

Regression Module Final Submission – Air BnB Ratings

Elisabeth Johnson

May 17, 2022

Abstract:

The aim of this project is to understand the features that may influence Air BnB Ratings and build a robust predictive model used to uncover a predicted rating from several predictors that may influence rating. For this project I worked with several data sources including Wikipedia, Kaggle and (or course) the Air BnB Website! After scraping data using Beautiful Soup and Selenium, I have constructed several visuals illustrating 1) the predictors most influential in impacting rating and 2) the model most fit for predicting rating given several predictor values and more.

Design and Data:

The Design of this project is for the Metis Bootcamp Regression Module and the Data consists of 16 features (some that were dropped, totaling to 12) and over 1,000 rows of data. The main model used to predict response rate was a Ridge Regression model.

Algorithms:

- 1) Scraping the Air BnB Website using Selenium and Beautiful Soup
- 2) Converting this data into a csv (for future storage) and a data frame after data cleaning
- 3) Creating Boolean values for predictors
- 4) Creating several models for which to understand which best fits our use case (Ridge Regression)

Models:

Linear Regression, Lasso Regression, Ridge Regression, Single and Multiple K-Fold approach in order to understand robustness and error.

Model Evaluation and Selection:

For this, I took into account several variables

- 1) Predictor p-value
- 2) Mean Absolute Error (MAE)
- 3) R^2 results

Tools:

- 1) Python- Numpy, Pandas, Scikit-learn, Beautiful Soup, Selenium
- 2) Matplotlib and Seaborn for visuals
- 3) Excel