**House Price Analysis**

**Background**

Using the House Prices data, determine the factors that influence the price of a home.

**Resources**

Use the data set SCM 651 Homework 1 House Prices spreadsheet.

**Assignment**

**What’s due:**

Submit a categorization, visualization, correlation, and regression analysis of house prices **before the live class in week 4**. Suggested length is 5 pages but should not exceed 10 pages, single-spaced, 12-point font.

This is a group assignment; each student should upload a copy of the assignment to the learning management system. The paper must be a Microsoft Word document. You should also submit the Excel spreadsheet with the visualization, correlation, and regression analysis. Name the file HW1\_Team# where # is your team number. Be sure to include the names of everyone on the team on the first page of the paper. Late assignments will not be accepted. Failure to follow directions will be penalized.

**Outline and grading criteria:**

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?

Table

Description automatically generated with medium confidence

Patterns:  **The West neighborhood has homes with the largest square footage and the highest prices. In addition, brick homes are higher priced than non-brick homes in every neighborhood.  The North area has home prices that are below the overall average of $130,427.34 as well as square footage also below the average of 2000.94.**

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%)

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. Perform a correlation analysis of all quantitative variables except ID.
   1. Which two variables have the strongest (largest magnitude) correlation?
      1. **Price and Sqft.**
   2. Which two variables have the weakest (smallest magnitude) correlation?
      1. **Bedrooms and Offers**
   3. What does the largest magnitude imply if we perform a regression analysis next?
      1. **It implies that price and square feet will move in a positive direction and be strongly correlated.  Price and square feet will have a low p-value and high R-squared.**
   4. Are there any negative correlations?
      1. **Yes, price and offers are negatively correlated.**
   5. Are these correlations intuitive?
      1. **Yes, as an example it makes intuitive sense that square feet and price are positively correlated because bigger homes are usually more expensive.  It also makes sense that the number of bedrooms and bathrooms have a weak correlation with the number of offers. It also makes intuitive sense that price and offers are negatively correlated because as price goes up, the number of offers would go dow**n.

If not, why not? N/A (20%)

A picture containing table

Description automatically generated

1. Perform an initial regression analysis of the quantitative variables excluding the ID. Do not include type of construction or neighborhood.
2. *Which variables are statistically significant?*
   1. **The intercept is not statistically significant because the p-value (.175) is greater than .05. All remaining variables are statistically significant with p-values less than .05.**

1. What does each coefficient mean in a real-world sense?
   1. **SqFt – for every additional square foot added, our price increases by $61.84**
   2. **Bedrooms- For every additional bedroom, our price increases by $9319.75**
   3. **Bathrooms – For every additional bathroom, our price increases by $12646.35**
   4. **Offers – For every additional offer, the price decreases by $13,601.01.**

1. *Are these coefficients intuitive?*
   1. **Bathroom, bedroom and sqft are intuitive because every addition of each would increase the price.  According to the output, for every additional offer, the price decreases by $13,601.01 and this is not intuitive because the price is established before the offers are made.**

1. *What does the R-squared mean?* (15%)
   1. **69.8% of the change in price is explained by changes in Sqft, Bedrooms, Bathrooms and offers.**

A picture containing line chart

Description automatically generated

1. Perform a second regression including variables from part 4 and dummy variables for type of construction and neighborhood.
   1. What does each coefficient mean in a real-world sense?
      1. **The coefficient of the intercept represents the fixed cost of a non-brick home in a western neighborhood. For every additional square foot, the price increases by $52.99. For every additional bedroom, price increases by $4,246.79. Every additional bathroom increases price by $7,883.28 and for every additional offer, price decreases by $8,267.49. The coefficients of the categorical variables are interpreted in relation to the intercept. On average, the fixed cost of a brick home in a western neighborhood is $17,297.39 more than a non-brick home in the west. On average, the fixed cost of a non-brick home in the east is $22,241.62 less than a non-brick home in the west. On average, a non-brick home in the north is $20,681.04 less than a non-brick home in the west.**

1. Are these coefficients intuitive? If not, why not?
   1. **All variables are intuitive except offers. Offers are non intuitive because in reality the price of a house is set before offers are made. An additional offer on a home will not decrease the price of the home.**

1. What does the R-squared mean? (10%)
   1. **The R-squared means that 86.9% of the change in price can be explained by the change in our variables.**

Graphical user interface, application, table, Excel

Description automatically generated

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%)

Chart

Description automatically generated with low confidence

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%)
   1. **The only variable that wasn’t intuitive is the “offers”.  This relationship’s intuitiveness is dependent upon how it is interpreted. We chose to interpret it as every additional offer decreasing the price. This is nonintuitive because offers do not affect the pricing of a house unless it is being auctioned; in which the price would increase for every offer. If we interpreted the relationship as higher priced homes resulting in less offers, we could say that it is in fact intuitive.  The additional data that would have helped to explain the 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.**

Justify your answers. Provide a snapshot of output from your analysis in your final paper.

**House Price Data**

ID unique identifier

Price price of home in dollars

SqFt square feet of area of home

Bedrooms number of bedrooms

Bathrooms number of bathrooms

Offers number of offers received on home before sale

Brick Yes/No on brick construction

Neighborhood location of home in east, west, or north quadrants of city