HOUSING PRICE ANALYSIS

BUSINESS ANALYTICS
HOMEWORK 1

TEAM 2

YODIT AYALEW, AMANDA NORWOOD, AMANDA N HIGGS, ELI RIVERA & BRETT BASTIANELLI

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

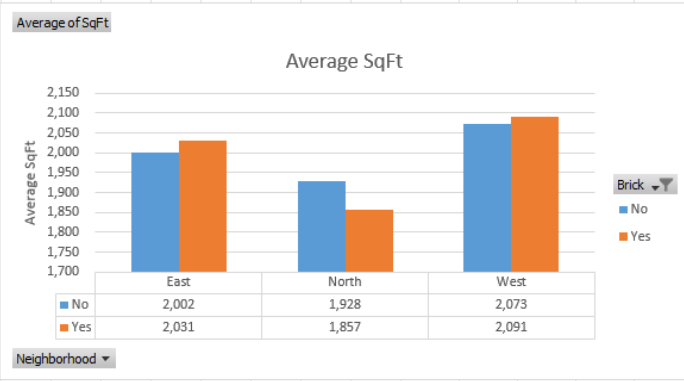
    - Combined with Question #2 below:

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

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

- The Pivot table above shows the average price of a home by neighborhood and building material.
- The Y axis represents the price of the homes. The X axis represents the neighborhoods that are grouped together by average cost by building material
- The chart displays two colors: blue and orange. Blue represents homes that were not built from brick, while orange represents homes that were built from brick.
    - "Yes" indicates brick homes and "No" indicates any other building material other than brick.
- Homes in the West neighborhood have the highest average cost, with or without brick construction. Brick homes are more expensive than other building materials.
- Homes with brick construction are more expensive than other construction materials in all three neighborhoods.

Average of SqFt	Column Labels		
Row Labels	No	Yes	Grand Total
East	2,002	2,031	2,014
North	1,928	1,857	1,917
West	2,073	2,091	2,081
Grand Total	1,989	2,025	2,001



- The Pivot table above depicts the average square feet of a home by the neighborhood and building material.
- The Y axis represents the average square feet, and the X axis represents the average square feet by neighborhood and construction material.
- The chart displays two colors: blue and orange. Blue represents homes that were not built from brick and orange represents homes that were built from brick.
    - "Yes" indicates brick homes while "No" indicates any other building material other than brick.
- Based on average square feet, the North neighborhood homes are smaller compared to East and West neighborhoods regardless of the type of construction.

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?

	Price	*SqFt*	*Bedrooms*	*Bathrooms*	*Offers*
Price	1				
SqFt	0.55298224	1			
Bedrooms	0.52592606	0.48380711	1		
Bathrooms	0.52325776	0.5227453	0.41455596	1	
Offers	-0.31363588	0.33692335	0.11427061	0.1437934	1

- As we can see from the above correlation, price and sq. ft. have a strong relationship. Price and sq. ft. are the two variables with largest magnitude correlation with a correlation of 0.552982243
- The weakest correlation is 0.11427061, which is the correlation between offers and bedrooms. Therefore, offers and bedrooms have the smallest magnitude correlation.
- If we run a regression with the largest magnitude variables, then result will be as below:

SUMMARY OUTPUT								
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	-10091.12991	18966.10399	-0.532061298	0.595620462	-47624.49215	27442.23233	-47624.49215	27442.23233
SqFt	70.22631824	9.42646968	7.449906553	1.30238E-11	51.57161187	88.88102462	51.57161187	88.88102462

- The regression confirms the strong correlation between price and sq. ft. as the p-value is very small (P-Value < 0.05), which implies sq. ft. has a strong impact on house prices. The correlation between price and offer is negative. A negative correlation normally describes the extent to which two variables move in opposite directions and it's not intuitive.

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?

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	0.175289936	-42535.52896	7840.775063	-42535.52896	7840.775063
SqFt	61.8399461	8.263773843	7.48325732	1.20211E-11	45.48231247	78.19757973	45.48231247	78.19757973
Bedrooms	9319.752602	2148.75444	4.33728137	2.97311E-05	5066.424938	13573.08027	5066.424938	13573.08027
Bathrooms	12646.34749	3109.662029	4.066791622	8.44849E-05	6490.962169	18801.7328	6490.962169	18801.7328
Offers	-13601.01141	1324.818659	-10.26631934	3.08843E-18	-16223.40872	-10978.61411	-16223.40872	-10978.61411

- Which variables are statistically significant?
  - Sq. ft., bedrooms, bathrooms and offers are statistically significant since their P-values are < 0.05. The intercept is not statistically significant since the P-value is > 0.05.

- What does each coefficient mean in a real-world sense?
    - For this analysis we will be looking the coefficients section:
    - SqFt: The price of home increase by $61 for every sq. ft.
    - Bedroom: The price of home will increase by $9,319 for every additional bedroom
    - Bathrooms: The price of home will increase by $12,646 for every additional bathroom.
    - Offers: The price of home will decrease by $13,601 for every additional offer received
- Are these coefficients intuitive? If not, why not?
    - Yes, sq. ft., Bedrooms and Bedrooms are intuitive. Generally, the bigger the home in sq. ft. and bedrooms, the more it costs to build.

    - As a group, we found that Offer can be both intuitive and not intuitive from a real-world aspect.
        - **Not Intuitive:** Offer is not intuitive, in real world, when a seller has more offers, they normally increase the price of the house, but the regression shows the opposite. It would make sense that more offers create a bidding war, thus driving up the price but that is not what is happening.
        - **Intuitive:** We felt on the contrary to offers not being intuitive, it can actually be intuitive based on knowing more about the housing market. If a house is getting more offers but at a consistently lower amount than the initial set price, the sellers may degrease the price of the home to better match the offers. This can happen if the housing market is crashing and the seller wants to get rid of the home. Another possible scenario is the house could be extremely expensive, making it less affordable to the general population. This would limit the amount of home buyers that can afford this house. Again, if the limited number of home buyers are making offers that are lower than initially anticipated, the seller may need to decrease the price.
- What does the R-squared mean?
    - R-squared is used to explain the degree to which input variables (predictor variables) explain the variation of output variables (predicted variables) so in this case we can say 69.8% of variation in price (predictor variable) can be explained by variation of the predicted variables (SqFt, Bedrooms, Bathrooms and Offers)

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

	Coefficients	Value	Coeffecent * Value
Intercept	-17347.37695	1	-17347.37695
SqFt	61.8399461	2000	123679.8922
Bedrooms	9319.752602	3	27959.25781
Bathrooms	12646.34749	2	25292.69497
Offers	-13601.01141	3	-40803.03424
		Total	**$ 118,781.43**

		SqFt											
	$118,781.43	1500	1700	1900	2100	2300	2500
Offer	1	$115,063.48	$ 127,431.47	$ 139,799.46	$ 152,167.45	$ 164,535.44	$ 176,903.43
	2	$101,462.47	$ 113,830.46	$ 126,198.45	$ 138,566.44	$ 150,934.43	$ 163,302.42
	3	$ 87,861.46	$ 100,229.45	$ 112,597.44	$ 124,965.43	$ 137,333.42	$ 149,701.41
	4	$ 74,260.45	$ 86,628.44	$ 98,996.43	$ 111,364.42	$ 123,732.41	$ 136,100.40
	5	$ 60,659.44	$ 73,027.43	$ 85,395.42	$ 97,763.41	$ 110,131.39	$ 122,499.38
	6	$ 47,058.43	$ 59,426.42	$ 71,794.40	$ 84,162.39	$ 96,530.38	$ 108,898.37

- The reason we chose SqFt and Offer to do the 2-way sensitivity analysis is because the offer has a negative coefficient and we wanted to further investigate the behavior.
- Based on the above sensitivity analysis, we got the highest price when our SqFt is 2500 but interestingly it has got only one offer. Also, as you can see above, the predicted price of a house with 3-bedroom, 2 bathrooms and 3 offers are $118,781

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?
    - Using the data provided, what could explain the intercept and offers being negative is that as more offers are made on a house below the asking price, the sellers may decrease their initial price point. The intercept would be negative because as more houses are on the market, it would create a bidding war between buyers, trying to maximize their utility and buy a house for the cheapest price point. As more houses enter market at a lower point, sellers will decrease their house price.
        - Demographic data of the area would greatly help explain the non-intuitive results. Information such as the average income, year the homes were built, or the geographic location would help explain the non-intuitive results. It is likely that the more expensive

the home, the less affordable it is to the general population, but we cannot clearly define the relationship between the two without the average income of the area.

- In addition, the current market or demand of the area would assist as well. We can see that the West neighborhood is distinctively more expensive regardless of being built with or without brick. If we had information on the geographic location or year built, then we could explain why the West neighborhood is more desirable.
- An additional set of statistics that would be beneficial to take note of would include the Crime rate, school system ratings, other homes sold in that area, demographic information, and a big data set. This give a clearer picture of the overall neighborhood where the homes are sold.