

CS6140 Project Proposal Presentation

Abhay Kasturia, Nakul Cammasamudram, Philip Parker

Overview

Airbnb is a popular online company through which property owners can short-term rent their space to consumers as an alternative to hotels. Using data on Airbnb property listings in Boston, we will develop and test methods for determining the optimal price per night an owner should set for their property.

Dataset: data source

The Inside Airbnb project by Murray Cox has collected public Airbnb listing data for 40+ popular international cities. We will use the Boston dataset containing ~5000 property listings and their features for the Boston area.

(<http://insideairbnb.com/get-the-data.html>)

Dataset: data features

- ▶ *super_host?* (categorical) [yes/no]
- ▶ *verified_host?* (categorical) [yes/no]
- ▶ *zip_code* (categorical) [...]
- ▶ *property_type* (categorical) [House, Apartment, etc.]
- ▶ *room_type* (categorical) [Shared Room, Private Room, etc.]
- ▶ *accomodates* (continuous)
- ▶ *bathrooms* (continuous)
- ▶ *bedrooms* (continuous)
- ▶ *beds* (continuous)
- ▶ *bed_type* (categorical) [Real Bed, Futon, etc.]
- ▶ *minimum_nights* (continuous)
- ▶ *cancelation_policy* (ordered categorical) [Flexible, Moderate, Strict]
- ▶ **price** (continuous)

Dataset: initial data analysis

Methods

- ▶ We will perform and compare a variety of regression techniques, including linear regression and kernel methods. The comparison between the results of these methods is straightforward.
- ▶ We will also consider approaching the problem from a classification point of view, dividing the prices into ordered categorical ranges. Doing this will allow us to investigate the use of classification methods such as logistic regression and tree-based approaches.
- ▶ We will research/develop a means of comparing the results of the regression and classification methods.