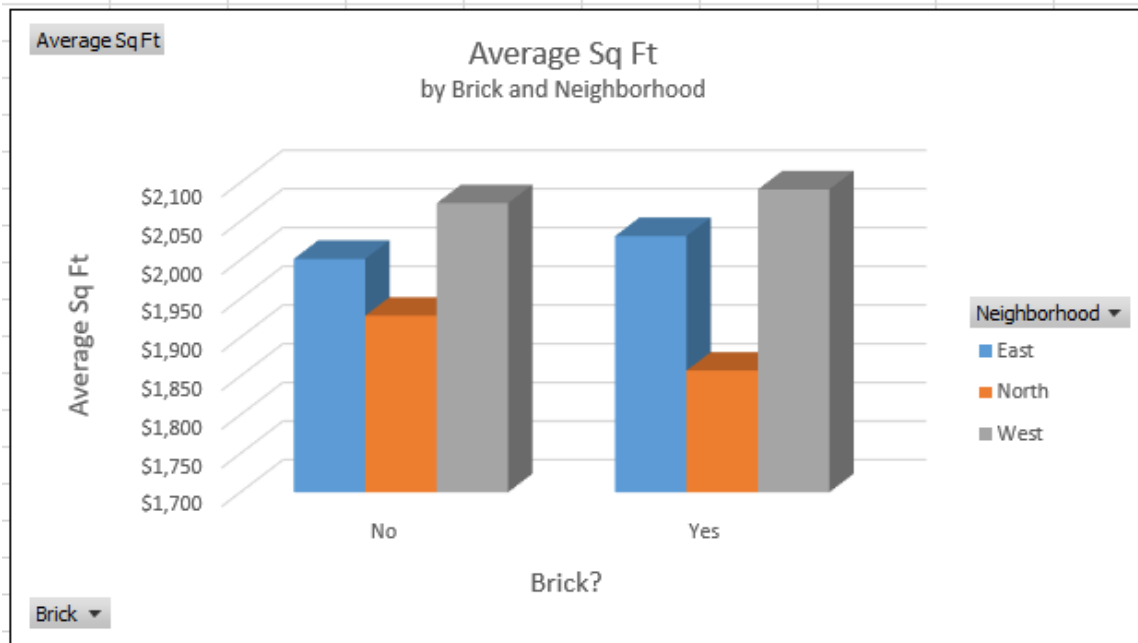
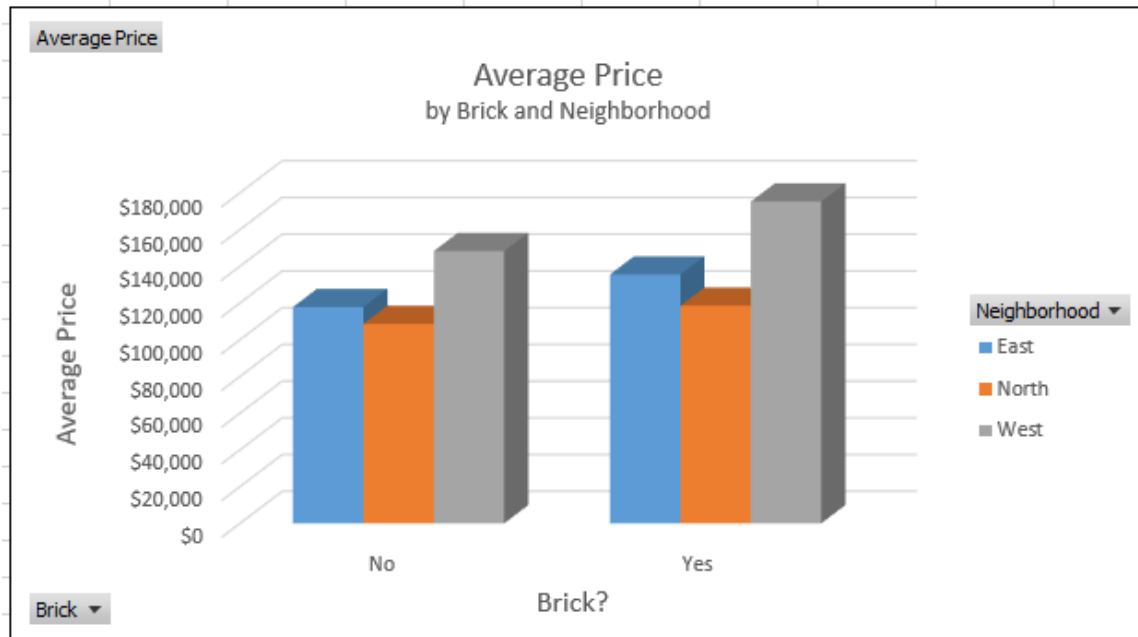


HW1_Team #92

1. Develop a categorization of your data using pivot tables. Develop two pivot tables: One pivot table of average price, varying type of construction (brick) and neighborhood as the two dimensions; a second pivot table of average square feet varying type of construction (brick) and neighborhood as the two dimensions (20%)

Average Price					Average Price by Brick and Neighborhood		
Brick?		East	North	West			
No		\$117,750	\$108,584	\$148,230			
Yes		\$135,468	\$118,457	\$175,200			
Average Sq Ft					Average Sq ft by Brick and Neighborhood		
Brick?		East	North	West			
No		\$2,002	\$1,928	\$2,073			
Yes		\$2,031	\$1,857	\$2,091			

2. Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood (10%)



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? (20%)

	<i>Price</i>	<i>SqFt</i>	<i>Bedrooms</i>	<i>Bathrooms</i>	<i>Offers</i>
<i>Price</i>	1				
<i>SqFt</i>	0.552982243	1			
<i>Bedrooms</i>	0.525926058	0.483807112	1		
<i>Bathrooms</i>	0.523257758	0.522745301	0.414555956	1	
<i>Offers</i>	-0.313635883	0.336923352	0.11427061	0.143793404	1

1st highest correlation: Price and SqFt = 0.553

2nd highest correlation: Bedrooms and Price = 0.526

1st lowest correlation: Bedrooms and Offers = 0.114

2nd lowest correlation: Bathrooms and Offers = 0.144

The independent variable with the largest magnitude is likely to be the best predictor for the dependent variable. In a regression model we would see that it also is the most statistically significance in the model.

Yes there is a negative correlation, Price and Offers at -0.314.

These correlations are intuitive for the most part, it makes sense that Price, Sq Ft, Bedrooms, and Bathrooms are coorelated with each other. For example, Sq footage tends to rise as bedrooms and bathrooms rise. That should goes without saying because more rooms means square footage, and consequently, price as well.

What is not as intuitive is the negative correlation between offer and price, one would initutive think that as offer increases, price might as well though there may be outside variables that could affect this intuition.

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? (25%)

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.835573066							
R Square	0.698182349							
Adjusted R Square	0.688367141							
Standard Error	14999.24552							
Observations	128							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	4	64012998276	16003249569	71.13270927	4.43749E-31			
Residual	123	27672216021	224977366					
Total	127	91685214297						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-17347.37695	12724.89631	-1.363262736	0.17528994	-42535.52896	7840.775063	-42535.52896	7840.775063
SqFt	61.8399461	8.263773843	7.48325732	0.00000000	45.48231247	78.19757973	45.48231247	78.19757973
Bedrooms	9319.752602	2148.75444	4.33728137	0.00002973	5066.424938	13573.08027	5066.424938	13573.08027
Bathrooms	12646.34749	3109.662029	4.066791622	0.00008448	6490.962169	18801.7328	6490.962169	18801.7328
Offers	-13601.01141	1324.818659	-10.26631934	0.00000000	-16223.40872	-10978.61411	-16223.40872	-10978.61411

Sq Feet, Bedrooms, Bathroom, and Offer are statistically significant because they are all less than 0.05 in p-value. The intercept or fixed cost has a higher p-value, making it less confident. Each coefficient is the amount that the dependent variable moves in response to an increase of one unit for the corresponding independent variable. In other words, if the coefficient were 61.8 (which is the case of the SqFt variable in the regression here), that means that for every increase of 1 for SqFt the estimated price of the home increases by 61.8. In other words, one square foot costs \$62 (rounded up from 61.8), one bedroom costs \$9,320, one bathroom costs \$12,646

Much like the correlation analysis, Price, Sq Ft, Bedroom, and Bathroom coefficients are intuitive, but for offer, it's not intuitive that for every 1 offer, the price decreases by \$13,601.

The R2 measures how good of predictors our variables are for predicting the outcome variable (price in this case). The highest possible value would be 1 meaning that the model is perfect. A value of .69 says that we can be confident with our model, but there is still some uncertainty. If I were a business analyst, the next thing I would want to do is see if adding in any other variables might increase that R2 a little bit more so we can have more confidence in the model.

5. Create a spreadsheet prediction of the model. Perform a two-way sensitivity analysis and use conditional formatting to highlight the results. (15%)

Full Prediction Formula						
$\hat{y} = -17347 + 62*(\text{SqFt}) + 9320*(\text{Bedrooms}) + 12646*(\text{Bathrooms}) + -13601*(\text{Offers})$						
	SqFt	2001				
	Bedroom	3				
	Bathroom	3				
	Offers	3				
	Price	131490				
		Bedrooms				
	131490	1	2	3	4	5
Sqft	1500	↓ 81868	↓ 91188	↓ 100508	↓ 109828	→ 119147
	1600	↓ 88052	↓ 97372	↓ 106692	→ 116012	→ 125331
	1700	↓ 94236	↓ 103556	↓ 112876	→ 122196	→ 131515
	1800	↓ 100420	↓ 109740	→ 119060	→ 128380	→ 137699
	1900	↓ 106604	→ 115924	→ 125244	→ 134564	→ 143883
	2000	↓ 112788	→ 122108	→ 131428	→ 140748	↑ 150067
	2100	→ 118972	→ 128292	→ 137612	→ 146932	↑ 156251
	2200	→ 125156	→ 134476	→ 143796	↑ 153116	↑ 162435
	2300	→ 131340	→ 140660	↑ 149980	↑ 159300	↑ 168619
	2400	→ 137524	→ 146844	↑ 156164	↑ 165484	↑ 174803
	2500	→ 143708	↑ 153028	↑ 162348	↑ 171668	↑ 180987

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? (10%)

The following scenarios and if provided this additional data would help explain the non-intuitive results:

- Home inspection outcome – depending on the result of a home inspection, the price can be skewed to be lower.
- Longevity of a house on market – how long the house was on the market, how many times has the price lower, etc.
- Seasonality – there could be more homes on sale or more buyers looking depending on the time of year.
- Specific location/school district – even though the data breaks down the geographic quadrant in relation to the city, additional detail such as school zoning, restaurants, etc. would make a difference in the data.
- The type of housing market, if it's a buyer's market meaning there is a greater supply of homes and a lower demand of buyers. This could explain why the price would drop based on number of offers.