

Airbnb Pricing Model

Supervised Learning

Capstone

Jacob Phillips
12/20/2020



Introduction

Is my Airbnb listing priced accurately??

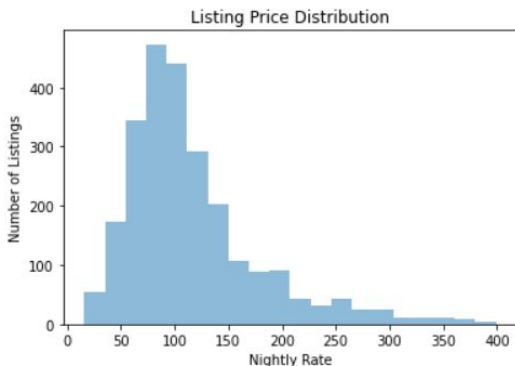
- Airbnb has listings in over 100,000 cities worldwide
- There are 660,000 listings in the United States alone
- Most Airbnb hosts are normal people who might not know how to price their listing
- Simply matching local hotel room prices usually won't work...so what does?



Data

- Source of the data: [Inside Airbnb](#)
 - An independent project that aggregates and analyzes publicly available Airbnb data
- Only used Seattle data for this model
- Utilized current listings to train a model that predicts the nightly price of a listing based on a number of features

Target Variable (Nightly Rate)



Features:

- Number of beds
- Number of bathrooms
- Total number of people that can stay
- Room type
- Host response time
- Neighborhood
- Superhost status
- Instant Book status
- Number of reviews
- Average reviews per month
- Review rating

Model Preparation

What data is left out of the model?

- Luxury listings (anything over \$400/night) \$\$\$
- Abandoned listings (no reviews from the past year)
- Hotel listings (some hotels list on Airbnb as well)



So what data is left?

- Still ~2500 rows of data remain
- Transformed some data from strings to floats
- Used one-hot encoding on categorical variables

Model Results

Model type: Gradient Boosting

Accuracy after tuning: 0.7

Most important features:

- Number of people that can stay
- Number of bathrooms
- Review rating

Other models types tried:

- Linear Regression (OLS)
- K-Nearest Neighbors
- Random Forest

Testing with Real Data



To the real world!



Our “real world” listing is a 1 bed/1 bath apartment in the Queen Anne neighborhood

Case 1: New listing (few reviews)

\$105 per night

Case 2: Superhost listing

\$135 per night

Next Steps

How can the model be improved?

- Add weights for bathroom types
- Take data from other cities
- Use text analysis to identify key phrases in the listing descriptions and reviews
- Adjust for missing data instead of eliminating it

Questions

