# Homework 1: House Price Analysis

# Team: A4

## **1.** Develop a categorization of your data using pivot tables. Develop two pivot tables of average price and average square feet by type of construction (brick) and neighborhood.

**Pivot Table of Average Price**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Average of Price** | **Neighborhood** |  |  |  |
| **Brick** | **East** | **North** | **West** | **Grand Total** |
| No | $ 117,750 | $ 108,584 | $ 148,230 | $ 121,958 |
| Yes | $ 135,468 | $ 118,457 | $ 175,200 | $ 147,769 |
| **Grand Total** | **$ 125,231** | **$ 110,155** | **$ 159,295** | **$ 130,427** |

**Pivot Table of Average Square**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Average of SqFt** | **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** |

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

**Pivot Table of Average Price**

**Pivot Table of Average Square**

## **3.** Perform a correlation analysis of all quantitative variables except ID.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Price* | *SqFt* | *Bedrooms* | *Bathrooms* | *Offers* |
| Price | 1 |  |  |  |  |
| SqFt | 0.552982243 | 1 |  |  |  |
| Bedrooms | 0.525926058 | 0.483807112 | 1 |  |  |
| Bathrooms | 0.523257758 | 0.522745301 | 0.414555956 | 1 |  |
| Offers | -0.313635883 | 0.336923352 | 0.11427061 | 0.1437934 | 1 |

**Which two variables have the largest magnitude correlation?**

Square Feet and Price have the largest magnitude correlation at 55%.

**Which two variables have the smallest magnitude correlation?**

Offers and Bedrooms have the smallest magnitude correlation at 11%.

**What does the largest magnitude imply if we perform a regression analysis next?**

The largest magnitude implies that a regression analysis will reveal *likely* a low P-value for these variables.

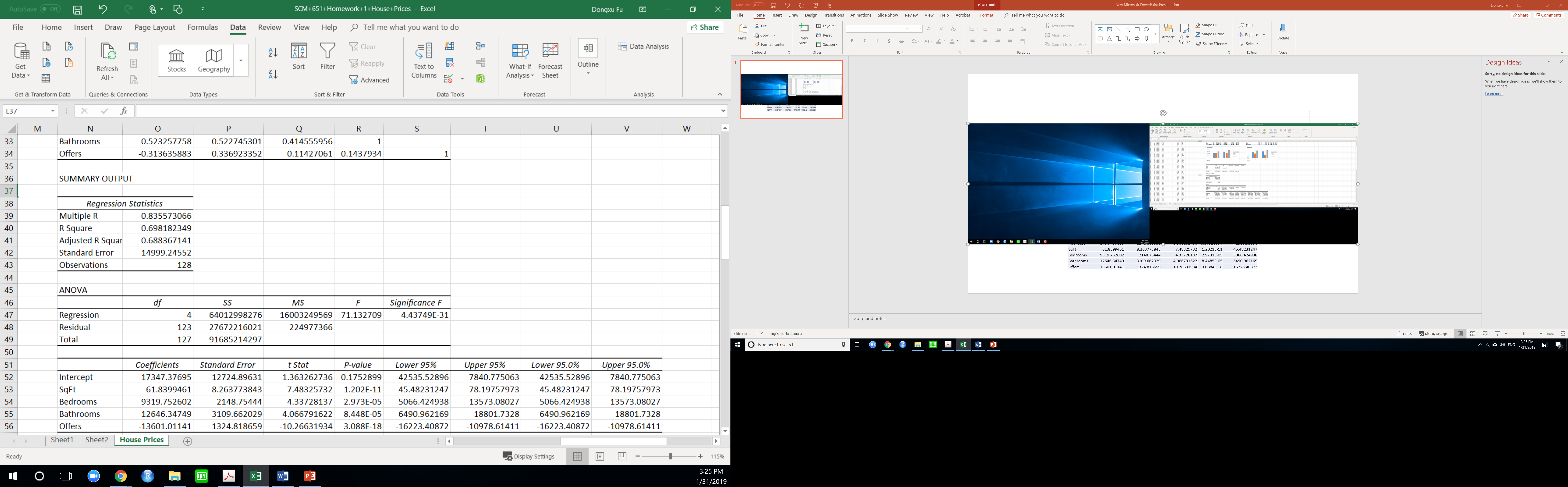
**Are there any negative correlations?**

Yes. Number of offers is negatively correlated with price at -31% correlation.

**Are these correlations intuitive?**

These correlations are mostly intuitive. The larger the house is by square feet, the more it will cost. This is to be expected for anyone in the market for a new home. Also, expensive homes can reasonably be expected to receive far fewer offers because the market of buyers is much smaller than that for more affordable homes.

## **4.** Perform an initial regression analysis of the quantitative variables excluding the ID.



### Which variables are statistically significant?

All variables, except for the intercept, are statistically significant with P-values under **.05 (95% confidence level).** The intercept may be misleading as at surface level because it seems to indicate a negative fixed cost; however, this figure may be ignored for our purposes because it is merely a continuation of the trends found in the data we have available. It is necessary to include merely for the integrity of our models.

### What does each coefficient mean in a real world sense?

The **coefficient for square feet**, 62, means that for every additional square foot a home has we can expect its price to go up by $62.

The **coefficient for number of bedrooms**, 9320, means that for every additional bedroom a home has we can expect the price of the home to go up by $9,320.

The **coefficient for number of bathrooms,** 12646, means that for every additional bathroom a home has we can expect the price of the home to go up by $12,646.

The **coefficient for number of offers**, -13601, means that for every additional offer made on a home has we can expect the price of the home to go down by $13,601.

**Are these coefficients intuitive? If not, why not?**

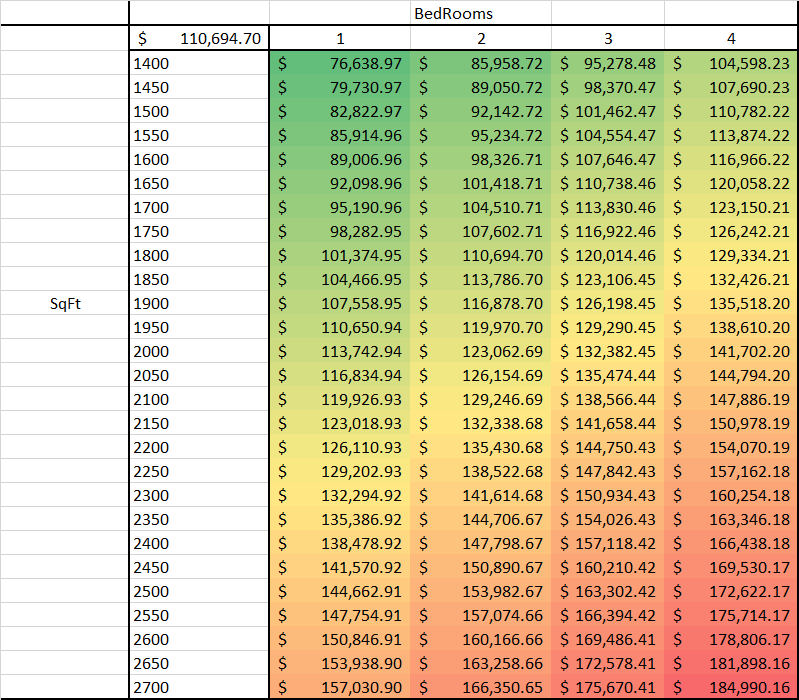
Every coefficient is intuitive except for offers. The negative correlation between the cost of a home and the number of offers made seems to indicate that more offers will actually drive down the cost of a home. However, this would imply a causal relationship between the two variables and would be an error in interpretation of the data. The reality is likely that homes with a lower cost are more likely to get more offers from a broader market than expensive homes.

**What does the R-squared mean?**

The R-squared indicates how much variability in cost can be explained by the other variables measured. In this case that's about 70%, which shows a high level of cost sensitivity to square feet, number of bedrooms & bathrooms, and offers made.

## **5.** Create a spreadsheet prediction of the model. Perform a two-way sensitivity analysis and use conditional formatting to highlight the results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Prediction Model** | |  | |  | |
| *Coefficients* | input | | output | |
| -17347.37695 | 1 | | **-17347.37695** | |
| 61.8399461 | 1800 | | **111311.903** | |
| 9319.752602 | 2 | | **18639.5052** | |
| 12646.34749 | 2 | | **25292.69497** | |
| -13601.01141 | 2 | | **-27202.02282** | |
|  | price | | **$ 110,694.70** | |



## **6.** What would explain non-intuitive results in your regression using the data which you were provided?

The relationship between offers received and the price of the home is the least intuitive; price seems to go up as number of offers goes down (with a negative correlation of -31%. One may speculate that as a property gets more offers the price should go up in a traditional supply and demand model. If we had more data such as the change in price per home over time along with offers made over time, we may be able to derive a more intuitive regression that shows demand intuitively driving the price up.

Worth noting is that West Neighborhood has the most expensive homes (average price: $159,294.87 and average 2.1 offers, maximum of 4 offers) and North has the cheapest (average price: $110,154.55 and average 3.1 offers, maximum of 6 offers). Thus, there is clearly a larger market for the more affordable homes in North Neighborhood. With more data we may be able to find more intuitive patterns here as well.

