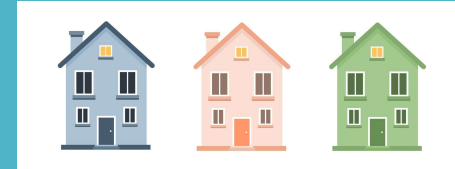# Home Prices Study: Ames, IA

By Rachael Friedman

Background

- The year is 2011. Homeowners are turning to the internet before putting their homes on the market.

- They want to gauge **how to price their homes** and have questions about **specific house features and home value**.

- A common google search that has spiked in popularity recently is "**does the kitchen sell the house?**"

- A real estate company in Ames, IA noticed this phenomenon and hired me to research this question more in depth for their community.

- In my analysis, I will use a regression model to predict housing prices based on home features and use this to determine if kitchen quality adds value to a home.

Kitchen Quality & Sales Price

Kitchen Quality Distribution
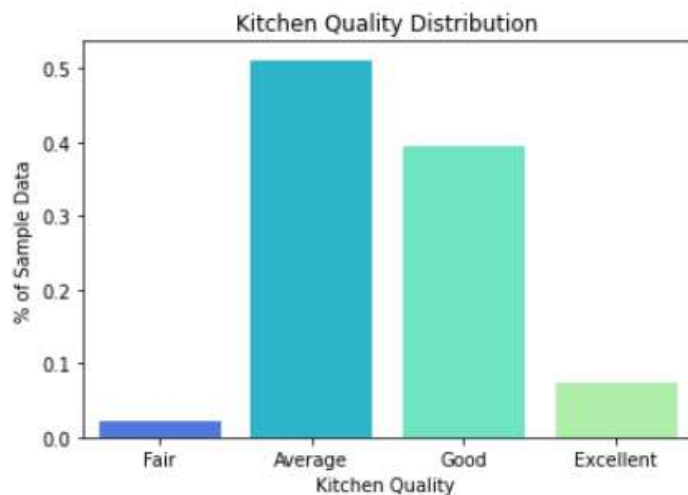
- Just under 10% of homes from our sample data have an excellent kitchen quality rating.

Kitchen Quality and Sale Price
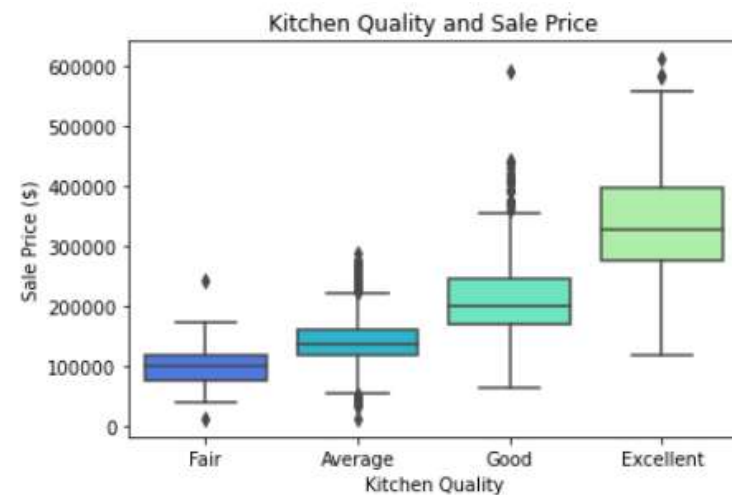
- The 25th, 50th, and 75th percentiles of home prices increase as the kitchen quality rating increases.

Methodology

- Dataset used for this analysis is **2006 – 2016 Ames, IA home prices** with about 2,000 entries and over 70 columns of different housing features

- Data is cleaned to prepare for modeling. Outliers are removed, features are transformed appropriately for categorical or ordinal data, and nulls are accounted for.

- Feature extraction and feature interaction is performed on cleaned data.

- EDA is used to explore relationships in x variables with each other and with y variable (sales price).

- Linear Regression, Ridge Regression, and Lasso Regression will be used to evaluate best prediction method for home prices based on home features.

- Kitchen quality coefficient will be interpreted in terms of best model to answer research question.

Regression Model Evaluation

- Various combinations of features and regression methods will be tested to find the best one based on the criteria below:

- Does the model meet the **LINEM assumptions**?

- How does the model's **RMSE** compare to the baseline model and other models?

- How do the R^2 **scores for training, cross validation training mean, and testing groups** compare to each other and other models?

- Is the model's coefficients **easily interpretable and statistically significant** (< 0.05 p value)?

- Does kitchen quality turn out to be a **significant feature** in the model?

Model Selection & Interpretation

- 7 different regression models were evaluated. Iterated through models starting with a kitchen sink linear regression, then adding and subtracting features, and lastly adding in regularization techniques.

- Results from my best model are shown to the right. The production model was chosen because it met the evaluation criteria the best.

- The production model ended up with 33 significant features (mix of regular features and a couple feature-engineered features.)

- The features covered house age, quality, number of rooms/bathrooms, building type, and neighborhood.

- Drawbacks of this model are slight multicollinearity and some features do not have a perfect linear relationship with price.

Predicted Values vs Residuals
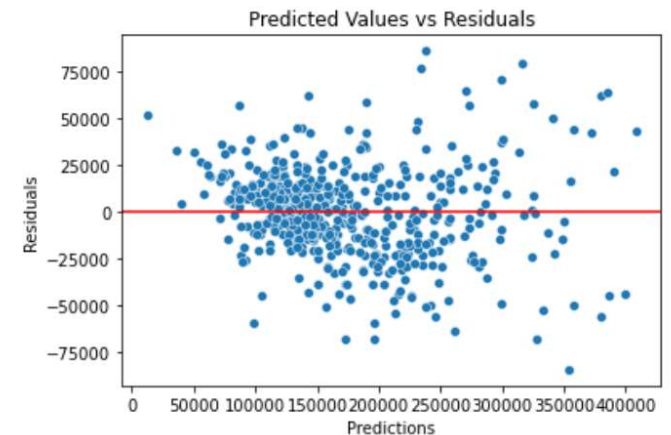
```
training r2 score: 0.8843740601067654,
cross val training r2 score mean: 0.8710889037671258,
test r2 score: 0.8837749463873775

mean squared error on test : 585055998.0610011,
root mean squared error on test: 24187.930834633233,

baseline mean squared error on test : 5107788971.020007,
baseline root mean squared error on test: 71468.7971846456
```

Key Findings

- The production model's coefficients and the corresponding p values can be used to infer the effect of home features on the selling price.

- The features that have the largest impact on the selling price in terms of this model are overall quality, the age of house and/or remodel, exterior quality, kitchen quality, basement exposure, functionality, total rooms, total bathrooms, proximity conditions, building types, and neighborhoods.

- The regression output for kitchen quality is shown in the table below. Since the p value on the coefficient is < 0.05, this is a statistically significant result that helps answer our research question.

- The coefficient for kitchen quality can be interpreted as: for every increase in the kitchen quality rating (1 – 5), the selling price will increase by $8,053.50.

	coef	std err	t	P>|t|	[0.025	0.975]
kitchen_qual	8,053.4944	1,492.945	5.394	0.000	5,125.602	1.1e+04

Recommendations

- Does the kitchen sell the home? A home's price is determined by many different features, but yes, **kitchen quality is one of the significant features that contributes to the home price.**

- As the kitchen quality rating increases, price increases. According to our model holding all else equal, sellers could improve their kitchen quality and increase home value by tens of thousands of dollars.

- **About 90% of homes sold in Ames, IA did not have a 5 star kitchen quality rating**. Knowing the potential increase in home price by increasing the kitchen quality rating is valuable information to all the homeowners with a less than excellent kitchen quality rating.

- The real estate company should capitalize on the success of this study. The company can now confidently advise their clients that increasing kitchen quality has a positive impact on home price and have an idea of what that impact is.

- The real estate company could also consider making the results of this study available for purchase to profit from this study.

References

- Ames Housing data from Kaggle:
  - https://www.kaggle.com/c/dsi-ames/data

- Article on the effect of kitchen quality on home sales from Fales Realty:
  - http://www.falesrealty.com/BLOG/ArticleID/1/Real-Estate-Secrets-Why-and-How-Kitchens-Sell-Homes