Renato Cruzalegui

John Bartlett

Becky Mattews-Pease

SCM 651

HW #1 April 18, 2020

|  |
| --- |
| 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 |
|  |

|  |
| --- |
| 1. Using the two pivot tables above, generate pivot charts for average price and average square feet by type of construction (brick) and neighborhood. |
|  |
|  |

|  |
| --- |
| 1. 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? |
| Performed correlation analysis on the quantitative variables in the house prices data sets, the two variables that show high magnitude are the square foot and price, in the same data set the two lowest are Offers and bedrooms. Conditional formatting is enabled on the grid showing highly correlated variables in green, next yellow and ultimately red.  Some of these correlations are intuitive some are not for example bedrooms and bathrooms; intuition would dictate that the more bedrooms you have in a house the more bathrooms you would need. The correlation between square footage and bedrooms/bathrooms is also intuitive, the more rooms in a house the larger the square footage will be and vice versa. A nonintuitive example would be square footage and offers, these categories are mildly corelated, so in plain English this means the higher the square footage the more frequent the offers, and vice versa (this could also imply that larger houses sell faster). |

|  |
| --- |
| 1. 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? |
| The RSquare at .69 means that the home value can be explained by the 4 variables, as the variables are statistically significant since their p value is less than .05. Also, to note is the F statisitcs show a high degree of confidence in the equation, in this example we have extremely high values.  In a multivariate regression model, you have more than one variable as an input to explain the rise of y. If we analyze the first variable, SqFT, the variable would be used in B1 in the following equation:  Y = Bo + B1X1 + B2X2 + B3X3 + … E  If we go in order the price will be driven by SqFt (Variable 1), Bedrooms(Variable 2), Bathrooms (Variable 3) and Offers (Variable 4).  We can interpret that for each X value, an increase in SqFt, Bedrooms, or Bathroom the home value will increase by the defined coefficient.  In the case of Offers (Variable 4) the value of the home will decrease again by value of $13,601.01 by the x factor. |

|  |
| --- |
| 1. Create a spreadsheet prediction of the model. Perform a two-way sensitivity analysis and use conditional formatting to highlight the results. |
|  |
| Based on the multivariate regression model we can estimate the home price based on the 4 variables afore mentioned. Based on the data shown, a home of 2 bedrooms and 1 bath with 3 offers its price will be $99,907. |

Additional Sensitive Analysis



|  |
| --- |
| 1. 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? |
| I would not think that the # of offers received on a house would have such dramatic impact on the price based on this data. The same house with used to create the baseline calculation with only one offer would be priced at 127,019. Based on the r square of the model 68% of the homes price can be explained by the variables previously described. Depending on the market , if it is a sellers’ market or buyers market the number of offers received could have impact on the price of the home. In order to improve the model it would be beneficial to have the dollar amount associated with the offers received. Additional data points could be neighborhood desirability, inventory of houses in the market, and other socio-economic metrics. |