|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Decorative | | | | |
|  |  |  | |  |
| Terro's Real Estate Agency | | |
| RIYA RAJ | | IMRIYA3099@GMAIL.COM |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Decorative | | | | |
|  |  |  | |  | |
|  | “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. | |  | |

|  |  |
| --- | --- |
|  |  |
| BACKGROUND Terro's real-estate agency utilizes an Auditor to assess house pricing by analyzing key geographic features such as pollution levels (NOX), crime rates, education facilities (pupil to teacher ratio), and connectivity (distance from highway). These factors contribute to the determination of property value, allowing the company to gauge the business worth of a given residence in a specific locality. The Auditor's role is crucial in identifying the most relevant features that influence house pricing within the real estate market. KEYFINDINGS \*\*Key Insights from Data Analysis:\*\*  1. \*\*Proximity to Employment Centers:\*\*  - Houses tend to be closer to employment centers, showing a positively skewed distribution.  2. \*\*Housing Size and Value:\*\*  - More houses have fewer rooms, and there's a concentration of lower median values, reflected in positively skewed distributions.  3. \*\*Price Distribution:\*\*  - Average prices are right-skewed, centered around 50, indicating a category of higher-priced houses. Some outliers have prices lower than the mode.  4. \*\*Variable Relationships:\*\*  - Positive covariance between crime rate and house age suggests a positive relationship.  5. \*\*Property Tax Rates:\*\*  - Left-skewed distribution indicates that more houses have lower property tax rates.  6. \*\*Variability and Accessibility:\*\*  - High variability in accessibility to highways and distance to employment centers, evident from wide ranges and high standard deviations.  7. \*\*Diversity in Housing Characteristics:\*\*  - Dataset reflects diversity in room numbers, median values, and distances, highlighting varied housing features.  8. \*\*Price Variability:\*\*  - Positive skewness, kurtosis, and outliers in average prices suggest significant variability, with some houses deviating notably.  9. \*\*Correlation Insights:\*\*  - While not explicitly discussed, exploring correlations between variables can unveil linear relationships, enhancing detailed analysis.  These concise insights provide a comprehensive overview of the dataset, emphasizing key trends and relationships in housing and neighborhood attributes. | |
|  | |  |
|  | |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Description | Description | Description |

|  |  |
| --- | --- |
|  |  |
| ANS 1  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Mean | 4.871976285 | Mean | 68.57490119 | Mean | 11.13677866 | Mean | 0.554695 | Mean | 9.549407 | Mean | 408.2372 | Mean | 18.45553 | Mean | 6.284634 | Mean | 12.65306 | Mean | 22.53281 | | Standard Error | 0.129860152 | Standard Error | 1.251369525 | Standard Error | 0.304979888 | Standard Error | 0.005151 | Standard Error | 0.387085 | Standard Error | 7.492389 | Standard Error | 0.096244 | Standard Error | 0.031235 | Standard Error | 0.317459 | Standard Error | 0.408861 | | Median | 4.82 | Median | 77.5 | Median | 9.69 | Median | 0.538 | Median | 5 | Median | 330 | Median | 19.05 | Median | 6.2085 | Median | 11.36 | Median | 21.2 | | Mode | 3.43 | Mode | 100 | Mode | 18.1 | Mode | 0.538 | Mode | 24 | Mode | 666 | Mode | 20.2 | Mode | 5.713 | Mode | 8.05 | Mode | 50 | | Standard Deviation | 2.921131892 | Standard Deviation | 28.14886141 | Standard Deviation | 6.860352941 | Standard Deviation | 0.115878 | Standard Deviation | 8.707259 | Standard Deviation | 168.5371 | Standard Deviation | 2.164946 | Standard Deviation | 0.702617 | Standard Deviation | 7.141062 | Standard Deviation | 9.197104 | | Sample Variance | 8.533011532 | Sample Variance | 792.3583985 | Sample Variance | 47.06444247 | Sample Variance | 0.013428 | Sample Variance | 75.81637 | Sample Variance | 28404.76 | Sample Variance | 4.686989 | Sample Variance | 0.493671 | Sample Variance | 50.99476 | Sample Variance | 84.58672 | | Kurtosis | -1.189122464 | Kurtosis | -0.967715594 | Kurtosis | -1.233539601 | Kurtosis | -0.06467 | Kurtosis | -0.86723 | Kurtosis | -1.14241 | Kurtosis | -0.28509 | Kurtosis | 1.8915 | Kurtosis | 0.49324 | Kurtosis | 1.495197 | | Skewness | 0.021728079 | Skewness | -0.59896264 | Skewness | 0.295021568 | Skewness | 0.729308 | Skewness | 1.004815 | Skewness | 0.669956 | Skewness | -0.80232 | Skewness | 0.403612 | Skewness | 0.90646 | Skewness | 1.108098 | | Range | 9.95 | Range | 97.1 | Range | 27.28 | Range | 0.486 | Range | 23 | Range | 524 | Range | 9.4 | Range | 5.219 | Range | 36.24 | Range | 45 | | Minimum | 0.04 | Minimum | 2.9 | Minimum | 0.46 | Minimum | 0.385 | Minimum | 1 | Minimum | 187 | Minimum | 12.6 | Minimum | 3.561 | Minimum | 1.73 | Minimum | 5 | | Maximum | 9.99 | Maximum | 100 | Maximum | 27.74 | Maximum | 0.871 | Maximum | 24 | Maximum | 711 | Maximum | 22 | Maximum | 8.78 | Maximum | 37.97 | Maximum | 50 | | Sum | 2465.22 | Sum | 34698.9 | Sum | 5635.21 | Sum | 280.6757 | Sum | 4832 | Sum | 206568 | Sum | 9338.5 | Sum | 3180.025 | Sum | 6402.45 | Sum | 11401.6 | | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 | Count | 506 |   Certainly! Let's focus on the \*\*Skewness\*\* values to identify which data has heavy skewness. Skewness measures the asymmetry of the distribution. Positive skewness indicates a longer right tail, while negative skewness indicates a longer left tail.  \*\*Insight:\*\*  - \*\*Column 5 (Distance):\*\*  - The skewness value for the "Distance" column is approximately 1.00.  - With a positive skewness, we can infer that the distribution of distances from employment centers has a longer right tail.  - This suggests that there are more houses that are relatively close to employment centers, while there are fewer houses that are farther away. It aligns with your interpretation that some houses are very far from highways and some are near.  \*\*Additional Observations:\*\*  - \*\*Column 8 (Number of Rooms):\*\*  - The skewness value for the "Number of Rooms" column is approximately 0.40.  - The positive skewness suggests a distribution with a longer right tail, indicating that there are more houses with fewer rooms and fewer houses with a higher number of rooms.  - \*\*Column 10 (Median Value of Owner-Occupied Homes):\*\*  - The skewness value for the "Median Value" column is approximately 1.11.  - This positive skewness indicates that the distribution of median values is right-skewed, suggesting that there are more houses with lower median values and fewer houses with higher median values.  \*\*Summary:\*\*  - \*\*Skewness Analysis:\*\*  - Column 5 (Distance) exhibits heavy skewness, indicating a concentration of houses closer to employment centers with a longer right tail.  - Columns 8 and 10 also show positive skewness, suggesting right-skewed distributions in the number of rooms and median values of owner-occupied homes. | |

|  |  |
| --- | --- |
|  |  |
| ANS 2 In summary, the dataset has a right-skewed distribution with a moderate amount of variability. There seems to be a concentration of houses with higher prices (around 50), and the presence of positive skewness and kurtosis indicates some degree of asymmetry and the presence of outliers in the distribution. | |
|  | |
|  | |
| |  |  | | --- | --- | |  |  | | ANS 3 In summary, the covariance matrix provides insights into the relationships and variability between different variables in the dataset. However, for a clearer understanding of the strength and direction of relationships, considering correlation coefficients and exploring the data visually may be helpful.   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *CRIME\_RATE* | *AGE* | *INDUS* | *NOX* | *DISTANCE* | *TAX* | *PTRATIO* | *AVG\_ROOM* | *LSTAT* | *AVG\_PRICE* | | CRIME\_RATE | 8.516147873 |  |  |  |  |  |  |  |  |  | | AGE | 0.562915215 | 790.7924728 |  |  |  |  |  |  |  |  | | INDUS | -0.110215175 | 124.2678282 | 46.97143 |  |  |  |  |  |  |  | | NOX | 0.000625308 | 2.381211931 | 0.605874 | 0.013401 |  |  |  |  |  |  | | DISTANCE | -0.229860488 | 111.5499555 | 35.47971 | 0.61571 | 75.66653 |  |  |  |  |  | | TAX | -8.229322439 | 2397.941723 | 831.7133 | 13.0205 | 1333.117 | 28348.62 |  |  |  |  | | PTRATIO | 0.068168906 | 15.90542545 | 5.680855 | 0.047304 | 8.743402 | 167.8208 | 4.677726 |  |  |  | | AVG\_ROOM | 0.056117778 | -4.74253803 | -1.88423 | -0.02455 | -1.28128 | -34.5151 | -0.53969 | 0.492695 |  |  | | LSTAT | -0.882680362 | 120.8384405 | 29.52181 | 0.48798 | 30.32539 | 653.4206 | 5.7713 | -3.07365 | 50.89398 |  | | AVG\_PRICE | 1.16201224 | -97.39615288 | -30.4605 | -0.45451 | -30.5008 | -724.82 | -10.0907 | 4.484566 | -48.3518 | 84.41956 |   The provided matrix appears to be a covariance matrix, which describes the covariance relationships between different variables. Here are some observations based on the given covariance matrix:  1. \*\*Diagonal Elements:\*\*  - The diagonal elements of the matrix represent the variances of each variable. For example, the variance of CRIME\_RATE is 8.52, the variance of AGE is 790.79, and so on. These values indicate the spread or variability of each individual variable.  2. \*\*Off-Diagonal Elements:\*\*  - The off-diagonal elements represent the covariances between pairs of variables. Covariance measures the degree to which two variables change together. Positive values indicate a positive relationship (as one variable increases, the other tends to increase), and negative values indicate a negative relationship.  3. \*\*Strength of Relationships:\*\*  - The magnitude of the covariance values can give an indication of the strength of the relationships. Larger magnitudes suggest stronger relationships, while values close to zero suggest weaker relationships.   |  |  | | --- | --- | |  |  | | ANS 4 In summary, the dataset has a right-skewed distribution with a moderate amount of variability. There seems to be a concentration of houses with higher prices (around 50), and the presence of positive skewness and kurtosis indicates some degree of asymmetry and the presence of outliers in the distribution.   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *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 | 0.611441 | 1 |  |  |  |  |  | | TAX | -0.016748522 | 0.506455594 | 0.72076018 | 0.668023 | 0.910228 | 1 |  |  |  |  | | PTRATIO | 0.010800586 | 0.261515012 | 0.383247556 | 0.188933 | 0.464741 | 0.460853 | 1 |  |  |  | | AVG\_ROOM | 0.02739616 | -0.24026493 | -0.391675853 | -0.30219 | -0.20985 | -0.29205 | -0.3555 | 1 |  |  | | LSTAT | -0.042398321 | 0.602338529 | 0.603799716 | 0.590879 | 0.488676 | 0.543993 | 0.374044 | -0.61381 | 1 |  | | AVG\_PRICE | 0.043337871 | -0.37695457 | -0.48372516 | -0.42732 | -0.38163 | -0.46854 | -0.50779 | 0.69536 | -0.73766 | 1 | | | | a) \*\*Top 3 Positively Correlated Pairs:\*\*  1. AVG\_ROOM and AVG\_PRICE: The correlation coefficient is 0.695, indicating a strong positive correlation. As the average number of rooms increases, the average price tends to increase.  2. INDUS and TAX: The correlation coefficient is 0.721, suggesting a strong positive correlation. This implies that areas with higher industrialization tend to have higher property tax rates.  3. NOX and INDUS: The correlation coefficient is 0.764, indicating a strong positive correlation. Higher levels of nitric oxides concentration (NOX) are associated with higher industrialization (INDUS).  b) \*\*Top 3 Negatively Correlated Pairs:\*\*  1. AVG\_PRICE and LSTAT: The correlation coefficient is -0.738, indicating a strong negative correlation. As the percentage of lower status of the population (LSTAT) increases, the average price tends to decrease.  2. INDUS and DISTANCE: The correlation coefficient is -0.595, suggesting a moderate negative correlation. Higher industrialization (INDUS) is associated with shorter distances to employment centers (DISTANCE).  3. AVG\_PRICE and INDUS: The correlation coefficient is -0.484, indicating a moderate negative correlation. Higher levels of industrialization (INDUS) are associated with lower average property prices.  It's important to note that correlation does not imply causation, and these observations indicate statistical relationships between variables. | | |  | | | |  |  | | --- | --- | |  |  | | ANS 5 In summary, the dataset has a right-skewed distribution with a moderate amount of variability. There seems to be a concentration of houses with higher prices (around 50), and the presence of positive skewness and kurtosis indicates some degree of asymmetry and the presence of outliers in the distribution.   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *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 | 0.611441 | 1 |  |  |  |  |  | | TAX | -0.016748522 | 0.506455594 | 0.72076018 | 0.668023 | 0.910228 | 1 |  |  |  |  | | PTRATIO | 0.010800586 | 0.261515012 | 0.383247556 | 0.188933 | 0.464741 | 0.460853 | 1 |  |  |  | | AVG\_ROOM | 0.02739616 | -0.24026493 | -0.391675853 | -0.30219 | -0.20985 | -0.29205 | -0.3555 | 1 |  |  | | LSTAT | -0.042398321 | 0.602338529 | 0.603799716 | 0.590879 | 0.488676 | 0.543993 | 0.374044 | -0.61381 | 1 |  | | AVG\_PRICE | 0.043337871 | -0.37695457 | -0.48372516 | -0.42732 | -0.38163 | -0.46854 | -0.50779 | 0.69536 | -0.73766 | 1 | | | | a) \*\*Top 3 Positively Correlated Pairs:\*\*  1. AVG\_ROOM and AVG\_PRICE: The correlation coefficient is 0.695, indicating a strong positive correlation. As the average number of rooms increases, the average price tends to increase.  2. INDUS and TAX: The correlation coefficient is 0.721, suggesting a strong positive correlation. This implies that areas with higher industrialization tend to have higher property tax rates.  3. NOX and INDUS: The correlation coefficient is 0.764, indicating a strong positive correlation. Higher levels of nitric oxides concentration (NOX) are associated with higher industrialization (INDUS).  b) \*\*Top 3 Negatively Correlated Pairs:\*\*  1. AVG\_PRICE and LSTAT: The correlation coefficient is -0.738, indicating a strong negative correlation. As the percentage of lower status of the population (LSTAT) increases, the average price tends to decrease.  2. INDUS and DISTANCE: The correlation coefficient is -0.595, suggesting a moderate negative correlation. Higher industrialization (INDUS) is associated with shorter distances to employment centers (DISTANCE).  3. AVG\_PRICE and INDUS: The correlation coefficient is -0.484, indicating a moderate negative correlation. Higher levels of industrialization (INDUS) are associated with lower average property prices.  It's important to note that correlation does not imply causation, and these observations indicate statistical relationships between variables. | | | **ANS 6**   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | SUMMARY OUTPUT |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  | | *Regression Statistics* | |  |  |  |  |  |  |  | | Multiple R | 0.737662726 |  |  |  |  |  |  |  | | R Square | 0.544146298 |  |  |  |  |  |  |  | | Adjusted R Square | 0.543241826 |  |  |  |  |  |  |  | | Standard Error | 6.215760405 |  |  |  |  |  |  |  | | 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 | 34.55384088 | 0.562627 | 61.41515 | 3.7E-236 | 33.44846 | 35.65922 | 33.44846 | 35.65922 | | LSTAT | -0.950049354 | 0.038733 | -24.5279 | 5.08E-88 | -1.02615 | -0.87395 | -1.02615 | -0.87395 | | | | The regression summary provides important information about the fitted regression model. Let's analyze the key aspects:  1. \*\*Multiple R (Correlation Coefficient):\*\*  - Multiple R is the correlation coefficient between the observed and predicted values. In this case, it is 0.738, indicating a moderate to strong positive correlation between the predictor variable(s) and the response variable (dependent variable).  2. \*\*R Square (Coefficient of Determination):\*\*  - R Square is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). Here, R Square is 0.544, suggesting that approximately 54.4% of the variance in the average price (AVG\_PRICE) is explained by the predictor variable(s).  3. \*\*Adjusted R Square:\*\*  - Adjusted R Square adjusts the R Square value for the number of predictors in the model. It is 0.543, providing a slightly more conservative estimate of the proportion of variance explained.  4. \*\*Standard Error:\*\*  - The standard error is a measure of the variability of the residuals, the differences between observed and predicted values. A lower standard error indicates a better fit. Here, it is 6.216.  5. \*\*ANOVA (Analysis of Variance):\*\*  - The ANOVA table provides information about the overall fit of the model. The F-statistic is 601.62 with a very low p-value (5.08E-88), indicating that the model is statistically significant. There is evidence that at least one predictor variable has a non-zero coefficient.  6. \*\*Coefficients:\*\*  - Intercept: The intercept is 34.55. This is the estimated average price when all predictor variables are zero.  - LSTAT Coefficient: The coefficient for LSTAT is -0.95. This represents the estimated change in the average price for a one-unit change in LSTAT while holding other variables constant.  \*\*Inferences:\*\*  - The model, with LSTAT as a predictor, explains a significant portion of the variance in AVG\_PRICE.  - The negative coefficient for LSTAT suggests that as the percentage of lower status of the population (LSTAT) increases, the average price tends to decrease.  - The intercept represents the average price when all predictors are zero, which might not have a practical interpretation in this context.  - The low p-value in ANOVA suggests that the model is statistically significant.  - The model's predictive ability is moderate to strong, as indicated by the R Square value.  The regression summary provides important information about the fitted regression model. Let's analyze the key aspects:  Multiple R (Correlation Coefficient):  Multiple R is the correlation coefficient between the observed and predicted values. In this case, it is 0.738, indicating a moderate to strong positive correlation between the predictor variable(s) and the response variable (dependent variable).  R Square (Coefficient of Determination):  R Square is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). Here, R Square is 0.544, suggesting that approximately 54.4% of the variance in the average price (AVG\_PRICE) is explained by the predictor variable(s).  Adjusted R Square:  Adjusted R Square adjusts the R Square value for the number of predictors in the model. It is 0.543, providing a slightly more conservative estimate of the proportion of variance explained.  Standard Error:  The standard error is a measure of the variability of the residuals, the differences between observed and predicted values. A lower standard error indicates a better fit. Here, it is 6.216.  ANOVA (Analysis of Variance):  The ANOVA table provides information about the overall fit of the model. The F-statistic is 601.62 with a very low p-value (5.08E-88), indicating that the model is statistically significant. There is evidence that at least one predictor variable has a non-zero coefficient.  Coefficients:  Intercept: The intercept is 34.55. This is the estimated average price when all predictor variables are zero.  LSTAT Coefficient: The coefficient for LSTAT is -0.95. This represents the estimated change in the average price for a one-unit change in LSTAT while holding other variables constant.    ANS 7   |  |  |  | | --- | --- | --- | | SUMMARY OUTPUT |  |  | |  |  |  | | *Regression Statistics* | |  | | Multiple R | 0.832978824 |  | | R Square | 0.69385372 | WITH ALL THE VARIABLES INCLUDING INTERCEPT | | Adjusted R Square | 0.688298647 |  | | Standard Error | 5.1347635 |  | | Observations | 506 |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Regression | 9 | 29638.8605 | 3293.206722 | 124.9045049 | 1.933E-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 | 0% | 19.7768278 | 38.70580267 | 19.77682784 | 38.70580267 |  | | CRIME\_RATE | 0.048725141 | 0.078418647 | 0.621346369 | 53% | -0.1053485 | 0.202798827 | -0.105348544 | 0.202798827 |  | | AGE | 0.032770689 | 0.013097814 | 2.501996817 | 1% | 0.00703665 | 0.058504728 | 0.00703665 | 0.058504728 |  | | INDUS | 0.130551399 | 0.063117334 | 2.068392165 | 4% | 0.00654109 | 0.254561704 | 0.006541094 | 0.254561704 |  | | NOX | -10.3211828 | 3.894036256 | -2.650510195 | 1% | -17.972023 | -2.670342809 | -17.97202279 | -2.670342809 |  | | DISTANCE | 0.261093575 | 0.067947067 | 3.842602576 | 0% | 0.12759401 | 0.394593138 | 0.127594012 | 0.394593138 |  | | TAX | -0.01440119 | 0.003905158 | -3.687736063 | 0% | -0.0220739 | -0.0067285 | -0.022073881 | -0.0067285 |  | | PTRATIO | -1.074305348 | 0.133601722 | -8.041104061 | 0% | -1.3368004 | -0.811810259 | -1.336800438 | -0.811810259 |  | | AVG\_ROOM | 4.125409152 | 0.442758999 | 9.317504929 | 0% | 3.25549474 | 4.995323561 | 3.255494742 | 4.995323561 |  | | LSTAT | -0.603486589 | 0.053081161 | -11.36912937 | 0% | -0.7077782 | -0.499194938 | -0.70777824 | -0.499194938 |  |   ANS 8   |  |  |  | | --- | --- | --- | | SUMMARY OUTPUT | |  | |  |  |  | | *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 | 4.80472862 | 6.124898157 | | | | 1.84597E-09 | | | | 19.98839 | | | 38.8685574 | | | 19.9883896 | | | 38.8685574 | | | | AGE | 0.03293496 | 0.01308705 | 2.516605952 | | | | 0.012162875 | | | | 0.0072222 | | | 0.058647734 | | | 0.00722219 | | | 0.058647734 | | | | INDUS | 0.130710007 | 0.06307782 | 2.072202264 | | | | 0.038761669 | | | | 0.0067779 | | | 0.254642071 | | | 0.00677794 | | | 0.254642071 | | | | NOX | -10.27270508 | 3.89084922 | -2.640221837 | | | | 0.008545718 | | | | -17.917246 | | | -2.628164466 | | | -17.917246 | | | -2.628164466 | | | | DISTANCE | 0.261506423 | 0.06790184 | 3.851242024 | | | | 0.000132887 | | | | 0.1280964 | | | 0.394916471 | | | 0.12809638 | | | 0.394916471 | | | | TAX | -0.014452345 | 0.00390188 | -3.703946406 | | | | 0.000236072 | | | | -0.0221186 | | | -0.006786137 | | | -0.0221186 | | | -0.006786137 | | | | PTRATIO | -1.071702473 | 0.13345353 | -8.030529271 | | | | 7.08251E-15 | | | | -1.3339051 | | | -0.809499836 | | | -1.3339051 | | | -0.809499836 | | | | AVG\_ROOM | 4.125468959 | 0.44248544 | 9.323400461 | | | | 3.68969E-19 | | | | 3.2560963 | | | 4.994841615 | | | 3.2560963 | | | 4.994841615 | | | | LSTAT | -0.605159282 | 0.0529801 | -11.42238841 | | | | 5.41844E-27 | | | | -0.7092519 | | | -0.501066704 | | | -0.7092519 | | | -0.501066704 | | | |  | Coefficients |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | 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 |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | a) Interpret the output of this model. | | |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | Overall Interpretation: | |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | The model, with an adjusted R-square of 0.6887, explains a substantial portion of the variance in AVG\_PRICE. | | | | | | | | | | | | | | | | |  | |  | |  | | |  | |  | |  | | | The signs of the coefficients indicate the direction of the relationship between each independent variable and AVG\_PRICE. | | | | | | | | | | | | | | | | | | | | |  | | |  | |  | |  | | | The p-values suggest that all variables are statistically significant in predicting AVG\_PRICE. | | | | | | | | | | | | | | |  | |  | |  | |  | | |  | |  | |  | | | The model can be used to make predictions for AVG\_PRICE based on the values of the independent variables. | | | | | | | | | | | | | | | | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | 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? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | COE IS NEG |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | d) Write the regression equation from this model. | | | |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | |  |  |  |  |  | | | |  | |  | | |  | |  | |  | |  | |  | | |  | |  | |  | | | Avg\_Price= 0.0329\*Age+0.1307\*Indus-10.2727\*Nox+0.2615\*Distance-0.0144\*Tax-1.0717\*Ptratio+4.1254\*Avg\_Room-0.6051\*LSTAT+29.4284 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | | The adjusted R-square values for the two models are as follows: | | | | | | | | |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Model 1: Previous model | | | | |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Adjusted R Square: 0.688298647 | | | | | |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Model 2: | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Adjusted R Square: 0.688683682 | | | | | |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Comparing these values: | | | | |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Model 1 has an adjusted R-square of approximately 0.6883. | | | | | | | |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Model 2 has a slightly higher adjusted R-square of approximately 0.6887. | | | | | | | | | | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | The adjusted R-square is a measure that considers the number of predictors in a model. It provides an indication of the proportion of the variance in the dependent variable that is predictable from the independent variables. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | In this comparison, Model 2 with the higher adjusted R-square (0.6887) suggests that it explains a slightly larger proportion of the variance in the dependent variable compared to Model 1 (0.6883). However, the difference is relatively small, and both models seem to be similar in terms of their explanatory power. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | | |  |  |  |  |  |  |  | | |  | |  | |  | |  | |  | |  | | |  | |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |   ANS 6   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 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 | 13638.49 | | | 444.3309 | | 7E-112 |  |  | | |  | | | | | | | | | | | | | | | | | | |  | | Residual | 503 | | | | 15439.3092 | 30.69445 | | |  | |  |  |  | | |  | | | | | | | | | | | | | | | | | | |  | | 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 | -0.4281 | | | 0.668765 | | -7.5919 | 4.875355 | -7.5919 | | | 4.875355 | | | | | |  | | | | | | | | | | | | | | | AVG\_ROOM | 5.094787984 | | | | 0.4444655 | 11.46273 | | | 3.47E-27 | | 4.22155 | 5.968026 | 4.22155 | | | 5.968026 | | | | | |  | | | | | | | | | | | | | | | LSTAT | -0.642358334 | | | | 0.043731465 | -14.6887 | | | 6.67E-41 | | -0.72828 | -0.55644 | -0.72828 | | | -0.55644 | | | | | |  | | | | | | | | | | | | | | |  |  | | | |  |  | | |  | |  |  |  | | |  | | | | | | | | | | | | | | | | | | |  | | SUMMARY OUTPUT | | | | | | | | | | | | | | | | |  | | |  | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | | | | | | | | | | | | | | | |  | | |  | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | *Regression Statistics* | | | | | | | | | | | | | |  | | | | | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Multiple R | | | | | | | | | | | | | | | | | 0.973885 |  | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | R Square | | | | | | | | | | | | | | | | | 0.948453 |  | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Adjusted R Square | | | | | | | | | | | | | | | | | 0.946366 |  | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Standard Error | | | | | | | | | | | | | | | | | 5.535767 |  | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Observations | | | | | | | | | | | | | | | | | 506 |  | | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | | | | | | | | | | | | | | | |  | |  | | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | ANOVA | | | | | | |  |  | | | |  | | |  | | | | | |  | | |  | | | | | | | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | | | | | | *df* | *SS* | | | | *MS* | | | *F* | | | | | | *Significance F* | | |  | | | | | | | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Regression | | | | | | | 2 | 284181.4 | | | | 142090.7 | | | 4636.712 | | | | | | 0 | | |  | | | | | | | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Residual | | | | | | | 504 | 15444.93 | | | | 30.64471 | | |  | | | | | |  | | |  | | | | | | | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | Total | | | | | | | 506 | 299626.3 | | | |  | | |  | | | | | |  | | |  | | | | | | | | | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | | | | | | | | | | | | | | | |  | | |  | | | | | | |  | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | *Coefficients* | | | | | | | | *Standard Error* | | | | | | | *t Stat* | | | | | | *P-value* | | | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* | |  | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | | Intercept | | 0 | | | | | | | | #N/A | | | | | | | #N/A | | | | | | #N/A | | | #N/A | #N/A | #N/A | #N/A | |  | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | | AVG\_ROOM | | 4.906906 | | | | | | | | 0.070193 | | | | | | | 69.90558 | | | | | | 1.6E-261 | | | 4.768998 | 5.044814 | 4.768998 | 5.044814 | |  | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | | LSTAT | | -0.65574 | | | | | | | | 0.030559 | | | | | | | -21.4585 | | | | | | 4.81E-73 | | | -0.71578 | -0.5957 | -0.71578 | -0.5957 | |  | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | | Regresion Equation | | | | | | | | | | | | | | | | |  | | |  | | | | | | | PRICE=4.9\*Room-\*LSTAT | | |  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  | | | | | | | | | | | | | | | | |  | | |  | | | | | | | 21.1852 | | | =4.9\*7-0.65574\*20 | | 30K |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |   b) 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 OVERCHARGING | | | | | | |  | | |  | | |  | | | |