

Airbnb Data Anylsis

Background:

Airbnb is a company based in San Francisco, CA, United States, who is an online marketplace for arranging or offering lodging, primarily homestays, or tourism experiences. Nowadays, it has become one of the world's largest housing sharing platform, which provided the service all around the world. Here we will focus on one particular city's airbnb data, on which develop the data analysis, exploration and prediction.

Goal:

The data consists of 7907 samples with 16 columns including three parts (host info, house location info and user reviews info). Our goal is to analyze the influence of features like location information and other circumstance factors on the price of local houses.

Finally, to use the Airbnb data to predict the house price given the related information. A pricing model will be built to match the housing price of this city. There will be a comparison between traditional statistical models and machine learning models.

Technical approach:

We are going to use the basic probability and statistics knowledge we learn from the course to analyze and visualize the housing data. Specially, by analyzing and statising the distribution of all kinds of features with respect to each other and to label, we can model those feature relevance with certain classic distribution model we've learnt from class like Exponential distribution, Gaussian distribution, etc. After that, we are planning to implement some statistical methods together with machine learning algorithms such as Tree models, Support Vector Machine and Neural Network working on the exploration part to get a deeper implicit understanding.

Group member:

Ruoqi Zhang A53275759

Yufei Li A53285858