Characterize Listings and Predict Price

Shreya Gupta, Jaya Ren

airbnb

Problem Statement

Cluster Airbnb Listings

What are the different kinds of properties?
What characterizes them?

Predict Airbnb Listing Price

Can we predict the price of a listing?

Data Overview

Property

property name, neighborhood, coordinates, license

Listing

room type, availability, price, review frequency

Host

host name, host id, number of listings

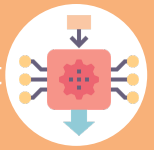
Why Do We Care?

Price Prediction

Is a tough challenge, especially for similar listings with minute differences

Traditional Model Limitation

Price prediction is a widely performed activity, but most applications still rely on traditional models (mostly linear regression), which has limited prediction power



Empower Hosts

Empower hosts to understand the key characteristics that make 2 properties similar and provide them a guidance on price setting

Methods

01

Stable Regression

- Train-validation split

02

Holistic Regression

- L1 regularization
- Sparsity
- Minimizing regularized squared residual errors
- Training R^2 17.5%, testing R^2 17.98%

03

XGBoost

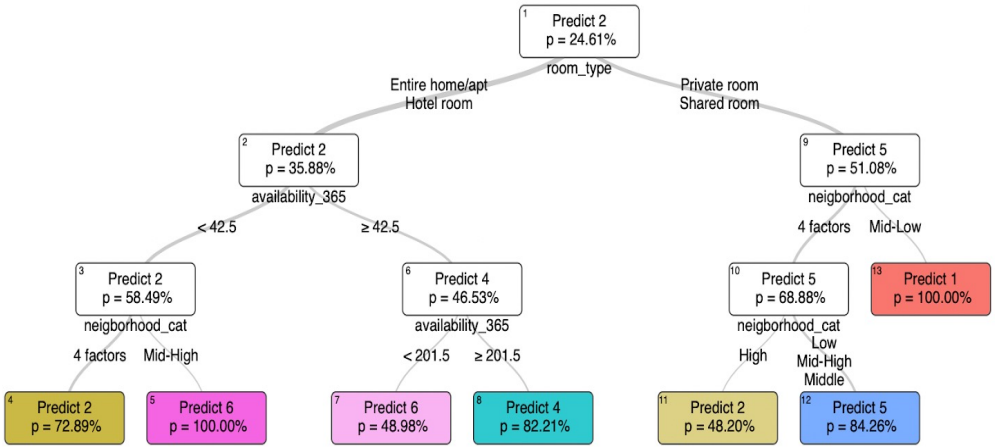
- Overall XGBoost
- Training R^2 52.8%, testing R^2 38.3%

04

Cluster + OCT

- K-means cluster k = 6
- XGBoost on each cluster
- OCT to characterize each cluster
- Avg cluster training R^2 42.17% avg cluster testing R^2 27.89%

Key Insights



- **Cluster 1:** Private/shared rooms in moderate neighborhood
- **Cluster 2:** Private or shared listings in most expensive neighborhoods; most reserved hotels/apartments in all neighborhoods other than mid-high
- **Cluster 3:** All other listings
- **Cluster 4:** Entire homes or hotel rooms that are mostly available
- **Cluster 5:** Private/shared rooms in moderately expensive or in low-cost neighborhoods (complement of cluster 1)
- **Cluster 6:** Hotel rooms/homes; most bookings in mid-high neighborhoods, mostly available in other neighborhoods