

Capstone #1 Proposal
Predicting AirBnB price and availability in Paris
Kseniya Kruchok
Brooklyn NY

Problem

Millions of people are travelling to Paris every year. They all need a good place to stay. Over the past years traditional hotels are sharing their guests with private hosts who are using AirBnB to list their properties. Often neither hosts nor guests know what is a good fair price for a certain apartment or room. The idea is to predict a market price for a listing based on property, location and host characteristics. Also it is not always easy to tell when and which type of property is available. Predicting availability at a certain time and at a certain budget can be a useful tool as well.

Client

The results of this project can be used by AirBnB itself to introduce a new feature or tool that can show how certain listing's price measures in comparison to a market one, to introduce "suggested price" tool when the host is adding a property to help the owner to price the listing correctly.

Travellers can also benefit from the findings of this project. They will be able to see if offered property is priced fairly. They will be able to see types of properties that they can expect to be available during certain time of the year.

Data

I will be using data from Inside AirBnB website. <http://insideairbnb.com/get-the-data.html> That has several data sets (Listings, Calendar and Reviews).

The main data set will be Listings one, that contains information on over 40,000 properties.

Solution

1. First, I will clean the data from irrelevant information and handle all missing values (either filling them or dropping).
2. I will perform EDA on available data sets to try to see what dependencies and correlations exist. Also it will help to understand data better. All analysis will be supported by visualizations.
3. I will combine availability information from Calendar data set and review info (possibly just indicating number of positive and negative listings) with the main

data set Listings. I will also introduce new columns indicating proximity of the listing to main attractions and sights.

4. I will perform machine learning to train the model to predict the price. Also will try to do clustering of the available listings to see how they cover the city.

Deliverable

Ideally, my deliverable will be a tool that allows to plug in property details and it will suggest it's price. I will provide the code used for analysis and machine learning along with corresponding graphs and visualizations.