Team 73: Clifford Forrester, Kathleen Fox, Mashundra Maclin, Courtney Smith, and Joyce Woznica

TEam 73  
Homework #1

House Price Analysis

Contents

[Introduction 3](#_Toc6487594)

[The Data 3](#_Toc6487595)

[#1 – Data Categorization 3](#_Toc6487596)

[Average Square Feet and Average Price in one Pivot Table 4](#_Toc6487597)

[#2 – Data Visualization 5](#_Toc6487598)

[#3 – Correlation Analysis 6](#_Toc6487599)

[Overall Correlation Conclusions 7](#_Toc6487600)

[# 4 – Regression Analysis 7](#_Toc6487601)

[#5 – Prediction Model and Sensitivity Analysis 9](#_Toc6487602)

[Prediction Model 9](#_Toc6487603)

[Sensitivity Analysis 10](#_Toc6487604)

[#6 – Results and Conclusions 10](#_Toc6487605)

List of Figures

[Figure 1: Average Housing Price by Construction Type and Neighborhood 3](#_Toc6487662)

[Figure 2: Average Square Footage by Neighborhood and Construction Type 4](#_Toc6487663)

[Figure 3: Average Square Footage by Construction Type and Neighborhood 4](#_Toc6487664)

[Figure 4: Average Housing Price and Square Feet by Construction Type and Neighborhood 4](#_Toc6487665)

[Figure 5: Average Housing Price and Square Feet by Neighborhood and Construction Type 4](#_Toc6487666)

[Figure 6: Average Housing Price and Square Feet by Neighborhood and Construction Type 5](#_Toc6487667)

[Figure 7: Average Square Feet by Construction Type and Neighborhood 6](#_Toc6487668)

[Figure 8: Initial Correlation of Housing Prices Data 6](#_Toc6487669)

[Figure 9: Regression on Quantitative Variables Only 8](#_Toc6487670)

[Figure 10: Prediction with Excel for Regression with Quantitative Variables Only 9](#_Toc6487671)

[Figure 11: Two-Way Sensitivity Analysis with Varying Square Footage and Number of Bedrooms 10](#_Toc6487672)

# Introduction

In this exercise, we will be using the house prices data provided in the associated Excel spreadsheet to determine the factors which influence the price of a home. To do this we will be executing the following:

* Data Categorization (Question #1)
* Pivot Tables for Visualization (Question #2)
* Correlation Analysis (Question #3)
* Initial Regression Analysis (Question #4)
* Prediction Model and Sensitivity Analysis (Question #5)
* Conclusions (Question #6)

## The Data

Our house price data provides the following variables about each house data point collected:

* ID – a unique identifier
* Price – price of the home in dollars
* SqFt – square foot area of the home
* Bedrooms – number of bedrooms
* Bathrooms – number of bathrooms
* Offers – number of offers received on the home before the sales
* Brick – Yes/No on if brick construction
* Neighborhood – location of the home in east, west, north quadrants of the city

# #1 – Data Categorization

To initially review the data, we created a pivot table to review the Price by type of construction (Brick: Yes or No) and Average Square Feet by designed neighborhood. Figure 1 represents reviewing prices within certain neighborhoods (East, West, North) based on the construction type.

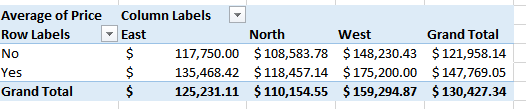


Figure 1: Average Housing Price by Construction Type and Neighborhood

In Figure 2, we created the pivot table, but then manipulated the result using a table in Microsoft Word to show a more aesthetically pleasing table. We also selected the neighborhood for the rows and the construction type for the columns in this table.

|  |  |  |  |
| --- | --- | --- | --- |
| *Average Square Feet of House* | Brick | |  |
| Neighborhood | *No* | *Yes* | **Grand Total** |
| *East* | 2001.54 | 2031.05 | 2014.00 |
| *North* | 1928.11 | 1857.14 | 1916.82 |
| *West* | 2073.48 | 2091.25 | 2080.77 |
| Grand Total | **1989.19** | **2025.00** | **2000.94** |

Figure 2: Average Square Footage by Neighborhood and Construction Type

Finally, we reflected the same information as that in the previous table, but modeled this using the structure of rows as construction type and columns as neighborhoods. The result is shown in Figure 3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Average Square Feet of House | Neighborhood | | |  |
| **Brick** | **East** | **North** | **West** | **Grand Total** |
| No | 2001.54 | 1928.11 | 2073.48 | 1989.19 |
| Yes | 2031.05 | 1857.14 | 2091.25 | 2025.00 |
| **Grand Total** | **2014.00** | **1916.82** | **2080.77** | **2000.94** |

Figure 3: Average Square Footage by Construction Type and Neighborhood

## Average Square Feet and Average Price in one Pivot Table

Out of curiosity, we created a more in-depth chart to represent both the average square feet and average housing price by neighborhood and construction type. The pivot table in the next figure includes the average square feet into the table reflecting the average square feet of the homes in the particular neighborhood areas and their average price by construction type.

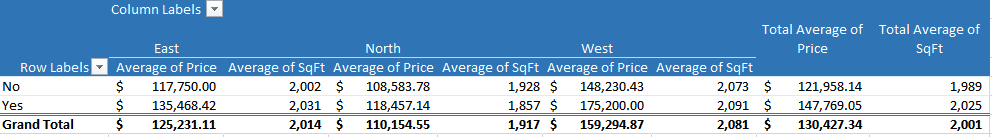


Figure 4: Average Housing Price and Square Feet by Construction Type and Neighborhood

The following table shows the same information as Figure 2, but listed by neighborhood on the left-hand side and then by Construction Type and the average house price and average square feet.

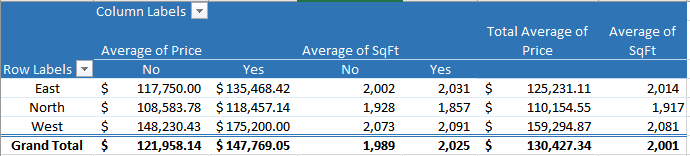


Figure 5: Average Housing Price and Square Feet by Neighborhood and Construction Type

# #2 – Data Visualization

Now that we can see the details about the housing prices in certain neighborhoods by both average square feet as well as type of construction in the pivot tables we created, we will create charts to show that in a more graphic visual.

The pivot chart shown in Figure 6, the average housing prices by neighborhood and construction type. You can see the different scale of the y-axis in this chart reflecting the average home price.

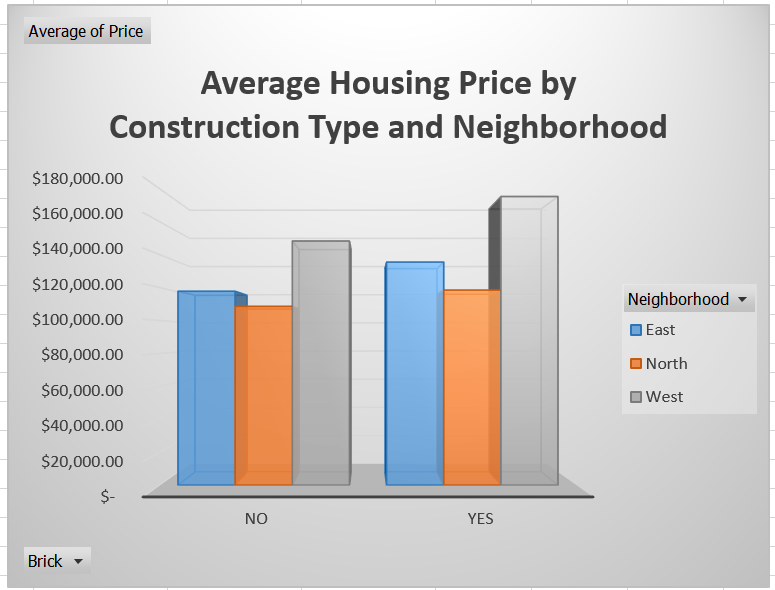


Figure 6: Average Housing Price and Square Feet by Neighborhood and Construction Type

In Figure 7, we have represented the average square foot of the homes by neighborhood and if the construction is brick (Yes) or not brick (No).

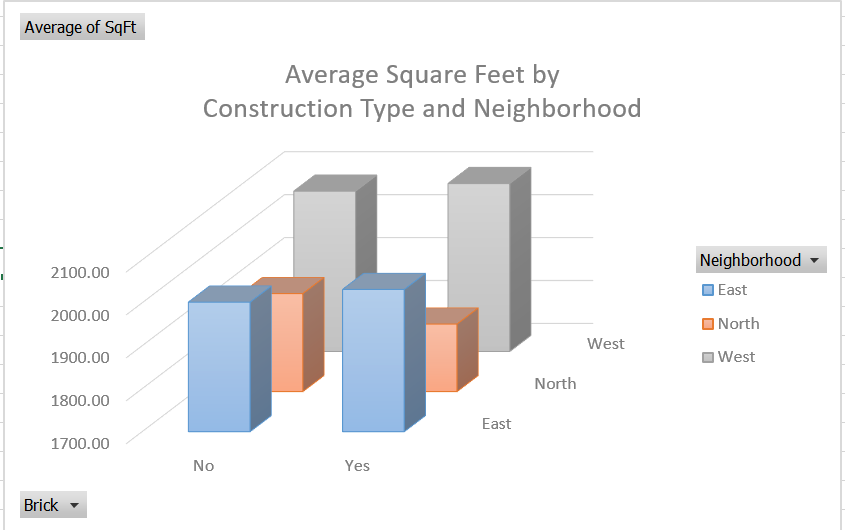


Figure 7: Average Square Feet by Construction Type and Neighborhood

# #3 – Correlation Analysis

It is important to review and determine if there are any correlations between the input quantitative variables (square feet, number of bedrooms, number of bathrooms, offers) and the dependent variable which is the housing price. We can make these determinations by looking at descriptive statistics on the data set provided.

We ran correlation statistics on only the quantitative variables which were the following:

* Square Feet
* Number of Bedrooms
* Number of Bathrooms
* Offers
* Price

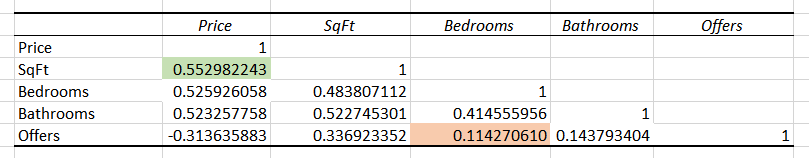


Figure 8: Initial Correlation of Housing Prices Data

As shown in the previous figure, the numbers in the table are our correlation coefficients. This information provides the relationship between two variables, but does not give us the statistical significance of that correlation (descriptive statistic) in the overall model.

To explain the data in this table, a perfect correlation is represented by 1 and no correlation would be represented by 0. The largest magnitude correlations are between square feet and price (0.552982243) highlighted in green in the figure which represents the correlation value farthest from 0 representing the highest correlation. The smallest magnitude correlation is between number of offers on the house and number of bedrooms (0.114270610) highlighted in orange in Figure 8 which is the number closest to 0 representing the least correlation.

There is a negative correlation related to the number of offers and the price of the home. This means that as the number of offers increase, the price of the home would decrease.

## Overall Correlation Conclusions

The information found as the result of running these correlations indicate that square footage and number of bedrooms has a large correlation on the price of the home. This, in our opinion, is intuitive, as we would expect to see that the higher the square footage, the more expensive the home. It is also intuitive that with more bedrooms are more desirable which is also indicated by the correlations. Alternatively, the number of bedrooms and bathrooms has very little to do with the number of offers made on this house indicated by their small correlations.

# # 4 – Regression Analysis

Now that we have done some initial correlation and descriptive statistics on this housing data, we need to look at the statistical significance of certain independent variables and how these affects the dependent variable (price).

The regression analysis was completed for the quantitative variables that were used in the correlation analysis in the previous section; however, here we have noted that the independent variables that we are trying to use to predict or determine the dependent variable which is the price of the home.

* Independent Variables
  + Square Feet
  + Number of Bedrooms
  + Number of Bathrooms
  + Offers
* Dependent Variable
  + Price

The output from the regression analysis done in Microsoft Excel is noted in the following figure.

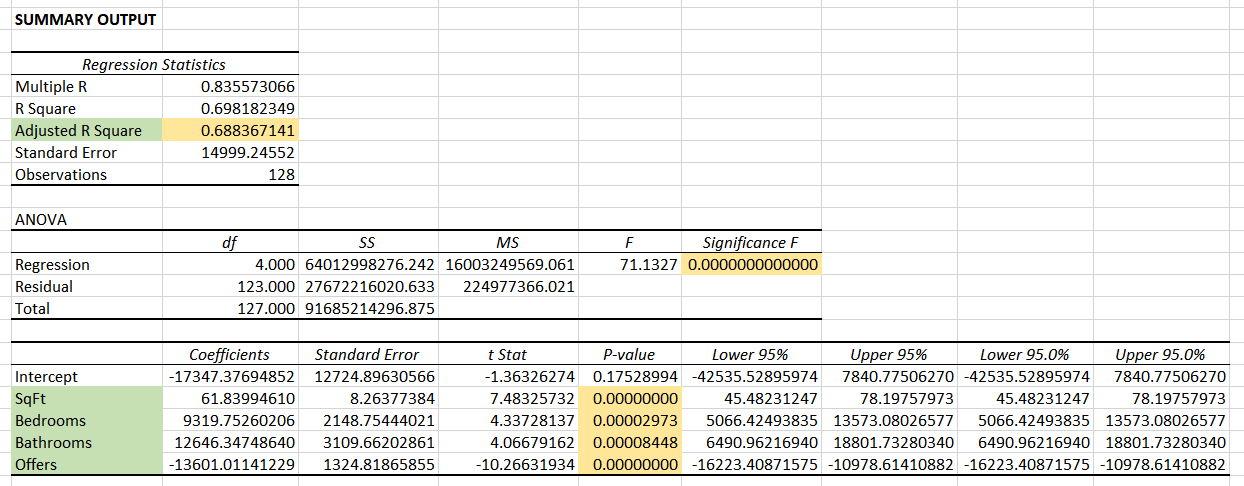


Figure 9: Regression on Quantitative Variables Only

Reviewing just the quantitative variables, we see a fairly strong Adjusted R Square (0.68836714) which is the appropriate R square to use if you have multiple independent variables. However, it is not as strong as we would like to see in a model. However, we do see that the model shows a very small *Significance F* (well below an α = 0.05) in the ANOVA model which represents that the model is considered statistically significant.

Since all of the independent variables show a *p-value* of below the desired α of 0.05, this would indicate that all variables are significant in the prediction of price. The number represented in the Coefficients column shows how the price is adjusted (up or down) based on the value of that variable. For example, previous offers on the house brings the overall price down where higher numbers of bathrooms, bedrooms and higher square footage increase the housing price. We believe these coefficients are very intuitive. It is clear that houses with more square footage and more bathrooms and/or bedrooms would have a higher price than those with less of these items. It also seems intuitive that as offers are made on the home, the price would decrease.

# #5 – Prediction Model and Sensitivity Analysis

## Prediction Model

To create the prediction model, we simply create an equation with the intercept and the coefficients for each independent variable. For example, the generic prediction model for four (4) independent variables and one dependent variable is shown below.

where:

* *Ŷ* is the predicted value
* *b0* is the y-intercept
* *b1* is the coefficient for the first dependent variable (*X1*) which is square feet
* *b2* is the coefficient for the second dependent variable (*X2*) which is number of bedrooms
* *b3* is the coefficient for the third dependent variable (*X3*) which is number of bathrooms
* *b4* is the coefficient for the fourth dependent variable (*X4*) which is number of offers

This means that our prediction model for the housing price based on quantitative variables only is as follows (note we have rounded so the equation would fit in this document, the actually numbers were not rounded for calculations):

To use this model, we can estimate something like the following:

* What is the housing price with the following house information?
  + 2500 square feet
  + 3 bedrooms
  + 2 bathrooms
  + 2 offers

The result would be as follows (note we have rounded so the equation would fit in this document, the actually numbers were not rounded for calculations):

Which yields a predicted housing price of $163,302.42.

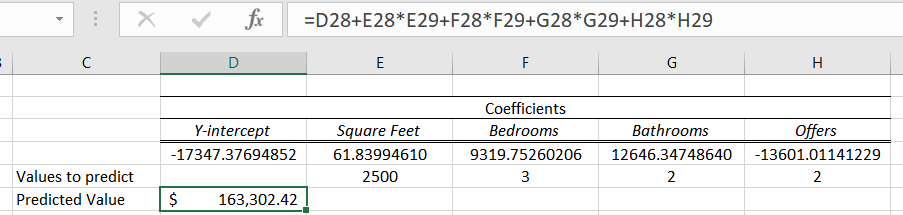
Doing this in Excel looks something like that shown in Figure 10. 

Figure 10: Prediction with Excel for Regression with Quantitative Variables Only

## Sensitivity Analysis

For the two-way sensitivity analysis, we selected to look at price when the square feet and the number of bedrooms varied. The resulting analysis is shown in Figure 11. Note that we have only ranged the square feet between 1,450 and 2,590 since those are within the minimum and maximum. For bedrooms, we have shown only 2 to 5 bedrooms for the same reason. We used the MIN and MAX Excel function to determine our ranges for this analysis.

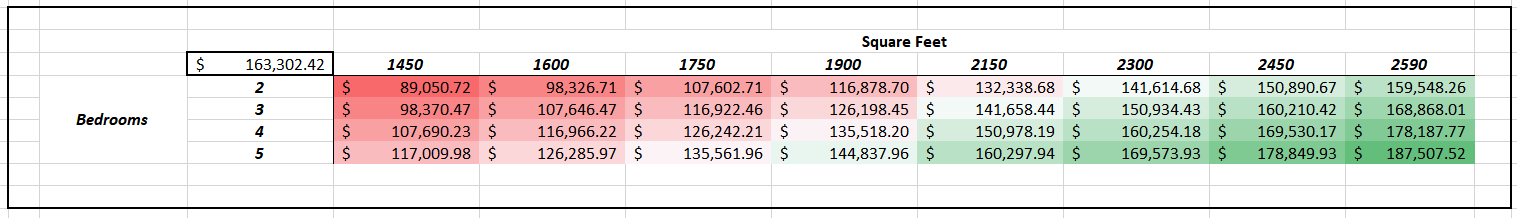


Figure 11: Two-Way Sensitivity Analysis with Varying Square Footage and Number of Bedrooms

As you can see in the conditional formatting and the sensitivity analysis, the prices are much higher as the number of bedrooms and square footage increases as shown by the darker red in the lower right-hand corner where these values are the largest.

# #6 – Results and Conclusions

For this model, we think it would be important to consider the non-quantitative variables like construction type and/or the neighborhood. Based on the pivot tables and charts done in the sections entitled #1 – Data Categorization beginning on page 3 and #2 – Data Visualization beginning on page 4, brick construction tends to make the home price higher as does living in the east or west neighborhoods. In addition, there is other information that would be useful to gather for a better model. For example, the current interest rates and the lot size could also influence the housing price. We believe adding these variables into the models could possibly yield a higher Adjusted R2 and a stronger correlation model.

If you consider the number of offers being a ‘non-intuitive’ result, then this could be removed from the model if we could include additional independent variables that may influence price. However, we believe that as the numbers of offers increase, the price going down is intuitive. If a home is for sale for a long time (a variable that is not being capture, for example, number of months on the market) and has had multiple offers, the price would likely decrease without knowing more about why those offers were not accepted. For example, maybe the offers were more reasonable that the listing price was too high or there is some reason(s) why an offer was withdrawn. Having this additional information would also strengthen the model and analysis.