House Price Analysis

Prepared by Team # 2

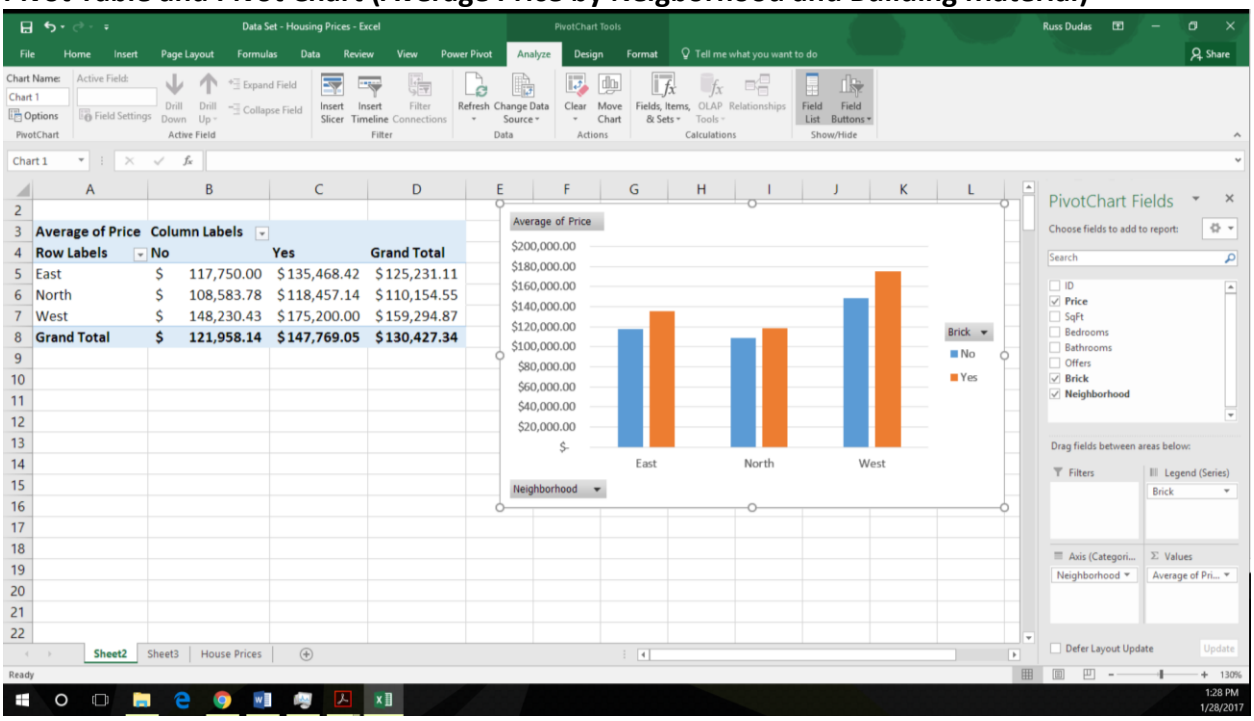
Russell Dudas
4/27/2017

Question 1 & 2

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

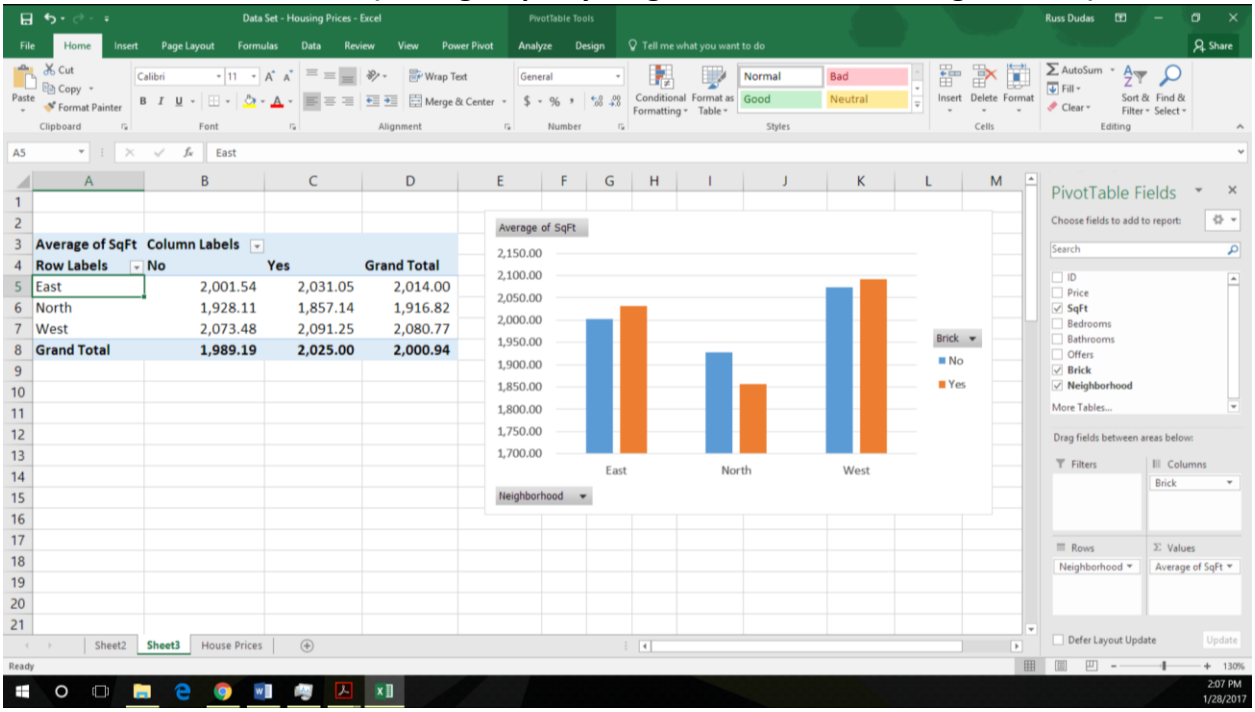
Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood

Figure 1
Pivot Table and Pivot Chart (Average Price by Neigborhood and Building Material)



- Figure 1 displays the Average price by neighborhood and building material.
- The Pivot table in Figure 1 shows that neighborhoods are listed in column A with the average cost by building material is listed across each row. (“Yes” indicating brick construction, and “No” indicating all other materials)
- The chart corresponding to the Pivot table in Figure 1 shows the home prices along the y- axis with the neighborhoods in the x-axis grouped together with their respective average cost by building material.
- The chart shows the chart colors indicate that the color orange corresponds with the homes built from Brick, and the blue corresponds with homes built of another material.

Figure 2
Pivot Table and Pivot Chart (Average SqFt by Neighborhood and Building Material)



- Figure 2 displays the Average SqFt of a home by neighborhood and building material.
- In Figure 2, we can see that the Pivot table again shows neighborhood in Column A and the average SqFt of each home along each row.
- In the corresponding chart for Figure 2 we can see the average SqFt of homes along the y-axis while neighborhood and building material variables are along the x-axis.
- Once again, we can see in this chart that the color orange corresponds with homes that are built of brick, and blue represents homes that are made with all other materials.
- In this section we can review both **Figure 1** and **Figure 2**, each is a Pivot Table, and Pivot Chart displaying the average price (Figure 1) and average SqFt (Figure 2) of homes in their respective neighborhoods, broken down via building material.

Question 3

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 sample of 128 homes, the price of a home and the total square footage show the greatest correlation.
- This correlation would be assumed, as the larger the home the more materials required, therefore increasing the overall price of the home. Interestingly, the total number of bedrooms and the total number of offers made show the lowest amount of correlation.
- The correlation analysis shows that the price of a home and the total number of offers are negatively correlated. This would be intuitive as the higher the price, the less affordable the home to the general population.

	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	0.143793404	1

Regression Statistics	
Multiple R	0.552982243
R Square	0.305789361
Adjusted R Square	0.300279752
Standard Error	22475.53365
Observations	128

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	28036363055	28036363055	55.50110765	1.30238E-11
Residual	126	63648851242	505149613		
Total	127	91685214297			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-10091.130	18966.104	-0.532	0.596	-47624.492	27442.232	-47624.492	27442.232
SqFt	70.226	9.426	7.450	0.000	51.572	88.881	51.572	88.881

- R Squared is only .305789361 indicating only 30.5% of the data fit the model.
- The P-value indicates statistical significance as it is well below .05
- The coefficient indicates that each unit increase in price, SqFt increases by 70.23

Question 4:

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?

We performed a multi-variate analysis of all quantitative variables and how they affect the overall price of a home. Below are the results:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.836
R Square	0.698
Adjusted R Square	0.688
Standard Error	14999.246
Observations	128

ANOVA

	df	SS	MS	F	Significance F
Regression	4	64012998276	1.6E+10	71.13	4.44E-31
Residual	123	27672216021	2.25E+08		
Total	127	91685214297			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-17347.38	12724.90	-1.36	0.175	-42535.53	7840.78	-42535.53	7840.78
SqFt	61.84	8.26	7.48	1.202E-11	45.48	78.20	45.48	78.20
Bedrooms	9319.75	2148.75	4.34	2.973E-05	5066.42	13573.08	5066.42	13573.08
Bathrooms	12646.35	3109.66	4.07	8.448E-05	6490.96	18801.73	6490.96	18801.73
Offers	-13601.01	1324.82	-10.27	3.088E-18	-16223.41	-10978.61	-16223.41	-10978.61

- From this analysis, we can conclude that the correlation between the output (price) and the variables are statistically correlated by looking at the Significance F statistic and it being relatively 0.

- Next, we look at the P-values for all the variables in this analysis. Square Foot, Bedrooms, Bathrooms, and Offers all have low P-values being relatively 0 which indicates that they do have an impact on the price and are statistically significant.
- In a real-world application, you would multiply the square footage, bedrooms, bathrooms, and offers by their respective coefficients. Add those values together and subtract the result by the intercept. This gives an approximate price of a home up to 69.8% of the cost (R Square statistic).
- The other 30.2% can be explained by outside variables not included in this model (location, additional building materials, etc.). What this means is that the following predictive formula can be applied to give an approximate estimate of the price of a home:

  Approx. Price = 61.84 SqFt + 9,319.75 Bed + 12,646.35 Bath – 13,601.01 Offers – 17,347.38

- The coefficients are intuitive in that one could reason that higher valued homes would likely have a greater square footage, more bedrooms, and bathrooms.
- We also concluded that offers have a negative correlation with the value of a home, potentially caused by more expensive homes being overpriced.
- Intuitively, lower priced homes will have more potential buyers, which is why lower priced homes receive more offers in our predictive formula.
- The R-squared value is a measure of fit and ranges between 0 and 1, where the greater the R-squared value, the closer our data points will fall to the regression line.

Question 5:

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

Sq Ft	1300	61.8399461			
Bedrooms	1	9319.752602		Incercept	-17347.38
Bathrooms	2	12646.34749			
Offers	3	-13601.01			
House Price	$ 56,853.97				

Bedrooms

Square Feet	$ 56,853	1	2	3	4	5
	1000	$ 38,301.98	$ 47,621.74	$ 56,941.49	$ 66,261.24	$ 75,580.99
	1050	$ 41,393.98	$ 50,713.73	$ 60,033.48	$ 69,353.24	$ 78,672.99
	1100	$ 44,485.98	$ 53,805.73	$ 63,125.48	$ 72,445.23	$ 81,764.99
	1150	$ 47,577.97	$ 56,897.73	$ 66,217.48	$ 75,537.23	$ 84,856.98
	1200	$ 50,669.97	$ 59,989.72	$ 69,309.48	$ 78,629.23	$ 87,948.98
	1250	$ 53,761.97	$ 63,081.72	$ 72,401.47	$ 81,721.23	$ 91,040.98
	1300	$ 56,853.97	$ 66,173.72	$ 75,493.47	$ 84,813.22	$ 94,132.98
	1350	$ 59,945.96	$ 69,265.72	$ 78,585.47	$ 87,905.22	$ 97,224.97

1400	$ 63,037.96	$ 72,357.71	$ 81,677.47	$ 90,997.22	$ 100,316.97
1450	$ 66,129.96	$ 75,449.71	$ 84,769.46	$ 94,089.22	$ 103,408.97
1500	$ 69,221.96	$ 78,541.71	$ 87,861.46	$ 97,181.21	$ 106,500.97
1550	$ 72,313.95	$ 81,633.71	$ 90,953.46	$ 100,273.21	$ 109,592.96
1600	$ 75,405.95	$ 84,725.70	$ 94,045.46	$ 103,365.21	$ 112,684.96
1650	$ 78,497.95	$ 87,817.70	$ 97,137.45	$ 106,457.21	$ 115,776.96
1700	$ 81,589.94	$ 90,909.70	$ 100,229.45	$ 109,549.20	$ 118,868.96
1750	$ 84,681.94	$ 94,001.69	$ 103,321.45	$ 112,641.20	$ 121,960.95
1800	$ 87,773.94	$ 97,093.69	$ 106,413.44	$ 115,733.20	$ 125,052.95
1850	$ 90,865.94	$ 100,185.69	$ 109,505.44	$ 118,825.19	$ 128,144.95
1900	$ 93,957.93	$ 103,277.69	$ 112,597.44	$ 121,917.19	$ 131,236.94
1950	$ 97,049.93	$ 106,369.68	$ 115,689.44	$ 125,009.19	$ 134,328.94
2000	$ 100,141.93	$ 109,461.68	$ 118,781.43	$ 128,101.19	$ 137,420.94
2050	$ 103,233.93	$ 112,553.68	$ 121,873.43	$ 131,193.18	$ 140,512.94
2100	$ 106,325.92	$ 115,645.68	$ 124,965.43	$ 134,285.18	$ 143,604.93
2150	$ 109,417.92	$ 118,737.67	$ 128,057.43	$ 137,377.18	$ 146,696.93
2200	$ 112,509.92	$ 121,829.67	$ 131,149.42	$ 140,469.18	$ 149,788.93
2250	$ 115,601.92	$ 124,921.67	$ 134,241.42	$ 143,561.17	$ 152,880.93
2300	$ 118,693.91	$ 128,013.67	$ 137,333.42	$ 146,653.17	$ 155,972.92
2350	$ 121,785.91	$ 131,105.66	$ 140,425.41	$ 149,745.17	$ 159,064.92
2400	$ 124,877.91	$ 134,197.66	$ 143,517.41	$ 152,837.16	$ 162,156.92
2450	$ 127,969.90	$ 137,289.66	$ 146,609.41	$ 155,929.16	$ 165,248.91
2500	$ 131,061.90	$ 140,381.65	$ 149,701.41	$ 159,021.16	$ 168,340.91
2550	$ 134,153.90	$ 143,473.65	$ 152,793.40	$ 162,113.16	$ 171,432.91
2600	$ 137,245.90	$ 146,565.65	$ 155,885.40	$ 165,205.15	$ 174,524.91
2650	$ 140,337.89	$ 149,657.65	$ 158,977.40	$ 168,297.15	$ 177,616.90
2700	$ 143,429.89	$ 152,749.64	$ 162,069.40	$ 171,389.15	$ 180,708.90
2750	$ 146,521.89	$ 155,841.64	$ 165,161.39	$ 174,481.15	$ 183,800.90
2800	$ 149,613.89	$ 158,933.64	$ 168,253.39	$ 177,573.14	$ 186,892.90
2850	$ 152,705.88	$ 162,025.64	$ 171,345.39	$ 180,665.14	$ 189,984.89
2900	$ 155,797.88	$ 165,117.63	$ 174,437.39	$ 183,757.14	$ 193,076.89
2950	$ 158,889.88	$ 168,209.63	$ 177,529.38	$ 186,849.14	$ 196,168.89
3000	$ 161,981.87	$ 171,301.63	$ 180,621.38	$ 189,941.13	$ 199,260.89
3050	$ 165,073.87	$ 174,393.62	$ 183,713.38	$ 193,033.13	$ 202,352.88
3100	$ 168,165.87	$ 177,485.62	$ 186,805.37	$ 196,125.13	$ 205,444.88
3150	$ 171,257.87	$ 180,577.62	$ 189,897.37	$ 199,217.12	$ 208,536.88
3200	$ 174,349.86	$ 183,669.62	$ 192,989.37	$ 202,309.12	$ 211,628.87
3250	$ 177,441.86	$ 186,761.61	$ 196,081.37	$ 205,401.12	$ 214,720.87
3300	$ 180,533.86	$ 189,853.61	$ 199,173.36	$ 208,493.12	$ 217,812.87
3350	$ 183,625.86	$ 192,945.61	$ 202,265.36	$ 211,585.11	$ 220,904.87
3400	$ 186,717.85	$ 196,037.61	$ 205,357.36	$ 214,677.11	$ 223,996.86
3450	$ 189,809.85	$ 199,129.60	$ 208,449.36	$ 217,769.11	$ 227,088.86
3500	$ 192,901.85	$ 202,221.60	$ 211,541.35	$ 220,861.11	$ 230,180.86

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?

Of the data provided, most results were as expected. However, certain factors such as the R-Squared values in questions 3 and 4, were less than expected. These lower R-Squared values suggested that other variables such as location and building materials play a greater role in determining home prices.

In addition, current market, demand and demographic data would help in explaining the non-intuitive results.