Company feature targets for remodeling profit

By Benjamin Mathis

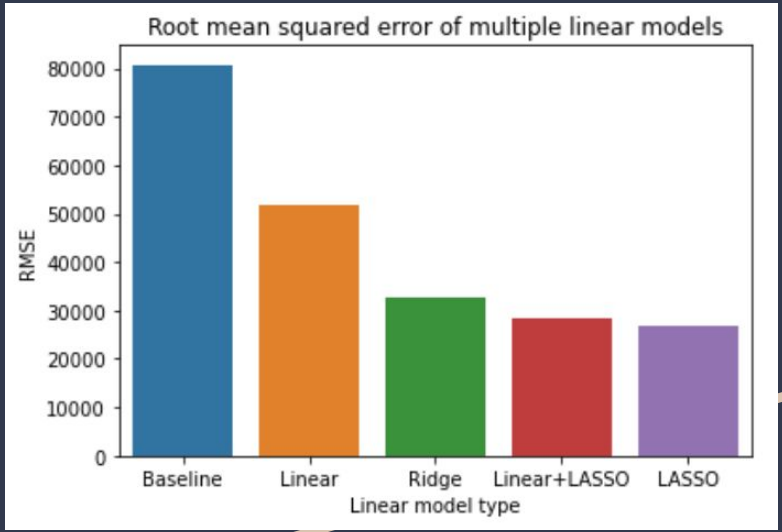
Problem Statement

We are a Data Scientist working for a company that buys housing with the intent of making cost-effective improvements and flipping them back onto the market. Our goal is to find the most effective features for this process in the context of the Ames, Iowa housing market.

Goals:

- Identify features most statistically likely to have an effect on housing sale price.
- Identify features that are the most cost-effective for remodeling and flipping.

Data Process



This dataset was pruned to better predict values for lower-cost homes.

We used two linear models in the final iteration of data processing:

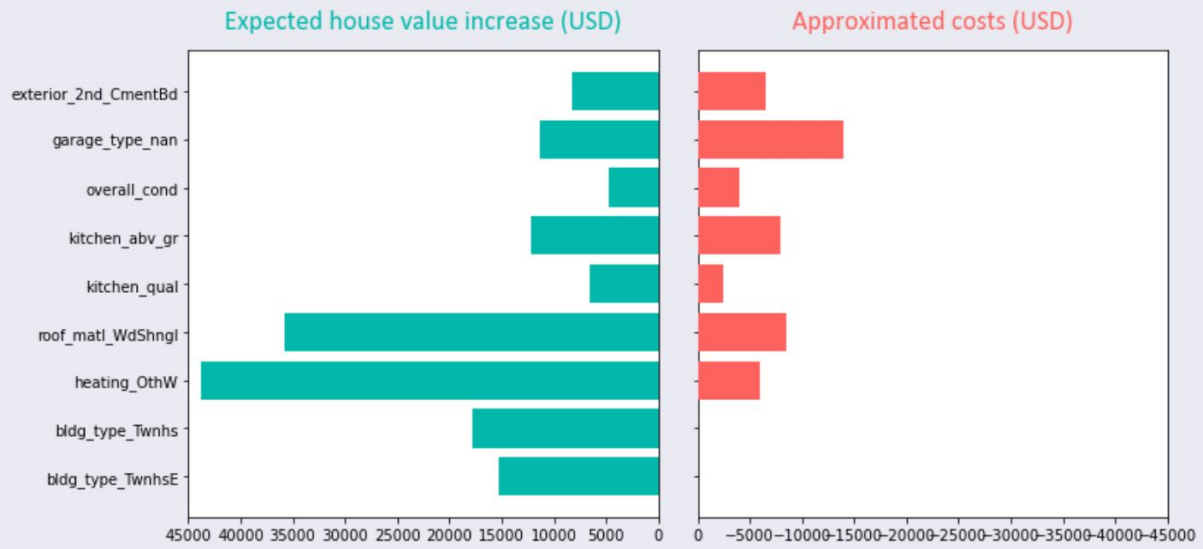
- The first used the LASSO model, which allowed us to filter out features that were not useful for predicting housing prices.
- The second used the remaining features identified by the first, and is used to determine an estimate of dollars gained by changing a feature.

Result:

The best 25 features to improve in a remodeling project were selected based on a combination of their effectiveness at predicting housing prices and the overall effect modifying them has on the price of a house.

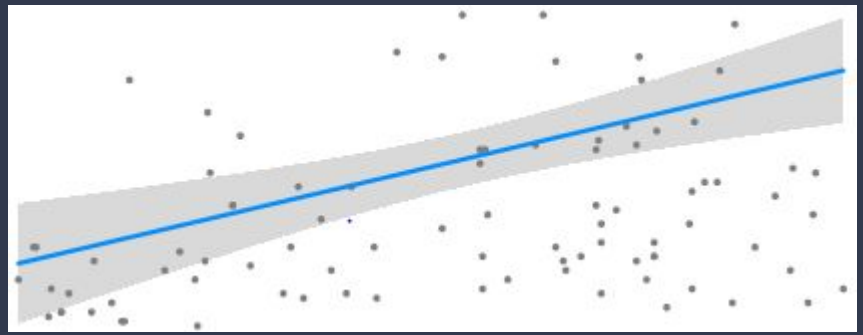
Feature Selection

Of the 25 features given, 16 were determined to be out of company scope, prohibitively expensive, or simply too vague to quantify.



The remaining features' expected profits are plotted here, with approximate costs of modifying the features alongside.

Conclusions



Several features have been identified as having potential for cost-effective remodeling:

- Adding cement board paneling
- Converting garages and kitchens to living space
- Painting the exterior
- Replacing roof with wooden shingles
- Purchasing better equipment for kitchens
- Adding a gas heating system
- Converting Townhouses to Condominiums

Early modeling suggests that these will turn profit for the company, but additional research is required. The accuracy of the model is bounded by its error, and the model does not hold strictly to necessary data assumptions, so these conclusions should be taken as recommendations only.

Credit

Final image taken from:
https://stats.stackexchange.com/questions/101318/understanding-shape-and-calculation-of-confidence-bands-in-linear-regression