|  |
| --- |
| SCM 651 Business Analytics |
| Home Price Analysis |
| Homework 1 |

|  |
| --- |
| Todd Varian, Dan Burke, Sugeun Chae, James Du, Paramdeep Singh, Kody Wolverton  2-4-2024 |

Contents

[Question 1 2](#_Toc157974612)

[Question 2 2](#_Toc157974613)

[Question 3 3](#_Toc157974614)

[Question 4 5](#_Toc157974615)

[Question 5 6](#_Toc157974616)

[Question 6 7](#_Toc157974617)

[Question 7 8](#_Toc157974618)

# Question 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%).  
What patterns do you see?***

A screenshot of a spreadsheet

Description automatically generated

Figure 1 Pivot Table One & Two

# Question 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%)  
Both price and square feet are greatest in western neighborhoods.***

* Both price and square feet are the lowest in northern neighborhoods.
* Mostly, the brick houses are expensive and bigger than non-rick houses.
* When looking at both pivot tables and charts, the square footage and home prices are highest in the West overall. Brick homes are smaller and cost the least in the Northern neighborhoods.

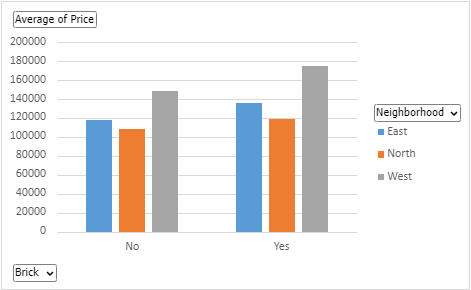


Figure 2 Pivot Chart One

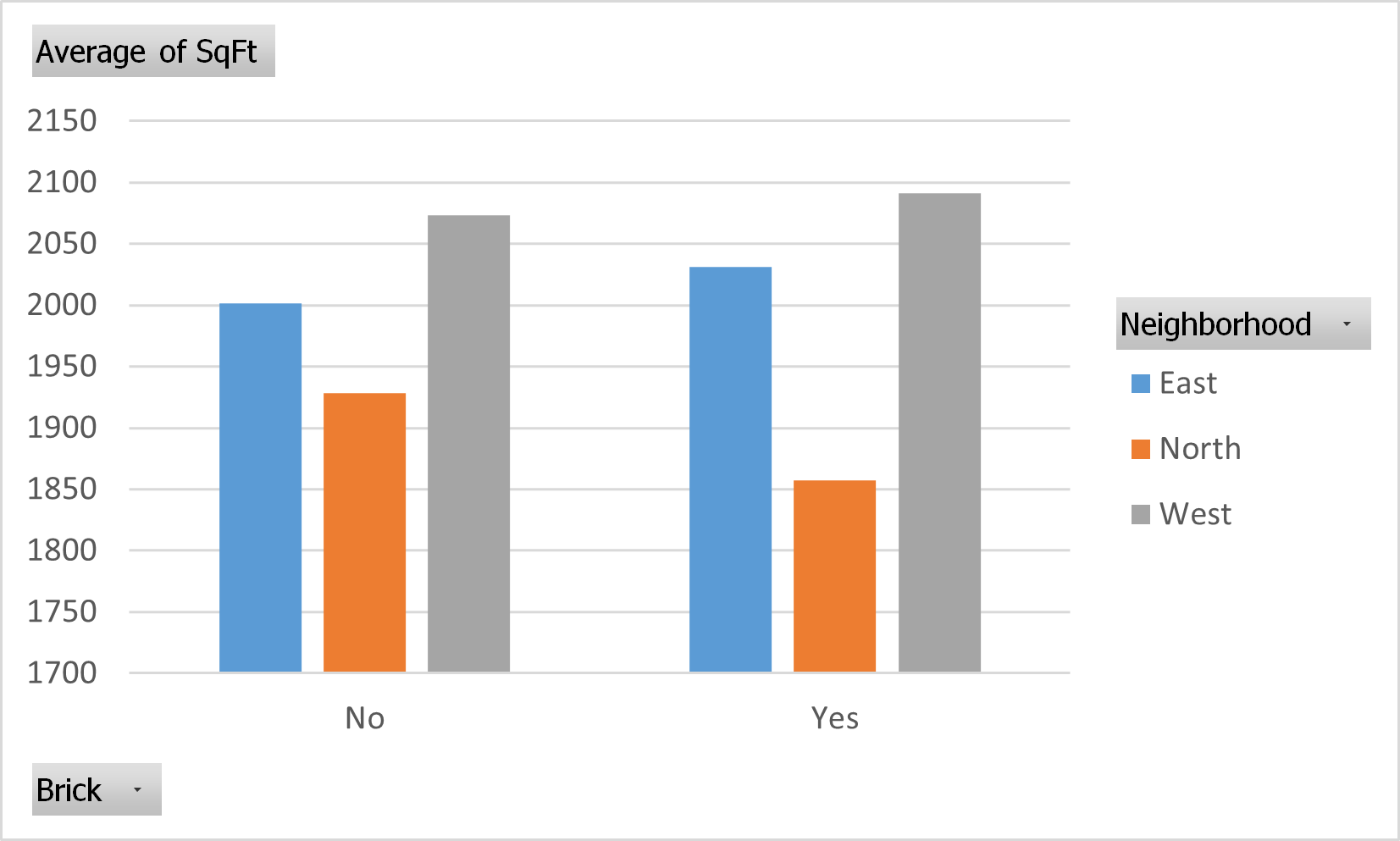


Figure 3 Pivot Chart Two

# Question 3

***Perform a correlation analysis of all quantitative variables except ID. Which two  
variables have the strongest (largest magnitude) correlation? Which two variables  
have the weakest (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%)***

A screenshot of a spreadsheet

Description automatically generated

Figure 4 Correlation Matrix

* Price and Square feet have the strongest correlation, the number of Bedrooms, then the number of Bathrooms a home has.
  + The largest magnitude implies these variables move in the same direction. In this instance, if price increases, it is expected square feet will also increase.
  + There is also some correlation between the square footage of a home and how many bedrooms and bathrooms it possesses, which is intuitive as larger homes will likely possess more rooms let alone, bedrooms/bathrooms.
* Bedrooms and offers have the weakest correlation.
* Price and number of offers made possess a negative correlation.
  + Due to this negative correlation, it is expected the price will not change based on the quantity of offers made.
* It is not intuitive. Because these numbers just indicate the direction of movement of two variables. Though it is not immediately intuitive, the negative correlation between price and offers may be explained by the affordability and increased demand for less expensive homes. This may be explored further if additional, higher fidelity data is available.

# Question 4

***Perform an initial regression analysis of the quantitative variables excluding the ID.  
Do not include type of construction or neighborhood. 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? (15%)***

A screenshot of a spreadsheet

Description automatically generated

Figure 5 Initial Regression Analysis  
Square Feet (“SqFt”), Bedrooms, Bathrooms, and Offers are all statistically significant.

* These Coefficients are further described by:
  + **Intercept:** Fixed price when all the variations are 0.
  + **SqFt:** With each increase in 1 SqFt, the price will increase by **$61.86**.
  + **Bedrooms:** The price will increase by **$9,320** per bedroom.
  + **Bathrooms:** Price will increase by **$12,646** per bathroom.
  + **Offers:** With each offer, the price will decrease about $13,601.

Square footage and price are significant and connected. The offers all made for the amounts mostly between $100,000 and $150,000 for these homes. As price increases on the X axis, SqFt on the Y axis increases. Interpret the data. Y = 0.0044x. As the price on the X axis increases by 1, the SqFt goes up by 0.0044 SqFt.

It is intuitive. But our independent variables have a small range except for the square feet to explain appropriate increment.

R-squared means 69.8% of the change in price is explained by the changes in square feet, bedrooms, bathrooms, and offers.

1433 is the intercept. It intercepts the Y axis at 1,433 sqft. R² = 0.3058. This is a range from 0 to 1. It is expressed as a percentage. 30.58% of the change in SqFt is explained by a change in price and how much we are spending.

# Question 5

***Perform a second regression including variables from part 4 and dummy variables for  
type of construction and neighborhood. What does each coefficient mean in a real-  
world sense? Are these coefficients intuitive? If not, why not? What does the R-  
squared mean? (10%)***

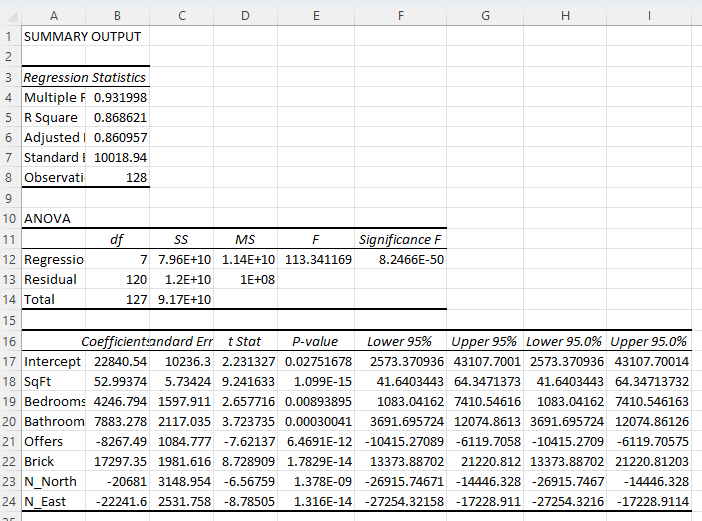


Figure 6 Second Regression

North and East are two of the coefficients to check. West is included in the intercept.

Square footage is very significant, as are the other values we use. The R2 value means that 86.86% of the changes in a home’s price are due to these coefficients we used.

* **Intercept:** It is the estimated price when the coefficients for all other variables are zero.
* **SqFt:** The coefficient of square feet is about 52.99, That means when 1SqFt increases, the price will increase by $52.99.
* **Bedrooms:** Each bedroom increases the price by approximately $4,247.
* **Bathrooms:** Each bathroom increases the price by approximately $7,883.
* **Offers:** Each additional offer decreases the price by approximately $8,267
* **Brick:** The property being made of brick increases the price by approximately $17,297.
* **N\_North:** The property not being in the Northern Neighborhood decreases the price by approximately $20,681.
* **N\_East:** The property not being in the Eastern Neighborhood decreases the price by approximately $22,241.

# Question 6

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

Two way sensitivity analysis, utilizing the variables (square feet, offers) which provide the highest R-squared value.

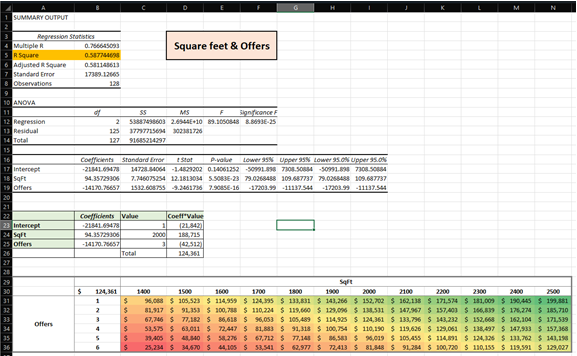


Figure 7 Sensitivity Analysis

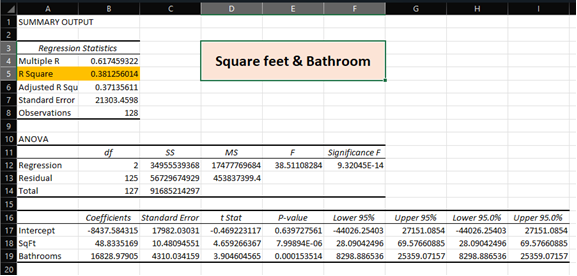


Figure 8 Square Feet & Bathroom Regression

Question 7

***What would explain nonintuitive results in your regression using the data that you  
were provided? What additional data would assist you in explaining the nonintuitive  
results? (10%)***

The data available for this initial analysis, though valuable and meaningful, is low fidelity. There are many different aspects which could affect the price of a home as well as explain correlations and relationships between the variables, and their respective coefficients; that may be used to predict a home’s value. Some of these include but are not limited to:

* Equivalent data for the “South” neighborhood
* County of Residence/Home
* Property Taxes
* Zoning laws
* “Walkability” score/rating
* Natural hazards (homeowners’ insurance may be a indicator)
  + Flood Zones
  + Wildfires and risk of occurrence
  + Hurricane, Tornado and other extreme weather circumstances and the risk thereof.
* “Bikeable” score/rating
* Climatic and other Weather conditions (harsh weather which increases home maintenance expenses)
* Available and state of local infrastructure/utilities
  + City Water
  + Electrical
  + Public Transport
  + Grocery Stores, Restaurants, Parks
  + Sewage/Septic Tank
* Local Property rental cost
* Local Real Estate Inventory
* Home Loan Interest Rates
* Renovations, if there have been any; it there has been, how many?
* MLS Listings; property turnover, any listing and delisting actions of the property within the MLS system.
* Home Energy Costs
* Property’s Assessed Value and the Assessment Year
* Land Tenure (if applicable ex. Hawaii)
* Year Built
* Number of Stories
* Exterior features such as paved or unpaved driveways, the presence of or lack of driveways as well as garages.