Mr. Mason,

Thank you for contacting DABAS LLC; we feel confident we can answer your questions about the Ames, Iowa housing market. We were able to track down the most recent 2,930 home sales in Ames as well as 79 characteristics of each house. Since you are only interested in three specific neighborhoods, North Ames (NA), Eastwood (EW) and Brookside (BS), and how square footage (SF) relates to their sale price (SP, in dollars), we will restrict our model to only focus on those variables (745 home sales).

By examining the scatterplot **(see Appendix [ENTER HERE SCATTERPLOT REFERENCE])**, we concluded we would need to take the log of the sales price and square footage to create a linear relationship. Also, the slope of the regression lines of different neighborhoods would be different, so we included interaction terms. This means our model would look like this:

ln(SP) = B0 + B1 \* ln(SF) + B2 \* BS + B3 \* EW + B4 \* ln(SF) \* BS + B5 \* ln(SF) \* EW

To test our model, we used a null hypothesis that all Bi are zero, or that there is no correlation between sales price and any of the other variables. And our alternative hypothesis is that at least one Bi is not zero, or that there is a correlation between at least one of the variables and sales price. We received a p-value well below 0.05 (it actually was below 0.001, **See Appendix [ENTER ANOVA TABLE REFERENCE]**), so we rejected the null hypothesis and concluded there was a relationship between the sales price and the other variables.

We then needed to test whether each individual variable was needed in our model. For each variable, our null hypothesis was that given the other variables contribution to the model, that particular variable did not help our model (or that the coefficient in front of that variable was 0). And our alternative hypothesis was that the variable contributed to the model (or that the coefficient in front of the variable was not 0). For each variable, including the interaction terms, we received p-value below 0.05 (**See Appendix [ENTER ESTIMATE TABLE REFERENCE]**), which means for each variable we rejected the null hypothesis and concluded that they were all beneficial to our model which is:

ln(SP) = 8.48 + 0.47 ln(SF) - 2.29 BS - 0.93 EW + 0.3 ln(SF) BS + 0.11 ln(SF) EW

This model has an r-squared of 0.4769 which means that 47.69% of the variation in sales price, can be explained the variation in neighborhood and square footage.

For this model, we had to assume independence even though that likely isn't completely true. Also, this model gave us relatively normal residuals (slightly curved, but with 745 observations, the Central Limit Theorem would help), but it did show some outliers and leverage points **(See Appendix [RESIDUAL PLOT REFERENCE]**. We concluded the outliers were all accurate measurements so we had to include them. The model created after removing the leverage point gave very similar results to our model, so we decided to keep our model and include the leverage points **(**all confidence intervals for the coefficients contain the coefficients from our full model. **See Appendix [REMOVED LEVERAGE TABLE REFERENCE]).**

We can simplify this model a little by separating it by neighborhood. This leaves us with three different equations you can look at depending on which neighborhood you are working with.

{ ln(SP) | Brookside } = 6.19 + 0.77 ln(SF)

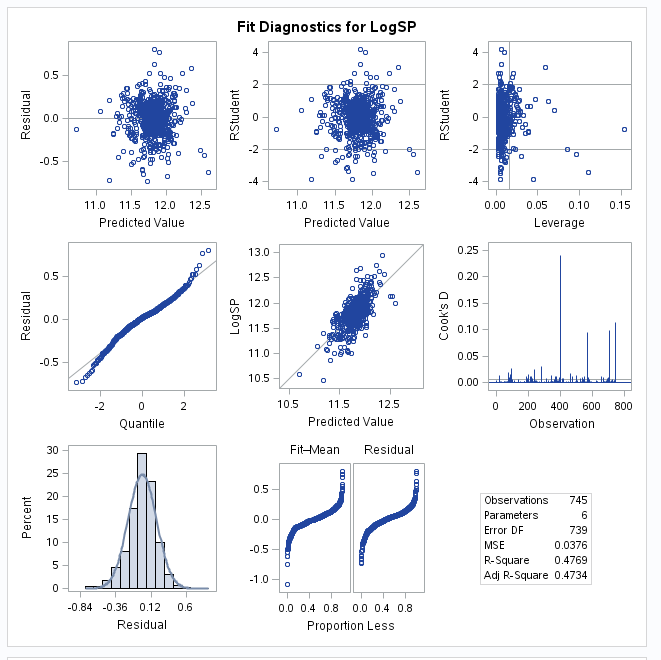
{ ln(SP) | Edwards } = 7.55 + 0.58 ln(SF)

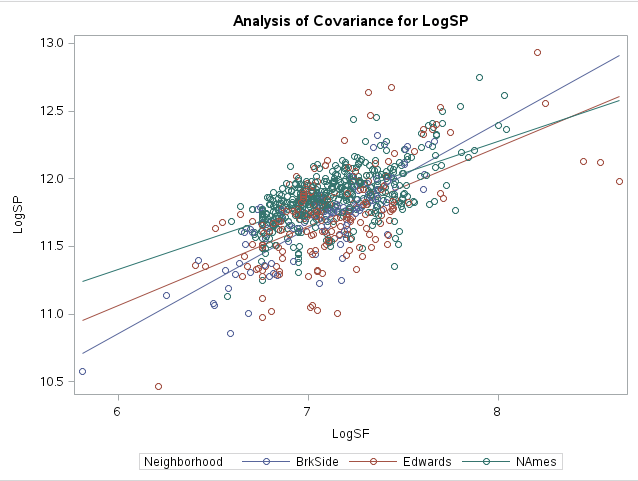
{ ln(SP) | North Ames } = 8.48 + 0.47 ln(SF)

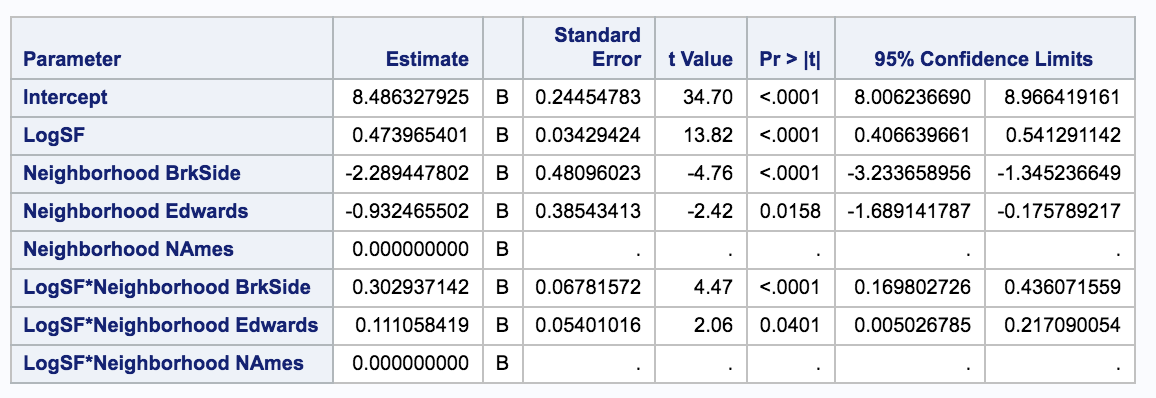
In general, when you double the size of a house in Brookside, the sale price increases by a factor of 2.16. When you double the size of a house in Edwards, the sale price increases by a factor of 1.79. And when you double the size of a house in North Ames, the sale price increases by a factor of 1.56.

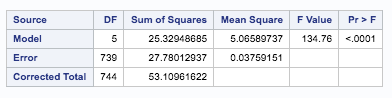
Due to the nature of this study (it was an observational study by nature), we cannot infer causation and say the square footage or the neighborhood causes a change in the sales price of the house; there are obviously many confounding variables that play a part. But, we can say there is a correlation between the sales price and the square footage and neighborhood.

APPENDIX STUFF:









Removed Leverage Points table:

