



HOMELESSNESS IN THE U.S.

LEIANA BEATRIZ OMINE

SEATTLE UNIVERSITY

DATA 3320 - SP'23

INTRODUCTION

- The Domain Problem: Identifying the relationship between local housing market factors and homelessness in the U.S.
- Additional Step Proposal: Investigating the differences between suburban areas and rural areas and their relationship with the increase of homelessness rates.



DATA USED



The U.S. Department of Housing and Urban Development (HUD) produced a report in 2019 [Market Predictors of Homelessness](#) that describes a model-based approach to understanding of the relationship between local housing market factors and homelessness.



The data consists of various demographic, housing, and economic, climate, and safety net factors.

QUESTIONS



What are some of the biggest factors on homelessness?



Does homelessness rates change based on location?

ANALYSIS METHODS

Train Test Split

Multiple Linear Regression

Lasso and Ridge Models

XGBoost

K-fold Cross Validation

MODEL PERFORMANCE

Lasso

MSE =
76.875

Ridge

MSE =
74.698

XGBoost

MSE =
8.831

MSE: Mean Squared Error

VARIABLES BY IMPORTANCE

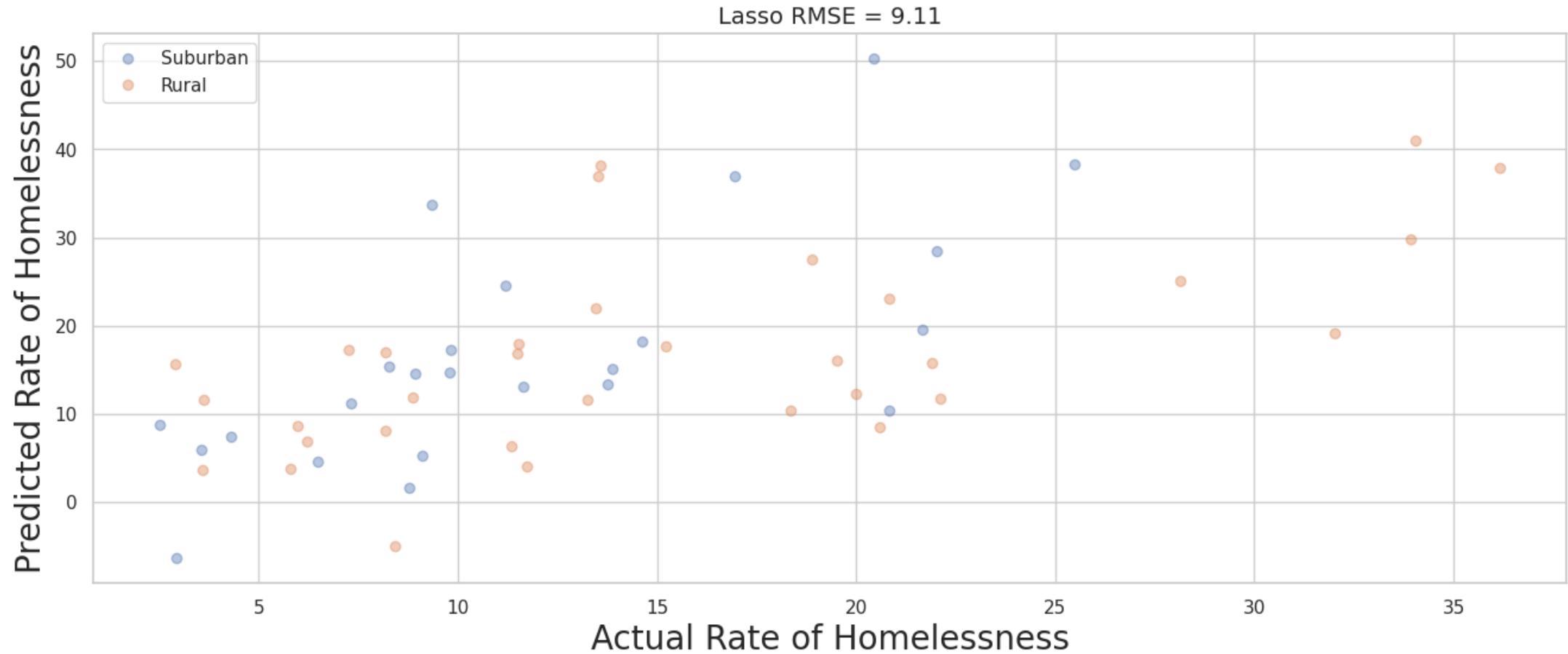
Lasso:

1. proportion_one_person_households
2. log_median_rent
3. share_overcrowded_units_2016
4. share_HUD_units
5. migration_4_year_change

Ridge:

1. proportion_one_person_households
2. log_median_rent
3. share_HUD_units
4. share_overcrowded_units_2016
5. average_Jan_temperature

XGBoost shared similar variable importance.



SUBURBAN VS RURAL HOMELESSNESS

CONCLUSION

- Lasso, Ridge, and XGBoost models had relatively similar performance. Through cross validation, the RMSE across each model were all similar at approximately 8.6.
- Each model had similar variable importance as well. All three models shared variables like `proportion_one_person_households`, `total_jan_precipitation`, and `share_overcrowded_units_2016`.
- Difficult to accurately investigate the relationship between the type of area (suburban vs rural) and rate of homelessness.