

# M0\_assignment

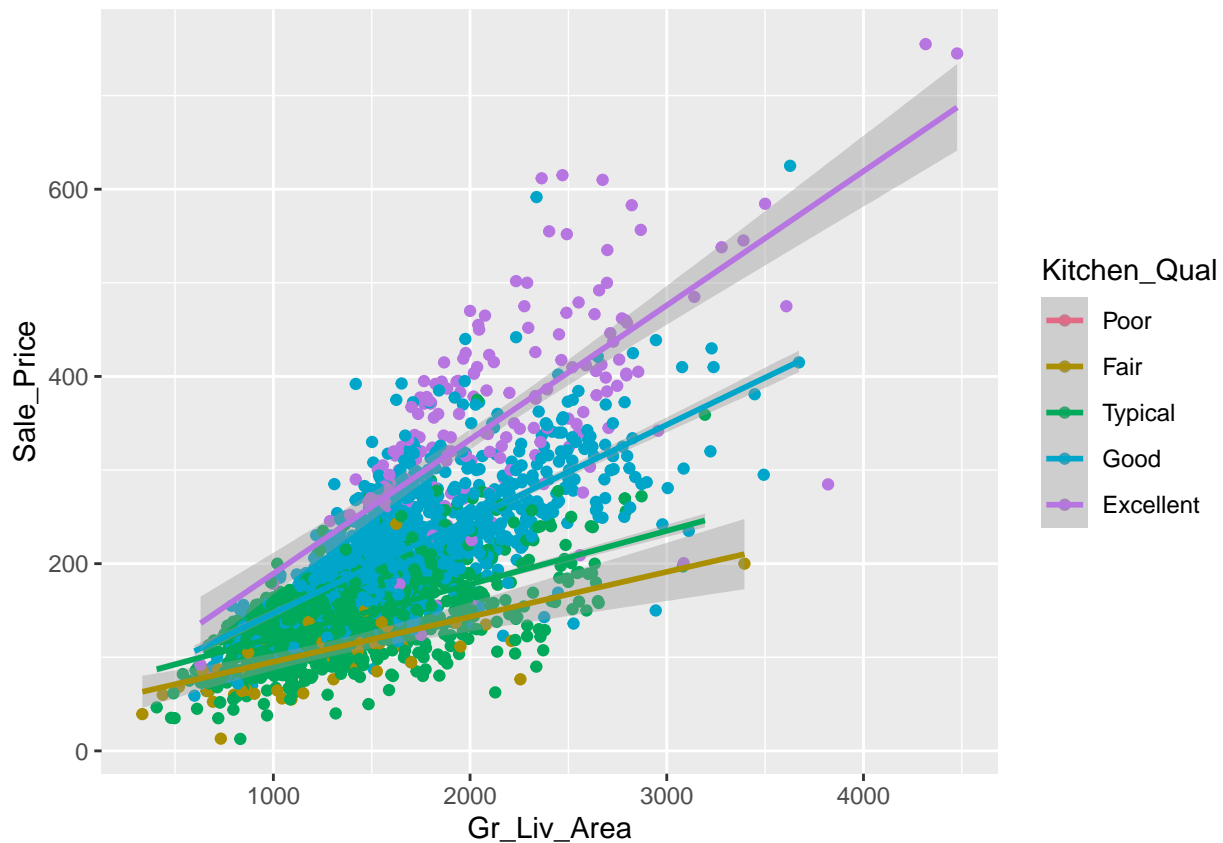
Andrew Estes

2022-08-10

*#Imagine you were consulting for a potential home seller. Write a technical paper  
#about whether the quality of a kitchen is important in the valuation of a home.  
#Discuss the chart and the regression models.*

Dear Potential House Seller,

Kitchen quality is a key indicator of sale price. As you can see in the graph, the kitchen quality is the key difference in sale price between similarly sized homes. There is a smaller sale price difference for homes with a smaller above grade living area square feet footprint. There is still a distinction but it is all relatively mixed together until we hit the 1500 above grade living area square feet, at which point the kitchen quality starts to make a more substantial impact.



The graph is a good eye-catcher, showing a clear relationship between kitchen quality and sale price. You may be asking what is the true impact, and that is where linear regression comes in. Regression analysis is a mathematical way of sorting out which variables have an impact and what the impact is.

The linear model shows there is actually a negative impact on sale price when the kitchen quality is only “fair”, meanwhile an “excellent” kitchen quality has a positive impact of 161.77. Above grade living area square feet has a positive impact of 0.09 per square foot – so 90 for every 1000 square feet in above ground living area.

Another statistic gleaned from linear regression is called the adjusted r-squared number. It indicates model accuracy. Above grade living area square feet by itself has an r-squared of 52.9, which is okay. Kitchen quality by itself has an r-squared of 48.9, which is also okay. But using both Kitchen quality and above ground living area to predict the sale price has an adjusted R-Squared of 71.0.

This means 71% of the output shown on the graph can be explained by kitchen quality and above grade living area square feet. It is a pretty good result (great considering only two variables were used). However, this is not an exhaustive list as other variables can impact sale price – neighborhood and home age spring to mind. A 100% accurate model is not possible so factors to consider before listing your home:

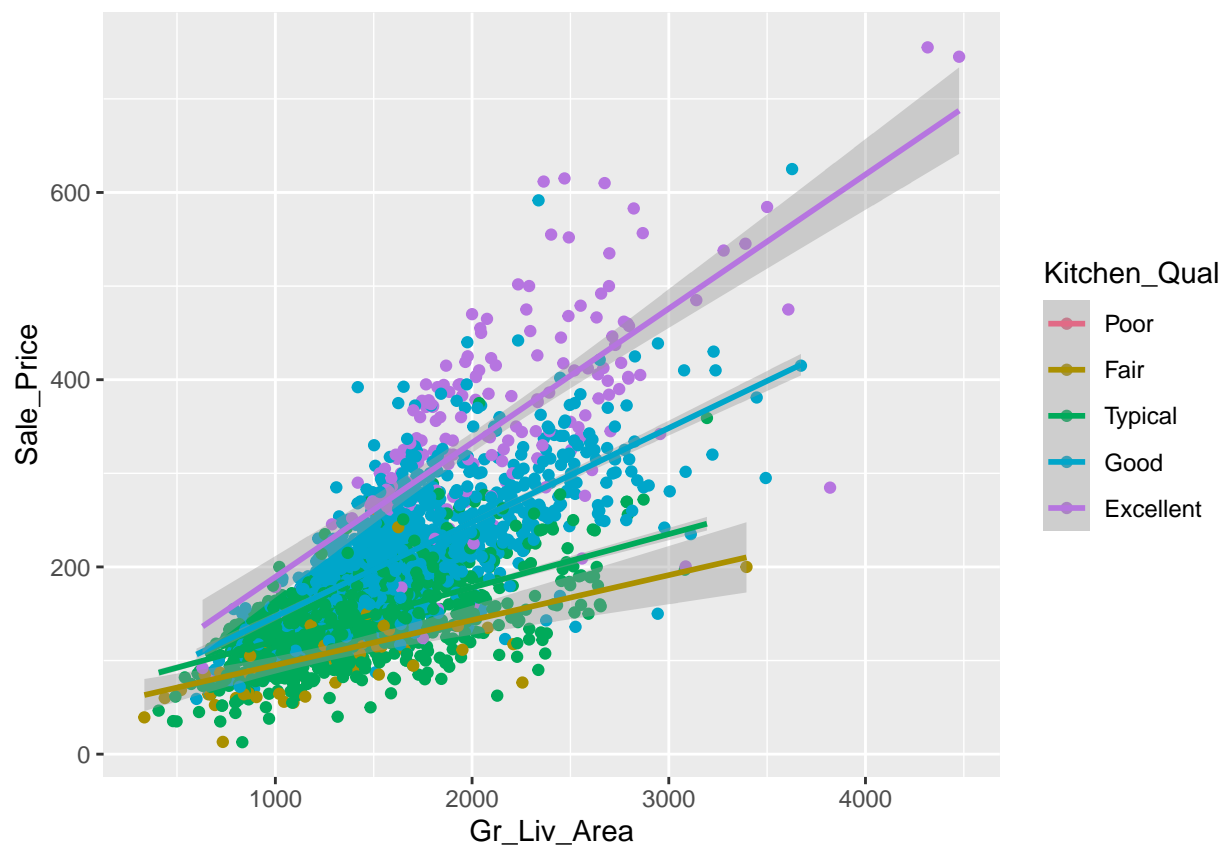
Size – should it be expanded? Can it be expanded? Kitchen quality – should it be updated? What other factors should be considered? Based upon your answers to these questions, we can consider increasing the above ground living area square footage, updating the kitchen, or sticking with the status quo if the ROI is not there.

*#Now, imagine you are consulting for a realtor in the area. Write a white paper  
#about the importance of the quality of a kitchen to the valuation of a home.  
#Discuss the chart and the regression models.*

Dear Potential Selling-Agent,

The goal of any house sale is to maximize the selling price. This is advantageous to both realtors as well as the home-seller. The question of how to best do that in Ames, Iowa is answered with a simple linear regression model.

There is a clear correlation that sale prices increase as above grade living area square footage increases. This makes logical sense – more space means more cost to build and correspondingly increases the price. Square footage is also something that is usually costly to change – removing or adding on to an existing structure requires extensive paperwork and construction work.



Another clear correlation is that sale price increases as kitchen quality increases, regardless of the above ground living space square footage. This distinction is exhibited in the graph below. Increasing the quality of the kitchen is a much more feasible project than increasing livable square footage. Due to continual changes in style, updating the kitchen before selling is a simple project to make it a turn-key sale.

```
##
## Call:
## lm(formula = Sale_Price ~ Gr_Liv_Area + Kitchen_Qual, data = ames_quick)
##
## Coefficients:
##           (Intercept)           Gr_Liv_Area           Kitchen_QualFair
##             3.00760             0.08537             -1.57710
## Kitchen_QualTypical      Kitchen_QualGood Kitchen_QualExcellent
##             23.32711             68.25500             161.77383
```

The coefficients of the linear regression model further exhibit that kitchen quality should be prioritized over above grade living area square feet. For every 1000 feet in above grade living area square feet, the coefficient is 90. Meanwhile, updating from a “fair” to “excellent” kitchen will have a positive coefficient of 163.36. Even updating from a “good” to “excellent” kitchen has a positive coefficient of 93.52.

To make this clearer, a potential customer is considering selling a 2000 above grade living area square feet space with a fair kitchen for 175k. They can increase their living space to 3400 and have sell price of 200k – an increase of 25k. Or they can update the kitchen from fair to excellent, leading to a sale price around 300k – an increase of 125k.

We should focus on kitchen renovation to increase house sale price.

*#Now, imagine you are presenting on this topic to fellow students in an mid-level #statistics course. Sketch a script outline explaining how the quality of a #kitchen influences the valuation of the home. Discuss the chart and the #regression models (you don't need to include them again).*

Dear fellow students (insert meme)

<https://www.theverge.com/2017/7/13/15966094/30-rock-buscemi-how-do-you-do-fellow-kids-meme-kill-it-please>

We ran a linear regression on the selling price of a home using both a quantitative, continuous variable in above grade living area square feet and a qualitative, descriptive variable in kitchen quality. We converted kitchen quality into a factor variable and limited the results to houses with central air AND less than 4500 square feet of above grade living area. The former was to present uniformity with responses – central air is a big deal with house purchases, and the latter is to avoid any outliers (huge houses) from skewing the data.

There were 2927 observations and 81 variables initially.

Looking at just two variables: the above grade living area square footage and the kitchen quality, we were able to present a model with an adjusted R-Squared of 71%. The Residual Standard Error is 43.

At the expense of additional work, we could have used stepwise regression or principal components analysis to perhaps find a model with a better adjusted R-Squared and smaller Standard Error. Further regression techniques, such as lasso regression, may present a better model.

*#Write a short reflection (about 300 words) explaining how the audience and #mode led to differences in the three reports.*

The third report, to fellow students, was the most straight-forward report. It was simply reporting the numbers without much explanation as to their importance. It was also enjoyable to throw a fun meme in there for a presentation affect.

The first report, written to the house seller, was the most difficult. It is hard to write about the topic without having a clear idea of who the audience is. Is the house seller a statistics professor or a painter? Not that they are mutually exclusive but the point remains that a report to an undefined audience is difficult to write.

The second report, the white paper, was the easiest. Perhaps because I had already written the first report, the technical paper. All I had to do was adjust the verbiage so it would transform into a white paper. And while realtors may not be statisticians or mathematicians, they do pay attention to economics and understand and utilize basic statistical principles on a frequent basis.