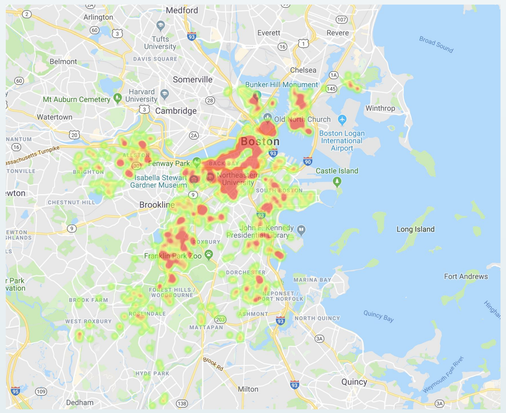
Airbnb Housing Recommendation

Haoxuan Jia U26349907, Jiahao Zhang U22220697
Boston University

Introduction

Since 2008, guests and hosts have used Airbnb to travel in a more unique, personalized way. Boston as a famous city for travel, the number of Airbnb houses has increased greatly these years, making Airbnb house choosing a time consuming thing to do. The purpose of our project is to pick out the best houses with low price as well as high review score for users and save their time of going through the whole list of Airbnb houses.



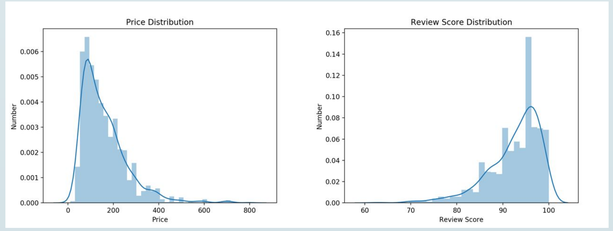
A heatmap distribution of Airbnb houses density in Boston

Data Analysis

Dataset we used:

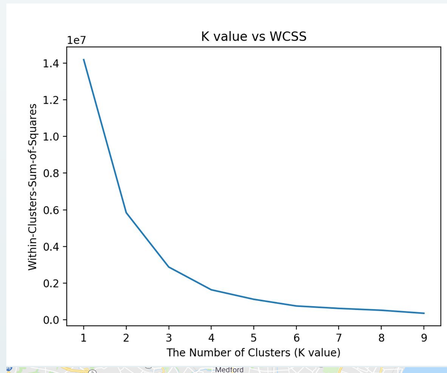
- Boston Airbnb Calendar
- Boston Airbnb Listing
- Boston Airbnb Reviews
- Boston Landmarks
- US Holidays

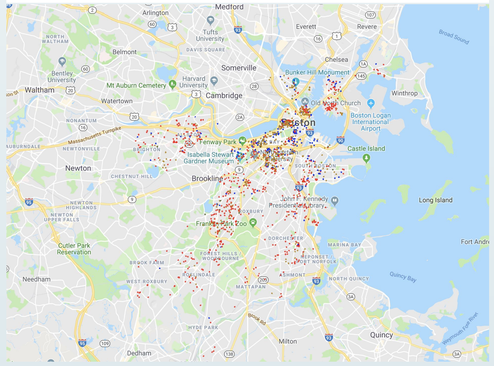
- There are total 3585 Airbnb houses, and the price range is from 0 to 800 and most is distributed below 200 dollars.
- The review score range is from 60 to 100, most of the data is distributed between 90 and 100.



Methodology and Results

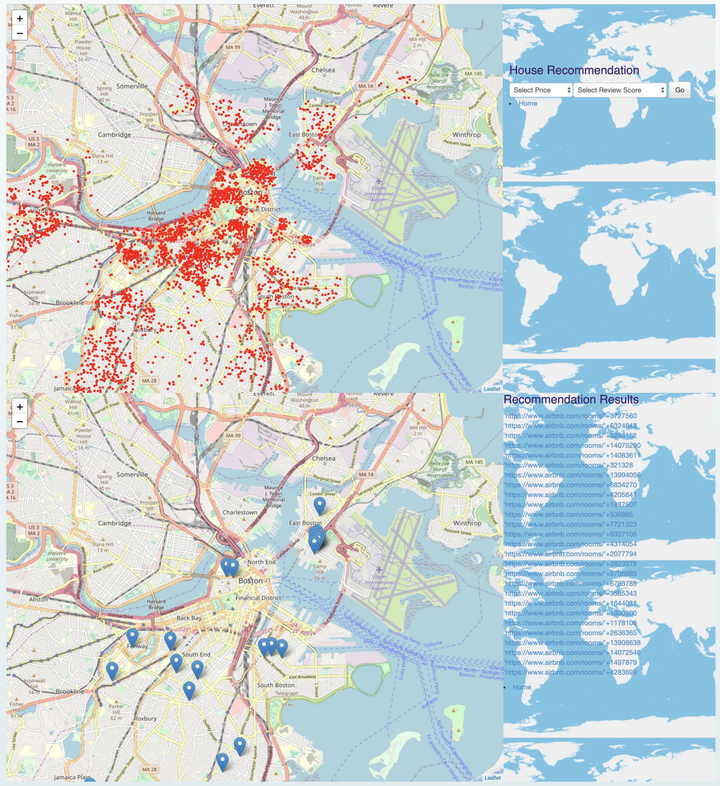
- K-Means
  - Calculated Within-Cluster-Sum-of-Squares (the sum of square deviations between each observation and its cluster center)
  - Plotted Within-Cluster-Sum-of-Squares vs the number of clusters (K-value)
  - Used Elbow Method to choose best K-value where the x value corresponding to the most significant changing point is best K
  - Plotted each cluster onto google map to reveal the housing pattern in clusters





- Results
  - Housing with highest recommendation
    - Allston, Brighton, East Boston and Mission Hill
  - Housing with second highest recommendation
    - Back Bay and Downtown area
  - Housing with second lowest recommendation.
    - North corner of Boston
  - Housing with lowest recommendation
    - Rare and no general pattern
- Improvements
  - K-Means++
    - Helped initialize centroids to be distant from each other
    - Helped speed up convergence
  - Mini Batch K-Means
    - The computation time was shortened
    - The performance was slightly worse than standard one

Web Visualization



- User Parameter Input Page
  - Select price range
    - $ 0 ~ 100
    - $ 100 ~ 200
    - $ 200 ~ 300
    - $ 300 +
  - Select review score
    - ! 85 -
    - !! 85 ~ 90
    - !!! 90 ~ 95
    - !!!! 95 ~ 100
- Register/Login Page
  - New user - register
  - Old user - login
- Recommendation Results Page
  - Marked in map using Leaflet
  - Urls listed as hyperlink
- Tool: Python, Flask, HTML, CSS, Javascript, Folium, Leaflet

Future work

- For features, we considered only prices and review scores in the projected. We will research into previous customers' comments and used modern NLP technique to pick out highly rated houses more accurately.
- For algorithm, K-Means is a naive classification algorithm. We can use more complicated classification methods.
- For web application, it is not very user-friendly right now. We can add some interesting function like multiple login ways, search lists, add friends, etc.