Alexa Mowbray, Ryan Cathcart, David Doman, & Jason Taylor

**House Price Analysis Case Study**

In order to glean more meaningful analyses on which factors had the greatest impact on the price of homes, our group decided to first analyze the quantitative and qualitative ways in which the data was categorized. This categorization included the price, construction type (brick or not), square footage, number of bedrooms or bathrooms, whether or not the home received an offer (and quantity thereof), and which neighborhood the home was located in. To make the data easier to conceptualize, the group first used pivot tables to condense hundreds of rows of data into something more meaningful. The first pivot table that was constructed analyzed the average price, square footage, and construction material of the home, while the second was geared toward understanding the relationship between square footage, neighborhood, and price.

**Pivot Table #1**

**Description:** Pivot Table 1 demonstrates the relationship between average price and the average square footage of a home. Providing more detail to the analysis was the addition of construction type. While the average square footage of each home type was similar with brick homes being on average only 1.7% larger, the average price of a brick home regardless of neighborhood was 17% higher than a home not constructed of brick. This leads us to conclude that brick construction is more desirable than non-brick, due either to the desirability of brick aesthetic or the durability of the construction type. It also leads us to conclude that average home prices are impacted by more than one variable.

|  |  |  |
| --- | --- | --- |
| **Brick** | **Average of Price** | **Average of SqFt** |
| No | $ 121,958.14 | 1,989.19 |
| Yes | $ 147,769.05 | 2,025.00 |
| **Grand Total** | **$ 130,427.34** | **2,000.94** |

**Pivot Table #2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Brick** | **Neighborhood** | **Average of SqFt** | **Average of Price** |
| **No** | East | 2,002 | 117,750 |
|  | North | 1,928 | 108,584 |
|  | West | 2,073 | 148,230 |
| **Non Brick Avg** |  | **1,989** | **121,958** |
| **Yes** | East | 2,031 | 135,468 |
|  | North | 1,857 | 118,457 |
|  | West | 2,091 | 175,200 |
| **Brick Avg** |  | **2,025** | **147,769** |
| **Average All** |  | **2,001** | **130,427** |

**Description of Pivot Table #2:** This pivot table provides more granular data concerning the location of each home type. In pivot table 1 we looked at averages of square footage and average price in totality, whereas here we were looking to understand if there was a difference in price by neighborhood.  Similar to the results we see in our first pivot table, average square footage does not vary greatly by construction type or neighborhood. It can also be noted once again that on average, brick homes are more expensive than non-brick homes regardless of neighborhood. In the same neighborhood, brick homes sold for:

* East: 13% greater than non-brick
* North: 8.3% greater
* West: 15.4% greater

However, we can also ascertain that homes in the West neighborhood are selling at a premium. Average home prices for brick construction in the West neighborhood were:

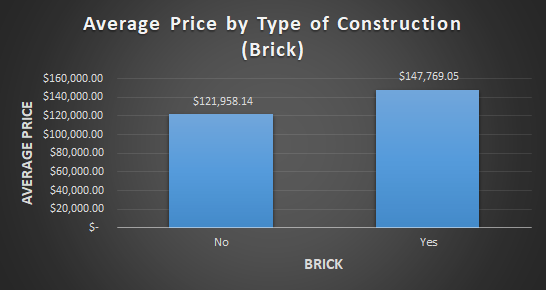
* 22.7% greater than in the East
* 32.4% greater than in the North

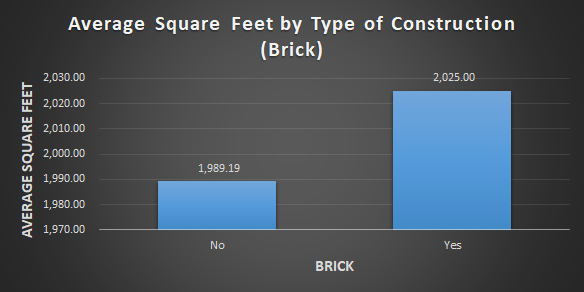
For non-brick construction, average home prices in the West neighborhood were:

* 20.1% greater than in the East
* 26.7% greater than in the North

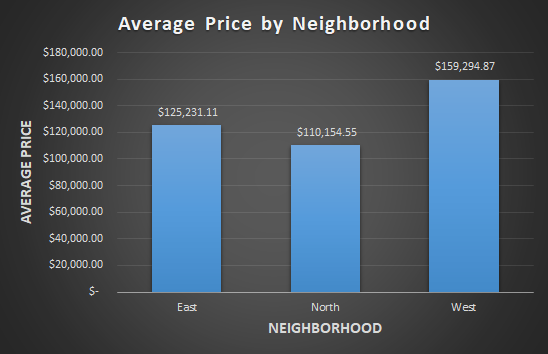
Thus, we can conclude that the West neighborhood is more desirable than either the East or North neighborhoods, and regardless of neighborhood brick construction was a more preferable home type.

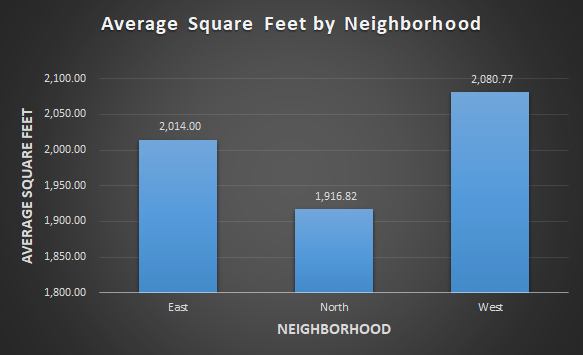
To aid in data visualization and communication, the team created pivot charts which illustrate the above conclusions





The above visualizations are based on pivot table 1, depicting construction type on the y axis, and square footage and average price on the x axis. Note that regardless of neighborhood, average square footage by construction type does not vary greatly by home type, with brick homes being on average 1.8% larger, whereas price does vary significantly.





The above visualizations are based on pivot table 2, square footage and average price on the y axes, and both neighborhood and construction type in the x-axis. Note that, while homes in the West neighborhood are slightly larger, the square footage does not vary greatly between them. However, there is a positive trend by construction type and neighborhood, with brick construction in the West being most expensive.

After the pivot table analyses which were largely based on qualitative data , we then performed a correlation analysis of quantitative variables: price, square footage, number of bedrooms, number of bathrooms, and number of offers received on home before sale.

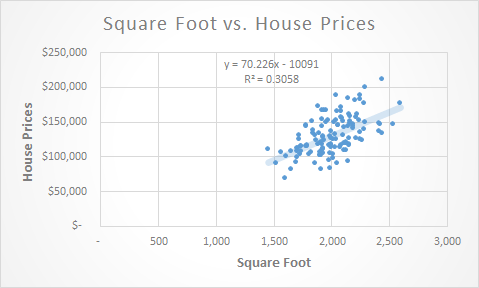
**Description of Correlation Analysis:**

Below is the output of our correlation analysis. To more fully understand which variables had the greatest impact on the price of the home, we looked at the magnitude of their correlation. The smallest magnitude correlation variable was the number of offers on the house that were received before the sale.

**Correlation Analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Price* | *SqFt* | *Bedrooms* | *Bathrooms* | *Offers* |
| Price | 1 |  |  |  |  |
| SqFt | 0.552982 | 1 |  |  |  |
| Bedrooms | 0.525926 | 0.483807 | 1 |  |  |
| Bathrooms | 0.523258 | 0.522745 | 0.414556 | 1 |  |
| Offers | -0.31364 | 0.336923 | 0.114271 | 0.143793404 | 1 |

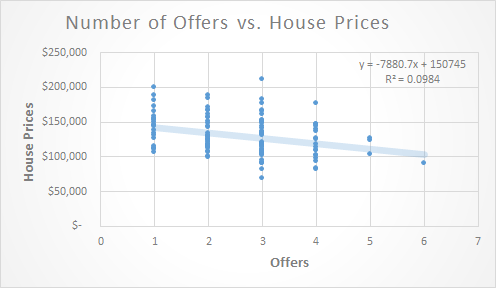
Based on the above data, we know that the value that is most correlated with price is square footage, due to its correlation coefficient being greater than any of the other variables. A visual representation of this is shown below.



The graph above shows an increasing trend in the data from left to right, thus demonstrating the relationship between square footage and price – i.e. as square footage increases, so too will price. The higher the correlation the closer the data points will be to the trend line. On the other hand, we can see that Offers and Bedrooms have the smallest correlation. This is determined by having the closest value to zero. A visual representation of these two variables would show data points all over the graph with no clear cut trend.

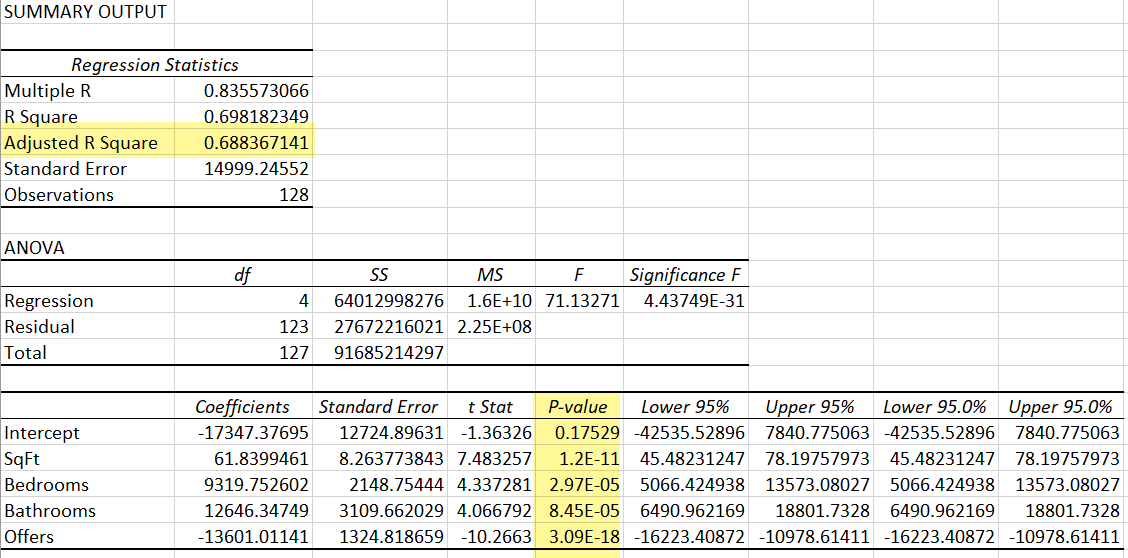
This analysis is associated with a linear regression analysis, which will model the association between a dependent variable, in this case home price, and one or more independent variables.

Present in the analysis was one negative correlation between the number of offers a house received and a price. This negative, or inverse, correlation illustrates to us that the two variables move in opposite directions: when the price of the home increases, fewer offers will be placed on the home. The illustration below shows this relationship.



The correlations here are very intuitive. Larger homes, with more bedrooms or bathrooms are desirable to buyers. Inversely, in terms of number of offers, as home prices increase, the size of the market of potential buyers decreases.

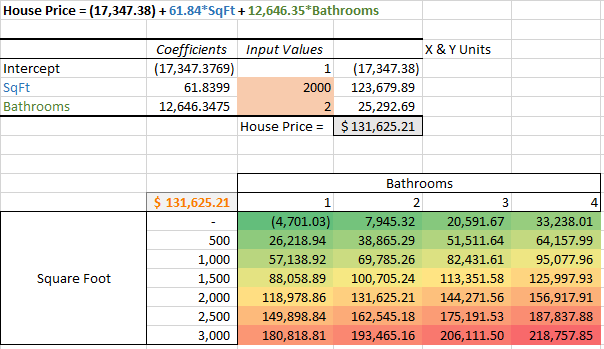
Strictly associated with correlation analysis, as noted above, is regression analysis. We performed this analysis using all quantitative variables. The data in the below table confirms our findings in the correlation analysis while also illustrating the relationship between bedrooms and the price of a home. The P-values of square footage, bedrooms, and bathrooms lead us to conclude that there is a strong relationship between these variables and the price of a home.



Another aspect of the regression analysis is R-squared. R-squared is a measure that calculates the proportion of a variable which can be explained by the independent variable. Because we are running a regression on multiple variables it is most appropriate to use the Adjusted R Square value which will tell us how much of the change in prices is due to the changes in our x variables - in this example, 68.8% of price variance can be explained by the x variables.

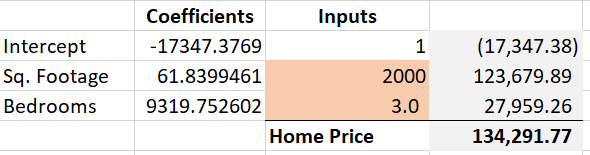
Finally, our analysis led us to a Sensitivity Analysis.

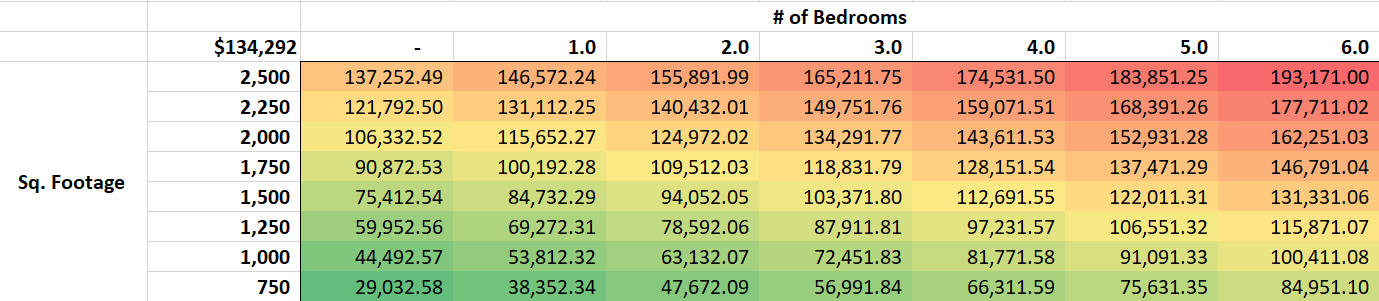
**Spreadsheet for Sensitivity Model**



**Description**: Two, two-way sensitivity analyses were performed based on our regression data. This type of analysis determines how independent variables, in this case square footage and number of bathrooms, affects a dependent variable, or price of the home. The above depiction is a spreadsheet of the analysis that was performed. To establish our home price, we multiplied our intercept, Sq. Ft. and bathroom coefficients by the input assumption values to determine the X & Y units, which in summation provides our home price = $131,625.21.

Conditional formatting applied to the data table, indicates higher house prices in red/orange and lower house prices in green/yellow.





**Description:** An additional sensitivity model was created using the number of bedrooms as a coefficient.

Our regression analysis yielded some interesting results for our overall look at housing prices. One thing that could explain any non-intuitive results could be where the house is located. Any demographic data could provide meaning to what we are seeing in our results. Other explanations for non-intuitive results could be things such as current housing market conditions and current state of the economy.