The purpose of this study is to investigate the factors that influence the price of a home. The dependent variable (y) in this study is the home price, measured in dollars. The independent variables (x) include lot size, age of the home, living area, number of bedrooms, number of bathrooms, and number of rooms, value of land, whether the home has a fireplace, whether the property is on a waterfront, whether the home is a new construction, and whether the home has central air conditioning.

Lot size, living area, number of bedrooms, number of bathrooms, and number of rooms are commonly known to have a positive effect on home price, while the age of the home is expected to have a negative effect. The value of land is also expected to have a positive effect on home price, as it reflects the location and desirability of the property.

The presence of a fireplace, being on a waterfront, being a new construction, and having central air conditioning are expected to have varying effects on home price. For example, having a fireplace may increase the home price for some buyers who value this feature, while it may not have any effect on others. Being on a waterfront is expected to increase the home price for most buyers, as waterfront properties are generally considered more desirable. Being a new construction may also increase the home price, as buyers may be willing to pay a premium for a newly built home. Having central air conditioning is also expected to have a positive effect on home price, as it is a desirable feature for many homebuyers.

Overall, this study aims to identify the relative importance of each independent variable in explaining the variation in home prices.

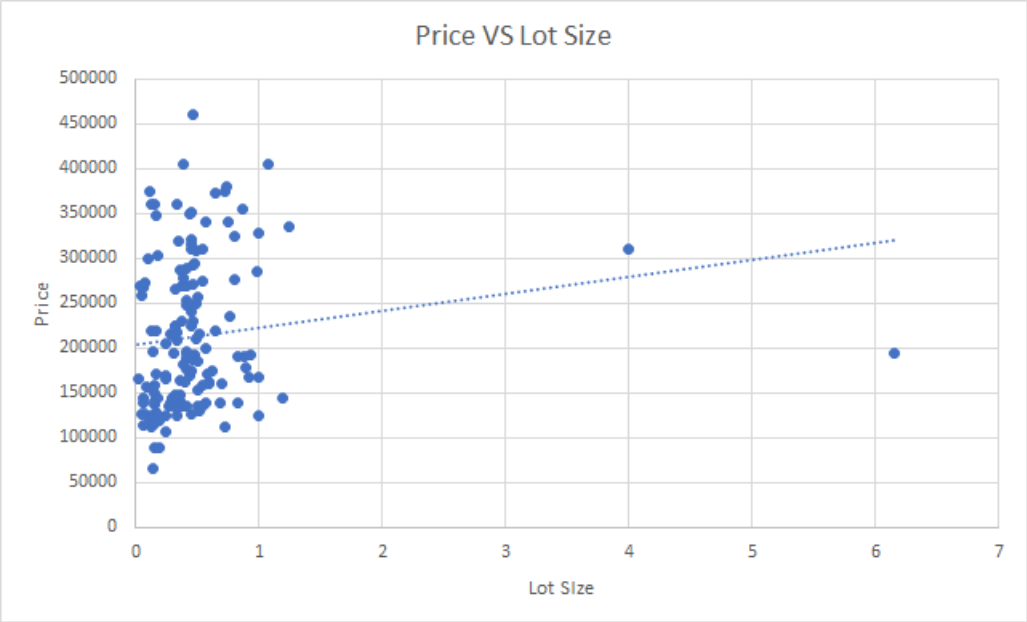
| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Price* | | *Lot.Size* | | *Age* | | *Land.Value* | | *Living.Area* | | *Bedrooms* | | *Bathrooms* | | *Rooms* | |
| 2 | | | | | | | | | | | | | | | | |
| 3 | Mean | 212209.3 | Mean | 0.479467 | Mean | 27.61333 | Mean | 32095.33 | Mean | 1786.12 | Mean | 3.306667 | Mean | 1.893333 | Mean | 6.88 |
| 4 | Standard Error | 6743.438 | Standard Error | 0.049564 | Standard E | 1.853826 | Standard E | 2531.92 | Standard E | 47.37759 | Standard E | 0.06342 | Standard E | 0.052581 | Standard E | 0.109471 |
| 5 | Median | 191250 | Median | 0.41 | Median | 22 | Median | 24150 | Median | 1668.5 | Median | 3 | Median | 2 | Median | 7 |
| 6 | Mode | 125000 | Mode | 0.46 | Mode | 33 | Mode | 15100 | Mode | 1480 | Mode | 3 | Mode | 2.5 | Mode | 6 |
| 7 | Standard Deviation | 82589.92 | Standard Deviation | 0.607037 | Standard D | 22.70464 | Standard D | 31009.56 | Standard D | 580.2546 | Standard D | 0.776731 | Standard D | 0.643988 | Standard D | 1.34074 |
| 8 | Sample Variance | 6.82E+09 | Sample Variance | 0.368494 | Sample Va | 515.5005 | Sample Va | 9.62E+08 | Sample Va | 336695.4 | Sample Va | 0.603311 | Sample Va | 0.41472 | Sample Va | 1.797584 |
| 9 | Kurtosis | -0.39999 | Kurtosis | 58.14796 | Kurtosis | 9.446472 | Kurtosis | 8.879589 | Kurtosis | -0.57617 | Kurtosis | 0.253421 | Kurtosis | -1.10954 | Kurtosis | -0.71783 |
| 10 | Skewness | 0.694625 | Skewness | 6.879673 | Skewness | 2.635526 | Skewness | 2.573547 | Skewness | 0.420755 | Skewness | 0.275389 | Skewness | 0.024346 | Skewness | 0.476077 |
| 11 | Range | 394000 | Range | 6.13 | Range | 138 | Range | 193700 | Range | 2568 | Range | 4 | Range | 2.5 | Range | 5 |
| 12 | Minimum | 66000 | Minimum | 0.02 | Minimum | 0 | Minimum | 1300 | Minimum | 768 | Minimum | 2 | Minimum | 1 | Minimum | 5 |
| 13 | Maximum | 460000 | Maximum | 6.15 | Maximum | 138 | Maximum | 195000 | Maximum | 3336 | Maximum | 6 | Maximum | 3.5 | Maximum | 10 |
| 14 | Sum | 31831402 | Sum | 71.92 | Sum | 4142 | Sum | 4814300 | Sum | 267918 | Sum | 496 | Sum | 284 | Sum | 1032 |
| 15 | Count | 150 | Count | 150 | Count | 150 | Count | 150 | Count | 150 | Count | 150 | Count | 150 | Count | 150 |
| 16 | | | | | | | | | | | | | | | | |

The average price of a home in the sample is $212,209.35, with a standard deviation of $82,589.92. The dataset contains a wide price range, with the highest price of a property being $460,000 and the lowest price being $66,000. The average lot size is 0.48 acres, with a standard deviation of 0.61 acres. The largest lot size available is 6.15 acres, while the smallest is 0.02 acres. Surprisingly, the sample's median lot size is 0.46 acres, indicating that this is the most common lot size. The average age of the houses is 27.61 years, with a standard deviation of 22.70 years. In the dataset, the maximum age of a house is 138 years, while the shortest age is 0. Given the huge disparity between the maximum and mean ages of the residences, we can see that the dataset contains a few outliers. The dwellings' average living space is 3,209.53 square feet, with a standard deviation of 31,009.56 square feet. A house in the dataset has a maximum living area of 195,000 square feet and a minimum living area of 1,300 square feet. The average number of bedrooms is 3.31, with a standard deviation of 0.77. The dataset contains a maximum of six bedrooms and a minimum of two bedrooms. Similarly, the average number of bathrooms is 1.89, with a standard deviation of 0.64 and a maximum of 3.5.

Finally, the average number of rooms in the houses is 6.88, with a 1.34 standard deviation. In the dataset, the greatest number of rooms in a house is 10, while the minimum number of rooms is 5.

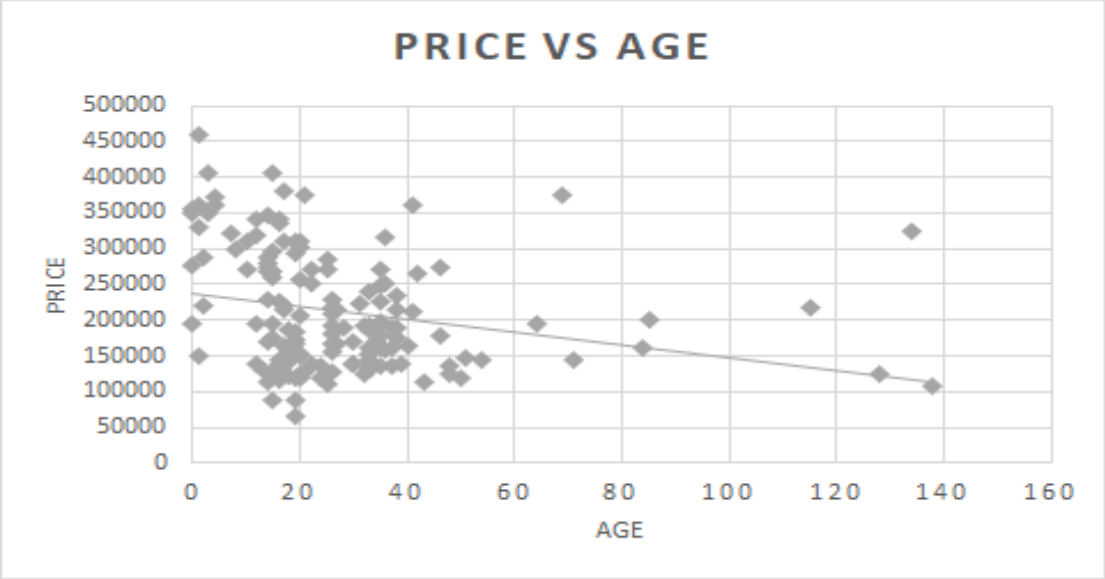
In Conclusion, we can see that the dataset contains a wide range of prices, lot sizes, living areas, and house ages. According to descriptive statistics, the most common lot size is 0.46 acres, and the most common number of rooms is seven. Based on these findings, it may be worthwhile to do additional research to discover the relationship between the various variables and property prices.

PRICE VS LOT SIZE:



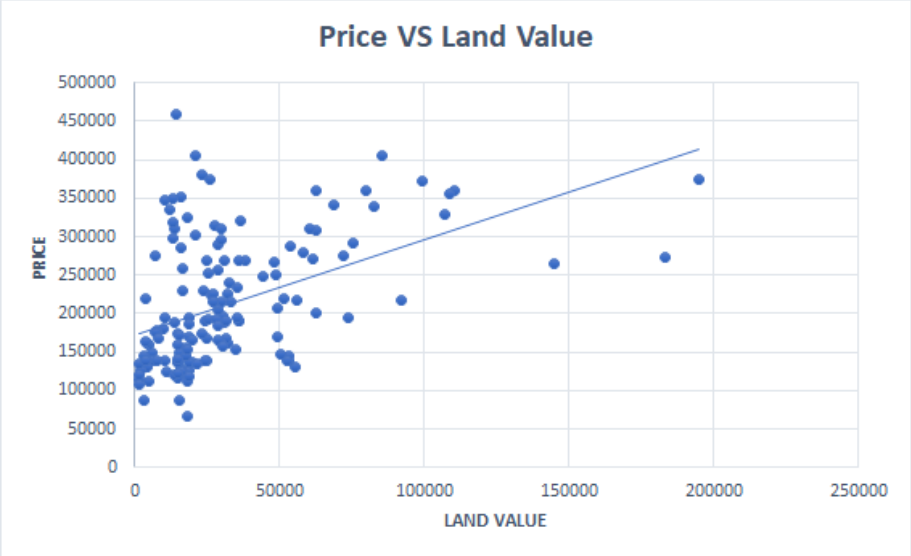

This relationship is non-linear and as the Lot Size increases the Price also increases as visible by the trend line as larger lots are more valuable hence greater price.

PRICE VS AGE:




This relationship is non-linear and as the Age increases the Price decreases as visible by the trend line as older houses require more maintenance and repair so people don't prefer them hence their prices are lower.

PRICE VS LAND VALUE:




Based on the scatter plot of Price vs. Land Value, the relationship appears to be somewhat linear, with a positive correlation between the two variables. As the value of land increases, so does the price of the property

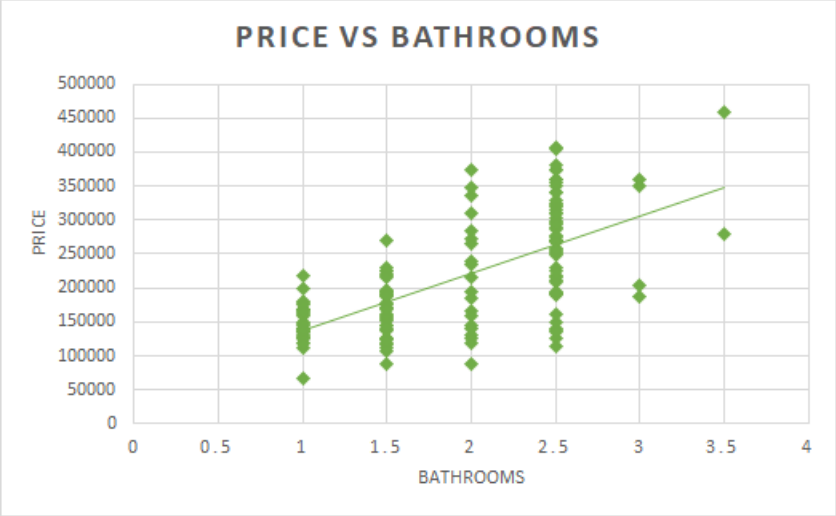
PRICE VS LIVING AREA:


Based on the scatter plot of Price vs. Land Value, the relationship appears to be somewhat linear, with a positive correlation between the two variables and as the living area increases the price also increases as larger homes are more valuable.

PRICE VS BEDROOMS:


The relationship between Price vs Bedrooms is linear as the number of bedrooms increases so does the price as house with more bedrooms are more valuable.

PRICE VS BATHROOMS:

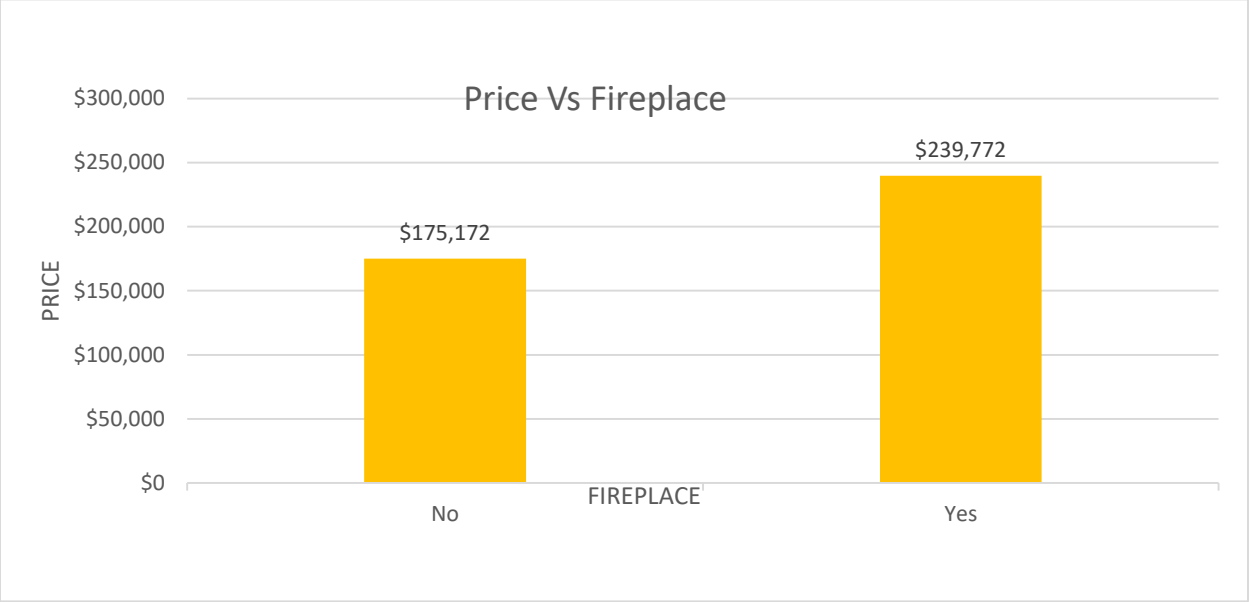



The relationship between Price vs Bathrooms is linear as the number of bathrooms increases so does the price as house with more bathrooms are more valuable.

PRICE VS ROOMS:

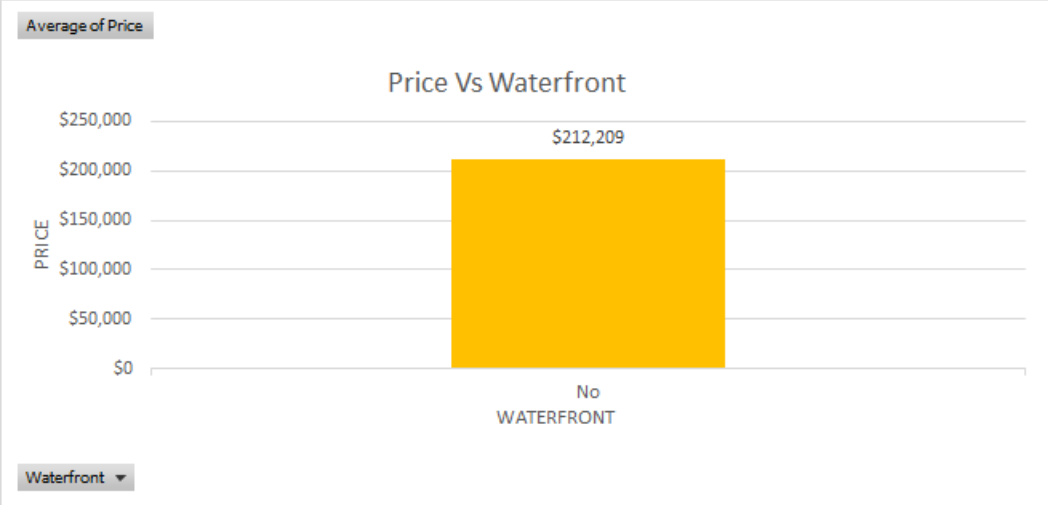

The relationship between Price vs rooms is linear as the number of rooms increases so does the price as house with more rooms are more larger and hence more valuable.

PRICE VS FIREPLACE:



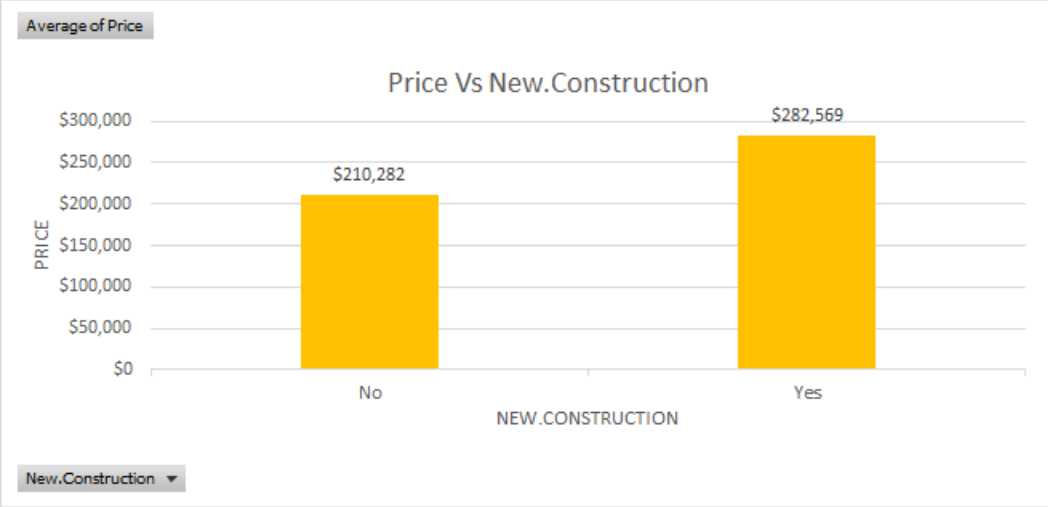

The relationship is not linear but if a fireplace is present then the average price is much higher when the fireplace is not present as it is an attractive feature which helps in the price to be increased.

PRICE VS WATERFRONT:



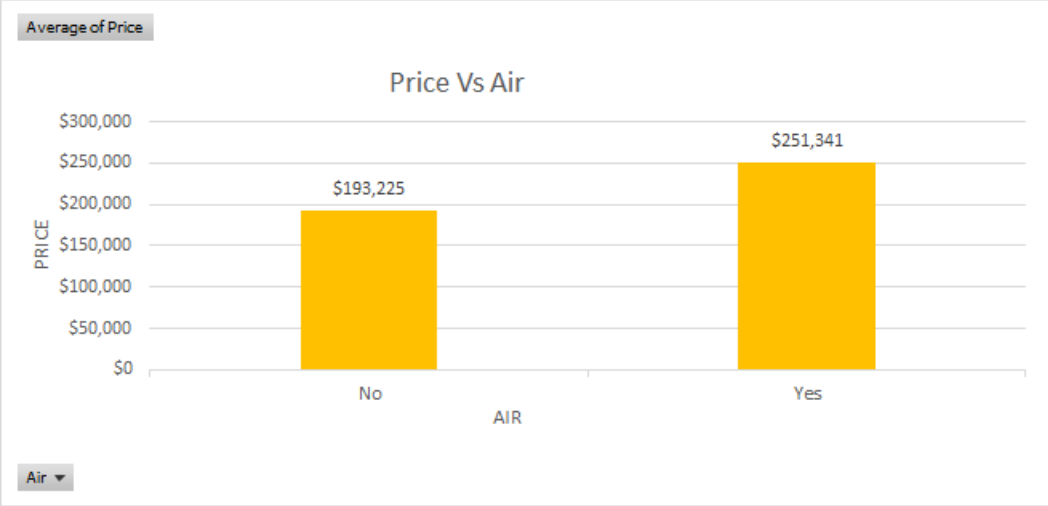

The relationship is not present as all the houses do not have any waterfront so there is no relationship among them.

PRICE VS NEW.CONSTRUCTION:




The relationship is not linear and if the house is new construction it is more expensive than the one which is not a new construction as new homes are more desirable than older homes.

PRICE VS AIR:




The relationship is non-linear and if Air conditioning is present the house is more expensive than if Air conditioning is not present because air conditioning is desirable feature that makes a house comfortable to live in.

ANS3C) The functional form of the regression model can be written as:

Price = β0 + β1Lot.Size + β2Age + β3Living.Area + β4Fireplace

In this equation, Lot.Size, Age, Living.Area are quantitative variables, and Fireplace is a dummy variable.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | SUMMARY OUTPUT | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *Regression Statistics* | | | | | | | | |
| 4 | Multiple R | 0.719906543 | | | | | | | |
| 5 | R Square | 0.518265431 | | | | | | | |
| 6 | Adjusted R Square | 0.504976201 | | | | | | | |
| 7 | Standard Error | 58108.55407 | | | | | | | |
| 8 | Observations | 150 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | ANOVA | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Significance F* | | | |
| 12 | Regression | 4 | 5.26735E+11 | 1.31684E+11 | 38.99890739 | 3.88923E-22 | | | |
| 13 | Residual | 145 | 4.89608E+11 | 3376604056 | | | | | |
| 14 | Total | 149 | 1.01634E+12 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| 17 | Intercept | 44195.08176 | 18363.67148 | 2.406658266 | 0.017357818 | 7900.027666 | 80490.13585 | 7900.027666 | 80490.13585 |
| 18 | Lot.Size | 5032.994795 | 8106.604007 | 0.620851196 | 0.535672226 | -10989.38021 | 21055.3698 | -10989.38021 | 21055.3698 |
| 19 | Age | -288.18029 | 219.7305364 | -1.3115168 | 0.191756099 | -722.4688099 | 146.1082299 | -722.4688099 | 146.1082299 |
| 20 | Living.Area | 94.45032059 | 9.721836716 | 9.715275348 | 1.70852E-17 | 75.23550344 | 113.6651377 | 75.23550344 | 113.6651377 |
| 21 | Fireplace | 8475.215357 | 11051.87063 | 0.766857995 | 0.444412724 | -13368.36004 | 30318.79076 | -13368.36004 | 30318.79076 |
| 22 | | | | | | | | | |

The estimated regression equation from the above output is:

Price = 44195.08176 + 5032.994795(Lot.Size) - 288.18029(Age) + 94.45032059(Living.Area) + 8475.215357(Fireplace)

R Square in the provided regression analysis is 0.5183, indicating that the independent variables (Lot.Size, Age, Living.Area, and Fireplace) in the model explain 51.83% of the variability in the dependent variable (Price). The model does not explain the remaining 48.17% of the variability in the dependent variable.

ANS3F)

The F-test is a statistical test used to assess the overall significance of a regression model. The F-statistic is 38.9989 in this situation, and the significance level is 3.88923E-22, which is less than the alpha level of 0.10. This means that the regression model is statistically significant and capable of making predictions. In other words, the regression model fits the data well, and the independent factors are linked to the dependent variable.

The intercept coefficient is 4419.08176. When all of the independent variables (X) are equal to zero, the estimated mean value of the dependent variable (Y) is reflected. When all other variables are zero, the estimated value of the house price is displayed. Because the p-value associated with the t-test is less than 0.10, we can argue that the intercept is statistically significant with = 0.10.

The coefficient for Lot.Size is 5032.994795. This coefficient shows the expected change in the value of the house price for a one-unit increase in the Lot.Size variable while keeping all other factors constant. Because the p-value associated with the t-test is greater than 0.10, we cannot conclude that this coefficient is statistically significant with = 0.10.

The Age coefficient is -288.18029. This coefficient shows the projected change in house price for a one-unit increase in Age while holding all other factors constant. We cannot conclude that this coefficient is statistically significant with = 0.10 because the p-value associated with the t-test is bigger than 0.10.

The coefficient for Living.Area is 94.45032059. While holding all other variables constant, this coefficient represents the estimated change in the value of the house price for a one-unit increase in the Living.Area variable. We can state that this coefficient is statistically significant with = 0.10 because the p-value associated with the t-test is less than 0.10.

The Fireplace coefficient is 8475.215357. This coefficient shows the projected change in house price for a one-unit increase in the Fireplace variable while all other variables remain constant. We cannot conclude that this coefficient is statistically significant with = 0.10 because the p-value associated with the t-test is bigger than 0.10.

ANS3H)

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | | *Lot.Size* | *Age* | *Living.Area* | *Fireplace* | | |
| 2 | Lot.Size | 1 | | | | | |
| 3 | Age | 0.121747 | 1 | | | | |
| 4 | Living.Area | 0.156475 | -0.250804937 | 1 | | | |
| 5 | Fireplace | 0.175467 | -0.101707031 | 0.486279133 | 1 | | |
| 6 | | | | | | | |
| 7 | | | | | | | |

There is no evidence of multicollinearity between the independent variables in your regression. The highest correlation coefficient is 0.156 between Lot.Size and Living.Area, which is relatively low

Multicollinearity occurs when two or more independent variables in a regression model are substantially associated with one another. This can be a problem for regression models since it makes determining the individual influence of each independent variable on the dependent variable difficult. The calculated coefficients for the impacted variables may be unstable and have substantial standard errors in the presence of multicollinearity. This can lead to a loss of statistical power and a reduction in the accuracy

of the regression model. Furthermore, multicollinearity can make it difficult to interpret the regression model's coefficients.

In conclusion, the presented regression analysis revealed that the independent variables, including Lot.Size, Age, Living.Area, and Fireplace, had a substantial impact on the dependent variable, Price. According to the computed regression equation, Lot.Size, Living.Area, and Fireplace were all favourably connected to Price, whereas Age was negatively related. The R-squared value of 0.5183 indicates that the independent variables explain 51.83% of the variability in the dependent variable, implying that other factors other than those included in the model influence the Price.

Furthermore, the F-test revealed that the regression model was statistically significant, with a significance level of 3.88923E-22, indicating that it could make reliable predictions. However, at the 0.10 alpha level, the coefficients for Lot.Size and Age were not statistically significant. Moreover, no indication of multicollinearity between the independent variables was found, implying that they were not significantly associated. The association between Lot.Size and Living.Area, on the other hand, was relatively high.

Based on the findings, it is recommended that the model incorporates more independent variables to improve its predictive potential. Furthermore, additional research might be conducted to find other elements that influence the price of a house but are not currently represented in the existing model. Inclusion of interaction and polynomial components in the model could also improve its accuracy. Furthermore, the model could be applied to a larger dataset to improve the accuracy of the results.