Housing Price Analysis Homework 1

4/24/2021

Team 71
Toby Anderson
Garen Moghoyan
Daniel Caley
Michael Johnson

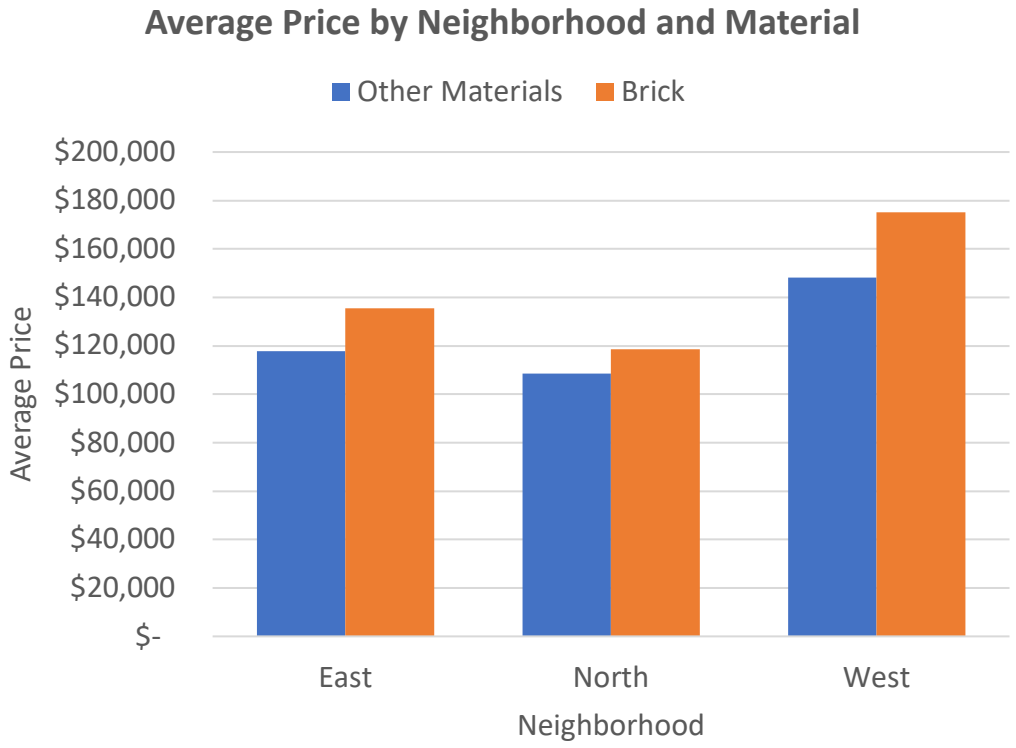
Outline and grading criteria:

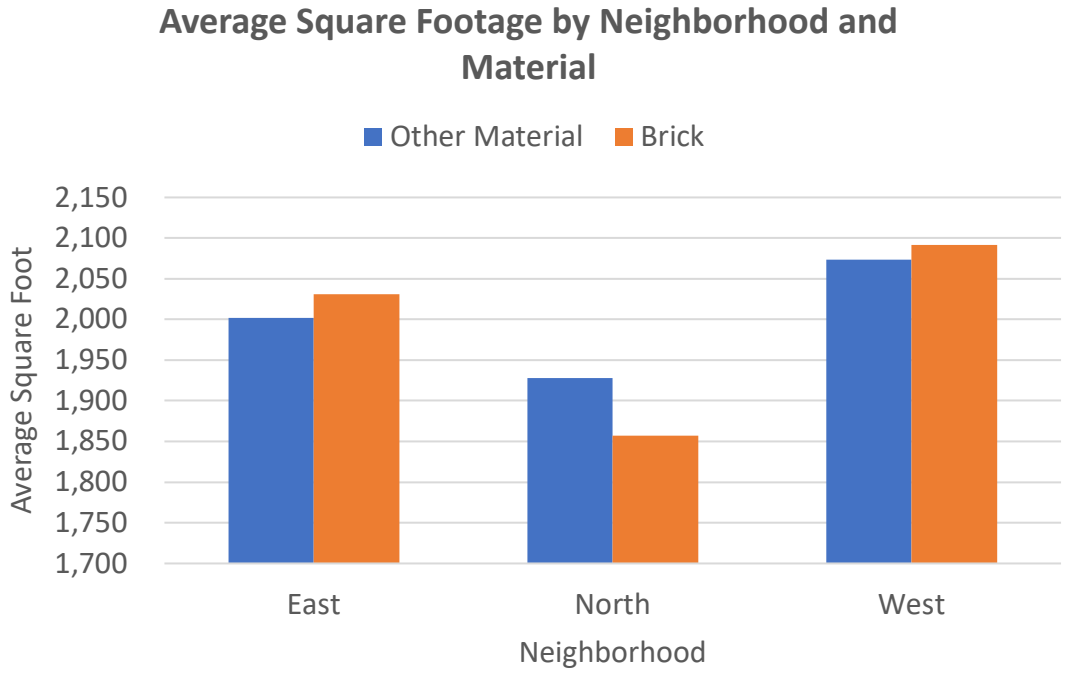
1. Develop a categorization of your data using pivot tables. Develop two pivot tables: One pivot table of average price, varying type of construction (brick) and neighborhood as the two dimensions; a second pivot table of average square feet varying type of construction (brick) and neighborhood as the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

Average of Price	Brick		
Neighborhood	Other	Brick	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**

Average of SqFt	Brick		
Neighborhood	Other	Brick	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**

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%)

	Price	*SqFt*	*Bedrooms*	*Bathrooms*	*Offers*
Price	100.0%				
SqFt	55.3%	100.0%			
Bedrooms	52.6%	48.4%	100.0%		
Bathrooms	52.3%	52.3%	41.5%	100.0%	
Offers	-31.4%	33.7%	11.4%	14.4%	100.0%

Correlation Analysis

Question	*Answer*
1. Which two variables have the largest magnitude correlation?	Square Feet and Price, with a 0.55 correlation.
2. Which two variables have the smallest magnitude correlation?	Bedrooms and Offers, with a 0.11 correlation.
3. What does the largest magnitude imply if we perform a regression analysis next?	We would expect Square Feet to be a significant positive factor in the Price formula.
4. Are there any negative correlations?	Offers and Price have a negative correlation of -0.31.
5. Are these correlations intuitive?	Everything is intuitive except the offer correlation.
a. If not, why not?	If we are predicting the final sales price of a home, we would expect an increase in offers to drive up the price.

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%)

5. Regression Statistics	
Multiple R	0.84
R Square	0.70
Adjusted R Square	0.69
Standard Error	14,999.25
Observations	128.00

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4.00	64012998276.24	16003249569.06	71.13	0.00
Residual	123.00	27672216020.63	224977366.02		
Total	127.00	91685214296.88			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-17,347.38	12,724.90	-1.36	0.18
SqFt	61.84	8.26	7.48	0.00
Bedrooms	9,319.75	2,148.75	4.34	0.00
Bathrooms	12,646.35	3,109.66	4.07	0.00
Offers	-1,3601.01	1,324.82	-10.27	0.00

Overall What does our data look like. Our Inputes to the linear regression model should fit in the below sample.

Price		*SqFt*		*Bedrooms*		*Bathrooms*		*Offers*	
Mean	$ 130,427	Mean	2,001	Mean	3	Mean	2	Mean	3
Standard Error	$ 2,375	Standard Error	19	Standard Error	0	Standard Error	0	Standard Error	0
Median	$ 125,950	Median	2,000	Median	3	Median	2	Median	3
Mode	$ 117,800	Mode	1,920	Mode	3	Mode	2	Mode	3
Standard Deviation	$ 26,869	Standard Deviation	212	Standard Deviation	1	Standard Deviation	1	Standard Deviation	1
Sample Variance	721,930,821	Sample Variance	44,763	Sample Variance	1	Sample Variance	0	Sample Variance	1
Kurtosis	(0.01)	Kurtosis	0	Kurtosis	(0)	Kurtosis	(1)	Kurtosis	(0)
Skewness	0.47	Skewness	0	Skewness	0	Skewness	0	Skewness	0
Range	$ 142,100	Range	1,140	Range	3	Range	2	Range	5
Minimum	$ 69,100	Minimum	1,450	Minimum	2	Minimum	2	Minimum	1
Maximum	$ 211,200	Maximum	2,590	Maximum	5	Maximum	4	Maximum	6
Sum	16694700	Sum	256,120	Sum	387	Sum	313	Sum	330
Count	128	Count	128	Count	128	Count	128	Count	128

Regression Analysis

Question	*Answer*
1. Which variables are statistically significant?	All of the variables have less than a 0.05 p-value, so they are all significant.
2. What does each coefficient mean in a real-world sense?	The coefficient means that a change in 1 unit of that variable will change the price of the house by that amount. For instance, adding 1 bathroom will increase the house price by $12,646.
3. Are these coefficients intuitive?	The Bedroom, Bathroom, and Square Feet are intuitive. However, the Offers are not. Meaning we can only do predictions inside of the range of the data we have. Extrapolating outside of that data is dangerous.
a. If not, why not?	Offers: We may have the causation between Offers and Price reversed. Making an offer on a house will likely not decrease its value, but multiple offers may actually increase the final sale price of the home due to bidding competition. However, a low price is likely to cause an increase in offers, so this would explain the reversed coefficient.
4. What does the R-squared mean?	The R-squared means that 70% of the change in the house price is explained by the change in the variables included in this regression model.

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

		Bedrooms				
	$ 131,149	1	2	3	4	5
SqFT	1,500	$ 69,222	$ 78,542	$ 87,861	$ 97,181	$ 106,501
	1,650	$ 78,498	$ 87,818	$ 97,137	$ 106,457	$ 115,777
	1,800	$ 87,774	$ 97,094	$ 106,413	$ 115,733	$ 125,053
	1,950	$ 97,050	$ 106,370	$ 115,689	$ 125,009	$ 134,329
	2,100	$ 106,326	$ 115,646	$ 124,965	$ 134,285	$ 143,605
	2,250	$ 115,602	$ 124,922	$ 134,241	$ 143,561	$ 152,881
	2,400	$ 124,878	$ 134,198	$ 143,517	$ 152,837	$ 162,157
	2,550	$ 134,154	$ 143,474	$ 152,793	$ 162,113	$ 171,433

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%)

Results

Question	*Answer*
1. What would explain non-intuitive results in your regression using the data which you were provided?	The non-intuitive result is that increasing the number of offers on a home would decrease its price. This could be explained if we are seeing the price that the home enters the market instead of the final selling price. The intercept could be intuitive if the closing costs are impacting the price of the home.
2. What additional data would assist you in explaining the non-intuitive results?	If we had both the initial price of the home and the final selling price, we could see if the number of offers impacted the delta between these two prices. We would be able to show that increased offers drove up the final price of the home. Location would be a great data element to have in order to help impact the price of the house. Not just region but state and areas with higher population density. Another piece of data that would be helpful is having the inventory of the area.

Bonus Section Linear Regression

Regression Statistics	
Multiple R	99.2%
R Square	98.3%
Adjusted R Square	97.5%
Standard Error	17,572
Observations	128

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	2.23084E+12	5.5771E+11	1806.166956	0%
Residual	124	38288870834	308781216.4		
Total	128	2.26913E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-	#N/A	#N/A	#N/A
SqFt	33.11	5	6.31	0.0%
Bedrooms	10,913.61	2,512	4.34	0.0%
Bathrooms	9,834.71	3,686	2.67	0.9%
brick	21,913.56	3,359	6.52	0.0%

As a bonus, we found evidence that the dataset may be fabricated. By forcing the intercept to zero, we were able to achieve an R-squared of 98%, which is very high for real data. If the dataset is indeed fabricated, this would explain the non-intuitive results found in the analysis above.