

AMES

Housing Price Prediction

2006 to 2010

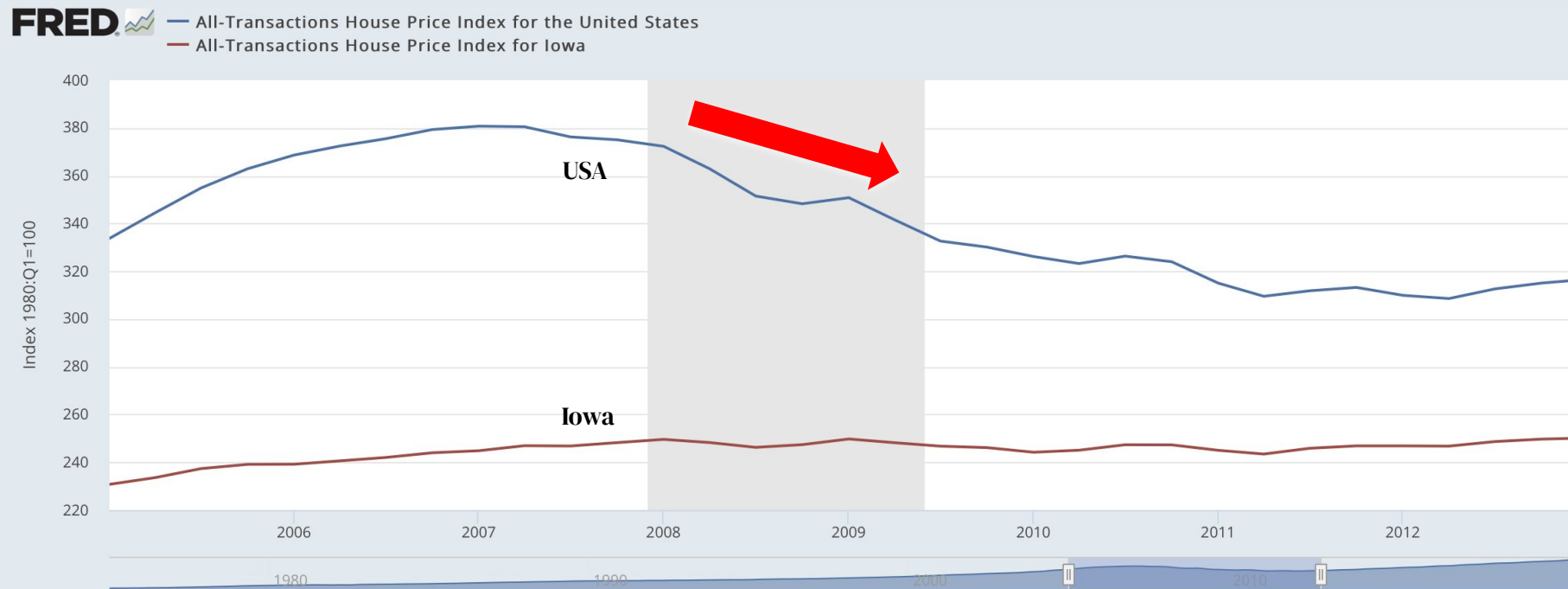


Problem Statement

- 1) Investigate how closely the linear regression **predicts** prices based on past data
- 2) Determine the most important **fixed** features in determining the sale price of a property.
- 3) Identify the features that can be **renovated** to improve prices.
- 4) Identify the features that are **detrimental** to prices



House Price Index



U.S. recessions are shaded; the most recent end date is undecided.

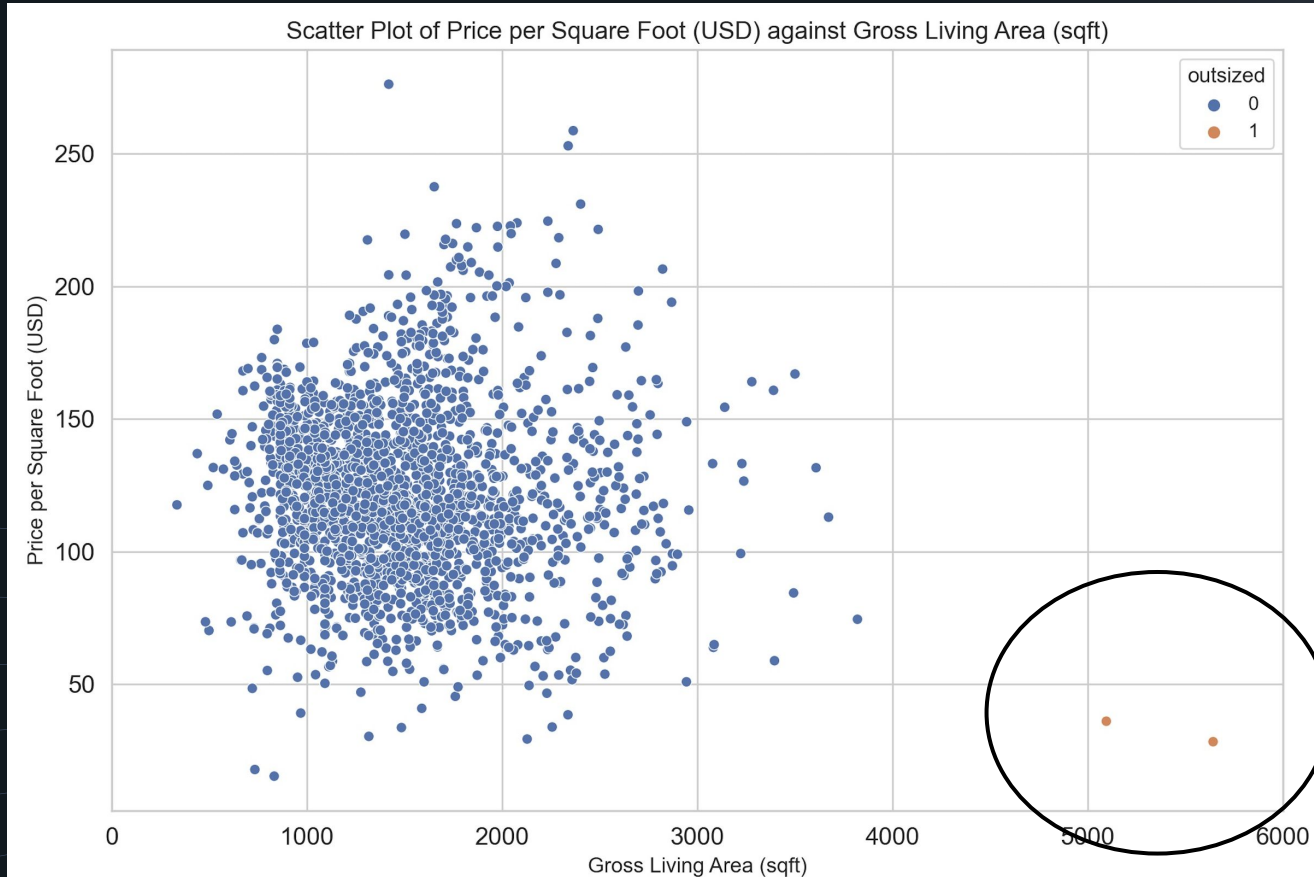
Source: U.S. Federal Housing Finance Agency

fred.stlouisfed.org



source: <https://fred.stlouisfed.org/series/USSTHPI#0>

Outliers



Preprocessing Methodologies

missing values

Impute with 'NA' or 0 if feature does not exist

ordinal discrete

map to integers accordingly to rank

nominal discrete

get dummies / one hot encode

continuous

add polynomial features

Missing Values: Lot Frontage



no frontage



frontage = 120 ft

Feature Selection and Engineering

43 variables

initial selection

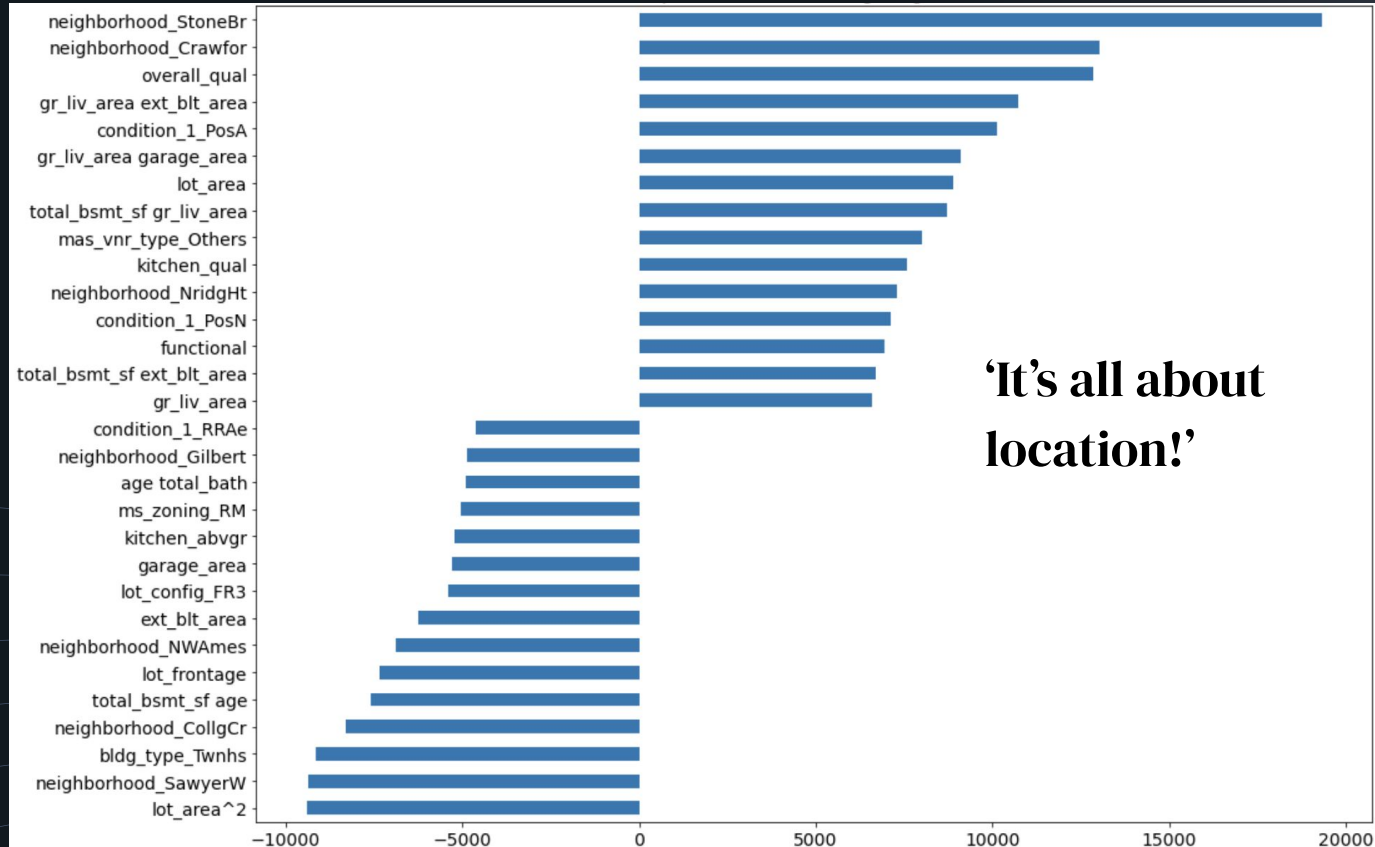
139 variables

after feature engineering

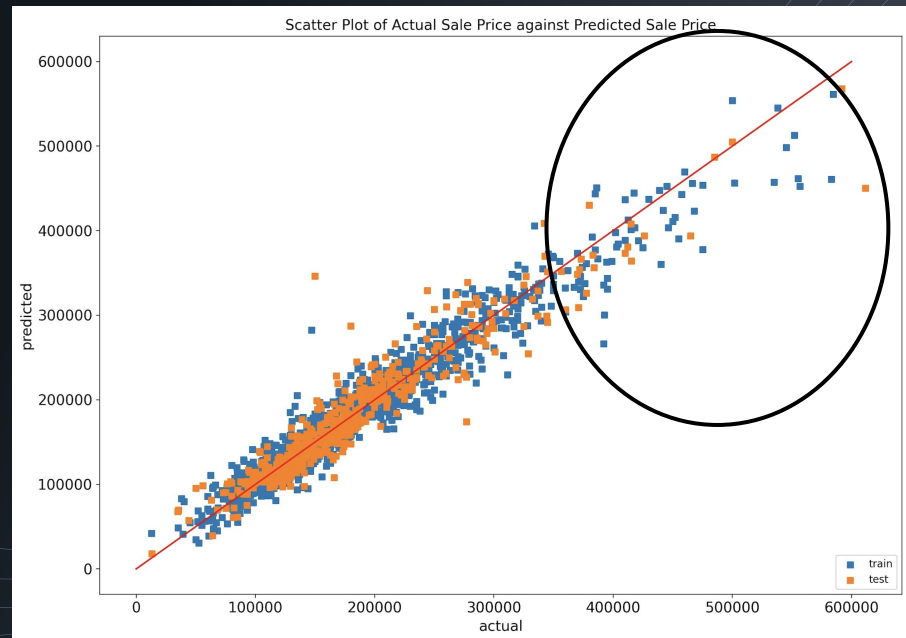
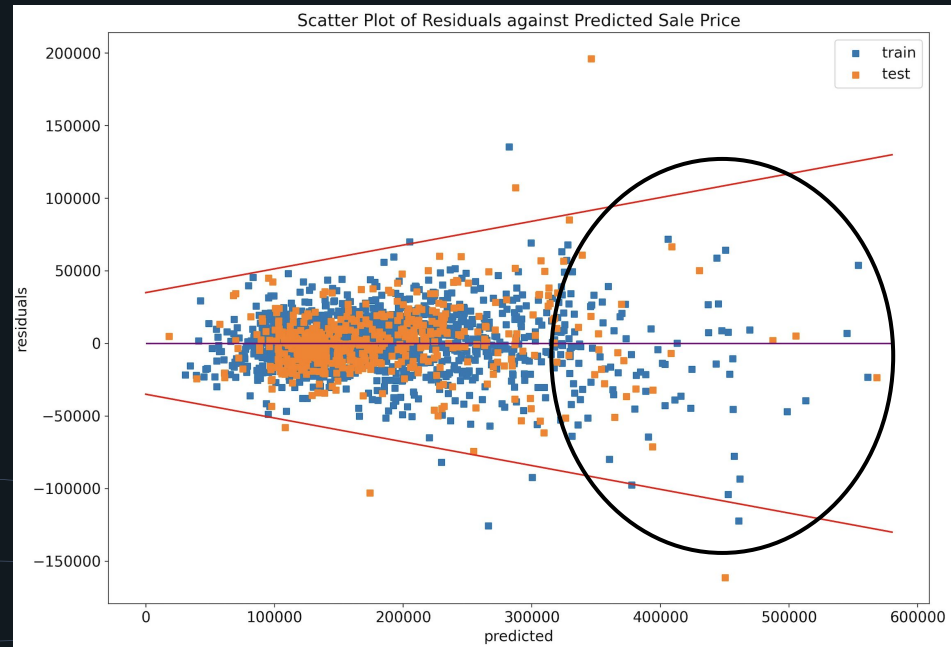
Final Model

Ridge Regression

Top Coefficients



How good are the predictions?



observation

residuals increase as price increases : heteroscedasticity

Improving the model

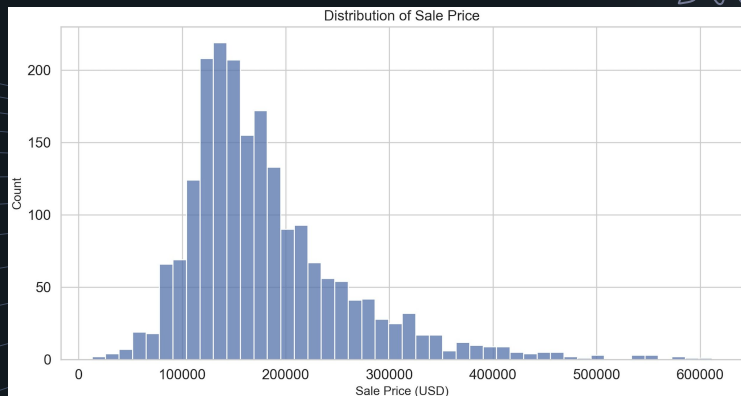
$$y = \text{price}$$

RMSE

24414

$$y = \log(\text{price})$$

18700



Conclusions

- 1) Stone Brook, Northridge Heights and Green Hills are priced the highest.
- 2) Properties nearest green belts and parks improve the sale price while those near railways do not do well.



Conclusions

- 3) Overall and in particular kitchen quality improves sale price.
- 4) Gross living area increases the sale price.



Recommendations



- 1) Improve the quality of **kitchen**, **garage** and **basement** as they do affect buyers' decision and increase sale prices.
- 2) Expand **gross living area** with extensions or additional floor.



Thank you.

The background features a series of thin, light blue lines that curve upwards from the bottom, creating a sense of depth and movement. A small, short orange horizontal dash is positioned below the 'Thank you.' text on the left side.