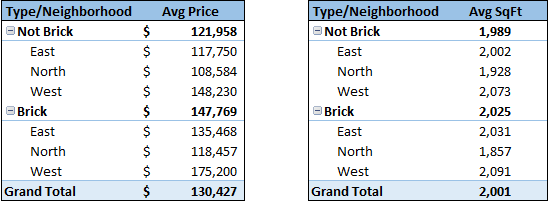
SCM 651

Homework #1

Team 4: Phoebe Sheahan, Emilio Monzalvo, James Eakins, Shaun Hall, J.R. Slouffman

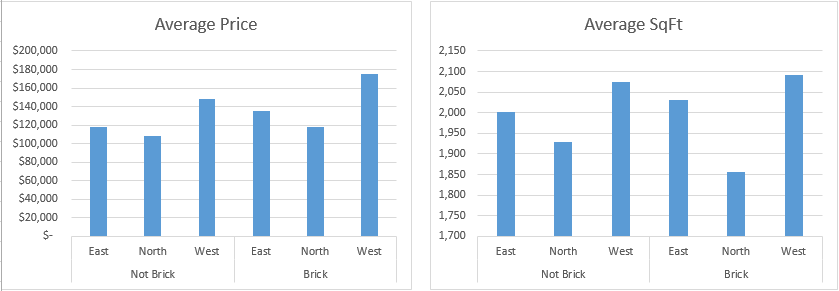
Question 1:

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:



Question 2:

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

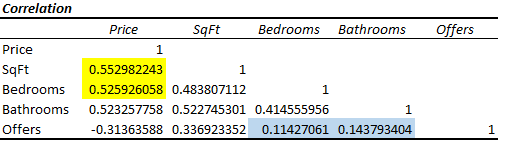


The average price for brick homes appears to be slightly more expensive. The histogram shows the relative comparison between neighborhoods. All of the neighborhoods’ bar magnitudes keep the same proportions when going from brick to no brick. West is always the most expensive, North is the cheapest, and East is middle of the road relative to pricing.

The average square feet of each house is relatively similar based on the neighborhood. The houses in the north always have the smallest square footage whether or not the house is brick or not. All houses in the west neighborhood have the most square footage whether or not they are brick or not. This shows that the housing square footage is more dependent on the houses location than brick being present.

Question 3:

Perform a correlation analysis of all quantitative variables except ID:



Which two variables have the largest magnitude correlation?

Highest Correlations - Price/SqFt and Price/Bedrooms

Which two variables have the smallest magnitude correlation?

Smallest Correlations - Bedrooms/Offers and Bathrooms/Offers

Are there any negative correlations?

Negative correlations - Price/Offers (makes sense, lowering the price should generate more offers)

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

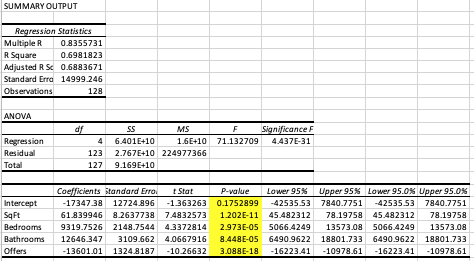
The largest magnitude implies that as price increases square footage and bedrooms would also increase. A regression analysis would depict the price increase relative to the square foot and bedrooms. More simply put, .553 (sq. ft) times variable X would produce a higher value unlike bathrooms and bedrooms which has a lesser correlation with offers thus producing a lower value in relative magnitude. The lower correlations would then not have as much of an impact relative to increasing the price.

Are these correlations intuitive? If not, why not?

These correlations are intuitive. As mentioned in negative correlations, lowering the price makes sense relative to an increase in offers. It also makes sense relative to square footage. An increase in square footage would logically drive the price up.

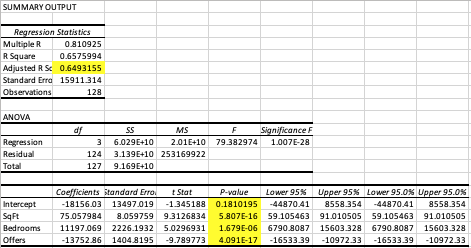
Question 4:

Perform an initial regression analysis of the quantitative variables excluding the ID:



In the above Regression model, we can identify the coefficients of the regression. In this example it tells us how much a one unit increase of each variable impacts the House Price. If we take a look at the Square Feet of the house, if we increase the square feet by one, then the price of the house will go up by $61.83. On the other hand, if the coefficient is negative like the variable Offers, then it will drive down the price. For every offer the house gets, the price will go down by $13601.01. One might expect the price of the house to go up if the demand goes up, but it does not seem to be the case in this regression model.

All of the variables have a p-value less than 0.05 which means that all of the variables are statistically significant in the regression model. Just to verify that this is true, since bathrooms has the smallest p-value, we can take it out in order to see the Adjusted R-Squared value goes down.

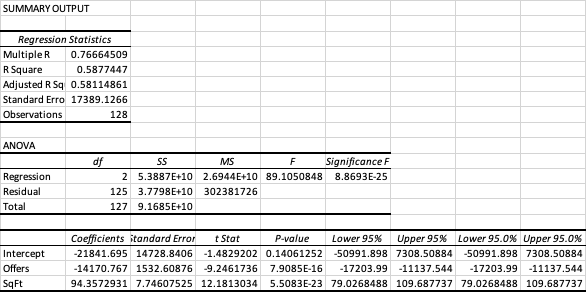


As seen above, the Adjusted R-Squared value did go down from 0.688 to 0.649, so it means that the regression model better predicts when the bathroom variable is included. In other words, the residual error is less if the Bathroom variable is included. It is also important to make sure we look at the Adjusted R-Squared instead of the R-Squared value for a Multiple Regression analysis since adding variables to the model usually drives the R-Squared value up, but the Adjusted R-Squared value is modified in order to be able to compare different Multiple Regression models.

Question 5:

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

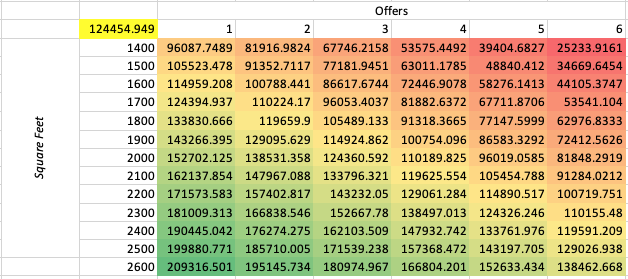
For the sensitivity analysis, we decided to pick the two variables that have the smallest p-values since they are the most statistically significant. As seen in question 4, Square Feet and Offers had the smallest p-values.

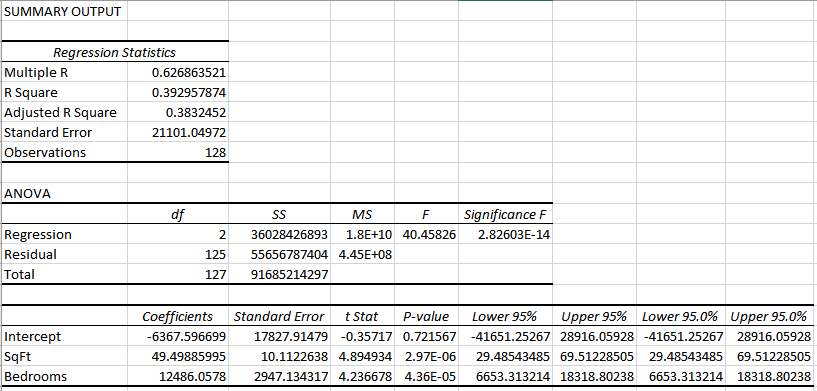


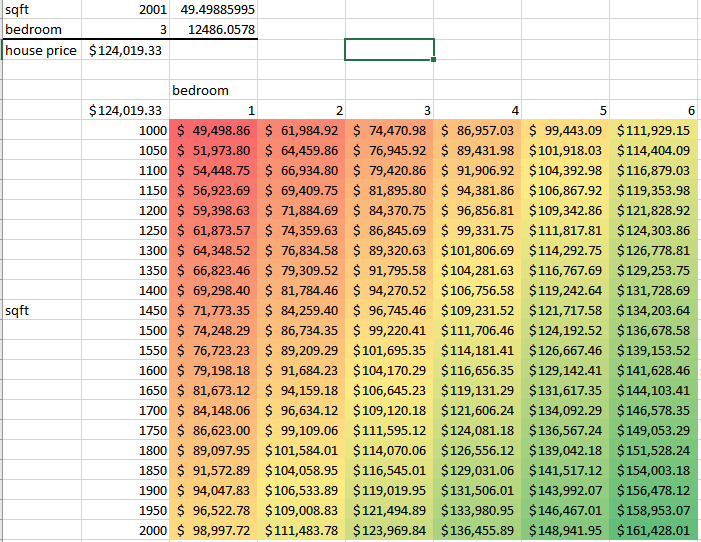
After the regression model was computed again using the two variables mentioned above, we took the coefficients in order to compute the two way sensitivity analysis. We used 2001 square feet since it is the mean value for the Square Feet and the mean of the offers which is 3.



$124,455 = -21841.695\*1+94.357\*2001+3\*(-14170.767)







$124,019.33 = (2001\*49.49885995) + (3 \* 12486.0578) - (12486.0578)

Question 6:

What would explain non-intuitive results in your regression using the data which you were provided? What additional data would assist you in explaining the non-intuitive results?

As mentioned in the fourth question, in a real world situation, you’d expect the demand to increase the price of the house. If there are many people putting up offers for the same house, you’d expect the offers to be going up. In this specific example, this does not seem to be the case. The number of offers is driving the house prices down. This could be a result of having a small dataset of 129 houses. Or this could also be a result of a house being overvalued, so they keep receiving a small offers that the seller is not willing to take.