

Home.Pl:

A Friendly Pricing Intelligence for Airbnb Hosts

Yiming Xu

yimingxu@mit.edu

Rong Liu

rongli@mit.edu

Shu Xu

shuxu@mit.edu

Abstract

Airbnb has allowed hosts to offer their houses and apartments for short term lodging, but new hosts may be lost regarding what price they should set for their listing. Thus, we want to build a product that utilizes past data to help Airbnb hosts understand the market, their competitors, and appropriate pricing options. More specifically, we want to use interactive visualizations to show users the distribution of listings around their own real estates, and the associated ratings, pricing, and booking status. We would also like to extract keywords from reviews for neighborhood real estates, in order to give recommendations on what Airbnb hosts should look into for positive ratings and reviews. At the end, we provide a pricing tool, which allows users to input features on their own houses/apartments and proposes pricing strategies. For this project, we limit our scope to Cambridge and Boston.

Introduction

We consider our project an extremely useful tool for new Airbnb hosts to price and improve their properties.

As an idea with huge business potential, there are lots of relevant existing work. Among all relevant products, the most official and widely-used pricing one is Airbnb's own *Smart Pricing Tool*, where you can enter some basic facts about your properties and get a suggested price according to the similar properties already in Airbnb's database. However, it has several drawbacks. First, it uses similar properties only as a reference to provide suggested price but ignores the fact that they could also be competitors. Second, it provides limited insight for new hosts to understand the suggested price and how they can make improvements on the properties to get a better deal.

Different from Airbnb's own Smart Pricing Tool, our tool wants to provide a more user-friendly pricing tool which not only provides the suggested price, but also with both static and interactive visualizations to help users understand the price we suggested and how they could improve.

According to our hypothesis, we think the suggested price is not only related to new host's own properties status (location, capacity etc), but also the status of competitors (total number, review), as well as some seasonal-trends. We will mainly focus on those aspects to build our pricing model.

Methodology

Our system is divided into two main parts, a pricing model and a front end webpage. Our front end handles user input and displays data visualization and model outputs, with RESTful

API as an interface between service and front end. We will create an interactive visualization with javascript and d3 that allows users to zoom, select and enter.

For the core pricing model, we will build and compare multiple regression methods including linear regression, logistic regression, random forest and k-nearest neighbors. KNN might be an intuitively good fit given the geographical information in the dataset, but we will also perform feature engineering to better suit other models. Our best model will be chosen by cross validation. To extract keywords from comments and reviews, we will rank phrases using TF-IDF scores. From there we can try and build a topic model if time permits.

We encapsulate our models into a Python service using Flask. The service handles API request from the front end and sends model output back.

Evaluation

Our project consists of two major parts. First, we can evaluate the predictive power of the underlying pricing model using a hold-out test set. Second, our deliverable, the interface that users see, is an HTML webpage. It should allow users to easily interact with the visualizations, and it should convey clear and insightful messages.

Data

[Inside Airbnb](#) offers public Airbnb data in various cities (including Boston and Cambridge). We utilize listing, calendar, and review datasets to extract necessary features, such as pricing, dates, geographic location, number of rooms and beds, etc.

Task-List and Timeline

Tasks	Person	Deadline
1. Data Integration and Cleaning	Rong	Midterm Report(11/04)
2. Static Visualization	Shu, Yiming	
3. Baseline Model	Rong	
1. Interactive Visualization	Shu, Yiming	Final Report(12/04)
2. Optimized Model	Rong, Yiming	
3. HTML Deliverable	Shu	
4. Poster and Report	All	
1. Presentation	All	Presentation(12/11)

Deliverable

- Final report: a web-page with interactive visualizations and pricing tool.
- Code: data cleaning and analysis modules, visualization scripts, and modeling scripts.