



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



VARIABLES BY IMPORTANCE

- proportion_one_person_households
- log_median_rent
- total_jan_precipitation





MODEL PERFORMANCE

Lasso

MSE =
76.875

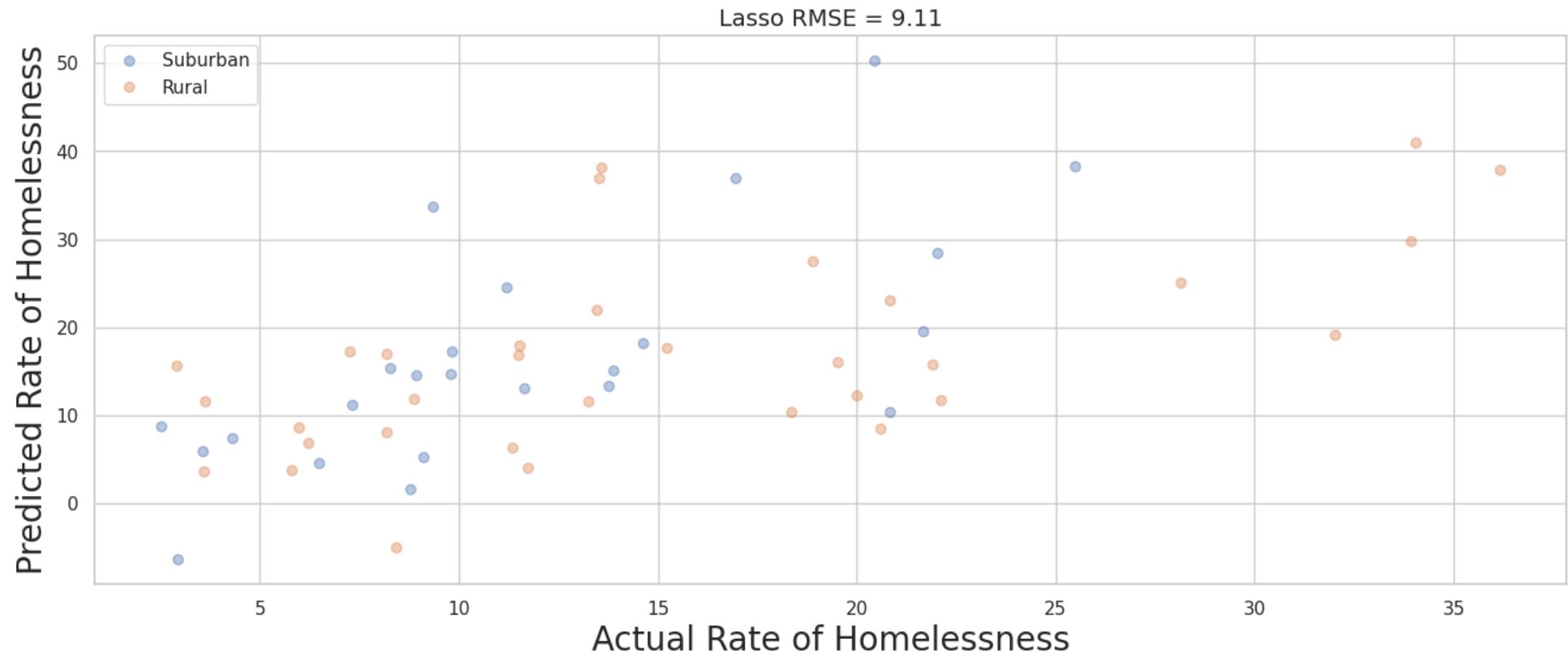
Ridge

MSE =
74.698

XGBoost

MSE =
8.831

MSE: Mean Squared Error





CONCLUSION

Out of the three models, XGBoost had the best performance with a mean squared error of 8.831.

All three models shared variables:

`proportion_one_person_households`, `total_jan_precipitation`,
and `log_median_rent`.

It was difficult to accurately investigate the relationship between the type of area (suburban vs rural) and rate of homelessness due to negative r-squared value.