Nox column.      |  |  |  | | --- | --- | --- | |  | Indicating largest value |  | |  | Indicating smallest Value |  | | | |  |  |  |  |  |  | | --- | --- | | **AGE** |  | |  |  | | Mean | 68.57490119 | | Standard Error | 1.251369525 | | Median | 77.5 | | Mode | 100 | | Standard Deviation | 28.14886141 | | Sample Variance | 792.3583985 | | Kurtosis | -0.967715594 | | Skewness | -0.59896264 | | Range | 97.1 | | Minimum | 2.9 | | Maximum | 100 | | Sum | 34698.9 | | Count | 506 |  |  |  | | --- | --- | | **NOX** |  | |  |  | | Mean | 0.554695059 | | Standard Error | 0.005151391 | | Median | 0.538 | | Mode | 0.538 | | Standard Deviation | 0.115877676 | | Sample Variance | 0.013427636 | | Kurtosis | -0.064667133 | | Skewness | 0.729307923 | | Range | 0.486 | | Minimum | 0.385 | | Maximum | 0.871 | | Sum | 280.6757 | | Count | 506 |        |  |  | | --- | --- | | **TAX** |  | |  |  | | Mean | 408.2371542 | | Standard Error | 7.492388692 | | Median | 330 | | Mode | 666 | | Standard Deviation | 168.5371161 | | Sample Variance | 28404.75949 | | Kurtosis | -1.142407992 | | Skewness | 0.669955942 | | Range | 524 | | Minimum | 187 | | Maximum | 711 | | Sum | 206568 | | Count | 506 |  |  |  | | --- | --- | | **AVG\_PRICE** |  | |  |  | | Mean | 22.53280632 | | Standard Error | 0.408861147 | | Median | 21.2 | | Mode | 50 | | Standard Deviation | 9.197104087 | | Sample Variance | 84.58672359 | | Kurtosis | 1.495196944 | | Skewness | 1.108098408 | | Range | 45 | | Minimum | 5 | | Maximum | 50 | | Sum | 11401.6 | | Count | 506 |         *2) Plot a histogram of the Avg\_Price variable. What do you infer?*   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Infer:** |  |  |  |  |  |  |  | | * This Histogram chart for the Average price shows that data is Positively Skewed. | | | | | | | |     *3) Compute the covariance matrix. Share your observations*    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 |   Observations:  From the above table,   * Age vs Tax and Indus vs Tax is directly related to each other. * Tax vs Average price and Age vs Average Price is inversely related to each other.   *4) Create a correlation matrix of all the variables (Use Data analysis tool pack).*  *a) Which are the top 3 positively correlated pairs and*  *b) Which are the top 3 negatively correlated pairs.*      **CORRELATION MATRIX**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | **CRIME\_RATE** | **AGE** | **INDUS** | **NOX** | **DISTANCE** | **TAX** | **PTRATIO** | **AVG\_ROOM** | **LSTAT** | **AVG\_PRICE** | | CRIME\_RATE | 1.00 |  |  |  |  |  |  |  |  |  | | AGE | 0.01 | 1.00 |  |  |  |  |  |  |  |  | | INDUS | -0.01 | 0.64 | 1.00 |  |  |  |  |  |  |  | | NOX | 0.00 | 0.73 | 0.76 | 1.00 |  |  |  |  |  |  | | DISTANCE | -0.01 | 0.46 | 0.60 | 0.61 | 1.00 |  |  |  |  |  | | TAX | -0.02 | 0.51 | 0.72 | 0.67 | 0.91 | 1.00 |  |  |  |  | | PTRATIO | 0.01 | 0.26 | 0.38 | 0.19 | 0.46 | 0.46 | 1.00 |  |  |  | | AVG\_ROOM | 0.03 | -0.24 | -0.39 | -0.30 | -0.21 | -0.29 | -0.36 | 1.00 |  |  | | LSTAT | -0.04 | 0.60 | 0.60 | 0.59 | 0.49 | 0.54 | 0.37 | -0.61 | 1.00 |  | | AVG\_PRICE | 0.04 | -0.38 | -0.48 | -0.43 | -0.38 | -0.47 | -0.51 | 0.70 | -0.74 | 1.00 |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | Indicating Top 3 positively pairs |  |  |  |  | |  | Indicating Top 3 negatively pairs |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Top 3 positively correlated pairs:** | | | | | | |  |  |  |  |  |  | | **Tax \*Distance** | 0.9102 |  |  |  |  | | **Nox\*Indus** | 0.7637 |  |  |  |  | | **Nox\*age** | 0.7315 |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Top 3 Negatively correlated pairs:** | | | | | |  |  |  |  |  | | **Avgprice\*PTRATIO** | -0.51 |  |  |  | | **LSTAT\*Avg Room** | -0.61 |  |  |  | | **Avg Price\*LSTAT** | -0.74 |  |  |  |   *5) Build an initial regression model with AVG\_PRICE as ‘y’ (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot.*  *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?*  Observations:   1. From the output summary the Variance, Coefficient, Intercept shows that Intercept and Average price   is positively related and LSTAT and Average price is negatively related.  From the residual plot we are unable to see any pattern. Hence it is known as  Homoskedasticity.   1. The LSTAT variable and Average price are positively correlated so, it is significant for our   analysis.  *6) Build a new Regression model including LSTAT and AVG\_ROOM together as Independent variables and AVG\_PRICE*  *as dependent variable.*   1. *Write the Regression equation. If a new house in this locality has 7 rooms (on an average) and has a value of*   *20 for L-STAT, then what will be the value of AVG\_PRICE? How does it compare to the company quoting a value of*  *30000 USD for this locality? Is the company Overcharging/ Undercharging?*   1. *Is the performance of this model better than the previous model you built in Question 5? Compare in terms of*   *adjusted R-square and explain.*   1. Regression Equation:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | y=5.0947\*(x1)-0.6423\*(x2)-1.3582 | | | | | | | | | | | | | | | | | |  | | |  | | |  | | |  | |  | |  |  |  | |  | | | | | | | | |  | | | |  | | |  | |  | | |  | | |  | | |  | |  | |  |  |  | | |  |  |  | | --- | --- | --- | | Intercept | 1 | -1.35827 | | AVG\_ROOM | 7 | 5.094788 | | LSTAT | 20 | -0.64236 | |  |  |  | |  |  |  | | Average Price is | | 21.45808 | | | | |  | | |  | | | |  | | | |  | | | |  | | |  | | |  | | |  | | |  | | | |  | | |  | | | |  | | | |  | | | |  | | |  | | |  | | |  | | |  | | | |  | | |  | | | |  | | | |  | | | |  | | |  | | |  | | |  | | |  | | | |  | | |  | | | |  | | | |  | | | |  | | |  | | |  | | |  | | |  | | | | | |  | | | |  | | | |  | | |  | | |  | | |  | | | |  | | | | | |  | | | |  | | | |  | | |  | | |  | | |  | | | | **Observation:** | | | | | | | | |  | | | |  | | |  | |  | | |  | | |  | | |  | |  | |  |  |  | |  | | | | | | | | |  | | | |  | | |  | |  | | |  | | |  | | |  | |  | |  |  |  | | * The company quoting a value of 30000 USD but the calculated value of Avg Price is lesser than the company   quotation. Hence we conclude that the company is **Overcharging.** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  | |  | | |  | | | |  1. Comparing Adjusted R Square Value  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Qtn 5 R Square Value** | |  |  |  | **Qtn 6 R square value** | | |  |  |  |  |  |  |  | | Regression Statistics |  |  |  |  | Regression Statistics |  | | Multiple R | 0.737663 |  |  |  | Multiple R | 0.7991 | | R Square | 0.544146 |  |  |  | R Square | 0.638562 | | Adjusted R Square | 0.543242 |  |  |  | Adjusted R Square | 0.637124 | | Standard Error | 6.21576 |  |  |  | Standard Error | 5.540257 | | Observations | 506 |  |  |  | Observations | 506 |   Observation:   * Therefore, When comparing R square value for both the questions, Qn.6 R square value is better than Qn.5.   *7) Build another Regression model with all variables where AVG\_PRICE alone be the Dependent*  *Variable and all the other variables are independent. Interpret the output in terms of adjusted*  *R square, coefficient and Intercept values. Explain the significance of each independent variable with*  *respect to AVG\_PRICE.*   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |  | | Intercept | 29.24132 | 4.81713 | 6.07028 | 2.5398E-09 | 19.77682784 | 38.7058 | 19.77683 | 38.7058 |  | | CRIME\_RATE | 0.048725 | 0.07842 | 0.62135 | 0.5346572 | -0.10534854 | 0.2028 | -0.10535 | 0.202799 |  | | AGE | 0.032771 | 0.0131 | 2.502 | 0.01267044 | 0.00703665 | 0.0585 | 0.007037 | 0.058505 |  | | INDUS | 0.130551 | 0.06312 | 2.06839 | 0.03912086 | 0.006541094 | 0.25456 | 0.006541 | 0.254562 |  | | NOX | -10.32118 | 3.89404 | -2.6505 | 0.00829386 | -17.9720228 | -2.67034 | -17.972 | -2.67034 |  | | DISTANCE | 0.261094 | 0.06795 | 3.8426 | 0.00013755 | 0.127594012 | 0.39459 | 0.127594 | 0.394593 |  | | TAX | -0.014401 | 0.00391 | -3.6877 | 0.00025125 | -0.02207388 | -0.00673 | -0.02207 | -0.00673 |  | | PTRATIO | -1.074305 | 0.1336 | -8.0411 | 6.5864E-15 | -1.33680044 | -0.81181 | -1.3368 | -0.81181 |  | | AVG\_ROOM | 4.125409 | 0.44276 | 9.3175 | 3.8929E-19 | 3.255494742 | 4.99532 | 3.255495 | 4.995324 |  | | LSTAT | -0.603487 | 0.05308 | -11.369 | 8.9107E-27 | -0.70777824 | -0.49919 | -0.70778 | -0.49919 |  |  |  |  | | --- | --- | | *Regression Statistics* | | | Multiple R | 0.832978824 | | R Square | 0.69385372 | | Adjusted R Square | 0.688298647 | | Standard Error | 5.1347635 | | Observations | 506 |   Observation:   * Comparing the R square value this model is better than the others. * While Comparing the Coefficients of Intercept, Crime rate, age, indus, distance, avg room are having the   direct relation and others having the inverse relation.   * P value for Crime rate is greater than 0.005.   8*) Pick out only the significant variables from the previous question. Make another instance of the Regression model*  *using only the significant variables you just picked and answer the questions below:*  *a) Interpret the output of this model.*  *b) Compare the adjusted R-square value of this model with the model in the previous question, which*  *model performs better according to the value of adjusted R-square?*  *c) Sort the values of the Coefficients in ascending order. What will happen to the average price if the value of NOX is*  *more in a locality in this town?*  *d) Write the regression equation from this model.*     |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* | | Intercept | 29.42847 | 4.804729 | 6.124898 | 1.85E-09 | 19.98839 | 38.86856 | 19.98839 | 38.86856 | | AGE | 0.032935 | 0.013087 | 2.516606 | 0.012163 | 0.007222 | 0.058648 | 0.007222 | 0.058648 | | INDUS | 0.13071 | 0.063078 | 2.072202 | 0.038762 | 0.006778 | 0.254642 | 0.006778 | 0.254642 | | NOX | -10.2727 | 3.890849 | -2.64022 | 0.008546 | -17.9172 | -2.62816 | -17.9172 | -2.62816 | | DISTANCE | 0.261506 | 0.067902 | 3.851242 | 0.000133 | 0.128096 | 0.394916 | 0.128096 | 0.394916 | | TAX | -0.01445 | 0.003902 | -3.70395 | 0.000236 | -0.02212 | -0.00679 | -0.02212 | -0.00679 | | PTRATIO | -1.0717 | 0.133454 | -8.03053 | 7.08E-15 | -1.33391 | -0.8095 | -1.33391 | -0.8095 | | AVG\_ROOM | 4.125469 | 0.442485 | 9.3234 | 3.69E-19 | 3.256096 | 4.994842 | 3.256096 | 4.994842 | | LSTAT | -0.60516 | 0.05298 | -11.4224 | 5.42E-27 | -0.70925 | -0.50107 | -0.70925 | -0.50107 |   a) Interpretation of the Significant variable regression output   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Regression Statistics |  |  | ANOVA |  |  |  |  |  | | Multiple R | 0.832836 |  |  | df | SS | MS | F | Significance F | | R Square | 0.693615 |  | Regression | 8 | 29628.68 | 3703.585 | 140.643 | 1.911E-122 | | Adjusted R Square | 0.688684 |  | Residual | 497 | 13087.61 | 26.33323 |  |  | | Standard Error | 5.131591 |  | Total | 505 | 42716.3 |  |  |  | | Observations | 506 |  |  |  |  |  |  |  |   Interpretation:   * This regression value is having greater accuracy.   b) Comparing Adjusted R square   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Qtn 5 | |  | |  | | Qtn 6 | |  | |  | |  | |  | |  | |  | | Regression Statistics | |  | |  | | Regression Statistics | |  | | Multiple R | | 0.737662726 | |  | | Multiple R | | 0.799100498 | | R Square | | 0.544146298 | |  | | R Square | | 0.638561606 | | Adjusted R Square | | 0.543241826 | |  | | Adjusted R Square | | 0.637124475 | | Standard Error | | 6.215760405 | |  | | Standard Error | | 5.540257367 | | Observations | | 506 | |  | | Observations | | 506 | | Qtn 7 |  | |  | | Qtn 8 | |  | | | |  |  | |  | |  | |  | | | | Regression Statistics |  | |  | | Regression Statistics | |  | | | | Multiple R | 0.832978824 | |  | | Multiple R | | 0.832835773 | | | | R Square | 0.69385372 | |  | | R Square | | 0.693615426 | | | | Adjusted R Square | 0.688298647 | |  | | Adjusted R Square | | 0.688683682 | | | | Standard Error | 5.1347635 | |  | | Standard Error | | 5.131591113 | | | | Observations | 506 | |  | | Observations | | 506 | | |  * Hence, while comparing the other Adjusted R Square Value this model performs more significant with   Average Price.  c) Sorting Coefficients value in ascending order.   |  |  | | --- | --- | |  | Coefficients | | NOX | -10.2727051 | | PTRATIO | -1.07170247 | | LSTAT | -0.60515928 | | TAX | -0.01445235 | | AGE | 0.03293496 | | INDUS | 0.130710007 | | DISTANCE | 0.261506423 | | AVG\_ROOM | 4.125468959 | | Intercept | 29.42847349 |   In this case,   * Nox is decreasing and the Average price is Increasing. So, if Nox value is more then Average   price gets decreased.    d) Regression Equation:  y= 0.0329\*(x1)+0.130\*(x2)-10.2727\*(x3)+0.261\*(x4)-0.01445\*(x5)-1.0717\*(x5)+4.125\*(x6)-0.60516 |