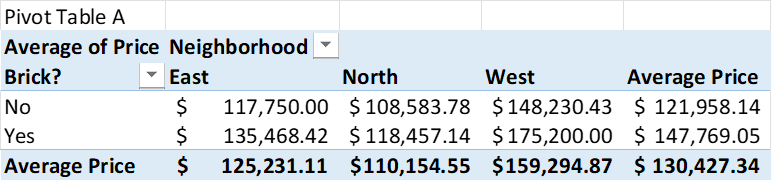
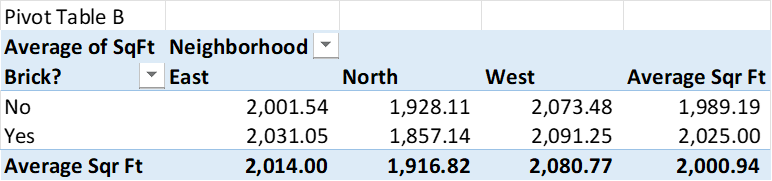
**Homework #1: Home Price Analysis Report**

**Team 1: Ava Hudak, Megan Rios, Kristan Rodriguez, Cynthia Scotton, Ceara Stewart**

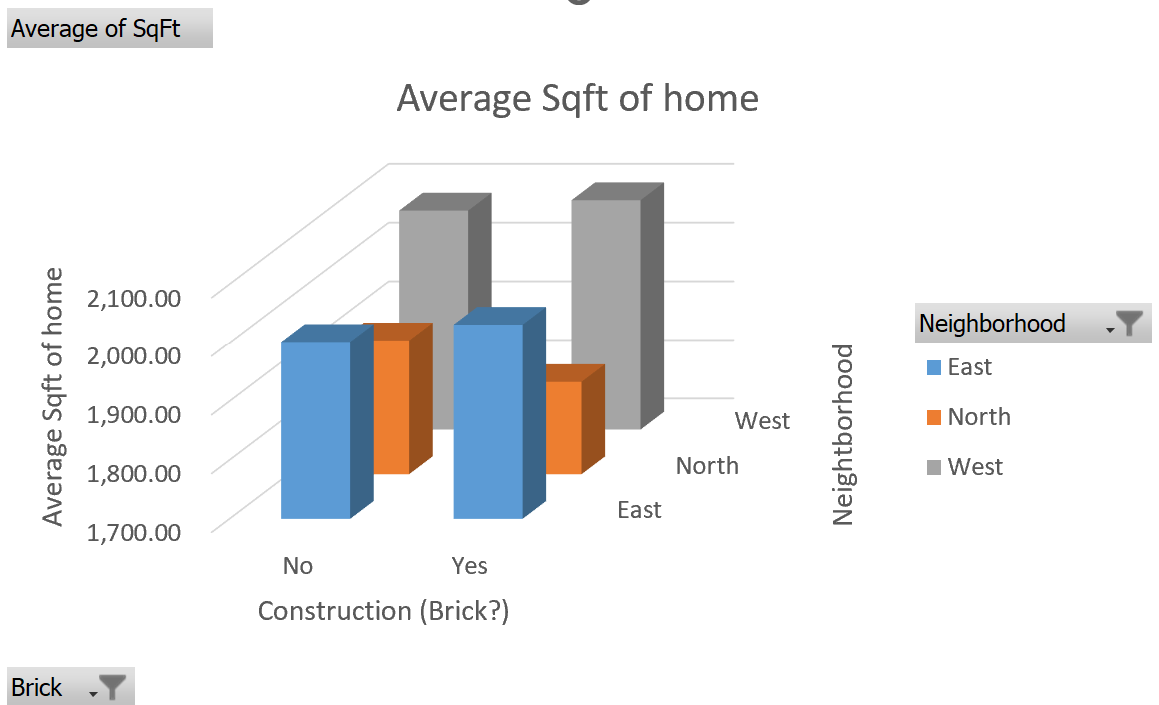
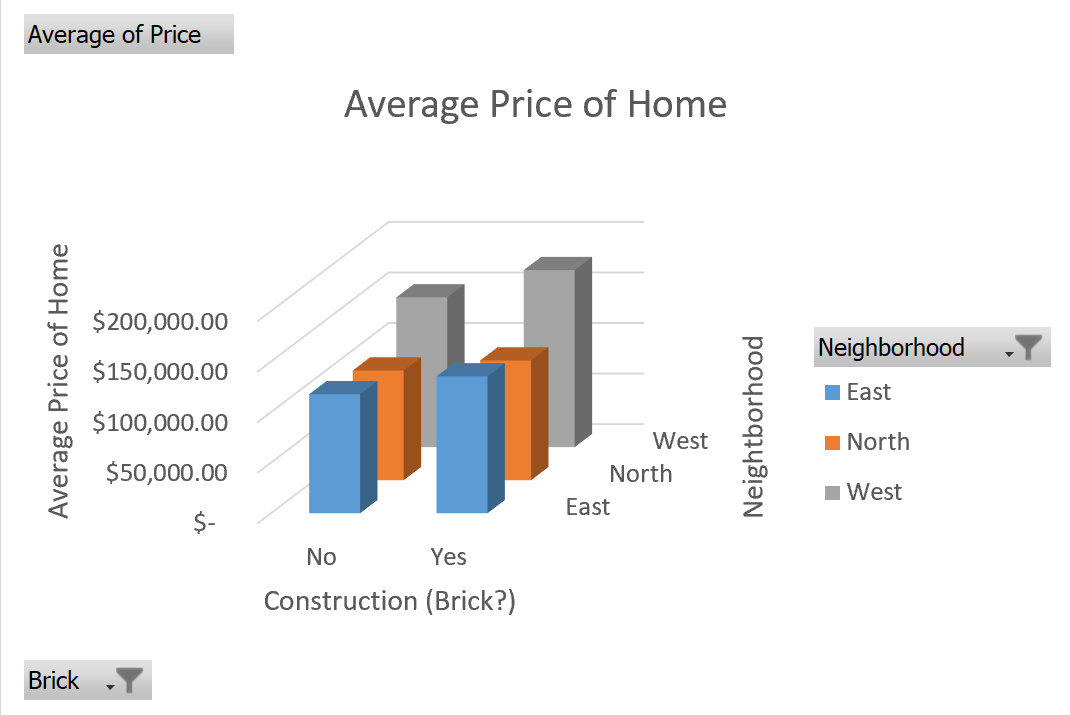
**Q1. Develop a categorization of your data using pivot tables. Develop two pivot tables: One pivot table of average price, varying type of construction (brick) and neighborhood as the two dimensions; a second pivot table of average square feet varying type of construction (brick) and neighborhood as the two dimensions (20%). What patterns do you see?**





Pivot Table A above shows the effect of the type of construction (brick) and neighborhood on the average price of a home. If the home is constructed of brick, the home is likely to cost more in all the neighborhoods. The average price of a home with brick is $147,769.05 and without brick is $121,958.14. Regardless of the construction, homes in the North typically are more inexpensive than those in the West neighborhood. Pivot Table B above, shows the same dimensions (construction and neighborhood) but on their effect of the size of the home (square feet). Homes in the North not only cost less on average, but they are also smaller. On average the homes in the West are larger. Unlike the price, homes constructed of brick are not larger across all neighborhoods. Overall, on average homes are larger if made of brick, but the North homes on average are larger when not constructed out of brick.

**Q2. Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood. (10%)**



The Pivot charts above represent the same data from the pivot tables in question 1, visually. The homes with brick have a higher average price represented by the bar graph in the first table. The size of the home is drastically larger in the West neighborhoods, based on the chart. The orange bars that represent the North are smaller in both charts.

**Q3. 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%)** Application, table

Description automatically generated

The two variables with the strongest correlation are price and square footage. As the square footage and size of a home increases, so does its price and perceived value to the customer. The characteristics of a home that have the weakest correlation, closest to zero, are bedrooms and offers. The characteristics with a negative correlation are offers and price, moving in opposite directions. The correlation does not measure the impact of changing variables, but only the direction of the linear relationship.

**Q4. 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%)**

**Graphical user interface, application, table, Excel

Description automatically generated**

Running the first regression, Significance F is less than 0.05, so we can interpret it’s impact. By running the regression, we can accurately forecast what the perceived value is as independent variables either increase or decrease. For all variables of square footage, bedrooms, bathrooms, and offers are statistically significant. R2 shows that 69.8% of the price of a home is dependent on these variables that are significant.

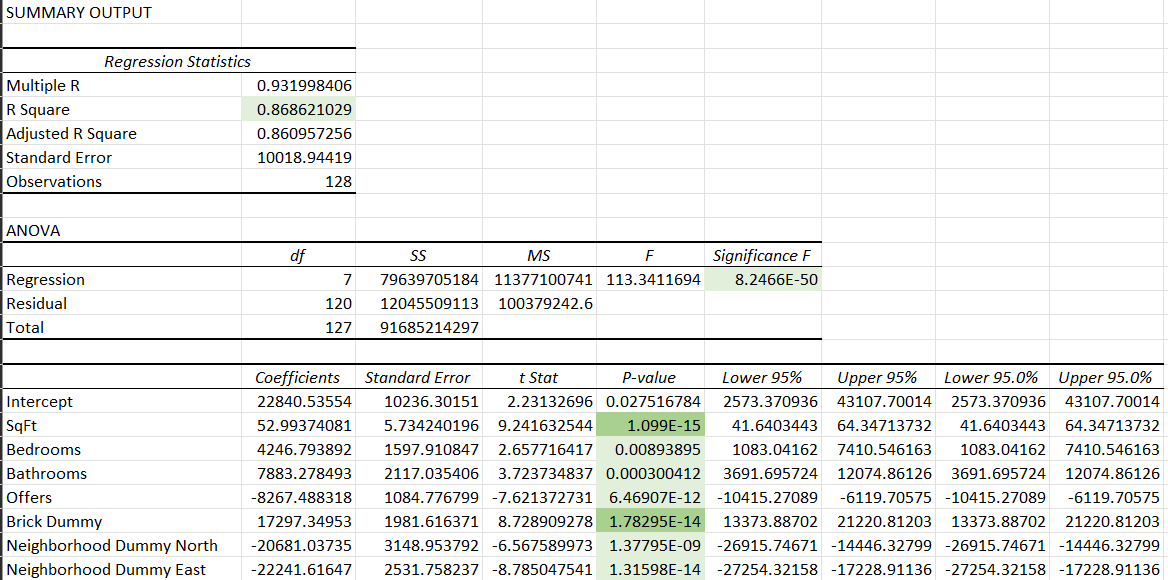
The intercept reflects fixed costs associated with pricing a home. For example, cost of permits or fixed cost associated with the realtor. The coefficients represent a cost of adding additional units of square feet, bedrooms, etc. The coefficient for square feet can be interpreted as cost of materials, for example (more flooring). The coefficient for bedrooms can include additional materials to build additional walls, doors, windows, and wiring for the additional rooms. The coefficient for bathrooms accounts for features such as plumbing, tiling, and fixtures not found in other rooms of the house.

The coefficients calculated are intuitive. The cost to add square footage is less than that of adding an additional bedroom or bathroom. The cost to add an additional bathroom is greater than adding a bedroom, because of the specialized features that go into building a bathroom, ie. Plumbing.



**Q5. 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%)**

Adding to the previous regression, we kept square feet, bedrooms, bathrooms and offers as they were statistically significant when running the multiple regression without dummy variables. We add to the multiple regression by adding dummy variables for if the house is brick and for which area the house is in, North, East or West. Since you always have one less dummy variable in the regression than listed, we added a dummy variable for if the house is made from brick, and if the house is in the North or the East. Running the regression in excel, we get the output below.



We get an R2 of 0.8686, meaning that 86.86% of the variability observed is explained by the regression model, or that that house prices are 86.86% dependent on the values of the intercepts. When looking at the F-value and its corresponding F significance value, we obtained an F significance smaller than 0.05, meaning our results from the regression model are significant and can be interpreted. We also looked at the p-values for the independent variables in relation to house prices. With everything else held constant, we know that changes in the independent variables result in significant changes in house prices. By looking at the intercept p-values for each of the variables used, every p-value was below 0.05 and thus can be interpreted.

In a real-world context adding the dummy variable for brick, holding everything else constant, that if the house was made of brick, house price would change by 17,297 dollars, and if the house was not made of brick, the house price would not change. The same idea can be used to talk about neighborhoods in a real-world context, where if we keep all the other values constant, and if a house is in the “North” neighborhood, house price would decrease by 20,681 dollars, while if a house were in the “East” neighborhood, house price would decrease by 22,241 dollars.

Overall, the model below is statistically significant and a good predictor of housing prices.

We can take this equation and utilize it directly in a “real-world scenario” (if our sample is a good representation of the population it was pulled from) by plugging in realistic house values and comparing house prices. Say we have a house that is 800 square feet, has 2 bedrooms, 1 bathroom, has 3 offers, is made of brick and is in the East, we would get a house price of, we can plug them into the above equation and solve for house price,

The dummy variables, like the values provided in question 4, are intuitive. Since different neighborhoods are deemed better than others, having a house in a neighborhood deemed “worse” would decrease the cost of a house in that neighborhood. Brick houses offer a sturdier foundation and require less upkeep compared to a vinyl sided house that could need constant painting, so a brick house has higher value and thus would result in a higher house price.

We also ran two other regression models where the neighborhood types were switched but those models produced statistically insignificant results, confirming the above equation as a good predictor model for house price.

**Q6. 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.**

Table

Description automatically generated

A prediction model allows us to estimate the outcome when our independent variables are manipulated. From this model, we can predict the estimated cost of a home as we adjust the values of our variables obtained from the regression model results. For every square foot, bedroom, and bathroom, the value of the homes increases by $52.99, $4226.80, and $7883.27, respectively. Each sales offer decreases the price by $8267.34.

Utilizing the predictive model equation:

We can predict the 2000 square feet 3-bedroom brick home cost in the East neighborhood to be $135,855.71.

Although all the variables impact the cost of homes in the North and East, for this analysis, we will use the information obtained for square footage and bedrooms in our calculations for the sensitivity analysis, regardless of neighborhood. These two variables have the larger ranges of values as shown on the regression analysis. Square feet of homes ranged from 1,700 to 2,100 and the number of bedrooms ranged between 2 to 5. Adding additional bedrooms will increase the size of the home. Larger homes are more desirable to larger families who understand that bigger homes will cost more than smaller ones.

Table

Description automatically generated

**Q7. 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%)**

* 1. Tax rates
  2. School districts
  3. Demographics (Age, Population, Median Income)
  4. Market Fluctuations; Time Value of Money
  5. Noise pollution (traffic, busy road, nearby trains)
  6. Accessibility (Shopping, grocery stores, highways)

There are many factors that influence what is perceived market value in the price of a home. As population and resources continue to change due to supply and demand, we can anticipate seeing other variables be included in forecasting metrics over time. Additional data that would assist in the nonintuitive results would be the age of the buyer, and what outside factors they need in a neighborhood. Including schools, families with small children, nearby parks, busy roads, and speed limits on the street. Other factors could include access to public transportation, nearby shopping, access to the highway, and one of the biggest factors is the rate of crime in the surrounding neighborhoods.