

# Ames Housing Market

What matters most when it comes to selling a home

Part of a team at a small real estate company in Ames Iowa, looking to develop our own inhouse methods for evaluating the market and improve our ability to advise our clients

---

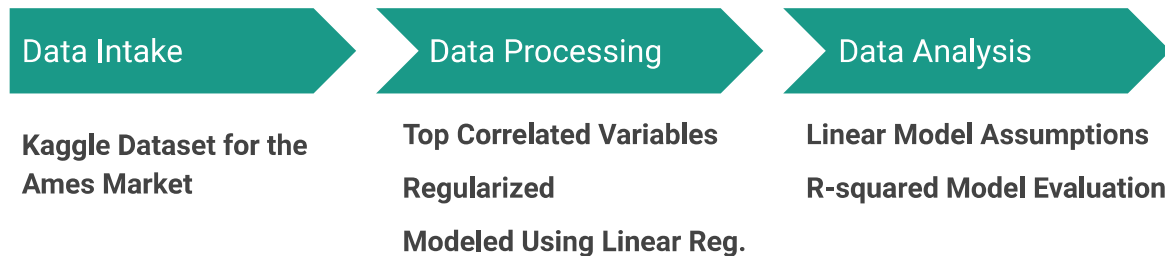
## Objectives:

- **Primary Factors in Home Value**
- **Model for Predicting Home Value**

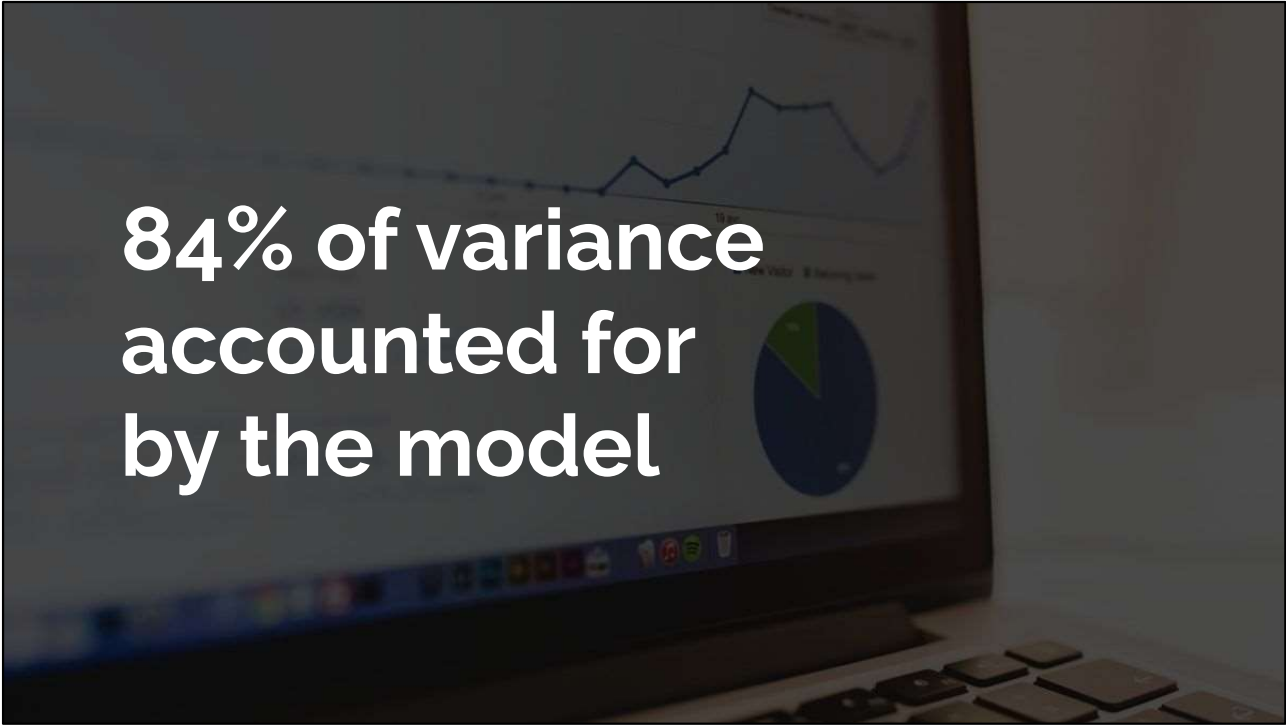
We want to determine the primary factors that affect home prices in our region.

We want to generate an effective model for predicting home value for our clients.

## Summary of Analysis



- For training data we used the extensive Kaggle Dataset for the Ames Market.
- That data was scrubbed for missing values and incorrect data types, then fed into our data processing strategy.
- We started by processing categorical variables along with numerical variables, and downselecting features of interest relative to correlation with sale price.
- Those features of interest were then combined using a polynomial feature generator, and then dominant features were identified using LASSO regularization.
- The top features were selected based on the LASSO output, and used in the linear model.
- In analyzing the model, we paid particular attention to the Linear Model assumptions of: Linearity, Independence, Normality of Errors, Equal Variance of Errors, and no Multicollinearity
- Model performance was evaluated using R2 error by way of cross-val scoring

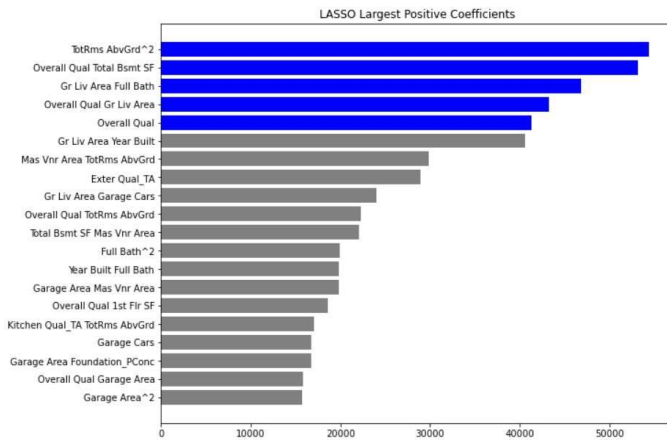


**84% of variance  
accounted for  
by the model**

Ultimately, the linear model we produced accounts for about 84% of the total variance between our predicted values and the true values in our test data. This amounts to predictions being within plus or minus approximately \$50,000 with 95% confidence.

While the model is a significant improvement over baseline, it underperforms peers in the market, and is difficult to ascertain exact impacts of features due to coefficients being obfuscated through LASSO regularization.

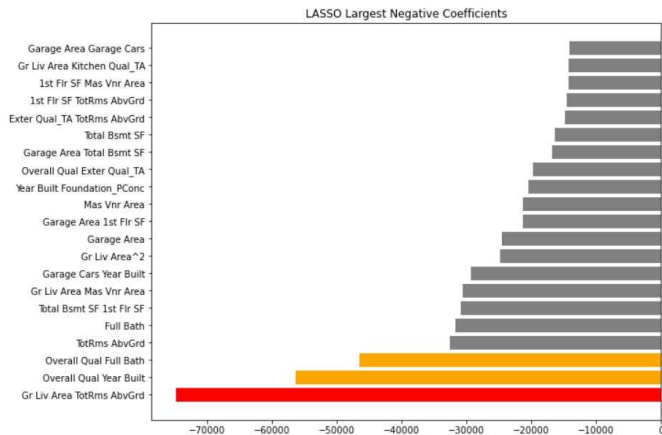
## Dominant Factors



Top 5 include either Above Ground Living Area, or the Overall Home Quality rating

The model does point towards recurring features that appear to have outsized effect, namely: Overall Quality Rating of the Home, Above Ground Living Area, Year Built, Number of Cars the Garage was built for, Year of Remodel, and finally, an inverse relationship with Number of Rooms compared with Above Ground Living Area. This is verified using a substantially simplified model, that while underperforming its more complex cousin, is far more legible and also outperforms the baseline significantly.

## Dominant Factors



Somehow...

This certainly makes sense, however.

Some of the most unusual conclusions of the LASSO regularization was that the combination of high overall home quality with the number of full bath, or year built, somehow had a strong negative effect on price.

On the other hand, the top negative coefficient was the combination of total above ground living area with the number of rooms above ground. It is better to have fewer, but larger rooms, compared to many smaller rooms.

## Recommendations:

- **Clients Should Know: Quality of Home is Key**
  - **Next Steps: Refine Model**
- 

Recommendations going forward:

1. As a starting position, we can advise clients who are selling their homes that the overall appearance of quality is of top importance in fetching a high price, and that recent remodels are also valuable. Additionally, when performing a remodel, it is advised to prioritize the size of rooms in the house over the total number of rooms.
2. We would like to go ahead to investigate further the effects of outliers on the model, and try to determine if there are high leverage features that have not been identified.

# Thank you!

---

Thank you for your time!