



# Ames Iowa House Price Prediction

DSI Program

Project\_2

By

Dereje Workneh



April 10, 2020

<https://d1ic4altzx8ueg.cloudfront.net/finder-au/wp-uploads/2016/05/house-graph-vector.jpg>



# Title of Presentation

- Introduction
- Methodology
- Result and discussion
- Conclusion
- Recommendation

# Introduction



- compiled by Dean De Cock
- and is commonly used in data science education,
- it has 2930 observations with 82 explanatory variables
- **train and test** dataset describing (almost) every aspect of residential homes in Ames, Iowa.
- This dataset is part of an ongoing [Kaggle housing price prediction competition](#) that challenges you to predict the final price of each home.



# Motivation

- To explore and analyze the house dataset to find the key features that influences the sales price and develop a model to predict in house price in IOWA.

# Work Flow



EDA + Pre-  
processing  
Understand &  
visualize  
Histogram,  
density and  
scatter plot, box  
plot



Feature  
Engineering  
Transforming and  
scaling features

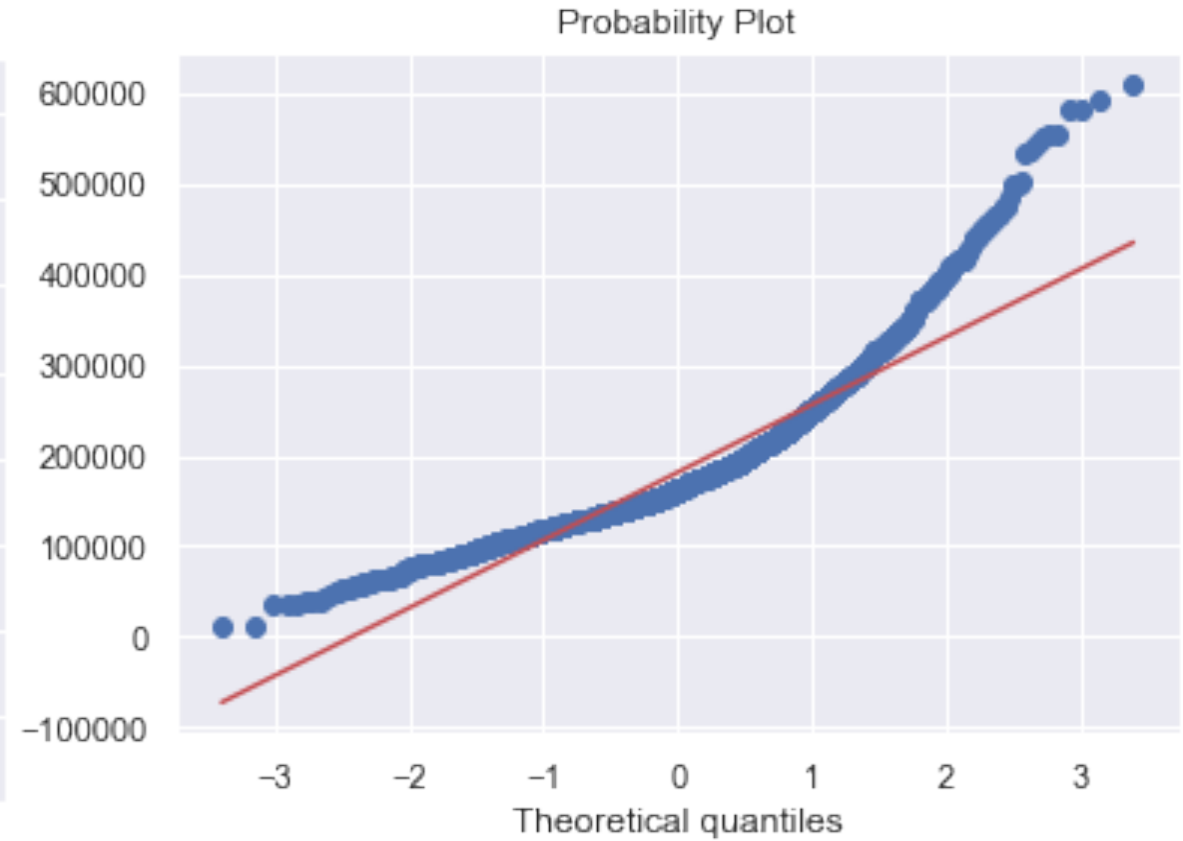


Modeling  
Ridge  
Lasso

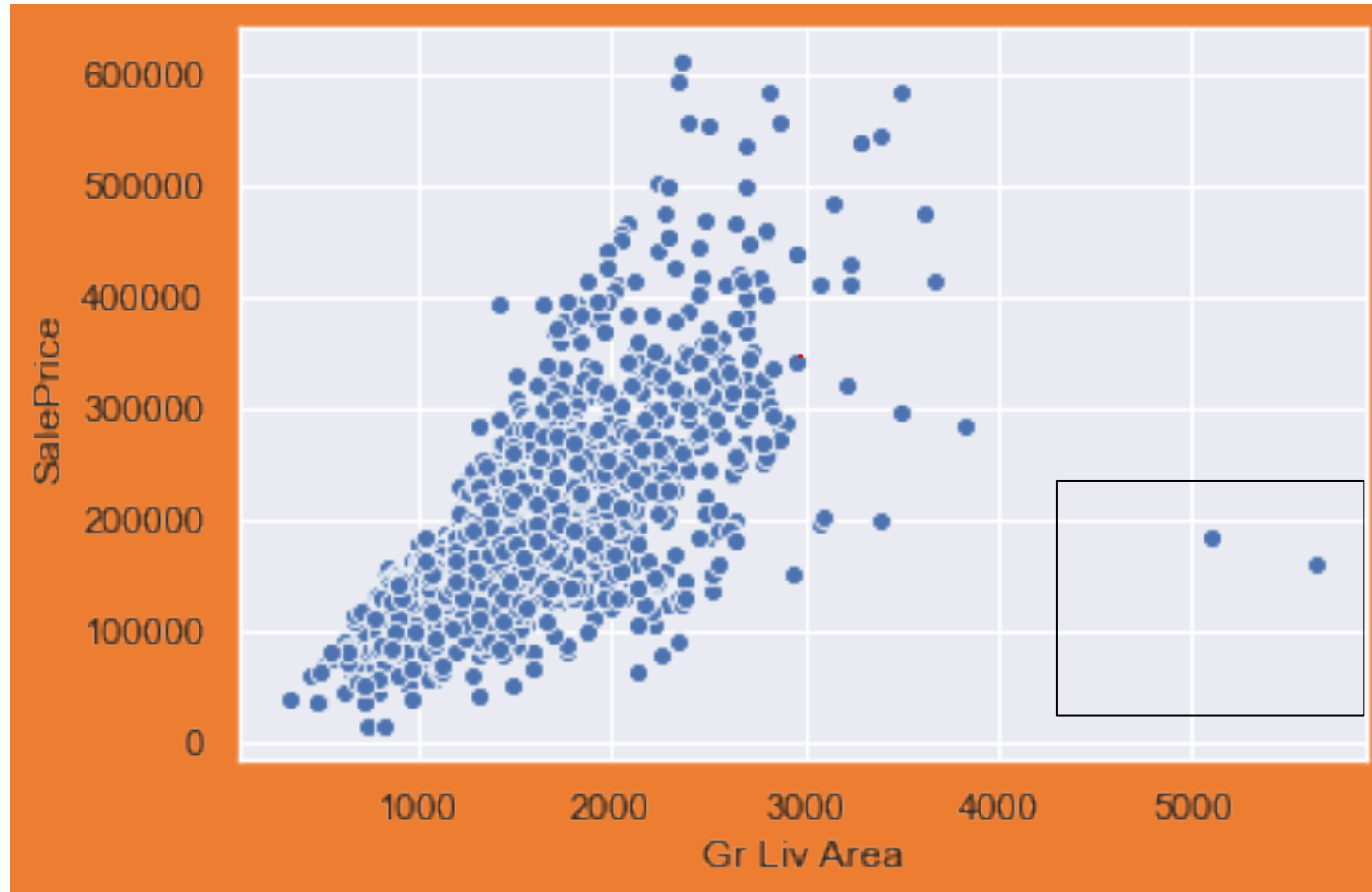


Hyperparamet  
er Tuning  
Adjusting  
hyperparameters  
based on grid  
search

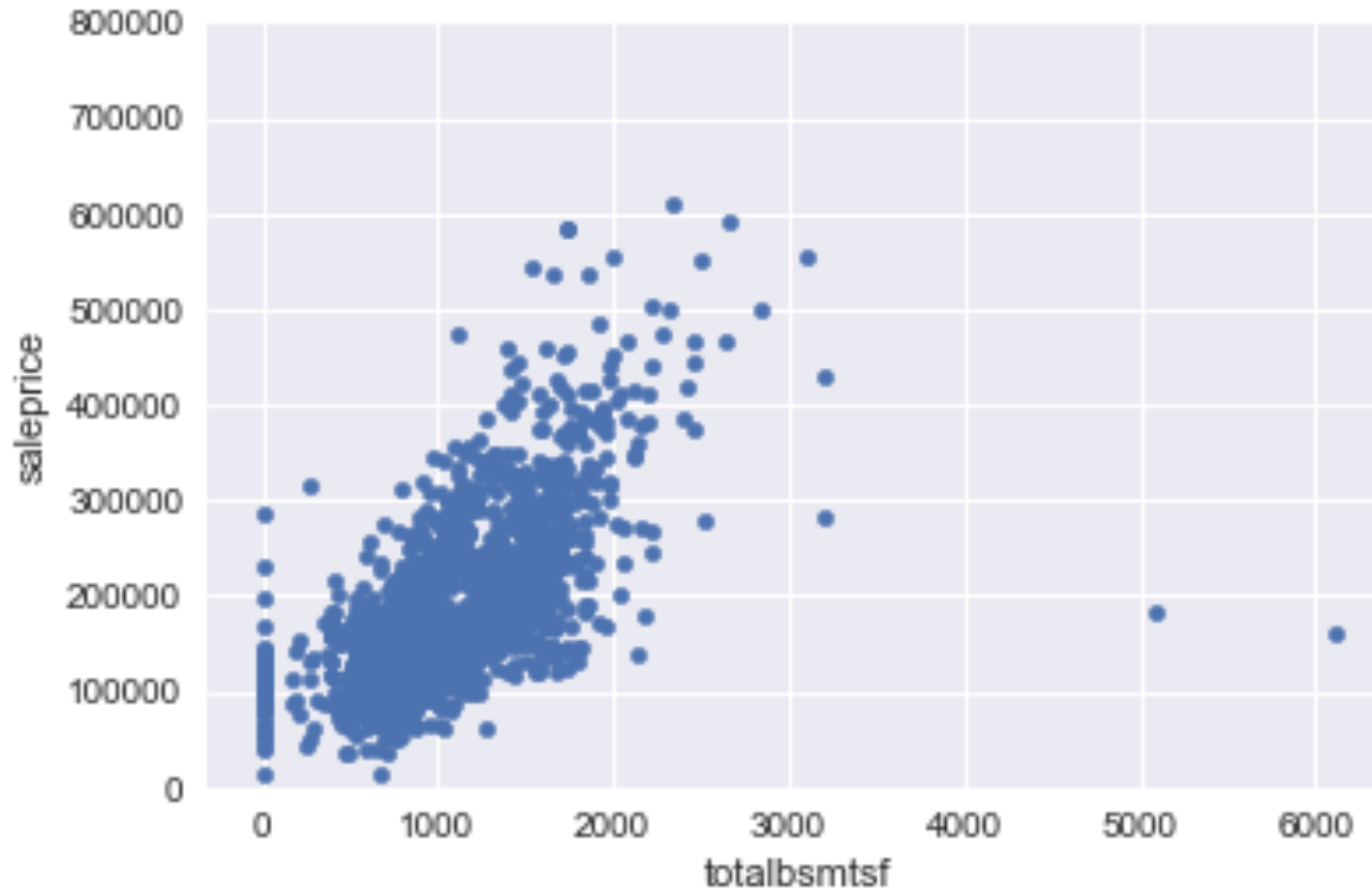
# Distribution of the target variables: SalesPrice



# Outliers: Relation Exploration for Few Numerical Variables

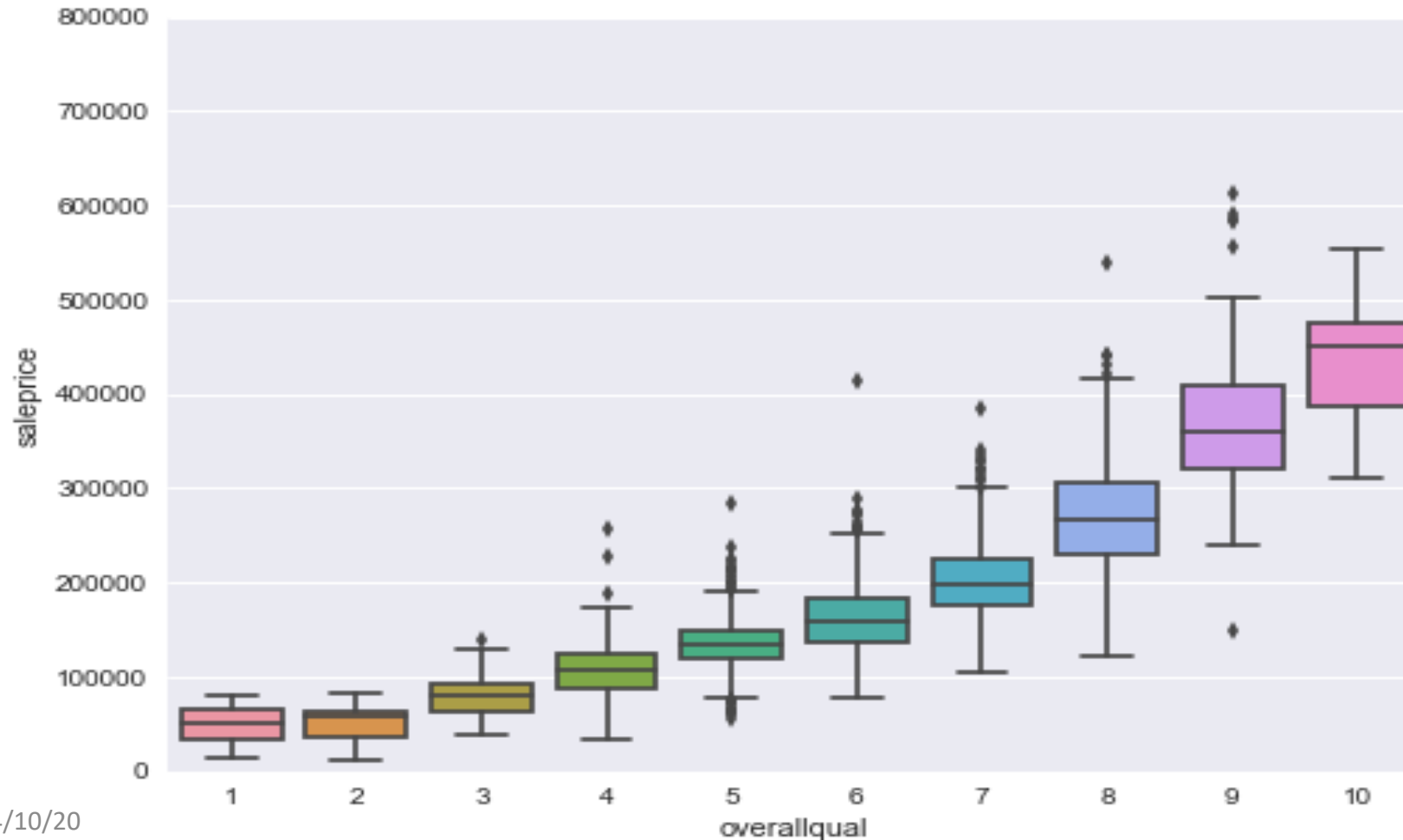


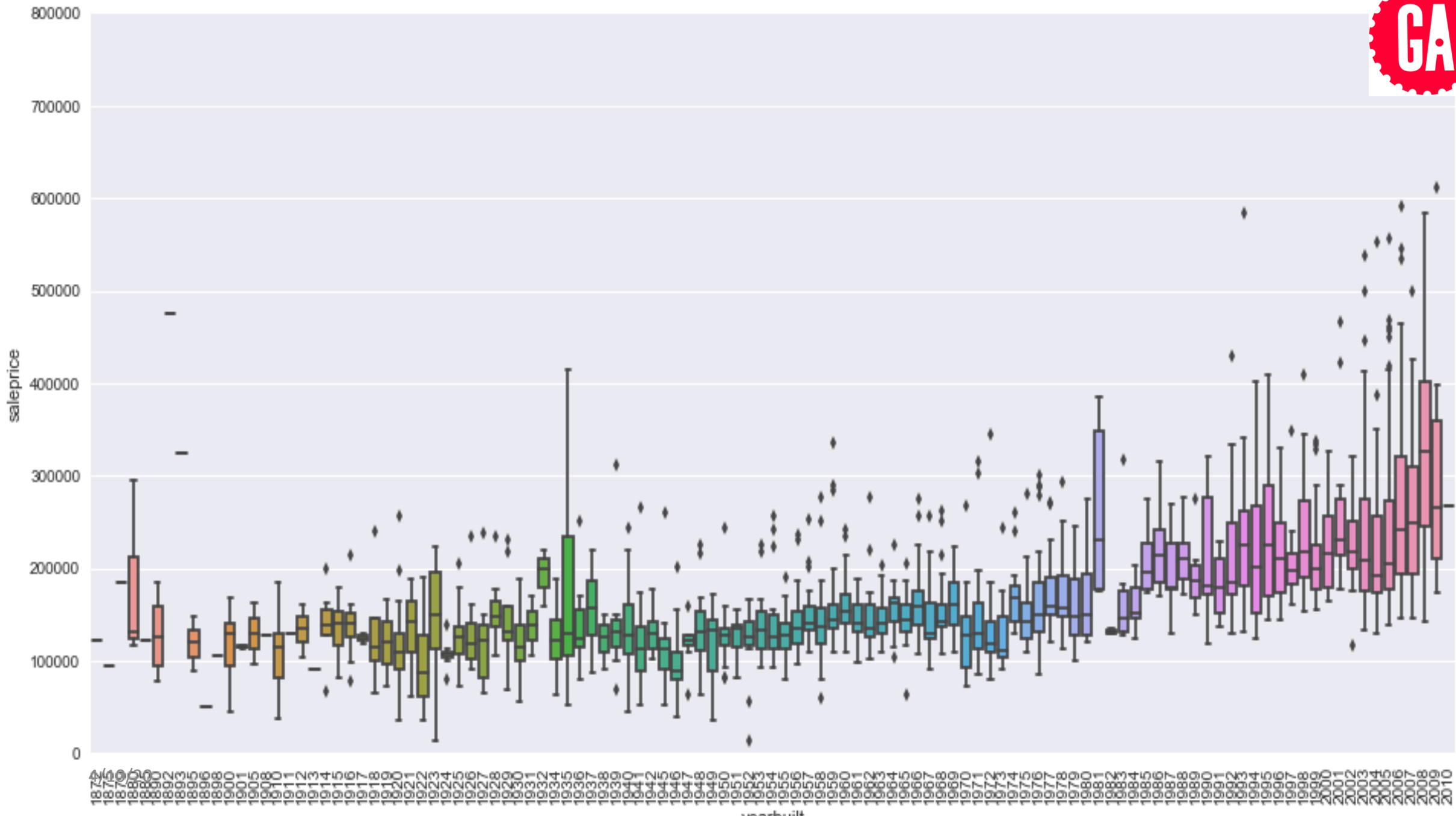
# Outliers: Relation Exploration for Few Numerical Variables



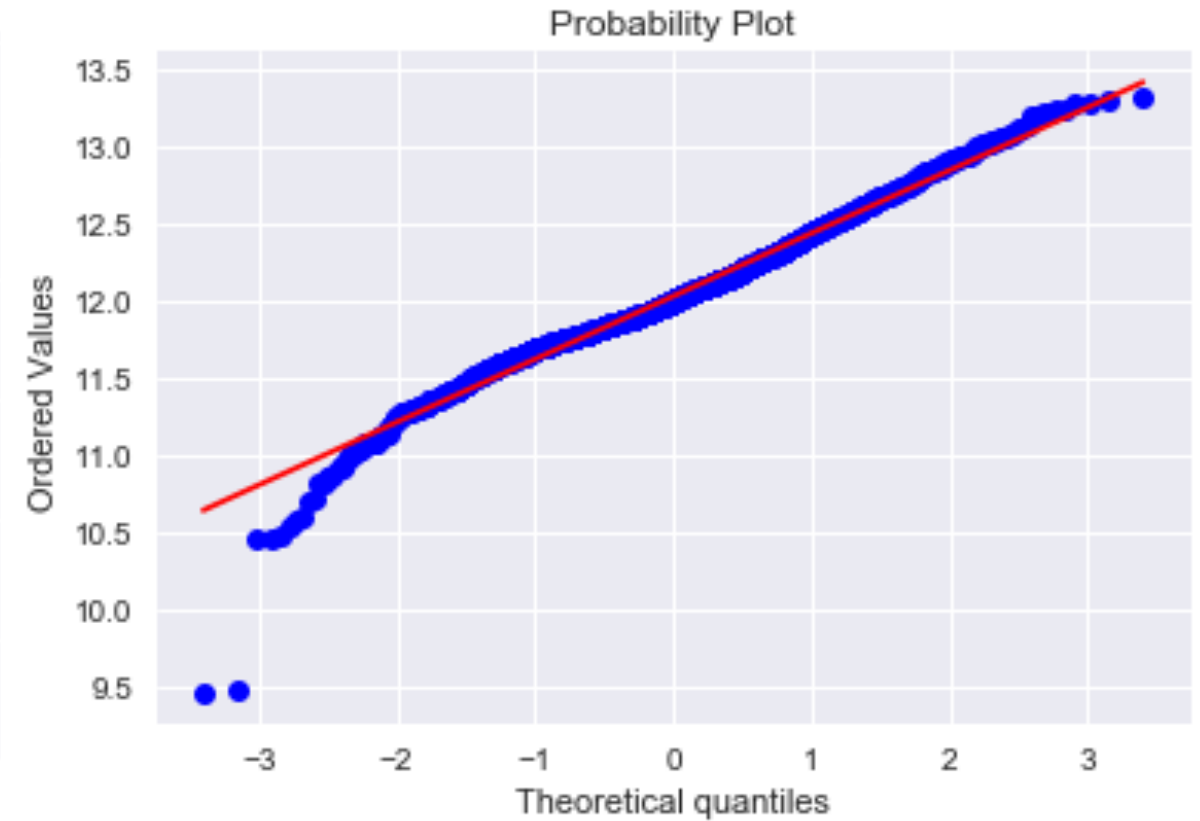
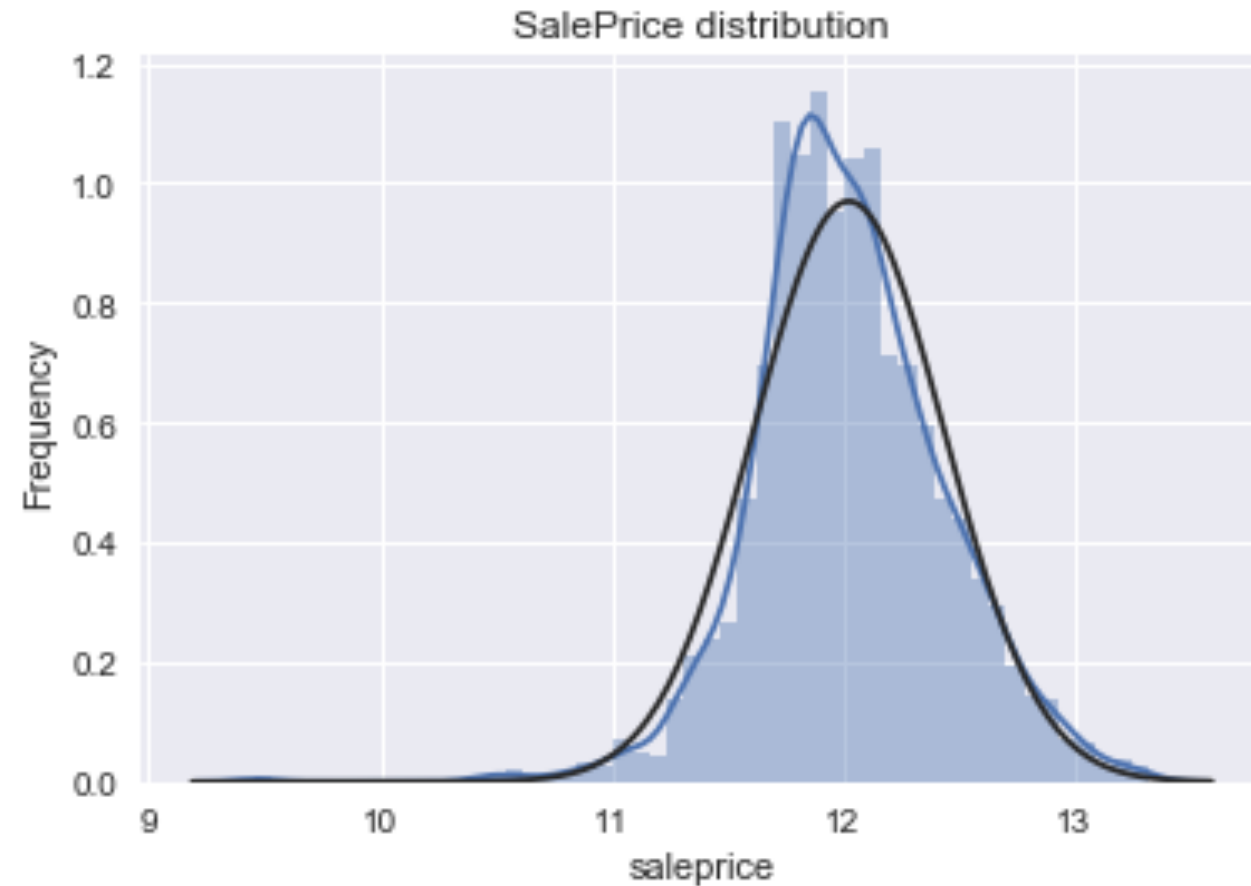


# Outliers: Relation Exploration for categorical features

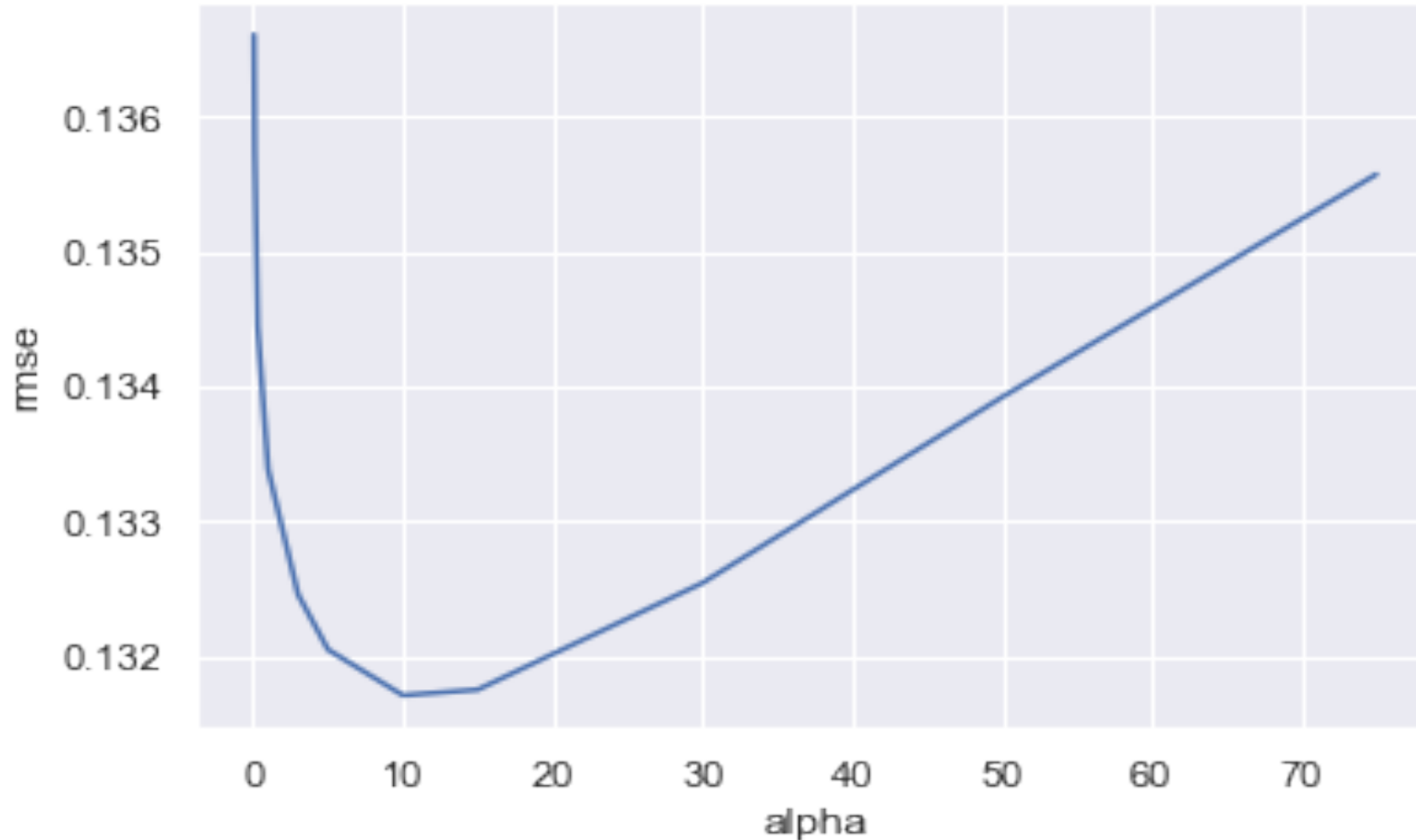




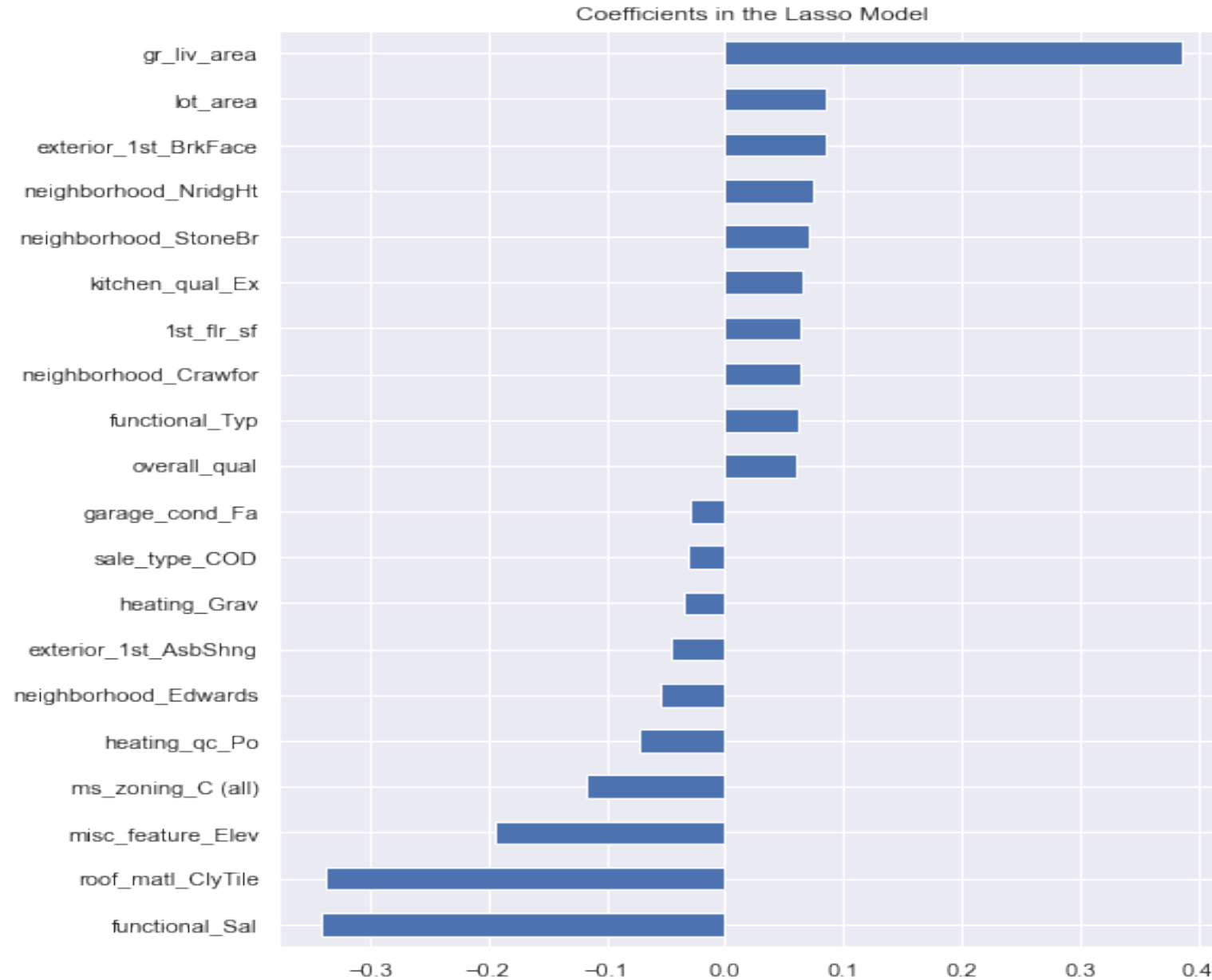
# Data Preprocessing: Log transform Distribution vs Frequency



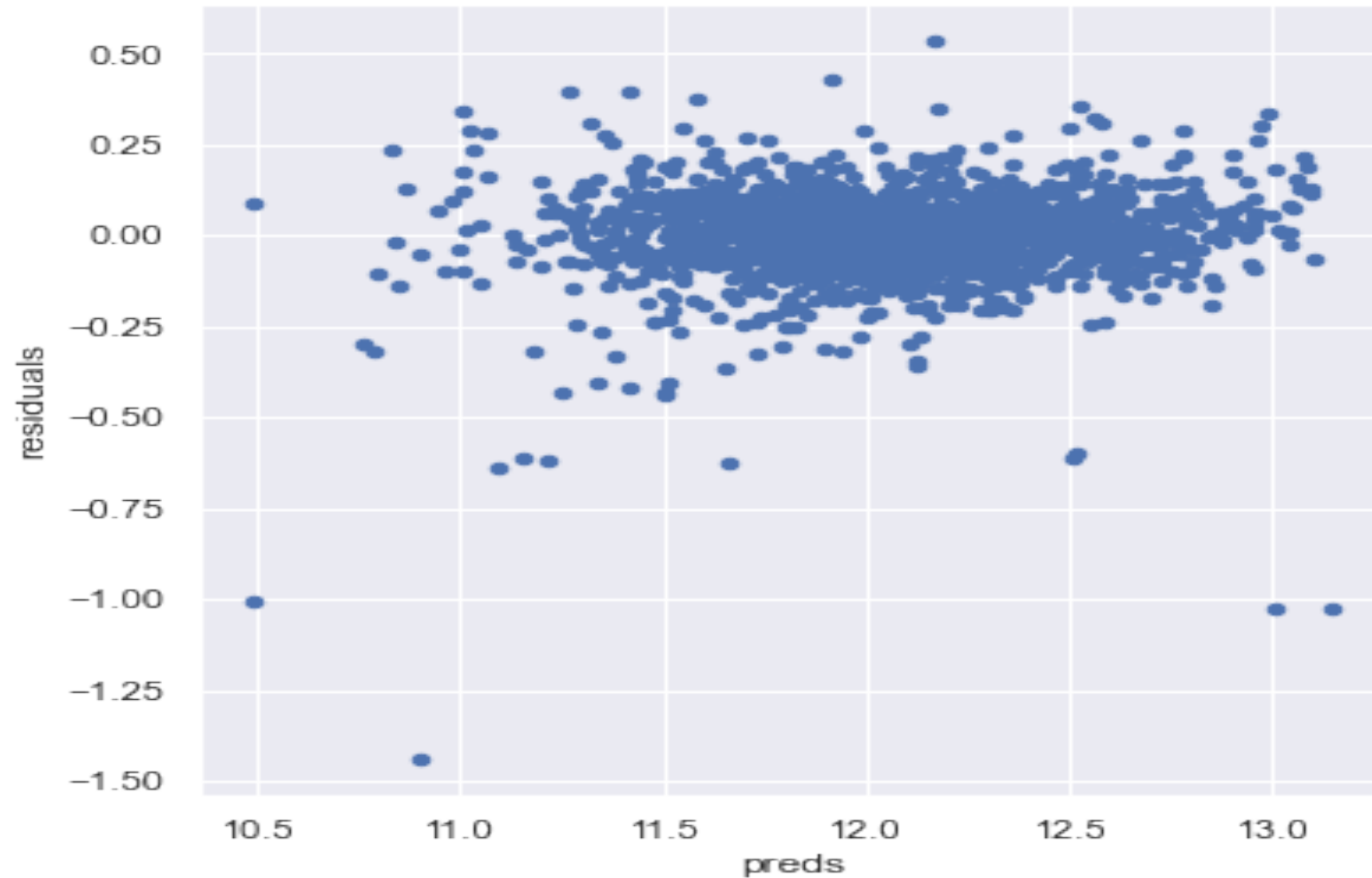
# Implementing Ridge Regression



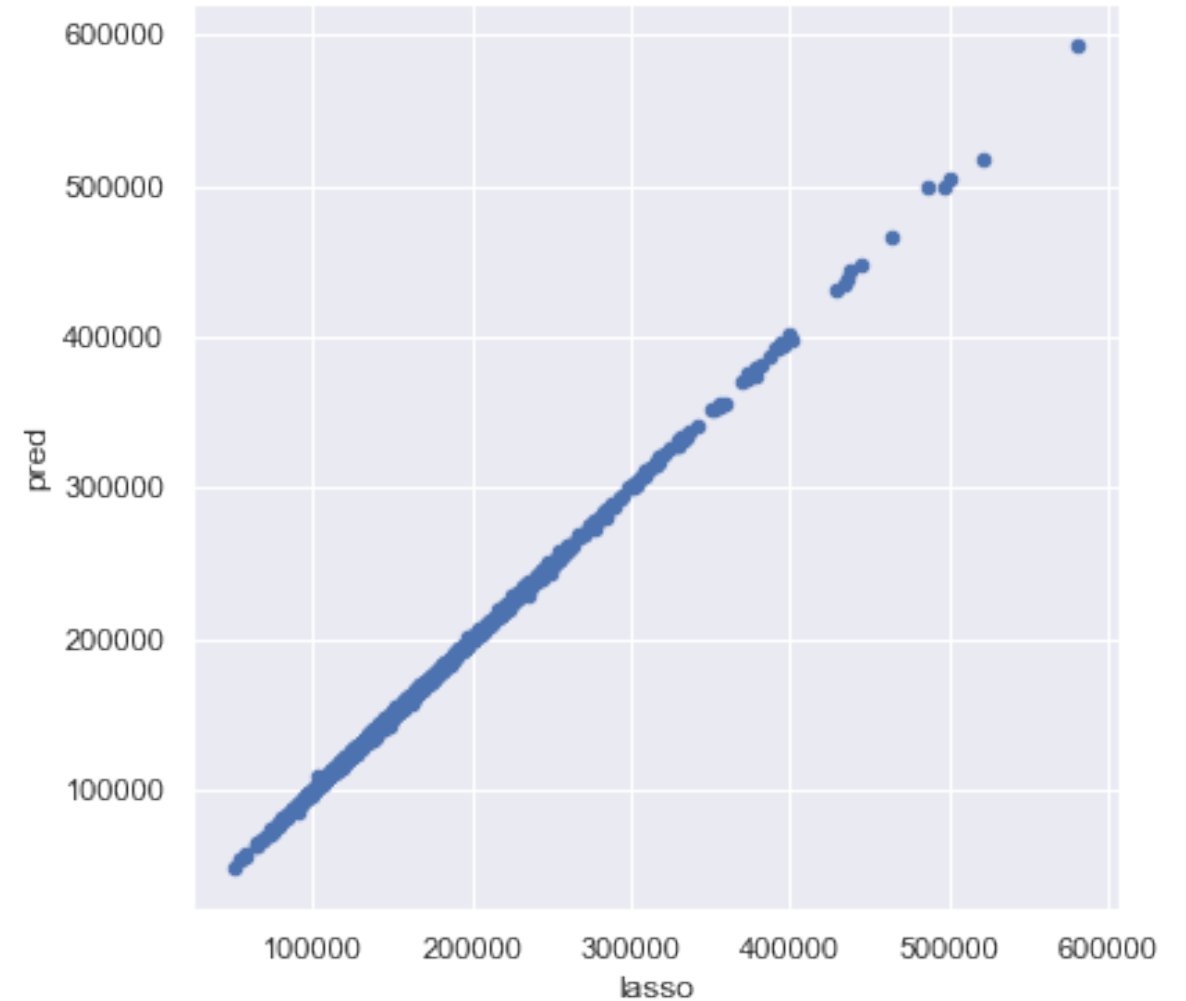
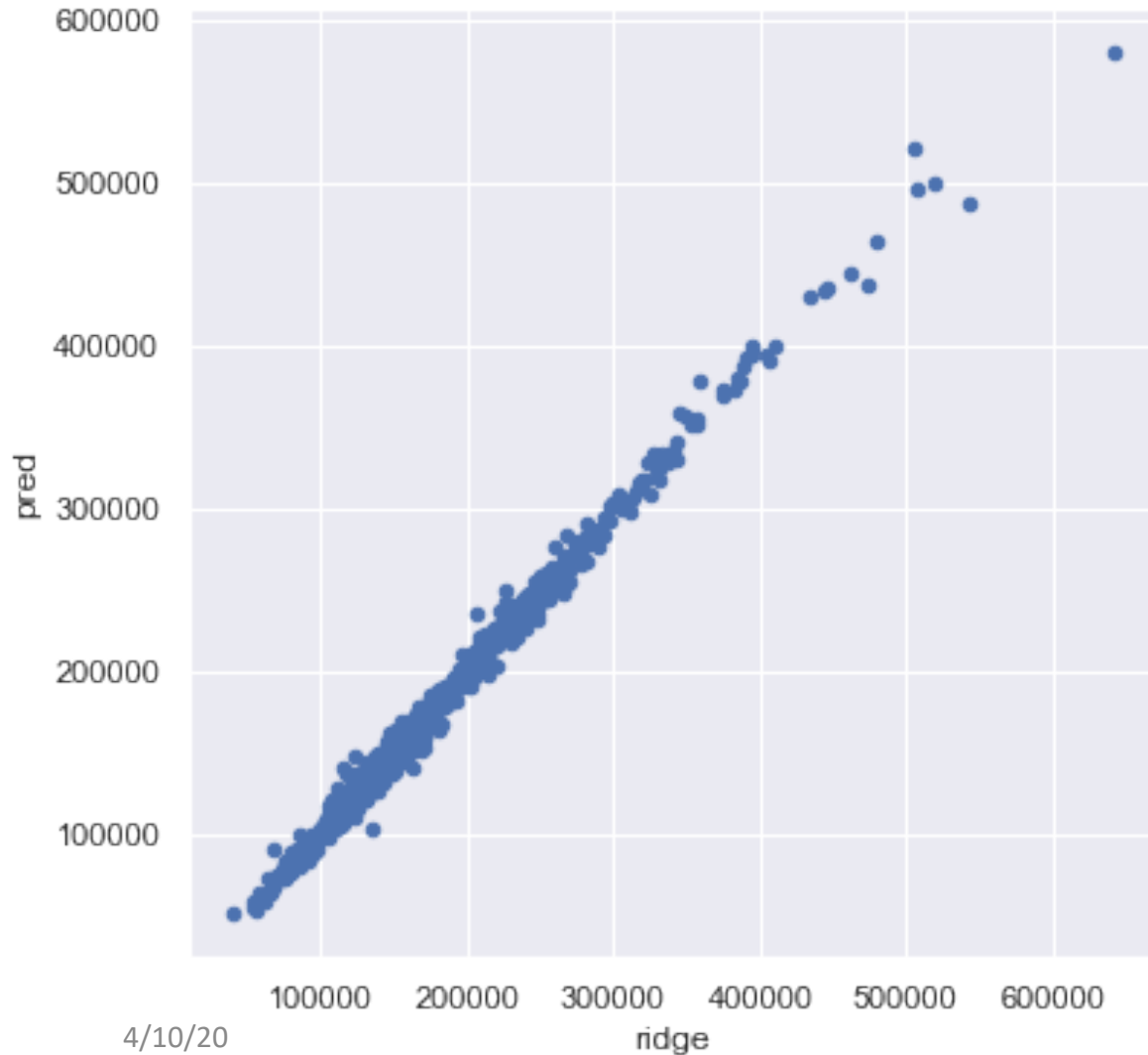
# Implementing Lasso Regression



# Residual vs Prediction



# Base Model Result: ridge & lasso



# Conclusions and Recommendation



- The most valuable feature in this case is GrLivArea
- To solve the regression problem, Linear models tend to outperform tree-based in terms of speed and score.
- Lasso helped to the feature selection because it shrinks a relatively unimportant coefficient to zero.



# Recommendation



- data involving local policy changes and economic trends in the housing market specific to Ames, Iowa.
- adding even more data such as school zoning or transportation and commercial information would produce models with more predictive power.
- time-series analysis to predict when is the best time to buy a house.
- Nation wide data may be important



**THANK YOU !**

**TIME TO QUESTIONS AND COMMENTS**