Team #3

Joel Bravo

Stephanie Gardner

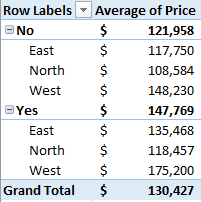
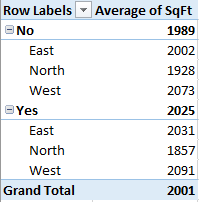
Kendra Danforth Osburn

Penematafele Seloti

Kenneth Smith

Homework #1

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.

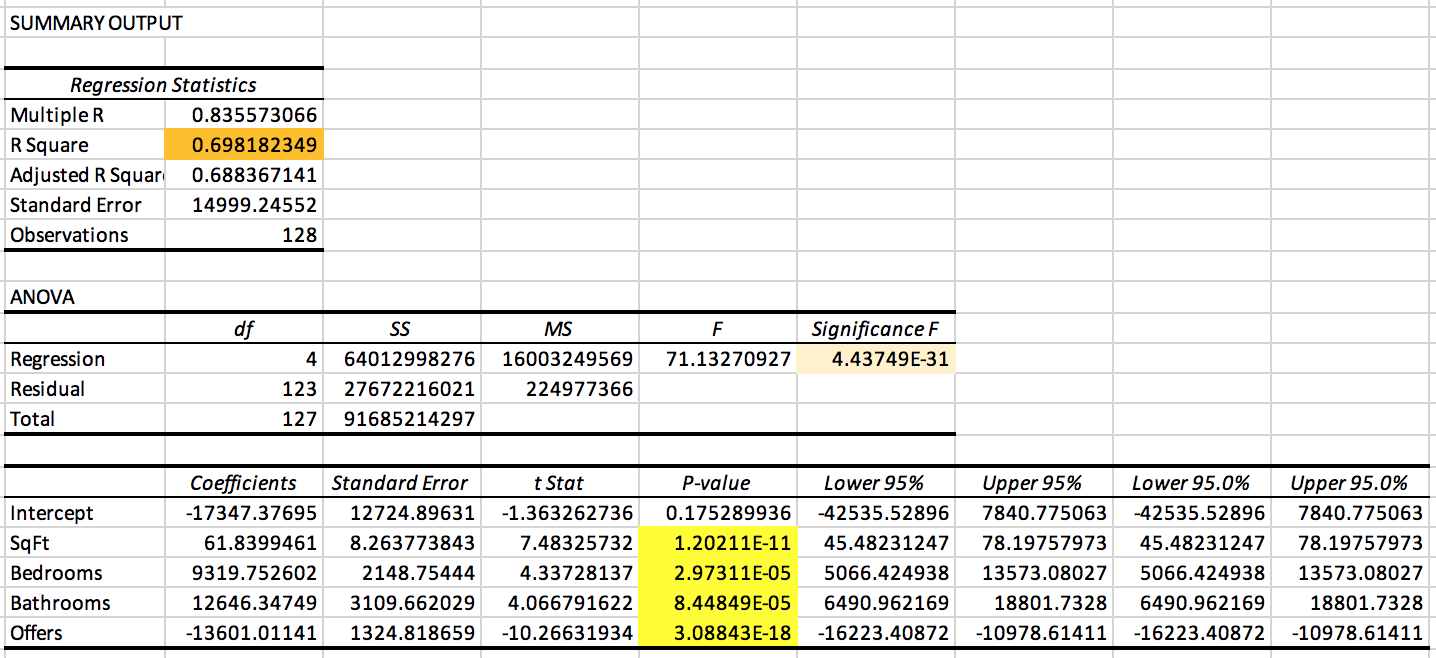
 

1. Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood.
2. Perform a correlation analysis of all quantitative variables except ID. Which two variables have the largest magnitude correlation? Which two variables have the smallest magnitude correlation? What does the largest magnitude imply if we perform a regression analysis next? Are there any negative correlations? Are these correlations intuitive? If not, why not?



* Based on the correlation, Price and SqFt have the largest magnitude correlation at .5529, while Bedrooms and Offers have the smallest magnitude correlation at .1142.
* The largest magnitude correlation, Price and SqFt, implies that as price goes up, so does the square footage of a home. Compared to all other variables, this combination has the strongest correlation, meaning these variables will increase more consistently compared to all other correlations.
* There is one negative correlation in the form of Price and Offers, which has a correlation of -.3136. In other words, as price increases, the number of offers generally decrease.
* All of these correlations are intuitive, as they more or less match up with expectations regarding house prices and offers. A larger home, measured in square footage, typically means the presence of additional bathrooms and bedrooms. Houses that are more expensive also typically are larger and have more bedrooms and bathrooms. Larger homes will typically attract more offers, but as a house increases in price the number of offers will typically drop.

1. Perform an initial regression analysis of the quantitative variables excluding the ID. Which variables are statistically significant? What does each coefficient mean in a real-world sense? Are these coefficients intuitive? If not, why not? What does the R-squared mean?

**INITIAL REGRESSION ANALYSIS:**

* + Square feet, bedrooms, bathrooms and offers (highlighted in yellow) are *all* statistically significant (assuming an alpha of 0.05).
  + In real world terms, the coefficients (holding everything else constant) indicate the average change in the response variable for one unit of change in the predictor variable. Essentially, this model is saying that the addition of a bathroom is worth roughly 12.6K, while the addition of a bedroom is worth 9.3K. The model is also saying that with each additional offer, the price is expected to decrease by 13.6K, which means that more people are putting offers on lower priced homes. Lastly, the model is saying that for every additional square foot, the price will increase roughly $62.
  + It's definitely intuitive that increasing square footage increases price. It’s also intuitive that increasing the number of bedrooms or bathrooms also increases the price. It’s less immediately intuitive that as offers increase, the price decreases. However, upon thoughtful examination, it is understandable that (in some situations) the less expensive homes would attract more buyers. See Question #6 for more detail.
  + The R squared means that the model explains almost 70% (69.8%) of the variability of the response data around the mean. In short, it provides an estimate of the relationship strength between the model and the response variables.

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





1. 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?
   1. Variables such as neighborhood or building quality will impact both number of offers and average price. For example, the average price for a home in the West neighborhood is 50K higher than a home in the North neighborhood and 35K higher than a home in the East neighborhood, yet the West had only 80 total offers while North had 135, and East had 115. Another thing to consider is the number of offers on non-brick houses (231) compared to the number of offers on brick houses (99). The North neighborhood had the lowest percentage of brick buildings, and the lowest average price but the highest number of offers by neighborhood. The non-intuitive results, namely that price decreases as offers increase, can be better understood by rephrasing the correlation to “as the price of the house increases, the number of people who can comfortably afford that type of house or that particular neighborhood decreases.”
   2. Better location data, neighborhood data, information on amenities (pools, schools, etc.) and condition of the house (new construction, renovation, foreclosure etc.) are all examples of additional data that would help us explain the non-intuitive results.