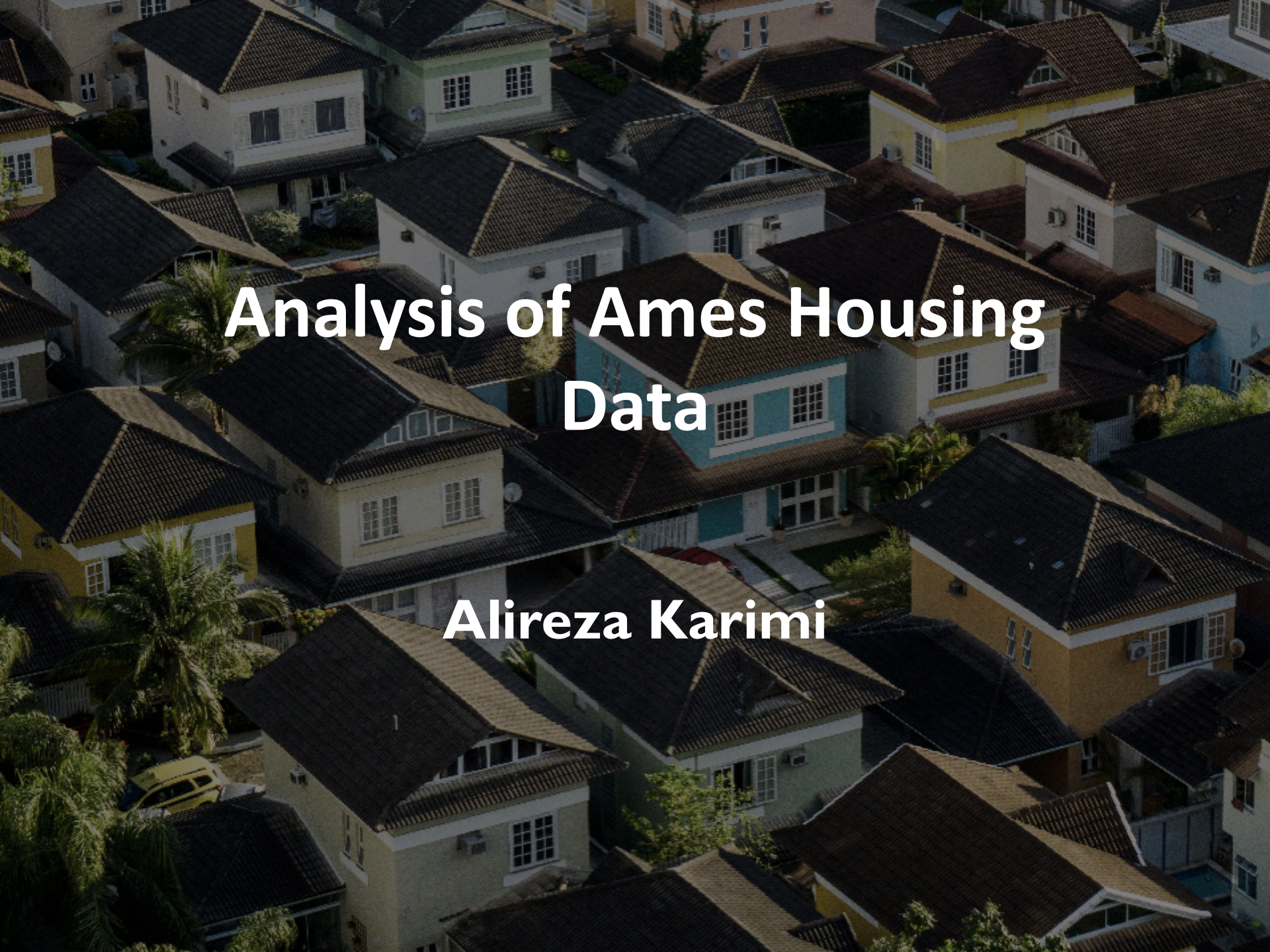


Analysis of Ames Housing Data

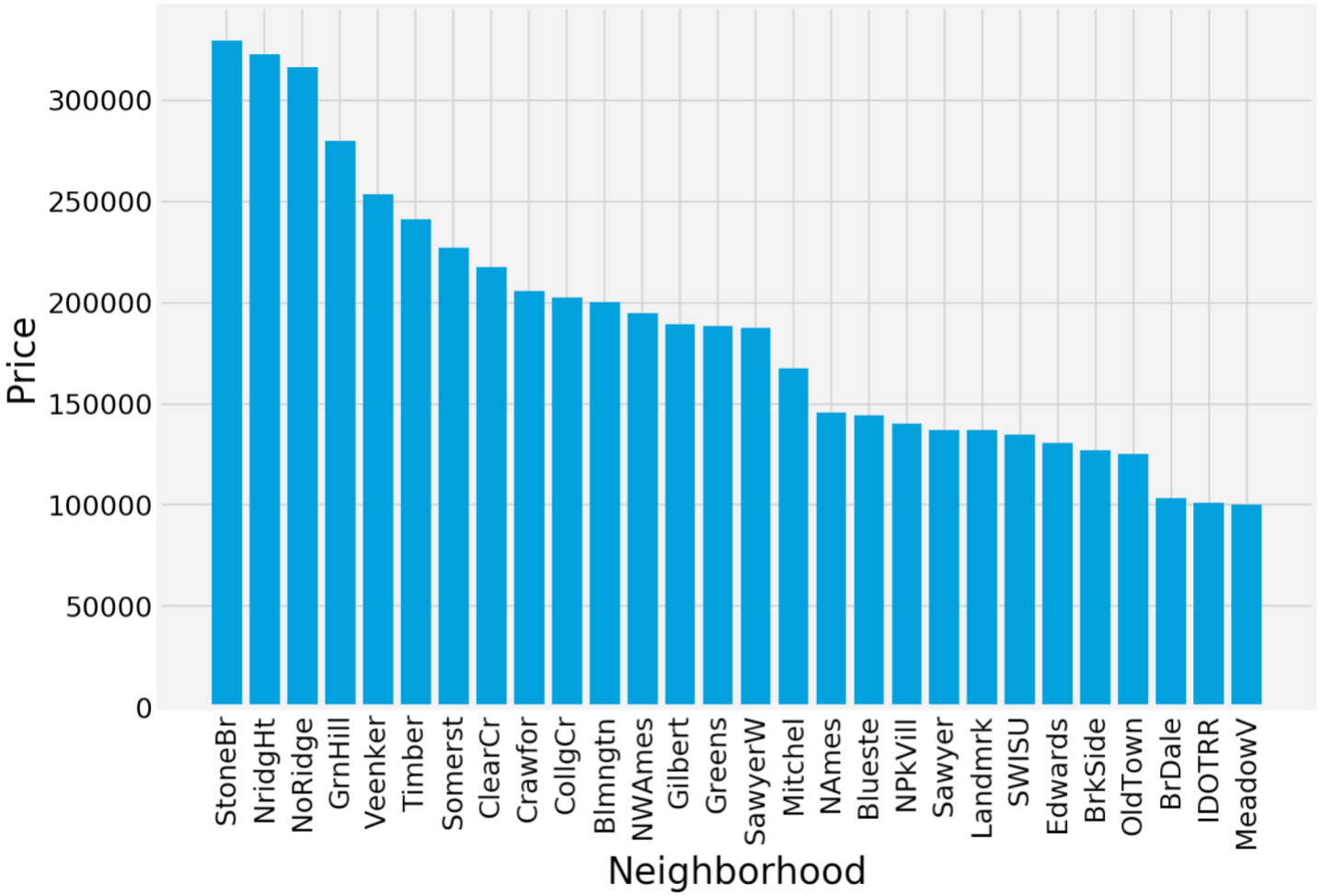
Alireza Karimi

Background

- The Ames Housing Dataset was introduced by Prof. Dean DeCock in 2011 as an alternative to the Boston Housing Dataset.
- It contains 2,930 observations of housing sales in Ames, Iowa between 2006 and 2010.
- There are 23 nominal, 23 ordinal, 14 discrete, and 20 continuous features describing each house's size, quality, area, age, and other miscellaneous attributes.
- The objective was to apply machine learning techniques to predict the sale price of houses based on their features.

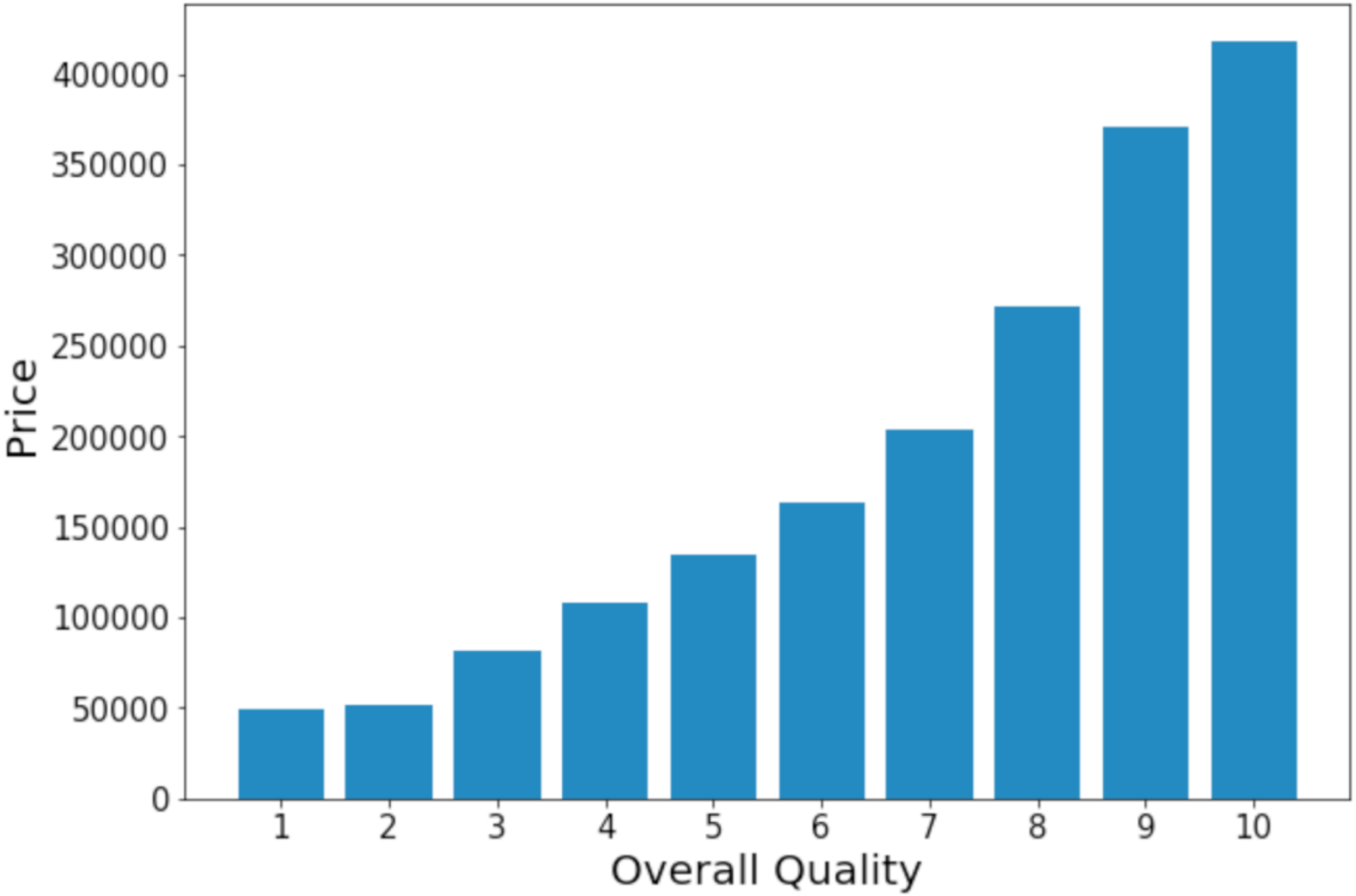
Exploratory Data Analysis

- Neighborhood has a definite impact on each property's sale price.



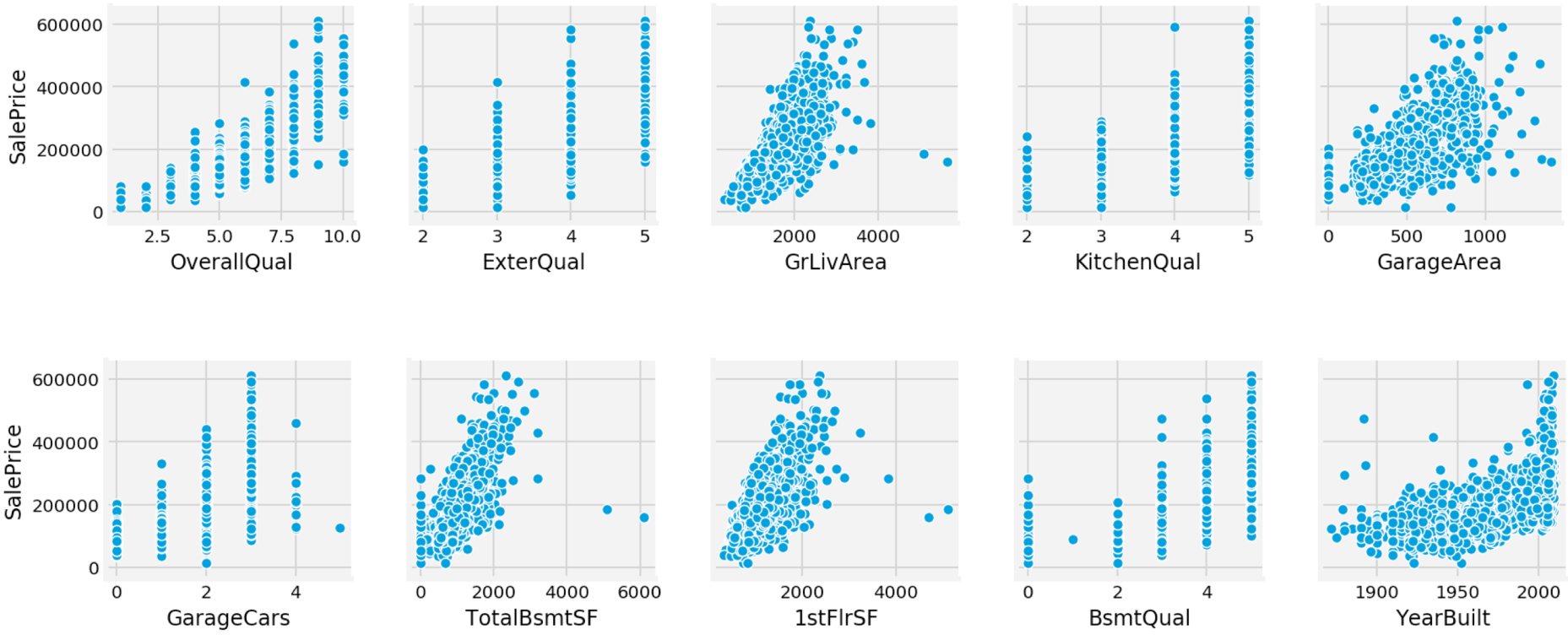
Exploratory Data Analysis

- The price tag grows steadily with the overall rating of the house.



Exploratory Data Analysis

- 10 features mostly correlated with the sale price



Feature Engineering

- All ordinal features are encoded as ordered numbers.

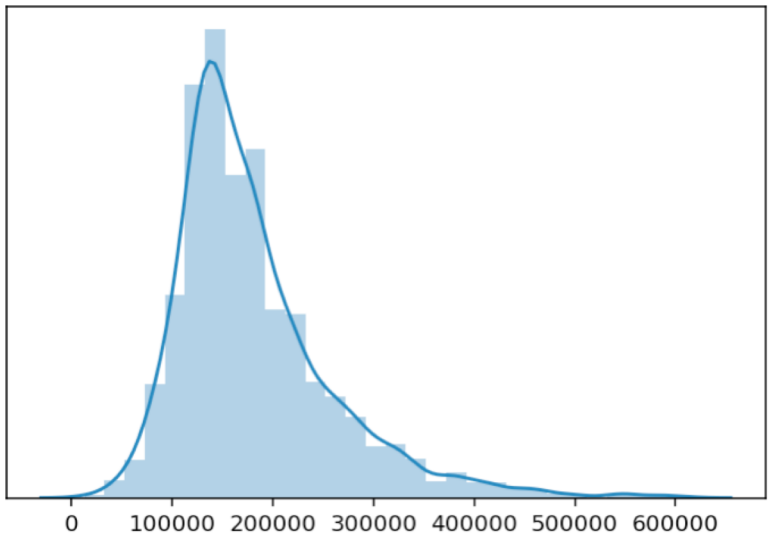
Basement Condition	Encoded Number
NA	0
Poor	1
Fair	2
Typical	3
Good	4
Excellent	5

- Dummy variables for all nominal categorical features are generated.

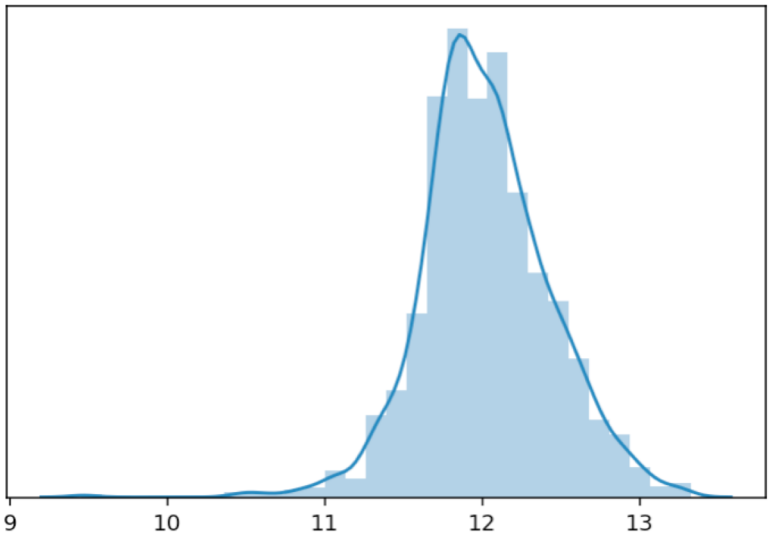
- Polynomial features of degree 2 (both interaction and quadratic terms) are generated for the first 40 parameters mostly correlated with the sale price.

Normalization of the target parameter

- The sale price shoes a positive skew which adversely impacts the quality of the model.
- A Log transformation is utilized to remedy this problem.

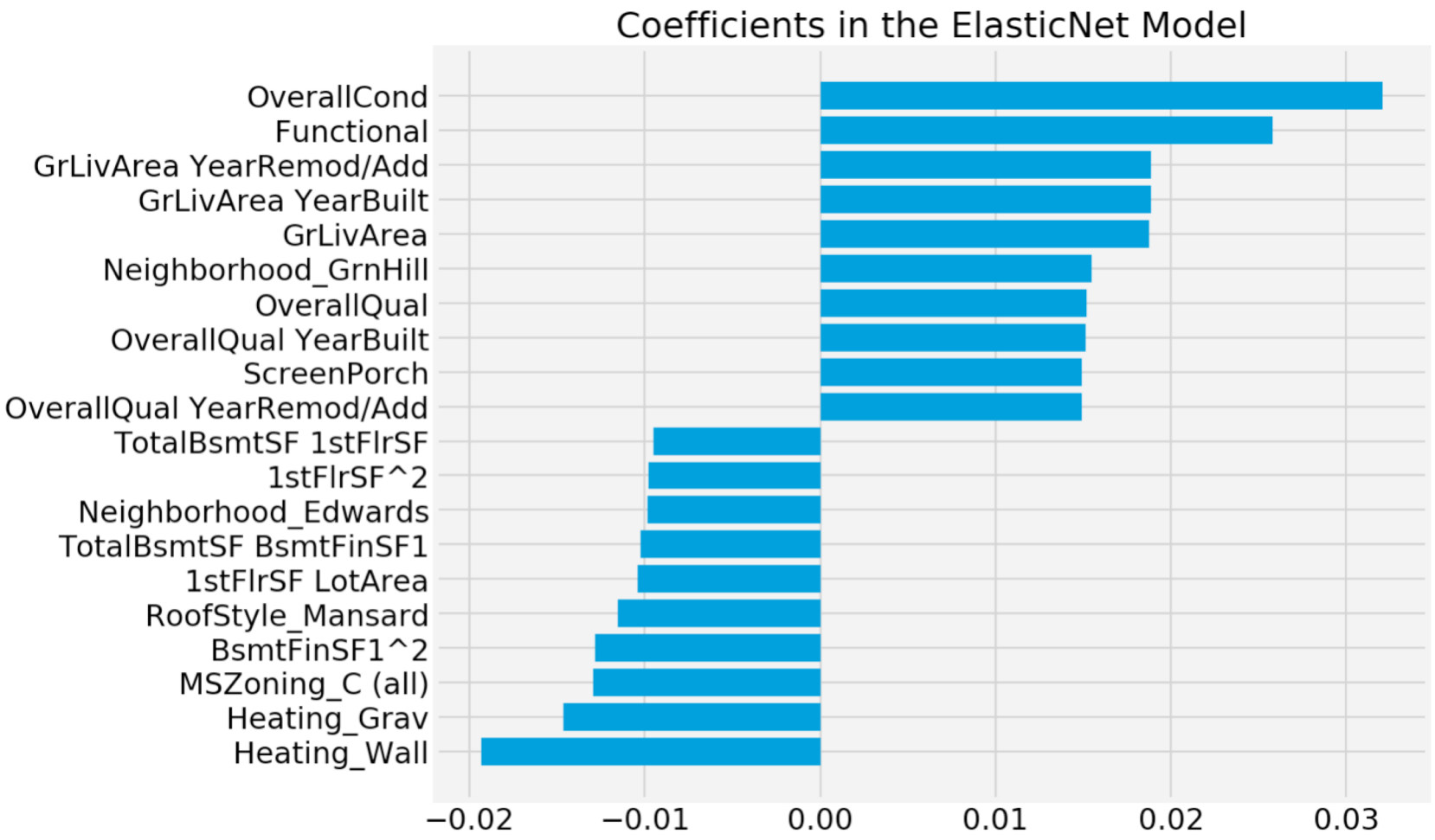


Log



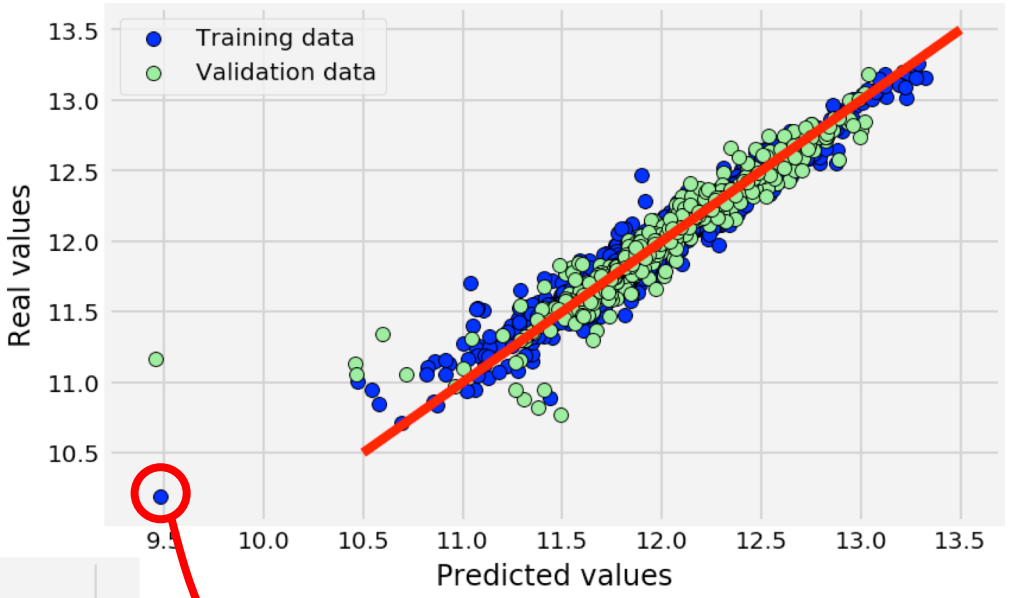
Elastic Net Regularization

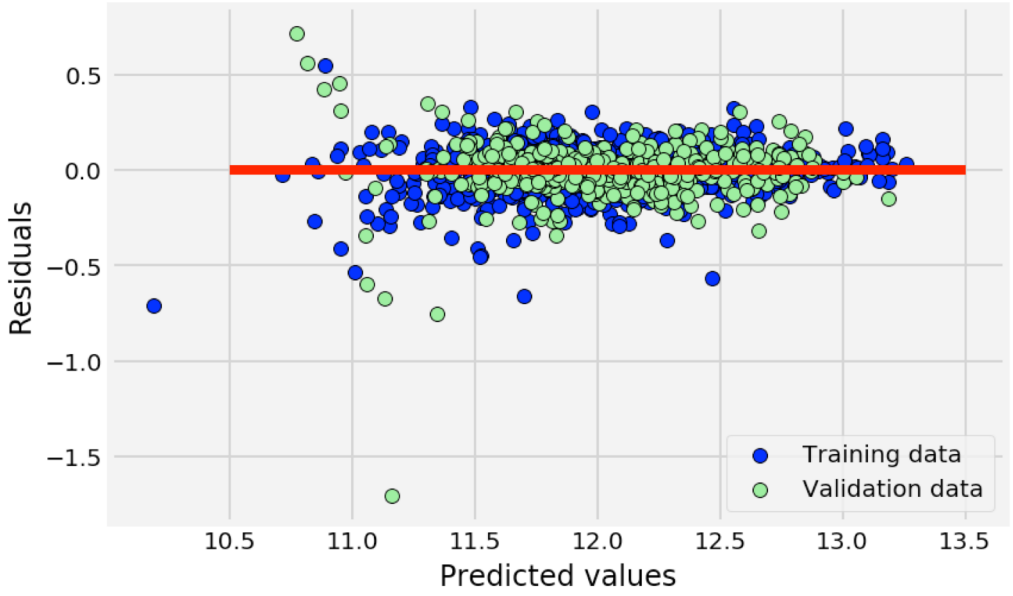
- Elastic Net picked 446 features and eliminated the other 590 features.



Assessing the Model

Data set	R^2 Score
Train	0.9475
Test	0.8859
Cross Validation (cv=5)	0.9030





Recommended by Prof. DeCock to be removed from the data set

Takeaways

Price Elevation

- Better overall condition
- Enhanced functionality
- Higher living area
- Newly constructed/remodeled
- Affluent neighborhood
- Larger screen porch

Price Depreciation

- Unfavorable type of heating
- Commercial zoning
- Unfavorable roof style
- Very large basements
- Poor neighborhood