Homelessness Project

DATA 3320
Hope Crosier

The Problem

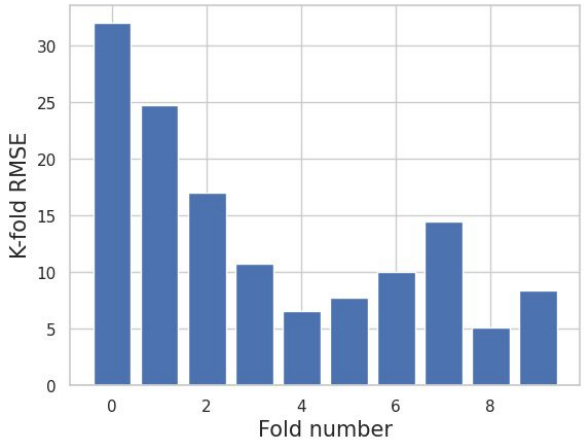
- We wanted to look into homelessness in the United States
- Homelessness is on the rise
  - We want to look at homelessness from the lens of housing market factors

The Data

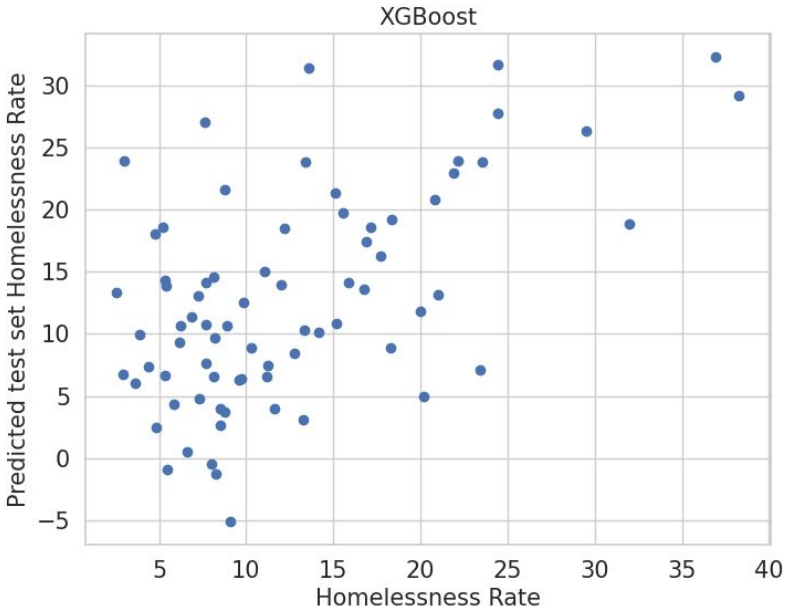
- The U.S. Department of Housing and Urban Development (HUD) has a report from 2019
  - Contains statistics for homelessness rates along with housing market factors

The Analysis

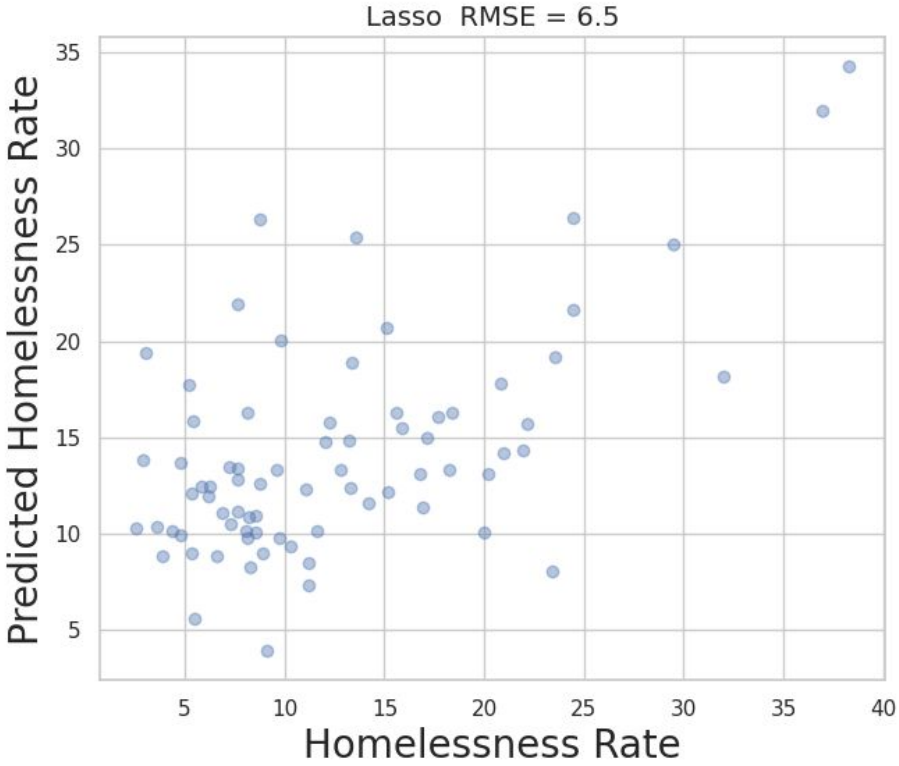
- We took a linear regression approach to analyzing this data
  - Lasso, ridge regression and XGBoost
- Wanted to find which ways of modeling and predicting the data gave us the best results
- Used cross validation (K-fold) along the way

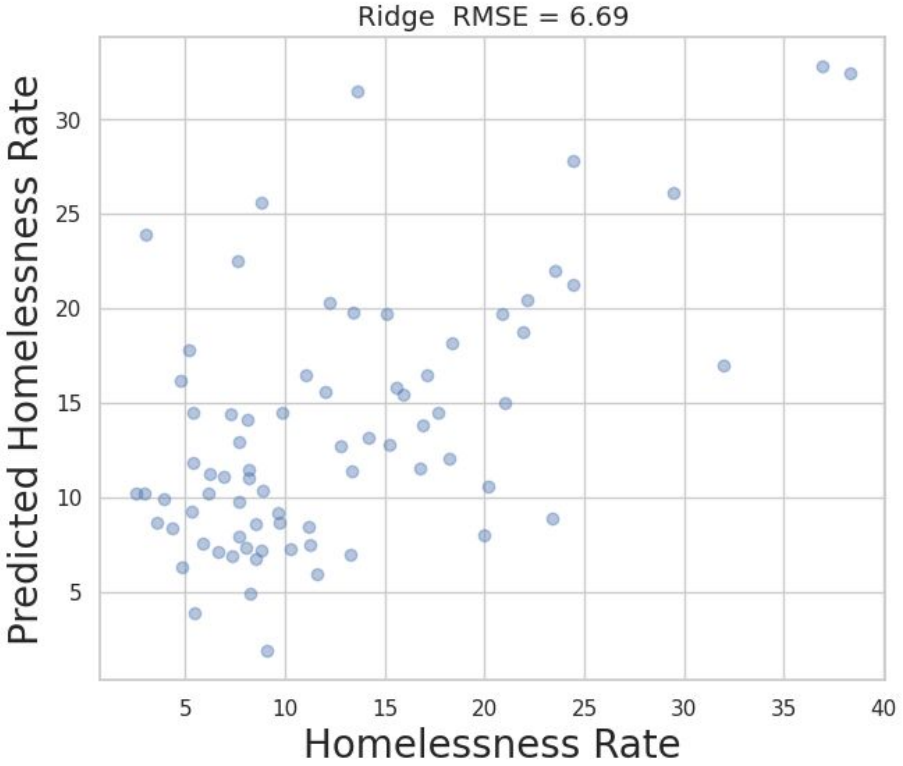


The Results

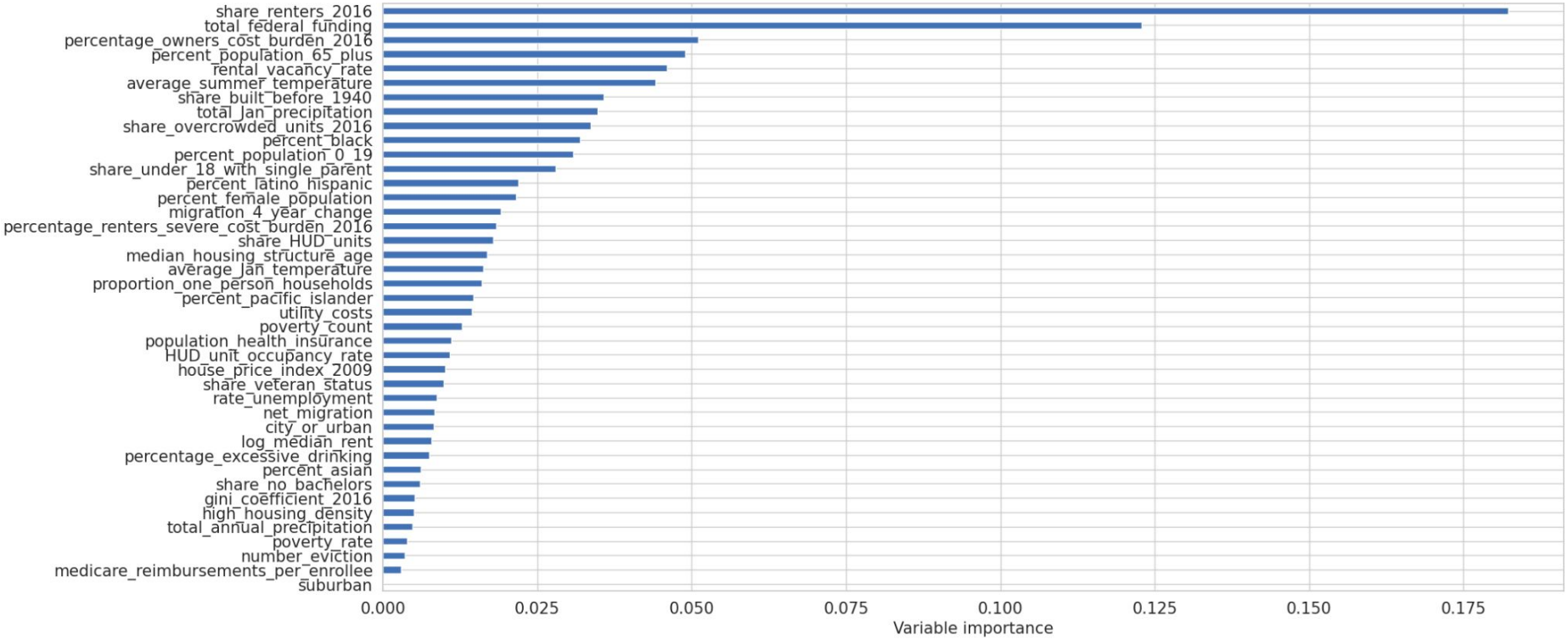


- Overall the graphical relationship between the actual data and the predicted data was somewhat linear
  - RMSE: 7.58
  - MAE: 5.9





Predictions using Lasso and Ridge Regression Techniques

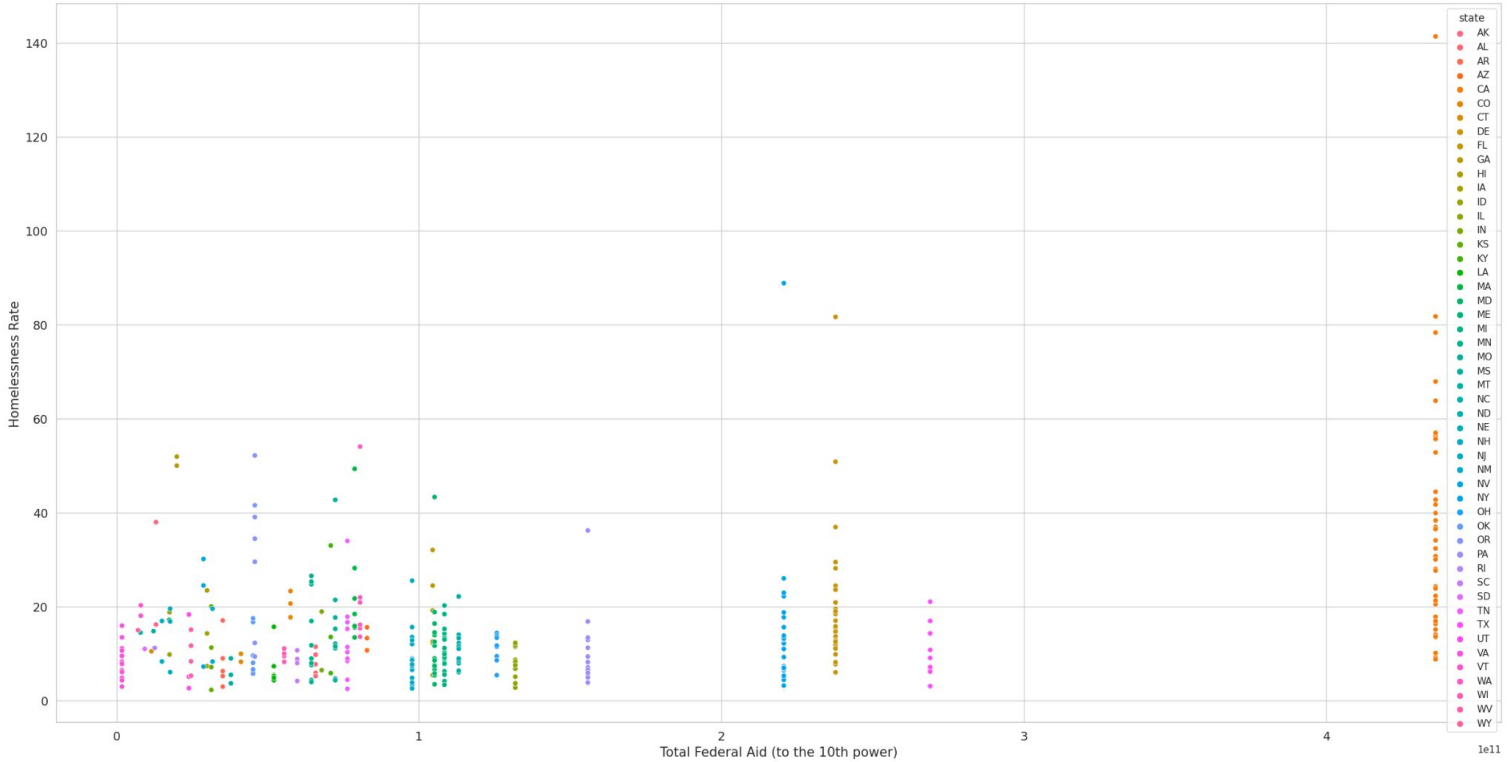


Importance of Predictors by Variable

Additional Step

- As an additional step, we looked into how much federal aid each state received from the government
- Each state might have more than one CoC, so we tried to look at it from a state based approach
- Found it to be one of the predictors with highest significance in our model

Additional Step



The Conclusion(s)

- Amount of federal aid received by state does seem to be a significant predictor in predicting homelessness rate
- We can predict the homelessness rate within around 6-7% accuracy of a CoC
- Share of renters, average summer temperature, number of vacant rentals, age of housing structure are some predictors that seem to be significant