Final Project Proposal: San Diego Airbnb Analysis

Zequn Che & Jiyu Feng

Data Source: <https://data.world/ajsanne/san-diego-airbnb>

* Background Introduction

Airbnb is the abbreviation of “AirBed and Breakfast”. It is an online service that helps tourists to arrange their travel experiences and homestays. Airbnb was founded in August 2008, in San Francisco. Users can use filters to find desirable houses and complete online booking processes through the internet or mobile apps. The platform only provides services as brokers and receives commissions from each booking. According to the public media, the company provides tens of millions of unique homestay choices in 192 countries and operates in 65,000 cities. In 2011, services from Airbnb increased 800% incredibly.

* Project Goal

For the project, we design to use two methods to solve two real-world problems. To the customer’s position, we will find patterns of the price and other attributes. To the host’s position, we will help them to find a price range for them to set if they want to join Airbnb in San Diego. The main goal is to perform analyses and find insights of the rental industry and sharing economy by using Airbnb data sources.

* Data Mining Task

1. Association rules:

We will use the Association rules to find the relationship between the price and other attributes. We will use them to understand the influence of the attributes and determine which attributes are significant or not.

1. SVMs and Decision Tree:

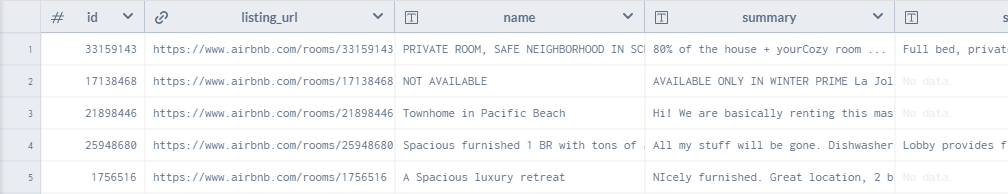
We will first use SVMs to classify the data for better interpreting the data and improve the readability. Then we will make the decision tree. The two methods mainly help us make the model to predict the price.

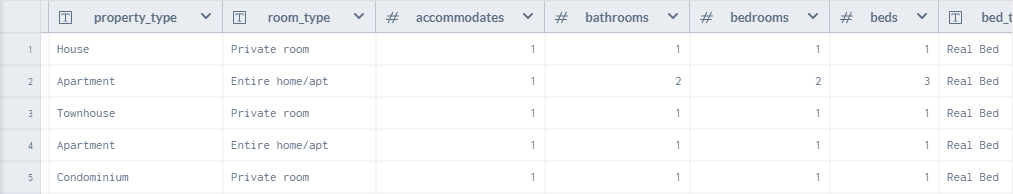
* Data Source Interpretation

Dataset for the project: San Diego Airbnb

The dataset contains all active Airbnb listings in San Diego, including information such as locations, room details and relative fees.

Sample tables:





(there are 75 attributes, so we only show some of them as the sample table)

Data Summary(significant attributes):

The attributes about address and description are not good for data analysis, so we are not using these attributes here.

The attribute “price\_per\_stay” will be our target attribute for both problems.

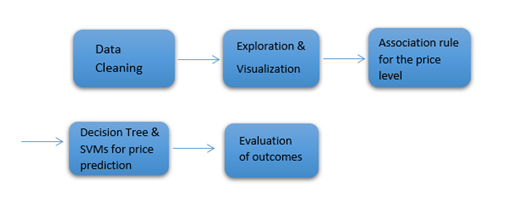
The attribute “host\_response\_time” and “host\_response\_rate” will be classified as different levels for better analysis.

The attribute “host\_is\_superhost” “host\_identity\_verified” “host\_has\_profile\_pic” “” will be converted to a dummy variable.

* Evaluation

To evaluate the result of the data mining project, we will do a moderate amount of tests with real data. The accuracy of the predictions is the best way to evaluate the result.

* Illustration



* References:

<https://en.wikipedia.org/wiki/Airbnb>