

Evaluation of Airbnb in Boston Area

Bohan Li, Han Xiao, Ningyi Xue, Yiran Pan

Introduction

As one of the most famous online marketplaces and hospitality service websites, Airbnb becomes more and more popular among travelers all around the world. However, do travelers really know the Airbnb housing they choose to stay? Unlike looking for a traditional hotel on travel agency websites, customers usually need to spend more time on finding a safe and convenient house that matches their expectation because Airbnb let hosts provide all the information, which could lead to the lack of credibility. In our project, we decide to let the data talks. Besides the original Airbnb reviews given by customers, we also evaluate other factors that could affect the qualities of Airbnb housings in Boston Area. The goal of our project is to help potential Airbnb customers to choose the best Airbnb of they needs.

Tools and Datasets

Programming Languages

- Python
- MongoDB

Data Sets:

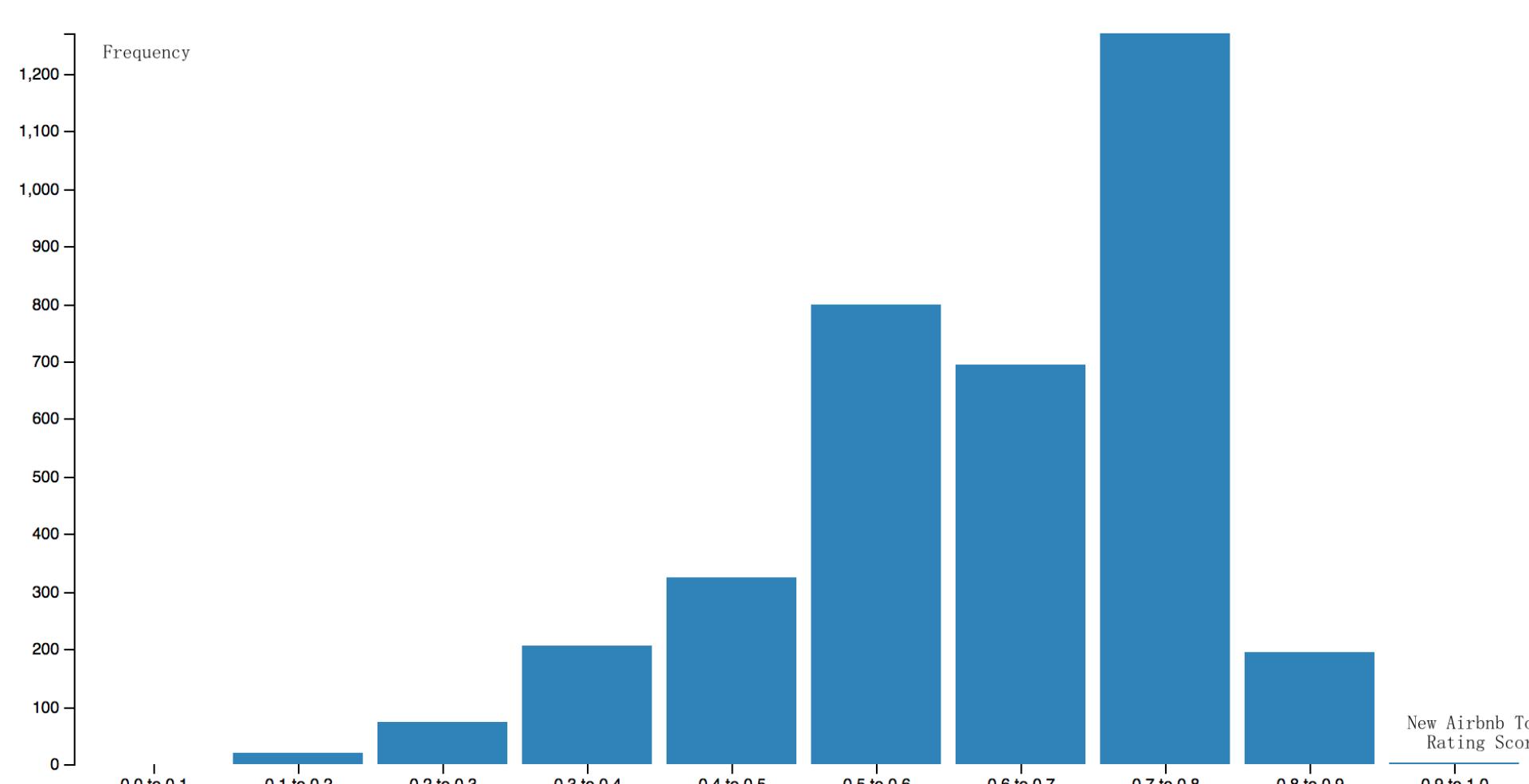
- City of Boston crime incident July 2012 - August 2015
- Active Food Establishment Licenses
- Food Establishment Inspections
- Entertainment Licenses
- Airbnb Boston
- MBTA Bus Stops

Algorithm: Scoring System

In this project, the original Airbnb reviews by customers only serve as a part of the evaluation. For each Airbnb housing, we also count and calculate safety level, surrounding dining options (safety and cleanliness), recreations, entertainments, transports (bus stops). Then we normalize and weight all scores evenly to get the final rating of Airbnb housings under our standard.

`finalScore = (reviews + safety + transportation + recreationAndDinning)/4`

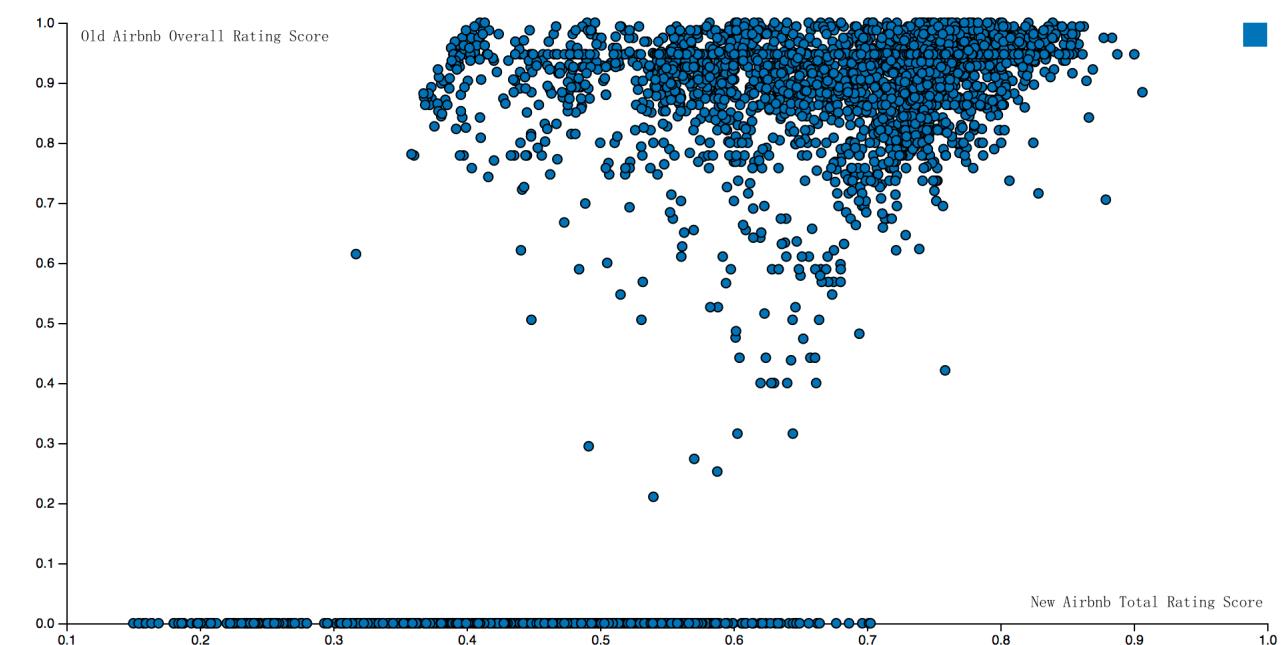
Frequency of Weighted Airbnb ratings



Correlation Analysis

We calculate correlation coefficients between the most concerned factors ([original Airbnb ratings](#), [weighted Airbnb scores](#), [safety level](#), [transportation](#)) and generate three plot graphs to show the relationships.

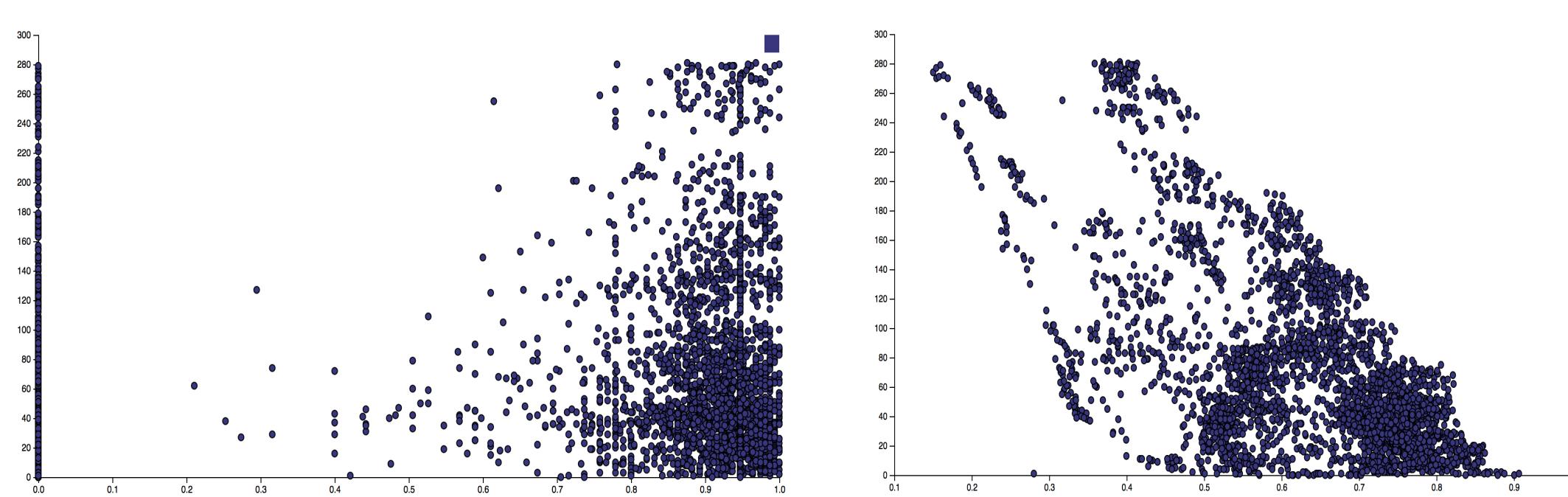
Original Airbnb Ratings(y-axis) vs. Weighted Airbnb Scores(x-axis)



Correlation coefficient: 0.66

Overall, The original Airbnb ratings is higher than the Weighted Airbnb scores, but they are strong positive related. The reviews from customers are relatively reliable. Also, for Airbnb housings that haven't been booked by any customer (dots on x-axis), there are still a few of them worth a try for they have relatively high weighted scores.

Original and Weighted Airbnb Scores(x-axis) vs Safety Level (y-axis)

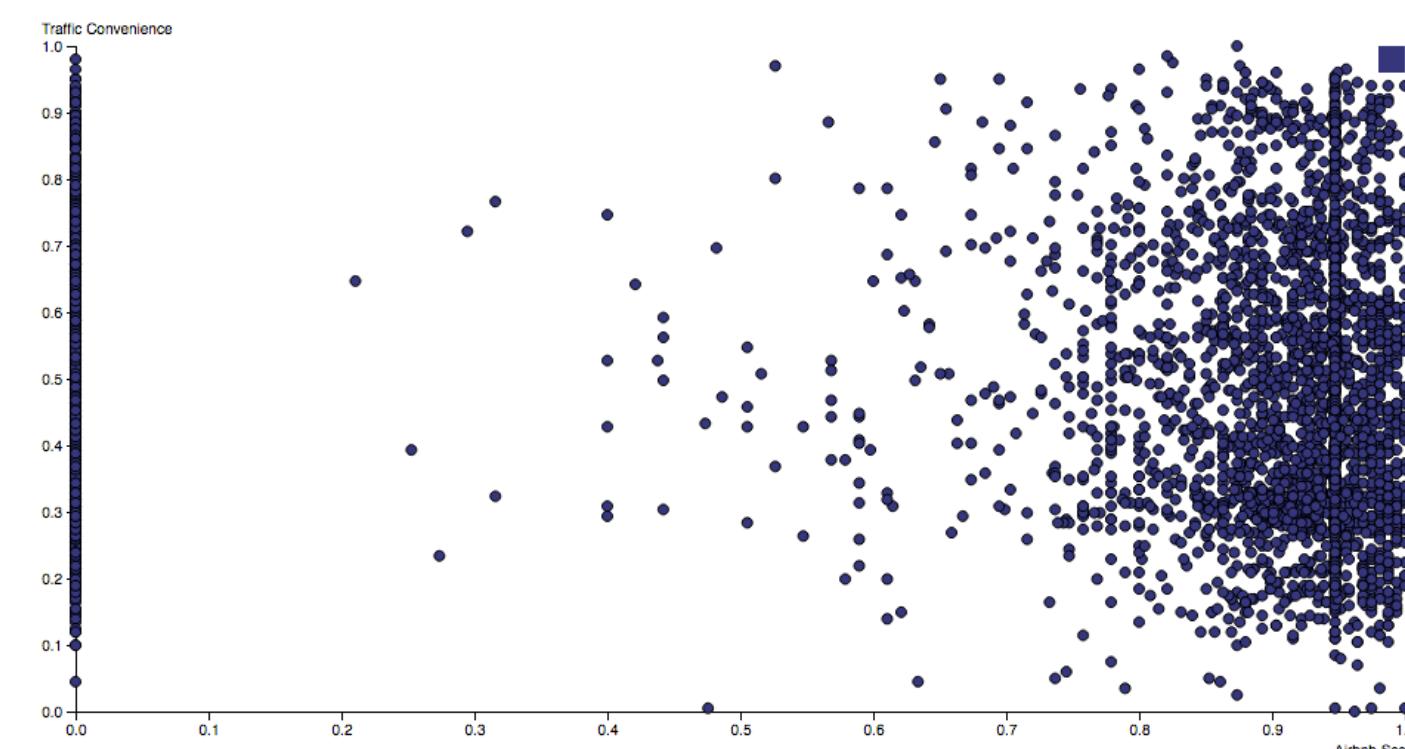


Correlation coefficient: 0.01 and -0.61

Above graph No.1 shows the weak positive relationship between original Airbnb scores and number of crimes happened 1km within of each Airbnb. Although the original ratings do not strongly connected to the number of crimes, we can observe that high rating Airbnb housings are mostly located in safe areas. Thus we think it is reasonable to take safety as one factor that may affect the evaluations.

Graph No.2 shows the strong negative relationship between weighted scores and number of crimes.

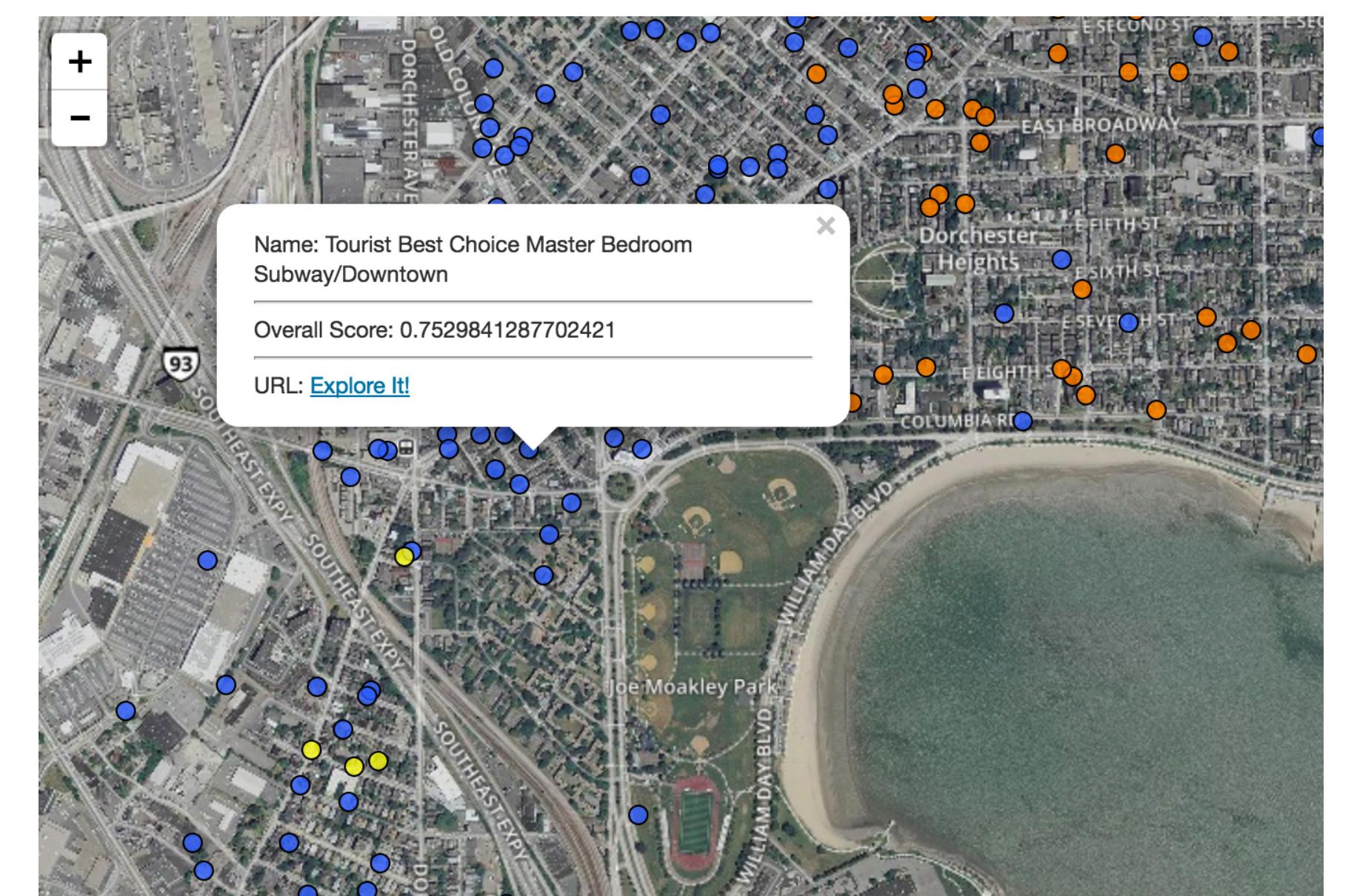
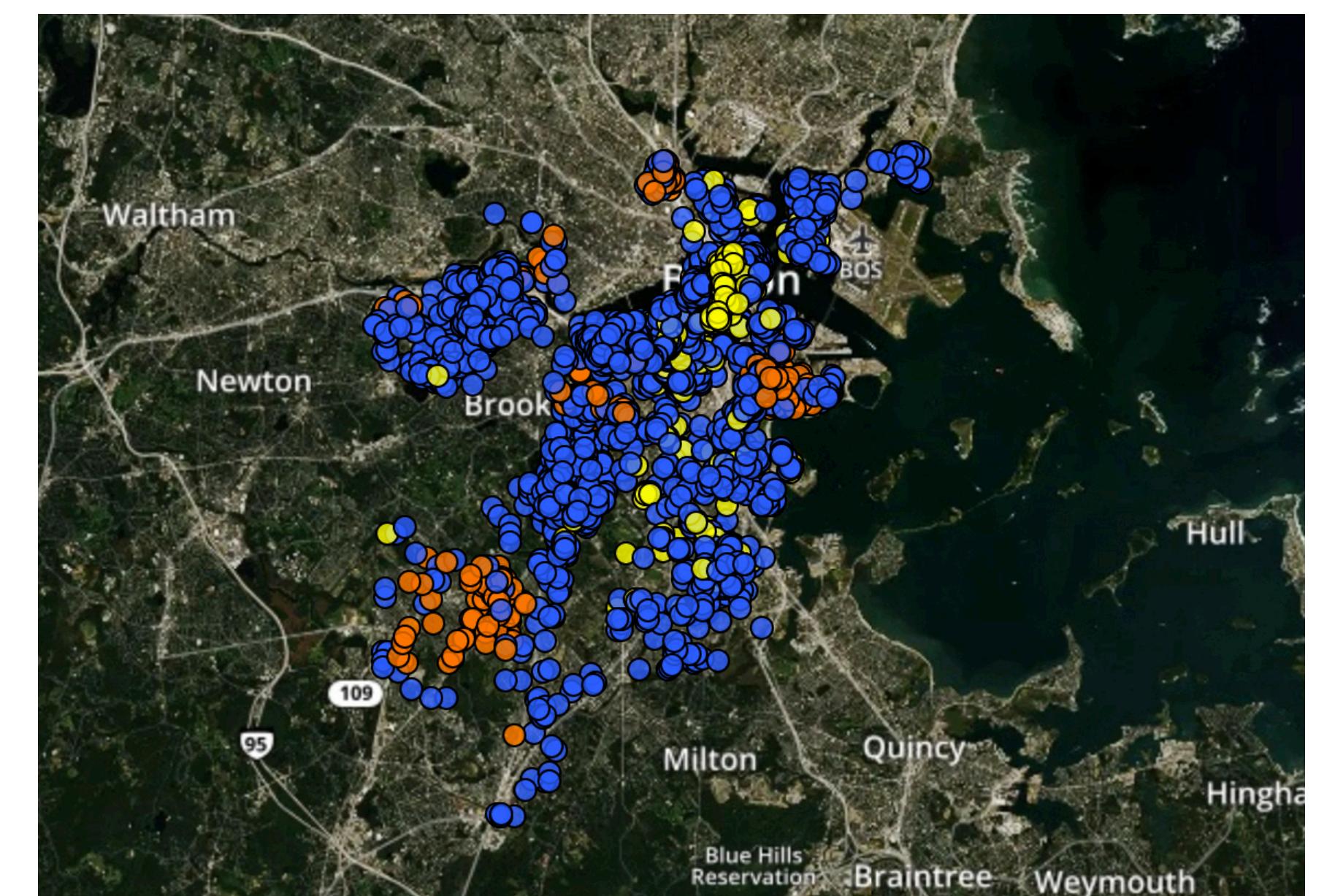
Original Airbnb Scores vs. Transportation Convenience



The relationship between weighted Airbnb scores and transportation convenience is almost not related

Visualization

We create a html page to visualize the Airbnb housings on a interactive map. For the visual convenience, we set a satellite map for Airbnb customers to locate their choices and help to make their best decisions. For each plot of houses on the map, overall score > 0.8 is orange, 0.5 < score < 0.8 is blue, and those below 0.5 are set to yellow. Users can move the mouse and hang over any of the points and the corresponding details (name of the Airbnb housing, weighted score and a URL leads to the Airbnb page of the house) will show up coordinately.



Further Extension

This project still has some space to extend in the future. Due to the limitation of the datasets that we can deploy, we only cover a small part of factors that may affect Airbnb ratings. Prices and evaluation of hosts should also be considered. We also plan to add a search engine that allows users to directly search their ideal or potential Airbnb housings for their trips. We could also allow searching by filters to satisfy various needs (ex. Highest rating, neighbors with most entertainments, safest neighbor). In fact, if we have enough information, we may be able to write a application that can be used in reality.

As for limitations, because we define the scoring system and algorithm by our own understandings, the scores may be biased.