

# *HOUSING PRICE ANALYSIS*

Team 5, HW #1

*Team 5: Chris Webster, Ryan Timbrook, David Eboh, Joseph Hill*

*Due: 7/25/2019*

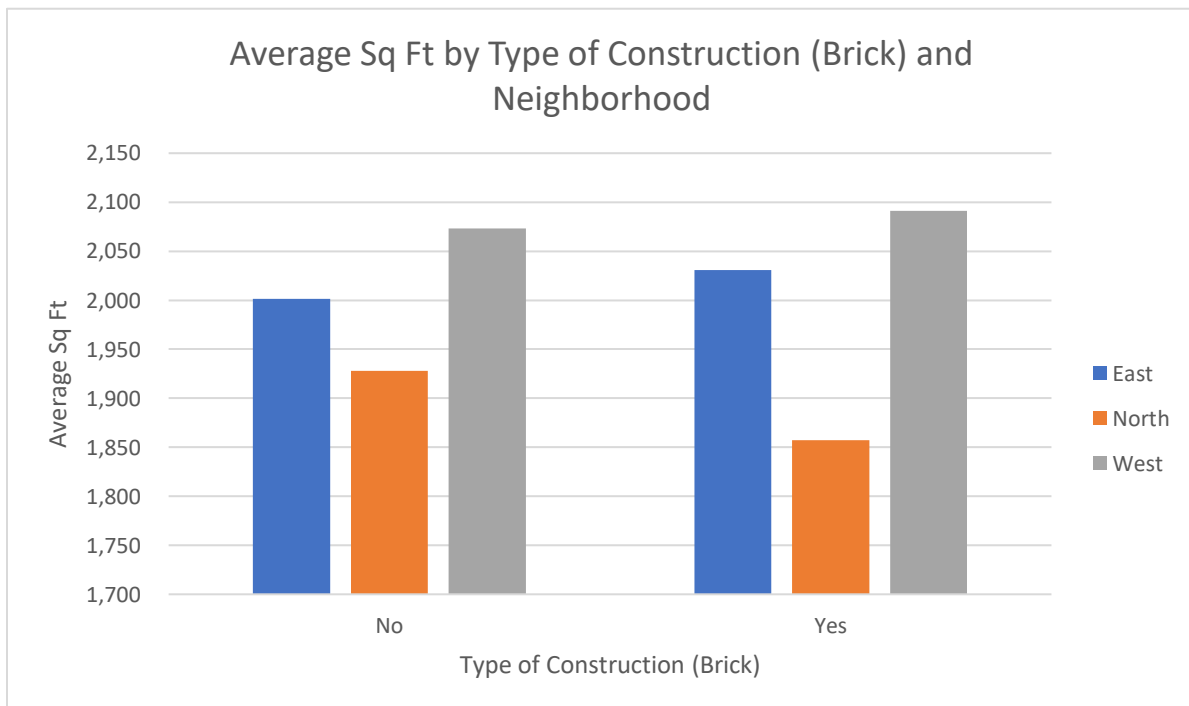
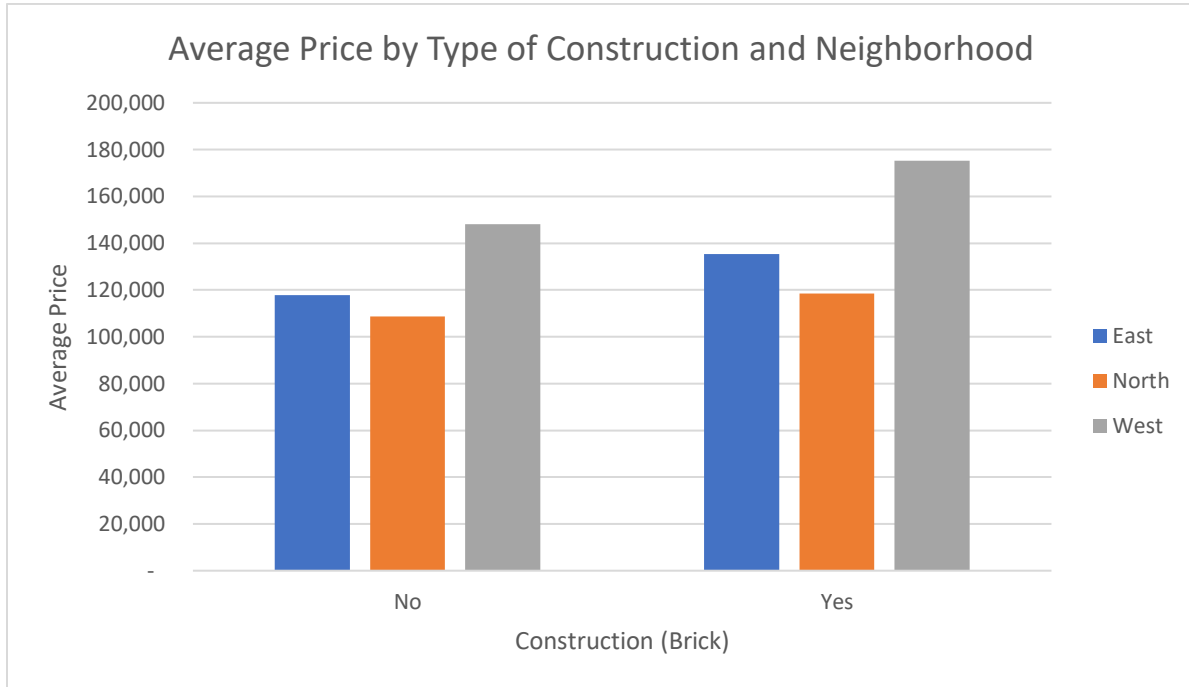
- 1) Develop a categorization of your data using pivot tables. Develop two pivot tables of average price and average square feet by type of construction (brick) and neighborhood (20%)

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

Average of SqFt	Column Labels			
Type of Construction	East	North	West	Grand Total
No	2,002	1,928	2,073	1,989
Yes	2,031	1,857	2,091	2,025
<b>Grand Total</b>	<b>2,014</b>	<b>1,917</b>	<b>2,081</b>	<b>2,001</b>

Team 5  
HW #1

- 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>
Price	1				
SqFt	0.55298224	1			
Bedrooms	0.52592606	0.48380711	1		
Bathrooms	0.52325776	0.5227453	0.41455596	1	
Offers	-0.3136359	0.33692335	0.11427061	0.1437934	1

The two variables with the largest magnitude correlation are Price and Sq Ft. This implies that there is a very strong relationship between the price and Sq Ft of a house.

The two variables with the smallest magnitude correlation are Bedrooms and Offers. This implies that there is very little relationship between the number of bedrooms and the number of offers for a house.

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.552982243							
R Square	0.305789361							
Adjusted R Square	0.300279752							
Standard Error	22475.53365							
Observations	128							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	28036363055	28036363055	55.50110765	1.30238E-11			
Residual	126	63648851242	505149613					
Total	127	91685214297						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
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

As the regression above shows, the P-value is very small. Since  $P < .05$ , there is a strong relationship between Price and Sq Ft.

The largest magnitude implies that Sq Ft will have the strongest effect on the price of a home. It will have the strongest relationship. There is one negatively correlation between Offers and Price. This is slightly counter intuitive, as increasing the number of offers would increase competition. With this increased competition, one would assume a higher price. Looking deeper at the data with a scatterplot of Price and number of offers would likely show a non-linear relationship.

- 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	6.4013E+10	1.6003E+10	71.1327093	4.4375E-31			
Residual	123	2.7672E+10	224977366					
Total	127	9.1685E+10						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-17347.37695	12724.8963	-1.3632627	0.17528994	-42535.529	7840.77506	-42535.529	7840.77506
SqFt	61.8399461	8.26377384	7.48325732	1.2021E-11	45.4823125	78.1975797	45.4823125	78.1975797
Bedrooms	9319.752602	2148.75444	4.33728137	2.9731E-05	5066.42494	13573.0803	5066.42494	13573.0803
Bathrooms	12646.34749	3109.66203	4.06679162	8.4485E-05	6490.96217	18801.7328	6490.96217	18801.7328
Offers	-13601.01141	1324.81866	-10.266319	3.0884E-18	-16223.409	-10978.614	-16223.409	-10978.614

All variables, but the intercept are statistically significant as they all have a p-value less than 0.05.

Sq Ft Interpretation: For each additional Sq Foot, the price of the home will increase by \$61.84

Bedrooms Interpretation: For each additional Bedroom, the price of the home will increase by \$9,319.75

Bathrooms Interpretation: For each additional Bathroom, the price of the home will increase by \$12,646.35

Offers Interpretation: For each additional Offer, the price of the home will decrease by \$13,601.01

Yes, all but the Offers coefficient is intuitive. One would expect that increasing the number of offers on a home would increase the price due to increased competition. However, the linear regression suggests the opposite effect. However, given the correlation matrix, it is expected that the coefficient for offers is negative.

Since this is a multivariate linear regression, the adjusted R-squared can be interpreted as: 68.8% of the variation in Price can be explained by the variation in SQ Ft, Bedrooms, Bathrooms, and Offers.

## HW #1

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

Variable	Coefficient	WhatIf (Input)																
Intercept	-17347.377	1																
Sq Ft	61.839946	2500																
Bedrooms	9319.7526	3																
Bathrooms	12646.347	2																
Offers	-13601.011	3																
Predicted Price	= Intercept + Sq Ft * Coefficient + Number of Bedrooms * Bedrooms Coefficient + Number of Bathrooms * Bathrooms Coefficient + Number of Offers * Offers Coefficient																	
Predicted Price	149701.41																	
			Sq Ft															
			\$149,701.41		500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500			
Number of Offers	1	\$	53,223.54	\$ 84,143.51	\$ 115,063.48	\$ 145,983.46	\$ 176,903.43	\$ 207,823.40	\$ 238,743.38	\$ 269,663.35	\$ 300,583.32	\$ 331,503.29	\$ 362,423.27					
	2	\$	39,622.53	\$ 70,542.50	\$ 101,462.47	\$ 132,382.45	\$ 163,302.42	\$ 194,222.39	\$ 225,142.36	\$ 256,062.34	\$ 286,982.31	\$ 317,902.28	\$ 348,822.26					
	3	\$	26,021.51	\$ 56,941.49	\$ 87,861.46	\$ 118,781.43	\$ 149,701.41	\$ 180,621.38	\$ 211,541.35	\$ 242,461.33	\$ 273,381.30	\$ 304,301.27	\$ 335,221.25					
	4	\$	12,420.50	\$ 43,340.48	\$ 74,260.45	\$ 105,180.42	\$ 136,100.40	\$ 167,020.37	\$ 197,940.34	\$ 228,860.31	\$ 259,780.29	\$ 290,700.26	\$ 321,620.23					
	5	\$	(1,180.51)	\$ 29,739.46	\$ 60,659.44	\$ 91,579.41	\$ 122,499.38	\$ 153,419.36	\$ 184,339.33	\$ 215,259.30	\$ 246,179.28	\$ 277,099.25	\$ 308,019.22					
	6	\$	(14,781.52)	\$ 16,138.45	\$ 47,058.43	\$ 77,978.40	\$ 108,898.37	\$ 139,818.35	\$ 170,738.32	\$ 201,658.29	\$ 232,578.26	\$ 263,498.24	\$ 294,418.21					
	7	\$	(28,382.53)	\$ 2,537.44	\$ 33,457.42	\$ 64,377.39	\$ 95,297.36	\$ 126,217.33	\$ 157,137.31	\$ 188,057.28	\$ 218,977.25	\$ 249,897.23	\$ 280,817.20					
	8	\$	(41,983.54)	\$ (11,063.57)	\$ 19,856.40	\$ 50,776.38	\$ 81,696.35	\$ 112,616.32	\$ 143,536.30	\$ 174,456.27	\$ 205,376.24	\$ 236,296.22	\$ 267,216.19					

According to the 2-way sensitivity analysis, the highest predicted price occurs when the home is 5500 Sq Ft and receives only 1 offer on the house.

If the home is 2500 Sq Ft, with 3 bedrooms, 2 bathrooms and 3 offers, the predicted price is \$149,701.41.

- 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%)

One possible explanation for the non-intuitive results in the regression would be if Price and Offers were not linearly related, but rather displaying a quadratic relationship. As the number of offers increase, price increases to a point. However, at higher prices, less people would be able to purchase the home, thus limiting the number of offers. This could drive price down as there is less competition.

A larger data set, with some additional data on how many homes have sold in the area recently? How is the school system? What is the crime rate? Comparable homes sold? Etc. This would help add additional clarity to the model and would hopefully increase our confidence in the output, as well as the amount of variation in prices explained by the variation in the inputs.