House Price Prediction in Ames, IA.

Hedonic Price Regression Model Construction and Analysis

Problem Statement:

Our client, Willoz, desires to break into the tech real-estate marketplace. In an effort to differentiate their product, they hope to provide industry leading home price prediction to their users.

They wanted to prioritize **predictive power** over i**nferential power.**

Ames , IA as Laboratory for Hedonic Regression

Hedonic Regression:

Values a property taking into account individual characteristics.

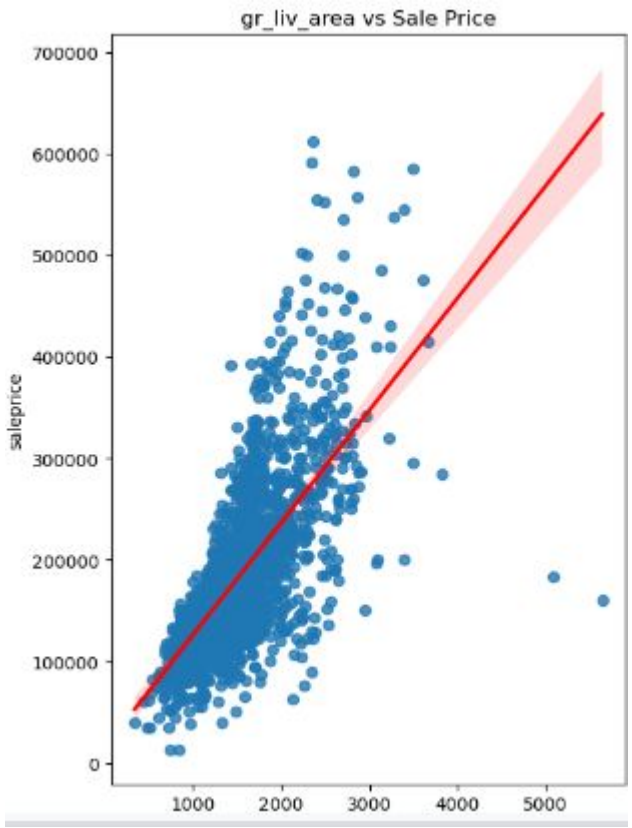
The Ames Dataset:

- Spans 2006-2010
- More than 80 features
- Covers everything from ceiling height in finished basements to the area of masonry and shingle material.

Correlation to Price

Most Important factors:

1. Gross Living Area
2. Overall Quality of construction/materials
3. Location

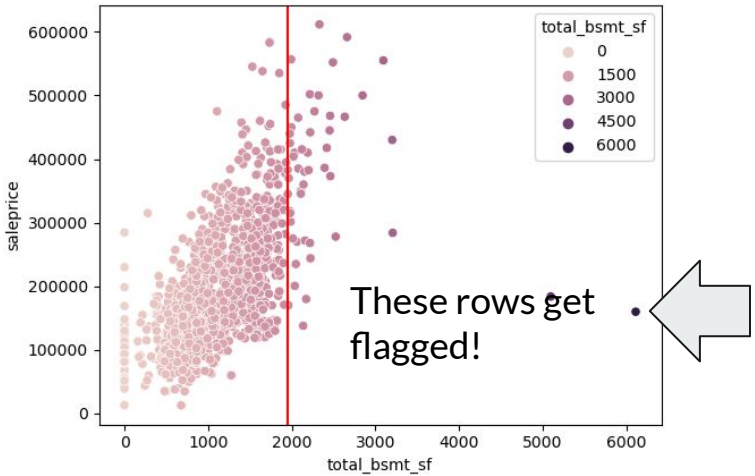


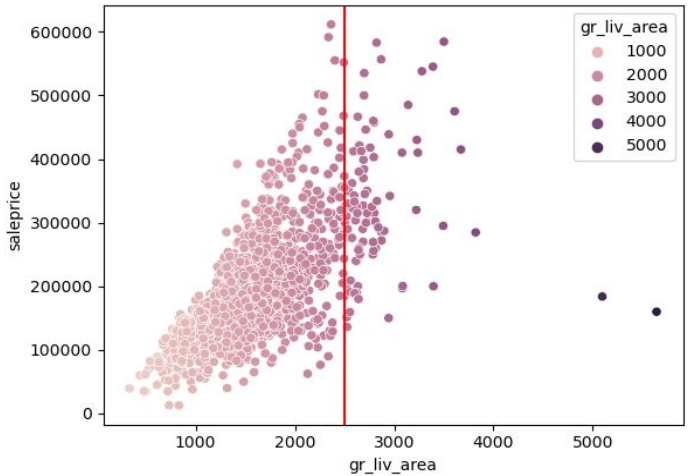
Feature Engineering

- Combined features (total square footage)
- Has a pool/fireplace/second story
- Proximity to 2007-2008 Housing Boom/Bust
  - **Avg. home sale price during bubble = $ 279,000**
  - **Avg. home sale price before bubble = $ 173,000**

Outliers

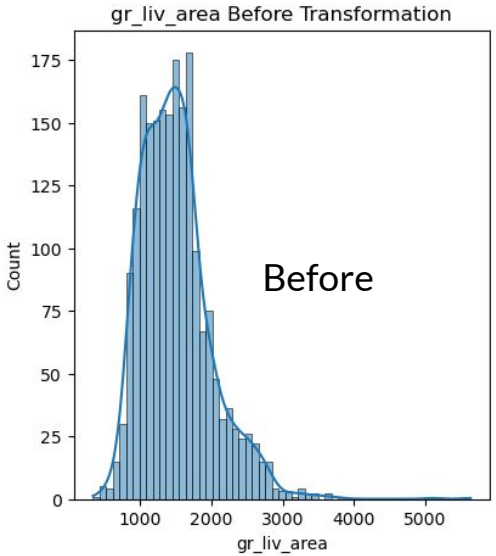
- Removed all homes > 4,000 square feet.
- Noticed Interesting patterns, created “flag” variables.

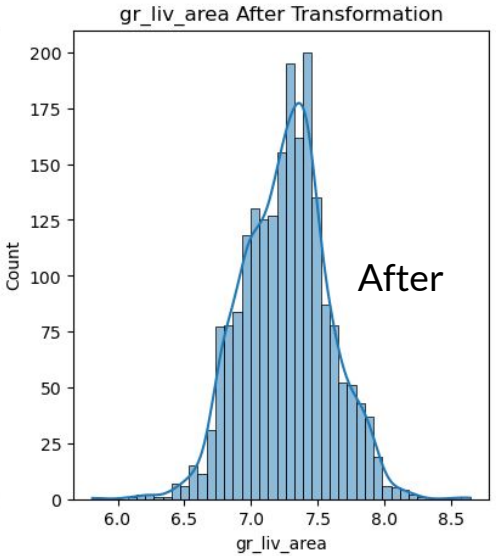


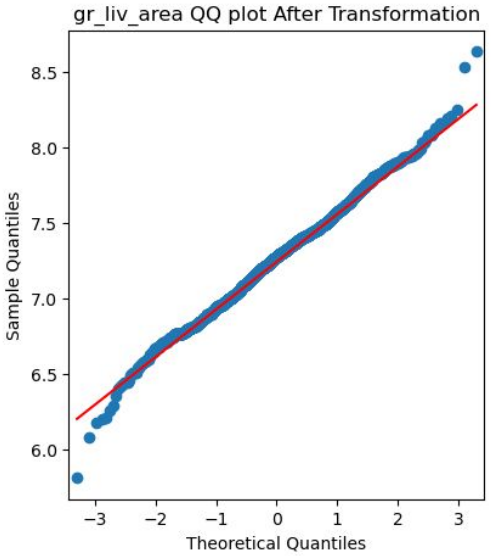


Transformations

- Normal Distributions in Disguise!

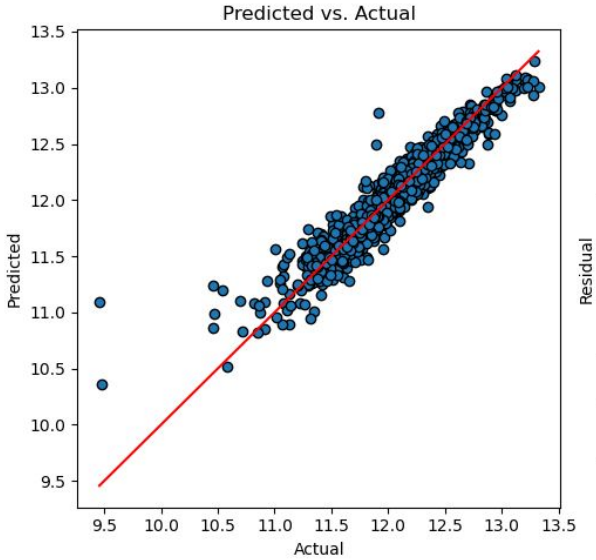


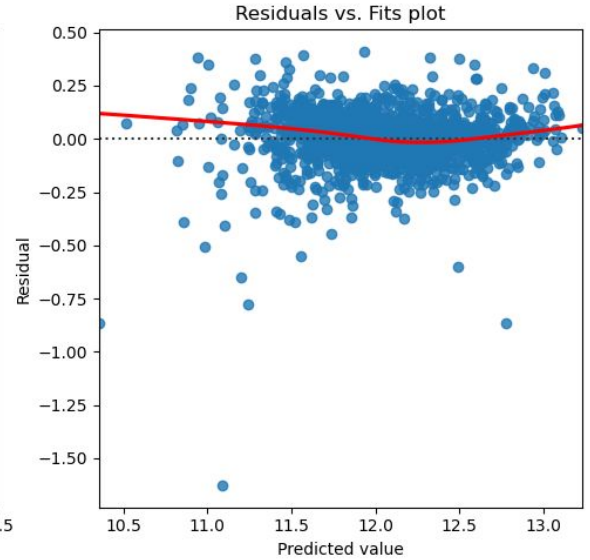


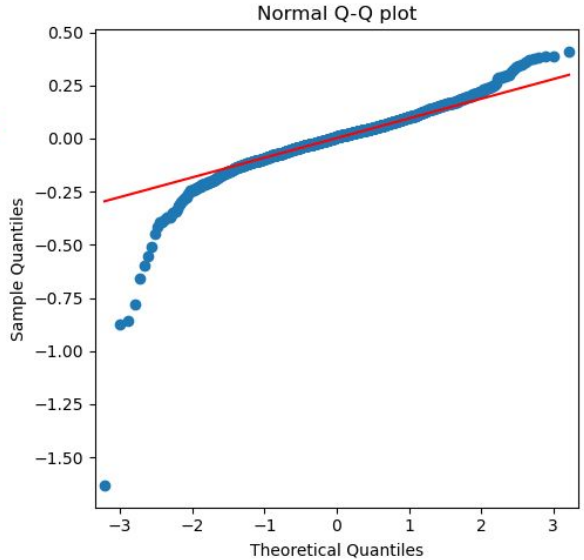


Modeling

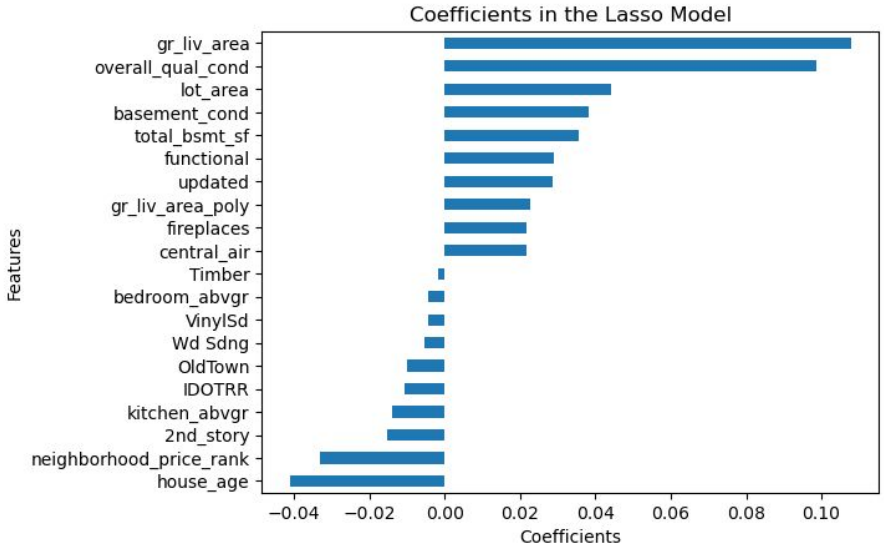
Metrics:
R-squared: **.92**
R.M.S.E : **$20,700**







Conclusions



1. Gross SF, home quality, age, neighborhood most important.
2. Transformations are likely required
3. With complex predictive power there is a loss of inferential power.

Thank You!

Questions?