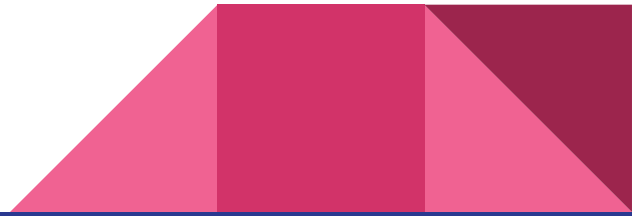
Housing Pricing Model: Ames Iowa

Scott Shapiro

Ames Housing Data:

- Data for 1450 residential properties
- 81 features
- Example features:
  - Neighborhood
  - Land/lot features
  - Internal features (beds/baths)
  - External features (material/foundation)
  - Utility features (heating/electrical)
  - Quality measures
  - Sales info (price/condition/timing)



Questions To Answer

1. Develop an algorithm to reliably estimate the value of residential houses based on fixed characteristics.
2. Identify characteristics of houses that the company can cost-effectively change/renovate with their construction team.
3. Evaluate the mean dollar value of different renovations.

Fixed Features and Assumptions

Features:

- Fixed Features (49 variables)
- Mix of continuous and categorical

Assumptions:

- Categorical missing data represents non-existence
- Continuous missing data
  - Lot Frontage, Masonry veneer area and garage year built
- Square foot breakdown is fixed
  - Ex. LowQualFinSF can’t be upgraded

Fixed Feature Model

Training: 2006-2009 House Sales

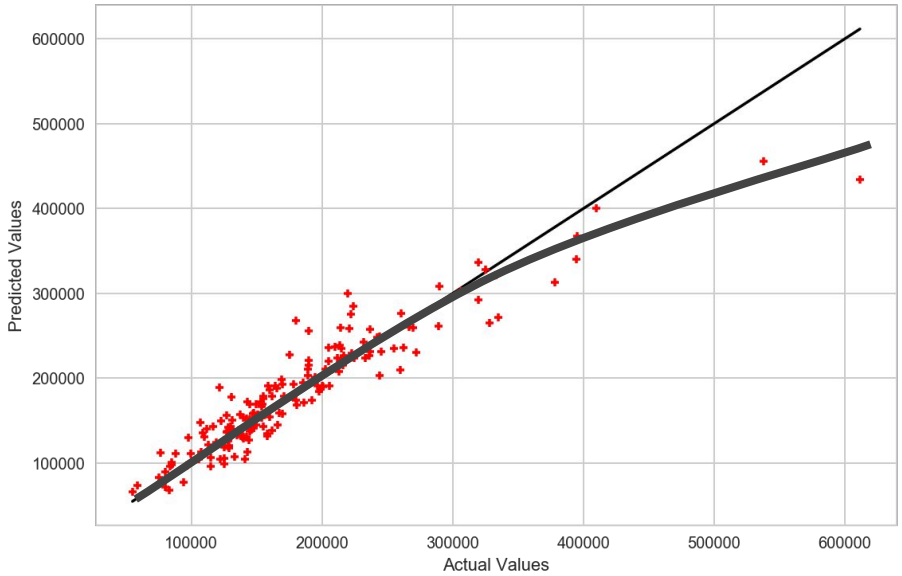
Test: 2010 House Sales

Features: 302 columns (36 continuous, 266 categorical)

Target: Sale Price

Model: Elastic Net Cross Validation

R^2 = 0.871470762104



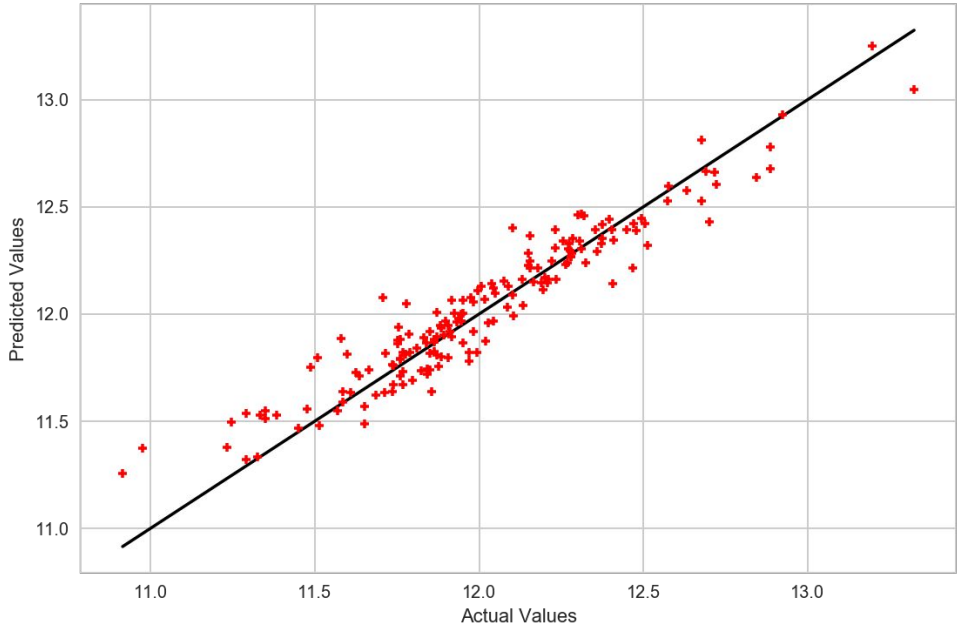
Logarithmic Scale

Model: Elastic Net Cross Validation

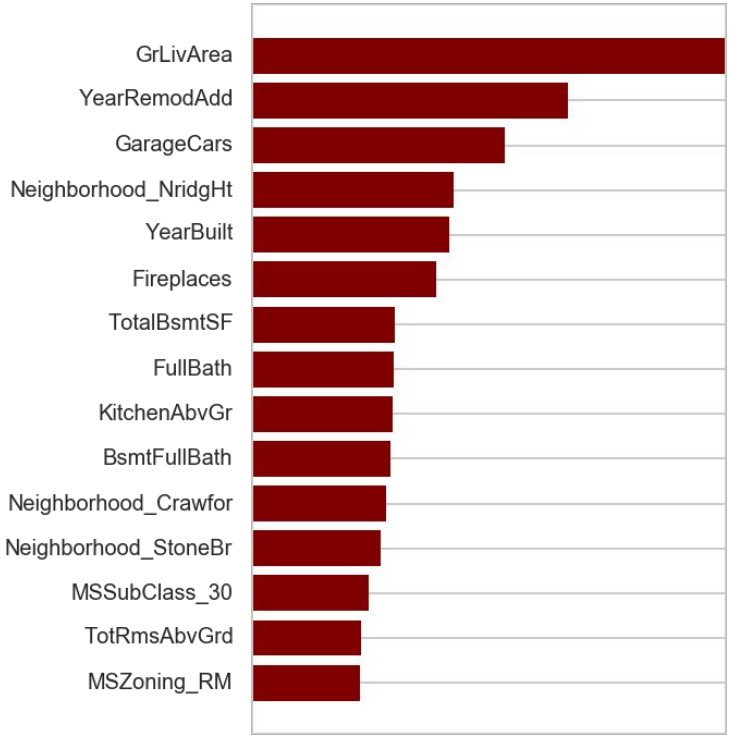
Target: Log(Sale Price)

R^2 = 0.903365868962

Features: 75 non-zero (20 continuous, 55 categorical)



Key Features



Top Features:

- **Above ground living area (sqft)**
- Remodel date
- Size of garage in car capacity
- Living in Northridge Heights
- Year Built

Variable Features

The residuals from the first model (training and testing) represent the variance in price unexplained by the fixed characteristics. Of that variance in price remaining, how much of it can be explained by the easy-to-change aspects of the property?

Residual = Price Sold - Predicted Price Sold

For this model, examples of things that **ARE renovate-able:**

- "Quality" metrics, such as kitchen quality
- "Condition" metrics, such as condition of garage
- Heating and electrical components

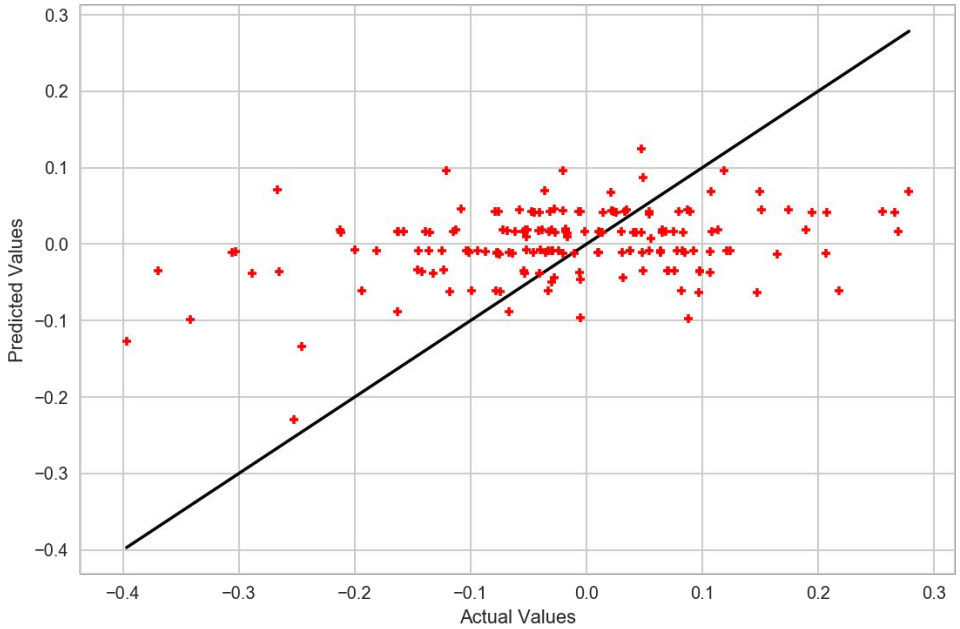
Variable Model

Model: Elastic Net Cross Validation

Target: Residuals

Features: 91

R^2 = 0.0796382056359



Targets For Renovations

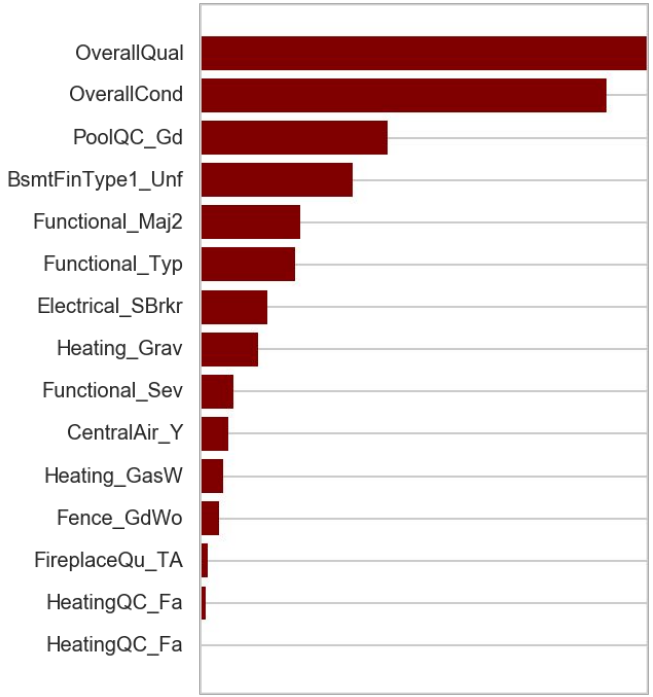
Inconclusive!

Top features: OverallQual, OverallCon

Need to know more about those metrics.

OverallQual: Rates the overall material and finish of the house

OverallCond: Rates the overall condition of the house

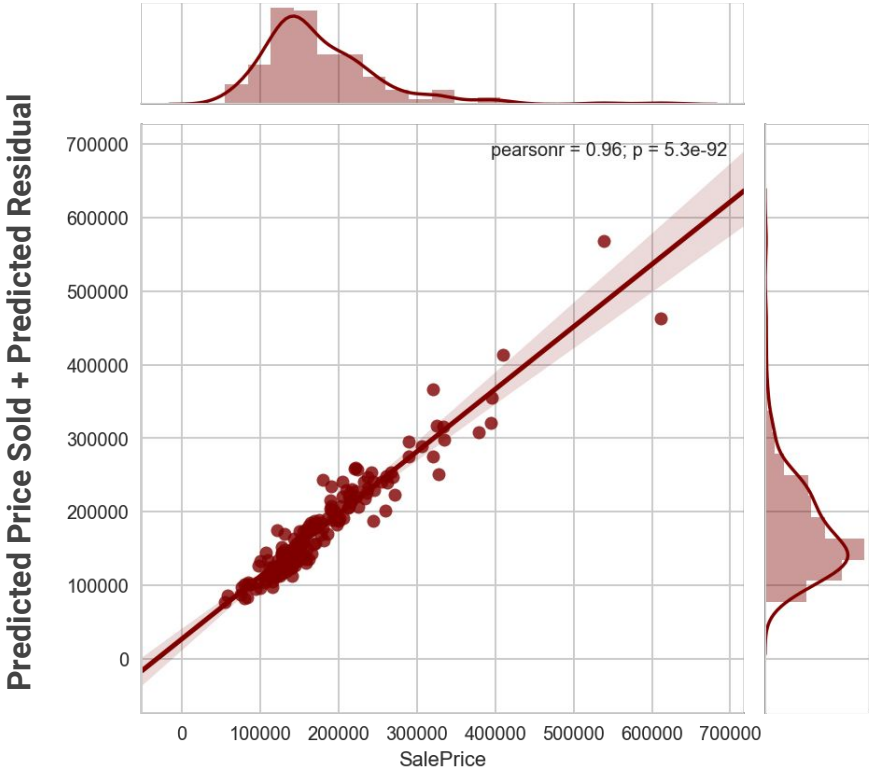


Final Results

While it does not provide insight on renovate-able elements, the residual model does contain added value for our predictions.

Actual Price Sold ~ Predicted Price Sold + Predicted Residual

Final Results



Correlation = 0.955440554433

Next Steps

- Learn more about OverallQual and OverallCond to find tangible renovation targets.
- Attempt to predict sales conditions - can we take advantage of foreclosures and other bargain sales?
- Start a consulting division with our hyper accurate model!