Assignment – Terro’s real estate agency

1. Generate the summary statistics for each variable in the table. (Use Data analysis tool pack). Write down your observation.

OBSERVATION:

The mean crime rate is approximately 4.871976, indicating that, on average, the crime rate is close to this value.

- The standard error of the crime rate is 0.12986, suggesting that there is a small amount of variability in the sample means.

- The median crime rate is 4.82, which is very close to the mean, indicating that the data is somewhat normally distributed.

- The mode of the crime rate is 3.43, implying that this value occurs most frequently in the data set.

- The standard deviation of the crime rate is 2.921132, indicating a relatively high amount of variability in the data.

- The sample variance of the crime rate is 8.533012, providing an estimate of the spread of the data.

- The kurtosis of the crime rate is -1.18912, suggesting that the data is moderately platykurtic, meaning it has thinner tails than a normal distribution.

- The skewness of the crime rate is 0.021728, indicating a slight right-skewness in the data distribution.

- The range of the crime rate is 9.95, which represents

|  |  |
| --- | --- |
| *CRIME\_RATE* |  |
|  |  |
| Mean | 4.871976 |
| Standard Error | 0.12986 |
| Median | 4.82 |
| Mode | 3.43 |
| Standard Deviation | 2.921132 |
| Sample Variance | 8.533012 |
| Kurtosis | -1.18912 |
| Skewness | 0.021728 |
| Range | 9.95 |
| Minimum | 0.04 |
| Maximum | 9.99 |
| Sum | 2465.22 |
| Count | 506 |

The mean age is approximately 68.57490119, indicating that the average age in the sample is around this value.

- The standard error of the age is 1.251369525, suggesting that there is a small amount of variability in the sample means.

- The median age is 77.5, which is higher than the mean, indicating that the data may be slightly negatively skewed.

- The mode of the age is 100, implying that this age occurs most frequently in the dataset.

- The standard deviation of the age is 28.14886141, indicating a relatively high amount of variability in the data.

- The sample variance of the age is 792.3583985, providing an estimate of the spread of the data.

- The kurtosis of the age is -0.967715594, suggesting that the data is slightly platykurtic, meaning it has thinner tails than a normal distribution.

- The skewness of the age is -0.59896264, indicating a slight negative skewness in the data distribution.

- The range of the age is 97.1, which represents the difference between the maximum and minimum values in the dataset

|  |  |
| --- | --- |
| *AGE* |  |
|  |  |
| Mean | 68.5749 |
| Standard Error | 1.25137 |
| Median | 77.5 |
| Mode | 100 |
| Standard Deviation | 28.14886 |
| Sample Variance | 792.3584 |
| Kurtosis | -0.96772 |
| Skewness | -0.59896 |
| Range | 97.1 |
| Minimum | 2.9 |
| Maximum | 100 |
| Sum | 34698.9 |
| Count | 506 |

The mean INDUS value is approximately 11.13677866, indicating that, on average, the proportion of non-retail business acres per town is around this value.

- The standard error of the INDUS is 0.304979888, suggesting that there is a small amount of variability in the sample means.

- The median INDUS value is 9.69, which is slightly lower than the mean, indicating a slightly right-skewed distribution.

- The mode of the INDUS is 18.1, implying that this value occurs most frequently in the dataset.

- The standard deviation of INDUS is 6.860352941, indicating a moderate amount of variability in the data.

- The sample variance of the INDUS is 47.06444247, providing an estimate of the spread of the data.

- The kurtosis of INDUS is -1.233539601, suggesting that the data has a platykurtic distribution with thinner tails than a normal distribution.

- The skewness of INDUS is 0.295021568, indicating a slightly right-skewed distribution.

|  |  |
| --- | --- |
| *INDUS* |  |
|  |  |
| Mean | 11.13678 |
| Standard Error | 0.30498 |
| Median | 9.69 |
| Mode | 18.1 |
| Standard Deviation | 6.860353 |
| Sample Variance | 47.06444 |
| Kurtosis | -1.23354 |
| Skewness | 0.295022 |
| Range | 27.28 |
| Minimum | 0.46 |
| Maximum | 27.74 |
| Sum | 5635.21 |
| Count | 506 |

The mean NOX value is approximately 0.554695059, indicating that, on average, the concentration of nitric oxides is around this value in the dataset.

- The standard error of the NOX is 0.005151391, suggesting that there is a small amount of variability in the sample means.

- The median NOX value is 0.538, which is slightly lower than the mean, indicating a slightly left-skewed distribution.

- The mode of the NOX is also 0.538, implying that this value occurs most frequently in the dataset.

- The standard deviation of NOX is 0.115877676, indicating a moderate amount of variability in the data.

- The sample variance of the NOX is 0.013427636, providing an estimate of the spread of the data.

- The kurtosis of NOX is -0.064667133, suggesting that the data has a mesokurtic distribution with approximately normal tails.

- The skewness of NOX is 0.729307923, indicating a slightly right-skewed distribution.

- The range of NOX is 0

|  |  |
| --- | --- |
| *NOX* |  |
|  |  |
| Mean | 0.554695 |
| Standard Error | 0.005151 |
| Median | 0.538 |
| Mode | 0.538 |
| Standard Deviation | 0.115878 |
| Sample Variance | 0.013428 |
| Kurtosis | -0.06467 |
| Skewness | 0.729308 |
| Range | 0.486 |
| Minimum | 0.385 |
| Maximum | 0.871 |
| Sum | 280.6757 |
| Count | 506 |

The mean distance is approximately 9.549407115, indicating that the average distance is around this value.

- The standard error of the distance is 0.387084894, suggesting that there is a small amount of variability in the sample means.

- The median distance is 5, which is significantly lower than the mean, indicating a positively skewed distribution.

- The mode of the distance is 24, implying that this value occurs most frequently in the dataset.

- The standard deviation of the distance is 8.707259384, indicating a relatively high amount of variability in the data.

- The sample variance of the distance is 75.81636598, providing an estimate of the spread of the data.

- The kurtosis of the distance is -0.867231994, suggesting that the data has a mesokurtic distribution with approximately normal tails.

- The skewness of the distance is 1.004814648, indicating a significant right-skewness in the data distribution.

- The range of the distance is 23, which represents the difference between the maximum and minimum values in the dataset.

|  |  |
| --- | --- |
| *DISTANCE* |  |
|  |  |
| Mean | 9.549407 |
| Standard Error | 0.387085 |
| Median | 5 |
| Mode | 24 |
| Standard Deviation | 8.707259 |
| Sample Variance | 75.81637 |
| Kurtosis | -0.86723 |
| Skewness | 1.004815 |
| Range | 23 |
| Minimum | 1 |
| Maximum | 24 |
| Sum | 4832 |
| Count | 506 |

The mean tax rate is approximately 408.2371542, indicating that, on average, the property-tax rate is around this value per $10,000.

- The standard error of the tax rate is 7.492388692, suggesting that there is a small amount of variability in the sample means.

- The median tax rate is 330, which is significantly lower than the mean, indicating a positively skewed distribution.

- The mode of the tax rate is 666, implying that this value occurs most frequently in the dataset.

- The standard deviation of the tax rate is 168.5371161, indicating a relatively high amount of variability in the data.

- The sample variance of the tax rate is 28404.75949, providing an estimate of the spread of the data.

- The kurtosis of the tax rate is -1.142407992, suggesting that the data has a platykurtic distribution with thinner tails than a normal distribution.

- The skewness of the tax rate is 0.669955942, indicating a right-skewed distribution.

- The range of the tax rate is

|  |  |
| --- | --- |
| *TAX* |  |
|  |  |
| Mean | 408.2372 |
| Standard Error | 7.492389 |
| Median | 330 |
| Mode | 666 |
| Standard Deviation | 168.5371 |
| Sample Variance | 28404.76 |
| Kurtosis | -1.14241 |
| Skewness | 0.669956 |
| Range | 524 |
| Minimum | 187 |
| Maximum | 711 |
| Sum | 206568 |
| Count | 506 |

The mean pupil-teacher ratio is approximately 18.4555336, indicating that, on average, there are around 18.46 students per teacher in the town.

- The standard error of the pupil-teacher ratio is 0.096243568, suggesting that there is a small amount of variability in the sample means.

- The median pupil-teacher ratio is 19.05, which is slightly higher than the mean, indicating a slightly left-skewed distribution.

- The mode of the pupil-teacher ratio is 20.2, implying that this value occurs most frequently in the dataset.

- The standard deviation of the pupil-teacher ratio is 2.164945524, indicating a moderate amount of variability in the data.

- The sample variance of the pupil-teacher ratio is 4.686989121, providing an estimate of the spread of the data.

- The kurtosis of the pupil-teacher ratio is -0.285091383, suggesting that the data has a relatively normal shape with a close-to-normal tail distribution.

- The skewness of the pupil-teacher ratio is -0.802324927

|  |  |
| --- | --- |
| *PTRATIO* |  |
|  |  |
| Mean | 18.45553 |
| Standard Error | 0.096244 |
| Median | 19.05 |
| Mode | 20.2 |
| Standard Deviation | 2.164946 |
| Sample Variance | 4.686989 |
| Kurtosis | -0.28509 |
| Skewness | -0.80232 |
| Range | 9.4 |
| Minimum | 12.6 |
| Maximum | 22 |
| Sum | 9338.5 |
| Count | 506 |

The mean number of rooms is approximately 6.284634387, indicating that, on average, there are around 6.28 rooms per dwelling.

- The standard error of the average number of rooms is 0.031235142, suggesting that there is a small amount of variability in the sample means.

- The median number of rooms is 6.2085, which is slightly lower than the mean, indicating a slightly right-skewed distribution.

- The mode of the average number of rooms is 5.713, implying that this value occurs most frequently in the dataset.

- The standard deviation of the average number of rooms is 0.702617143, indicating a relatively low amount of variability in the data.

- The sample variance of the average number of rooms is 0.49367085, providing an estimate of the spread of the data.

- The kurtosis of the average number of rooms is 1.891500366, suggesting that the data has a leptokurtic distribution with heavier tails compared to a normal distribution.

- The skewness of the average number of rooms is 0.403612133,

|  |  |
| --- | --- |
| *AVG\_ROOM* |  |
|  |  |
| Mean | 6.284634 |
| Standard Error | 0.031235 |
| Median | 6.2085 |
| Mode | 5.713 |
| Standard Deviation | 0.702617 |
| Sample Variance | 0.493671 |
| Kurtosis | 1.8915 |
| Skewness | 0.403612 |
| Range | 5.219 |
| Minimum | 3.561 |
| Maximum | 8.78 |
| Sum | 3180.025 |
| Count | 506 |

The mean percentage of lower status population is approximately 12.65306324, indicating that, on average, around 12.65% of the population falls under the lower status category.

- The standard error of the percentage of lower status population is 0.317458906, suggesting that there is a small amount of variability in the sample means.

- The median percentage of lower status population is 11.36, which is slightly lower than the mean, indicating a slightly left-skewed distribution.

- The mode of the percentage of lower status population is 8.05, implying that this value occurs most frequently in the dataset.

- The standard deviation of the percentage of lower status population is 7.141061511, indicating a relatively high amount of variability in the data.

- The sample variance of the percentage of lower status population is 50.99475951, providing an estimate of the spread of the data.

- The kurtosis of the percentage of lower status population is 0.493239517, suggesting that the data has a slightly leptokurtic distribution compared to a normal distribution.

- The skewness of the

|  |  |
| --- | --- |
| *LSTAT* |  |
|  |  |
| Mean | 12.65306 |
| Standard Error | 0.317459 |
| Median | 11.36 |
| Mode | 8.05 |
| Standard Deviation | 7.141062 |
| Sample Variance | 50.99476 |
| Kurtosis | 0.49324 |
| Skewness | 0.90646 |
| Range | 36.24 |
| Minimum | 1.73 |
| Maximum | 37.97 |
| Sum | 6402.45 |
| Count | 506 |

The mean average price of homes is approximately 22.53280632, indicating that, on average, the price of homes is around $22,533.

- The standard error of the average price of homes is 0.408861147, suggesting that there is a small amount of variability in the sample means.

- The median average price of homes is 21.2, which is slightly lower than the mean, indicating a slightly left-skewed distribution.

- The mode of the average price of homes is 50, implying that this value occurs most frequently in the dataset.

- The standard deviation of the average price of homes is 9.197104087, indicating a relatively high amount of variability in the data.

- The sample variance of the average price of homes is 84.58672359, providing an estimate of the spread of the data.

- The kurtosis of the average price of homes is 1.495196944, suggesting that the data has a leptokurtic distribution with heavier tails compared to a normal distribution.

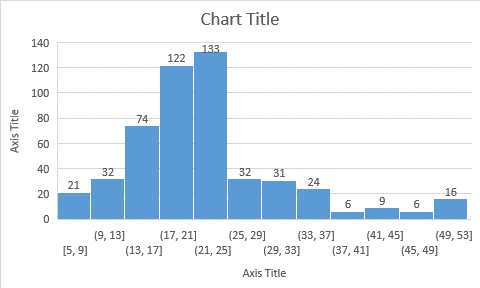
- The skewness of the average price of homes is 1.108098408, indicating a significant

|  |  |
| --- | --- |
| *AVG\_PRICE* |  |
|  |  |
| Mean | 22.53281 |
| Standard Error | 0.408861 |
| Median | 21.2 |
| Mode | 50 |
| Standard Deviation | 9.197104 |
| Sample Variance | 84.58672 |
| Kurtosis | 1.495197 |
| Skewness | 1.108098 |
| Range | 45 |
| Minimum | 5 |
| Maximum | 50 |
| Sum | 11401.6 |
| Count | 506 |

1. Plot a histogram of the Avg\_Price variable. What do you infer?

OBSERVATION:

HISTOGRAM AVERAGE PRICE (21,25)



1. Compute the covariance matrix. Share your observations.

OBSERVATION:

1. POSITIVE CORRELATION BETWEEN AGE AND TAX
2. NEGATIVE CORRELATION BETWEEN AVERAGE PRICE AND TAX

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *CRIME\_RATE* | *AGE* | *INDUS* | *NOX* | *DISTANCE* | *TAX* | *PTRATIO* | *AVG\_ROOM* |
| CRIME\_RATE | 8.516148 |  |  |  |  |  |  |  |
| AGE | 0.562915 | 790.7925 |  |  |  |  |  |  |
| INDUS | -0.11022 | 124.2678 | 46.97143 |  |  |  |  |  |
| NOX | 0.000625 | 2.381212 | 0.605874 | 0.013401 |  |  |  |  |
| DISTANCE | -0.22986 | 111.55 | 35.47971 | 0.61571 | 75.66653 |  |  |  |
| TAX | -8.22932 | 2397.942 | 831.7133 | 13.0205 | 1333.117 | 28348.62 |  |  |
| PTRATIO | 0.068169 | 15.90543 | 5.680855 | 0.047304 | 8.743402 | 167.8208 | 4.677726 |  |
| AVG\_ROOM | 0.056118 | -4.74254 | -1.88423 | -0.02455 | -1.28128 | -34.5151 | -0.53969 | 0.492695 |
| LSTAT | -0.88268 | 120.8384 | 29.52181 | 0.48798 | 30.32539 | 653.4206 | 5.7713 | -3.07365 |
| AVG\_PRICE | 1.162012 | -97.3962 | -30.4605 | -0.45451 | -30.5008 | -724.82 | -10.0907 | 4.484566 |

|  |  |
| --- | --- |
| *LSTAT* | *AVG\_PRICE* |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| 50.89398 |  |
| -48.3518 | 84.41956 |

4) Create a correlation matrix of all the variables (Use Data analysis tool pack). (5 marks)

a) Which are the top 3 positively correlated pairs and

b) Which are the top 3 negatively correlated pairs.

OBSERVATION:

1. the top 3 positively correlated pairs

|  |
| --- |
| AGE IS 0.73147 |
| INDUS IS 0.763651 |
| DISTANCE IS 0.910228   1. top 3 negatively correlated pairs.  |  | | --- | | LSTAT IS 0.73766 | | AVGBROOM IS 0.61381 | | PTRATIO IS 0.50779 | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *CRIME\_RATE* | *AGE* | *INDUS* | *NOX* | *DISTANCE* | *TAX* | *PTRATIO* | *AVG\_ROOM* | *LSTAT* | *AVG\_PRICE* |
| CRIME\_RATE | 1 |  |  |  |  |  |  |  |  |  |
| AGE | 0.006859463 | 1 |  |  |  |  |  |  |  |  |
| INDUS | -0.005510651 | 0.644779 | 1 |  |  |  |  |  |  |  |
| NOX | 0.001850982 | 0.73147 | 0.763651 | 1 |  |  |  |  |  |  |
| DISTANCE | -0.009055049 | 0.456022 | 0.595129 | 0.611441 | 1 |  |  |  |  |  |
| TAX | -0.016748522 | 0.506456 | 0.72076 | 0.668023 | 0.910228 | 1 |  |  |  |  |
| PTRATIO | 0.010800586 | 0.261515 | 0.383248 | 0.188933 | 0.464741 | 0.460853 | 1 |  |  |  |
| AVG\_ROOM | 0.02739616 | -0.24026 | -0.39168 | -0.30219 | -0.20985 | -0.29205 | -0.3555 | 1 |  |  |
| LSTAT | -0.042398321 | 0.602339 | 0.6038 | 0.590879 | 0.488676 | 0.543993 | 0.374044 | -0.61381 | 1 |  |
| AVG\_PRICE | 0.043337871 | -0.37695 | -0.48373 | -0.42732 | -0.38163 | -0.46854 | -0.50779 | 0.69536 | -0.73766 | 1 |

1. Build an initial regression model with AVG\_PRICE as ‘y’ (Dependent variable) and LSTAT variable as Independent Variable. Generate the residual plot. (8 marks)

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?

|  |  |
| --- | --- |
| *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 | 35.65922 |
| LSTAT | -0.95005 | 0.038733 | -24.5279 | 5.08E-88 | -1.02615 | -0.87395 | -1.02615 | -0.87395 |

1. Build a new Regression model including LSTAT and AVG\_ROOM together as Independent variables and AVG\_PRICE as dependent variable. (6 marks)

a) 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?

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.

OBSERVATION:

REGRESSION EQUATION AVERAGE AND LSAT

=$P$16+$P$17\*H2+$P$18\*I2

Avg-Price is 21460 USD and company quoting a value of 30000 USDcompany is overcharging household

R-square 0.638562 of this model better than the previous model

|  |
| --- |
|  |
|  |

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.7991 |
| 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 | -0.4281 | 0.668765 | -7.5919 | 4.875355 | -7.5919 | 4.875355 |
| AVG\_ROOM | 5.094788 | 0.444466 | 11.46273 | 3.47E-27 | 4.22155 | 5.968026 | 4.22155 | 5.968026 |
| LSTAT | -0.64236 | 0.043731 | -14.6887 | 6.67E-41 | -0.72828 | -0.55644 | -0.72828 | -0.55644 |

1. 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 Rsquare, coefficient and Intercept values. Explain the significance of each independent variable with respect to AVG\_PRICE.

OBSERVATION:

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.832979 |
| R Square | 0.693854 |
| Adjusted R Square | 0.688299 |
| Standard Error | 5.134764 |
| Observations | 506 |

|  |  |
| --- | --- |
| R Square | 0.693854 |

Is positive model

Coefficient is positive increase mean is vice verse

Intercept p value 0.05 lesser is positive

0.05 greater is negative

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 9 | 29638.86 | 3293.207 | 124.9045 | 1.9E-121 |
| Residual | 496 | 13077.43 | 26.3658 |  |  |
| Total | 505 | 42716.3 |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 29.24132 | 4.817126 | 6.070283 | 2.54E-09 | 19.77683 | 38.7058 | 19.77683 | 38.7058 |
| CRIME\_RATE | 0.048725 | 0.078419 | 0.621346 | 0.534657 | -0.10535 | 0.202799 | -0.10535 | 0.202799 |
| AGE | 0.032771 | 0.013098 | 2.501997 | 0.01267 | 0.007037 | 0.058505 | 0.007037 | 0.058505 |
| INDUS | 0.130551 | 0.063117 | 2.068392 | 0.039121 | 0.006541 | 0.254562 | 0.006541 | 0.254562 |
| NOX | -10.3212 | 3.894036 | -2.65051 | 0.008294 | -17.972 | -2.67034 | -17.972 | -2.67034 |
| DISTANCE | 0.261094 | 0.067947 | 3.842603 | 0.000138 | 0.127594 | 0.394593 | 0.127594 | 0.394593 |
| TAX | -0.0144 | 0.003905 | -3.68774 | 0.000251 | -0.02207 | -0.00673 | -0.02207 | -0.00673 |
| PTRATIO | -1.07431 | 0.133602 | -8.0411 | 6.59E-15 | -1.3368 | -0.81181 | -1.3368 | -0.81181 |
| AVG\_ROOM | 4.125409 | 0.442759 | 9.317505 | 3.89E-19 | 3.255495 | 4.995324 | 3.255495 | 4.995324 |
| LSTAT | -0.60349 | 0.053081 | -11.3691 | 8.91E-27 | -0.70778 | -0.49919 | -0.70778 | -0.49919 |

1. 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

OBSERVATION:

2) R-Square value of this model with previous model better then previous one because R-Square of this model is higher than previous model R-Square

3) Sort the values of the Coefficients in ascending order

Nox is negative it increase then price down

|  |  |
| --- | --- |
| NOX | -10.2727 |
| PTRATIO | -1.0717 |
| LSTAT | -0.60516 |
| TAX | -0.01445 |
| AGE | 0.032935 |
| INDUS | 0.13071 |
| DISTANCE | 0.261506 |
| AVG\_ROOM | 4.125469 |
| Intercept | 29.42847 |

4)REGRESSION EQUATION:

=$P$4+$P$5\*B2+$P$6\*C2+$P$7\*D2+$P$8\*E2+$P$9\*F2+$P$10\*G2+$P$11\*H2+$P$12\*I2

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.832836 |
| R Square | 0.693615 |
| Adjusted R Square | 0.688684 |
| Standard Error | 5.131591 |
| Observations | 506 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 8 | 29628.68 | 3703.585 | 140.643 | 1.9E-122 |
| Residual | 497 | 13087.61 | 26.33323 |  |  |
| Total | 505 | 42716.3 |  |  |  |