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?

|  |  |  |  |
| --- | --- | --- | --- |
| Average Home Price  by Building Material and Geography | | | |
| Brick? | East | North | West |
| No | $117,750 | $108,584 | $148,230 |
| Yes | $135,468 | $118,457 | $175,200 |

|  |  |  |  |
| --- | --- | --- | --- |
| Average Home Square Footage  by Building Material and Geography | | | |
| Brick? | East | North | West |
| No | $2,002 | $1,928 | $2,073 |
| Yes | $2,031 | $1,857 | $2,091 |

* Patterns: Brick does not imply the size of a home, but it definitely has a positive impact on the price of a home. Also, the average price of a brick home is higher in each 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. (10%)

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. 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%)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Price | SqFt | Bedrooms | Bathrooms | Offers |
| 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 |

* Square footage and Price have the strongest correlation of .5529
* Offers and Bedrooms have the weakest correlation of .1142
* Correlation relates to the strength and direction of the relationship between two variables. At minimum, the R-square =(.55298224)^2, or 30.5%, of the variability of Price can be attributed to Square Footage.
* There is one negative correlation between Price and Offers of -.3136
* Most of these correlations are intuitive. In the case of Square Footage, bedrooms and bathrooms this makes sense since the bigger a house is the more expensive it likely will be. It’s not a perfect correlation because other factors like location would likely have an effect as well; a big house in bad neighborhood will likely sell for less than a comparably sized houses in a nice neighborhood.
* The negative correlation between Offers and Price is not intuitive. Supply and demand says that more interest/demand in a house (i.e. offers) will drive up pricing. The regression here says the opposite – that as a house receives more offers its impact on Price is negative. One possible explanation for this relationship could be the impact of time in the market. If a house sits on the market too long, it may receive offers over time but never sell. The longer a house sits unsold on the market the more likely it is to experience a price cut. It would be informative to see how the offers were received over time. Offers that are clustered around the time a house hits the market should provide a boost to the sale price (aka a bidding war) while offers that are spread out over many weeks or months we would expect to have a negative effect on sale price.

1. 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%)

Table

Description automatically generated

Table

Description automatically generated

* The significance of F is less than .05 so the model is statistically significant
* Each independent variable is statistically significant with p-values <.05 except the Intercept which is not significant. A negative intercept means that if all the independent variables were fixed at zero, the home value would be negative $17,347. Theoretically this scenario would imply the cost of the land. A negative cost of land doesn’t make sense; however, it makes sense in the context of this model since our dataset does not include any land sales, only home sales.
* The coefficients correspond to the price increase/decrease per unit change in the variable house:
  + Increases by $61.84 for every additional square foot,
  + Increases by $9,319.75 for every additional bedroom,
  + Increases by $12,646.35 for every additional bathroom,
  + And decreases by $13,601.01 for every additional offer placed on the house.
* The coefficients are intuitive. For example, each bathroom in a house will add $12,646 of value to a house all else being equal.
* R-squared: 70% of the change in home price is explained by the independent variables

1. 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%)

Table

Description automatically generated

Table

Description automatically generated

* The coefficients correspond to the price increase/decrease per unit change in the variable
* The coefficients are intuitive. We now have information about the impact of where a home is located with the addition of East/North/West. With the added variables we can see that homes in the East and North Neighborhood decrease the value of that home by $22,241.61 and $20,681.04 respectively.
* R-squared: 86% of the change in home price is explained by the independent variables

1. 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%)

Initial inputs:

A picture containing text, receipt, screenshot

Description automatically generated

Table

Description automatically generated

1. 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%)

* A possible explanation for the negative relationship between price and offers could be market timing - house sits too long on the market and buyers are less inclined to bid or bid aggressively. This could result in the seller having to drop his price.
* Additional data that might be useful is seasonality data and interest rate data. Because many home buyers are financing their home purchases, they are very sensitive to interest rates. If higher rates have a strong enough effect this will reduce the buyer base, negatively impacting demand for a house. This in turn would put pressure on a seller to account for the increased borrowing costs and bring his home value down. Additionally, seasonality may play a part in the housing market. Families with children in school are more likely to move in the summer while the kids are not in school. This could create local hot spots in the housing market in areas with desirable school districts.
* Additionally, we also would expect to learn more about the impact offers has on the sale price if we had the initial listing price.