# Predicting House Sale Prices Using Regression

Kyle Ness, RE Consultants Co.

# Problem

Homeowners may often find getting an accurate valuation for their home difficult.

- Many factors at play - some obvious, while others are more abstract or subjective
- What features are the most significant for price?
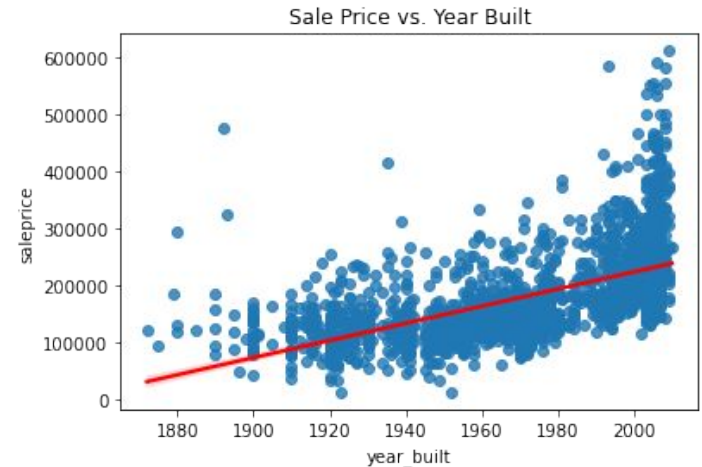- Solution: find your price using our model

# Analysis - Summary

Analysis was conducted on a dataset of residential RE sales in Ames, Iowa

- 2,051 observations on 80 independent variables for sales ranging from 2006 to 2010.
- Two multiple regression models (unscaled data), one Lasso regression, and one Ridge regression model explored.
- Engineered features: total square foot area, combined overall score, combined basement score, and more.

# Processing

# Strong Relationships


Sale Price vs. Total Sq. Ft. Area of Home


Sale Price vs. Overall Quality of Home


Sale Price vs. Year Built

Others: exterior quality, garage capacity, 1st floor area, fireplaces

# Best Model

The four models described were fit and evaluated using coefficient of determination ($R^2$).

- Multiple regression on 284 independent variables (including categorical), no scaling required.
- Only very slightly outperformed Lasso cross-validation.
- $R^2$ on test data = ~0.93

# Conclusion

- This model can be used by homeowners online for predictions that aren't a wild guess.
- Insights into what matters most
- Future:
  - Evaluate models with other metrics
  - Research the ROI of specific renovations - what are the best value adds for a renovator?