1. Develop a categorization of your data using pivot tables. Develop two pivot tables: One pivot table of average price, varying type of construction (brick) and neighborhood as the two dimensions; a second pivot table of average square feet varying type of construction (brick) and neighborhood as the two dimensions (20%). What patterns do you see?

Based off the results returned from my pivot table, it seems as though the “West” Neighborhood has both the most expensive average price, as well as largest brick homes in square feet in this data set. The data also confirms market behavior of the bigger the house, the more expensive the home will be at resale. This may suggest that the Western neighborhood is an affluent community based of these initial findings.

1. Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood. (10%)

Pivot Chart – Average Price

Pivot Chart – Average Square Feet

1. Perform a correlation analysis of all quantitative variables except ID. Which two variables have the strongest (largest magnitude) correlation? Which two variables have the weakest (smallest magnitude) correlation? What does the largest magnitude imply if we perform a regression analysis next? Are there any negative correlations? Are these correlations intuitive? If not, why not? (20%)

Based off the results I generated, the two variables with the strongest correlation are “Price” and “Square Feet”. The correlation value between these two variables was 0.553 and further highlights that as either value increases or decreases, the other value will also behave similarly.

The variables with the weakest correlation value between them was between price and offers with a value of -0.314.

The variables with the smallest magnitude was Bedrooms and offers

If we conduct a regression analysis against Price and Square Feet, the largest magnitude value returned from correlation analysis would confirm the behavior that as square feet of a home increases, the price of the home will increase by some value.

There were negative correlation values returned during correlation analysis. The correlation value between Price and Offers was -0.314 and suggest that there is a very weak relationship between the two variables.

The variables of square feet, number of Bedrooms and number of bathrooms related to price is an intuitive correlation to me. All these variables justify an increase in home value if any of the variables are increased.

1. Perform an initial regression analysis of the quantitative variables excluding the ID. Do not include type of construction or neighborhood. Which variables are statistically significant? What does each coefficient mean in a real-world sense? Are these coefficients intuitive? If not, why not? What does the R-squared mean? (15%)

After performing an initial regression analysis on the quantitative variables of the dataset, I was able to determine the variables of square feet, bedrooms, bathrooms and offers were statistically significant and that interpretation was allowed. The only value that was not statistically significant was the intercept value provided by regression analysis function.

Based off my understanding of the data, the coefficients in general, describe how a variable will change and by which magnitude. Below are my interpretation of the coefficient values:

* As the square footage Increase, the price for the home increases by $61.84 per square foot
* As number of Bedrooms available Increase, the price for the home increases by $9319.75
* As number of Bathrooms available Increase, the price for the home increases by $12,646.35
* Not intuitive

During my analysis, there is one independent variable that was not intuitive. The coefficient for offers was -13601.01 with a p-value of 3.08E-18. The p-value states that interpretation is allowed. However, the meaning of the coefficient does not explain the behavior naturally on how offers influences the change in home price.

In general, the R-Squared value in regression analysis determines the proportion of variance in the dependent variable that can be explained by independent values. In this scenario, the dependent value is Home Sale price and the independent variables are Square Feet, Number of bedrooms, Number of Bathrooms and amount of offers. The R-Squared value for this regression analysis resulted with 0.698. This can be translated as 68.9% of the change in price of homes can be determined by the behavior of square foot, bedrooms, bathrooms and current offers.

1. Perform a second regression including variables from part 4 and dummy variables for type of construction and neighborhood. What does each coefficient mean in a real-world sense? Are these coefficients intuitive? If not, why not? What does the R-squared mean? (10%)

In this scenario, there are three additional independent variables that can affect the price of a home: Brick Construction, and location of the home. Below are the explanations of each coefficient that can be interpreted:

* As the square feet value of the house increases, the price of the home will increase by $52.99
* As the number of bedrooms available increase, the price of the home will increase by $4246.79
* As the number of bathrooms increase, the price of the home will increase by $7883.28
* As the number of current offers increase, the price of the home will decrease by $8267.49
* If the home is constructed of brick, the price of the home will increase by $17297.35
* If the home is in the west neighborhood, the price of the home will increase by $20681.04

In this scenario, there were two independent variables that were not statistically significant: Intercept and East Neighborhood. It is worth mentioning that the p-value of East neighborhood variable was 0.5152 and was just over the threshold for interpretation.

In this scenario, the dependent value is home sale price and the independent variables are Square Feet, Number of bedrooms, Number of Bathrooms and amount of offers. The R-Squared value for this regression analysis resulted with 0.698. This can be translated as 86.9% of the change in price of homes can be determined by the independent variables of this scenario.

1. Create a spreadsheet prediction of the regression model from part 5. Perform a two-way sensitivity analysis and use conditional formatting to highlight the results. (15%)

Regression Analysis 1:

Table

Description automatically generated

Regression Analysis 2 – With Dummy

Table

Description automatically generated

1. What would explain nonintuitive results in your regression using the data that you were provided? What additional data would assist you in explaining the nonintuitive results? (10%)

I determined that offers was the only non-intuitive variable in this dataset. The additional data that would have helped to explain the non-intuitive results are: 1. Comparable sales of other homes 2. Additional features of the home such as a pool, lot size and age of home and 3. How many offers fell through before closing. This may have explained why the regression results conveyed that for every offer made, the price of the home decreased. This may have been explained by offers that fell through and did not close.