

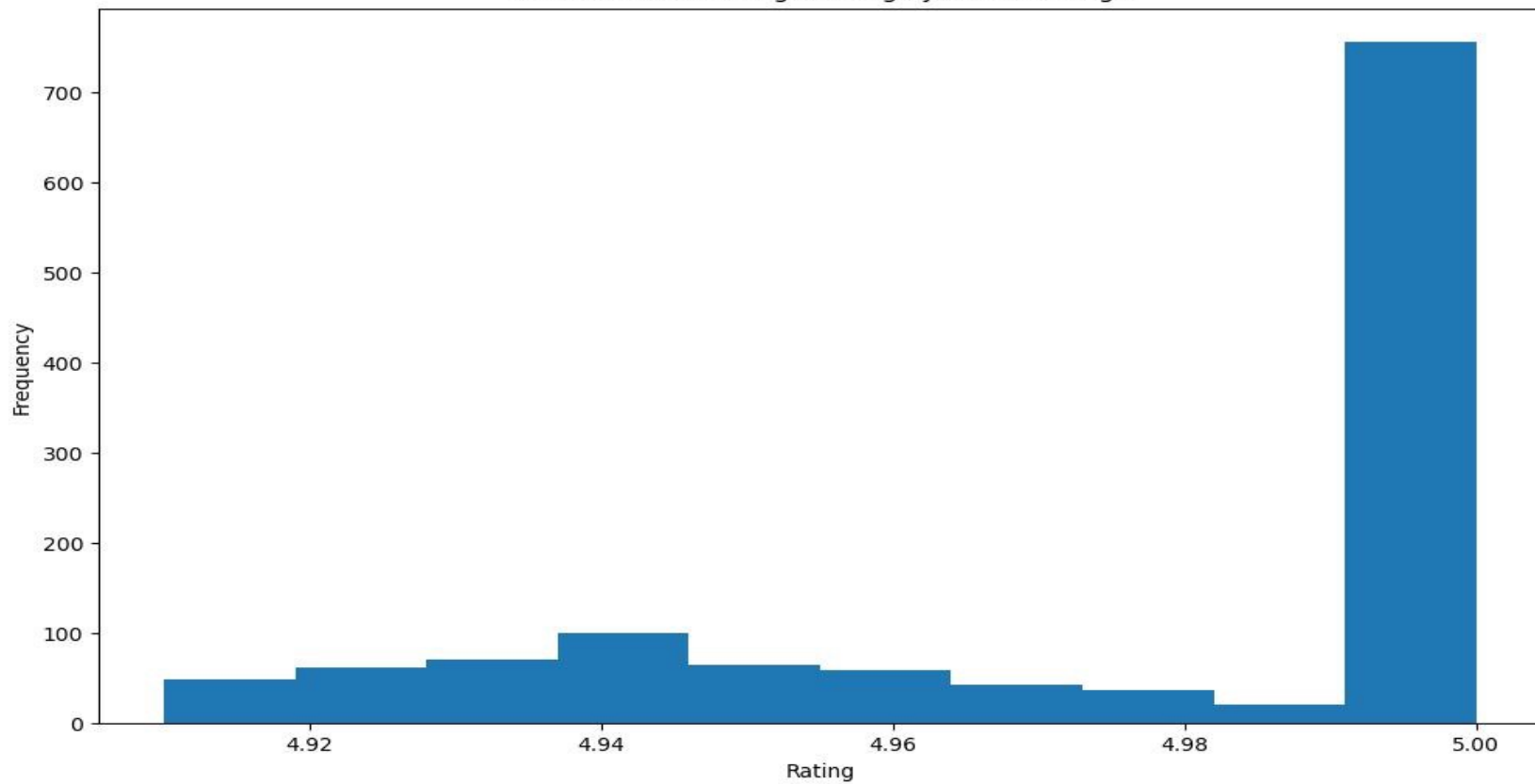
Predicting Airbnb Rating Type

Mya Carrizosa
General Assembly
Data Science Immersive

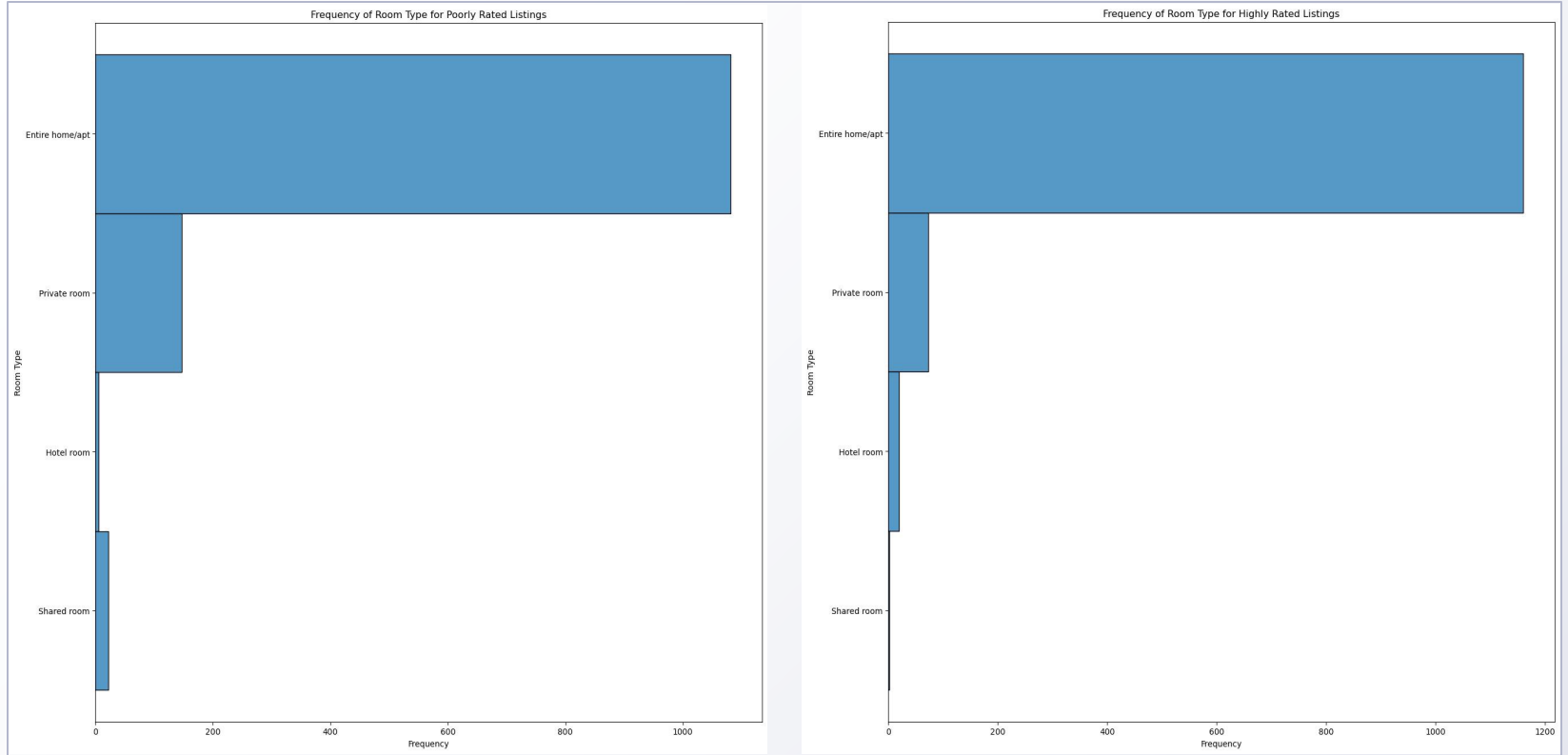
Problem Statement

To inform the Airbnb quality assurance division and provide feedback to hosts, we want to build classification models to predict—based off either property features or prior reviews—whether a property is poorly or highly rated.

Distribution of Ratings for Highly Rated Listings

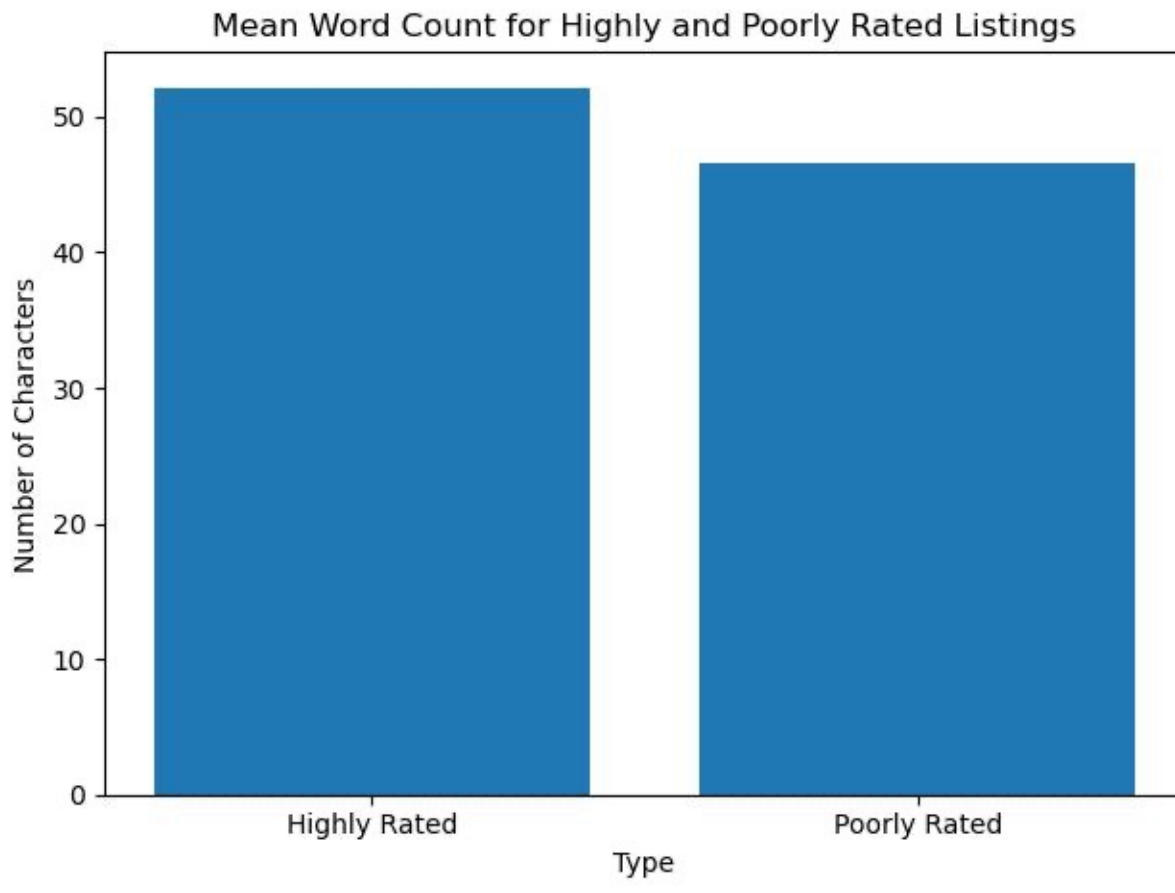


EDA



Data Review

- ▶ Typical
- ▶ Language
- ▶ Creation below
- ▶ Separation most

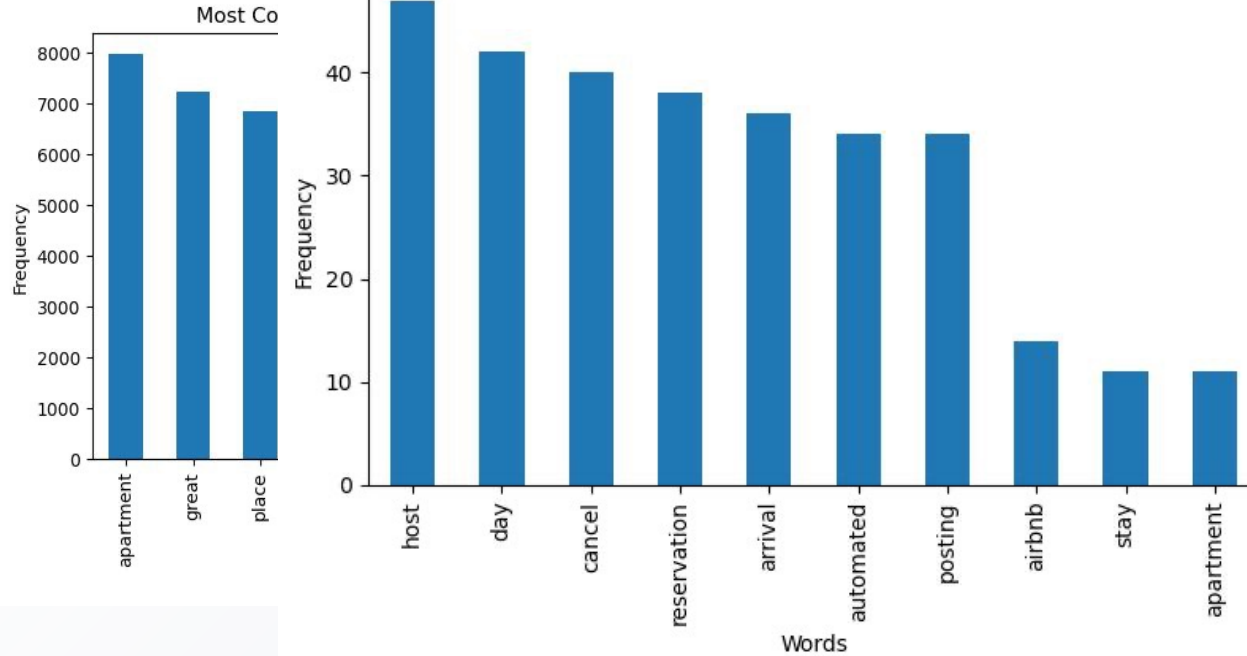


ratings

count,

EDA

Most Common Words in Very Poorly Rated Listings(< 2 Stars)



Models Selected

- ▶ Logistic Regression
- ▶ KNN
- ▶ Decision Tree
- ▶ Random Forest
- ▶ AdaBoost
- ▶ Gradient Boost
- ▶ Bagging Classifier
- ▶ Neural Networks (for reviews data only)

**Baseline
Accuracy:
0.50**

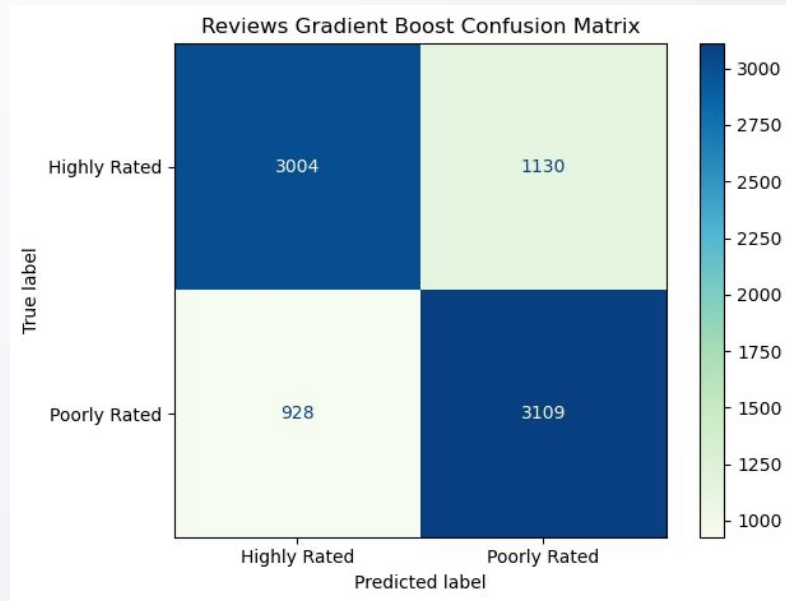
Best Listing Model

- ▶ Decision Tree
- ▶ 0.76 train accuracy
- ▶ 0.72 test accuracy
- ▶ Sensitivity: 0.67
- ▶ Specificity: 0.78
- ▶ Precision: 0.76



Best Reviews Models

- ▶ Gradient Boost
- ▶ 0.80 train accuracy
- ▶ 0.75 test accuracy
- ▶ Sensitivity: 0.77
- ▶ Specificity: 0.73
- ▶ Precision: 0.73



Conclusions

- ▶ Models built on reviews data are slightly more predictive
- ▶ Neural network models performed the worst
- ▶ We can predict with above baseline accuracy whether an Airbnb listing has a rating over 4.9 or under 4.5

Future Directions

- ▶ Check LINE assumptions for inference
 - ▶ BIG unanswered question...



Citations

- ▶ Inside Airbnb
- ▶ General Assembly Data Science Immersive Bootcamp Lessons 5.03, 5.04, 6.05
- ▶ Stack Overflow
 - <https://stackoverflow.com/questions/32464280/converting-currency-with-to-numbers-in-python-pandas>