



# Evaluation of Boston Hotels

Desheng Zhang / Haotian Wu/ Wenjun Shen  
Department of Computer Science/Boston University  
College of Art & Science



## Overview

- Boston is a popular destination for vacations and sightseeings, and there are numerous hotels located in the city of Boston for travelers to choose.
- We believe there are certain factors that affect the quality of the hotel and we want to find out those factors and analyze each hotel with a quantitative score. Eventually, we can calculate a potential location to build a hotel with high score, and we can also recommend hotels to customers based on their preferences.

## Datasets

- TripAdvisor Hotel Scores in Boston
- Boston Crime Incident Report
- MBTA Stops
- Boston Food Establishment Licenses
- Boston Public Gardens

## Statistical Analysis

- **Custom Score System:**
  - We obtain the original hotel scores in Boston from TripAdvisor website.
  - In this project, we calculate the number of gardens, crimes, MBTA stops, foods near each hotels(within certain distance). We use the formula below to calculate the normalized. score of each factor.

$$Score_{normalized} = \frac{Score - Score_{min}}{Score_{max} - Score_{min}}$$

- Then We use the normalized score of each individual factor to calculate the custom score. Since lower number of crime incidents indicates safer environment. We use 1 - Crime Score..

$$S_{custom} = \frac{S_{original} + S_{garden} + S_{food} + S_{original} + (1 - S_{crime})}{5}$$

- **Correlation Coefficient:**
  - Calculate the correlation coefficient between each factor and custom score of each hotel using the formula below.

$$corr(S_i, S_{custom}) = \frac{cov(S_i, S_{custom})}{std(S_i) - std(S_{custom})}$$

- **Result**
  - We find out that the coefficient between crime and custom hotel score is extremely close to zero, and the p-value is relatively large when comparing to other factors. Thus, we conclude the crime factor is not significant and no longer consider crime as a factor at next step.

	Coefficient	P Value
Crime	-0.09448652255976357	0.3984638884028119
MBTA Stops	0.5511472545817303	8.065664653031119e-8
Gardens	0.8712154440427992	1.9353053027735632e-26
foods	0.6383325374109514	1.1093179778057194e-10

## K-means and Clustering

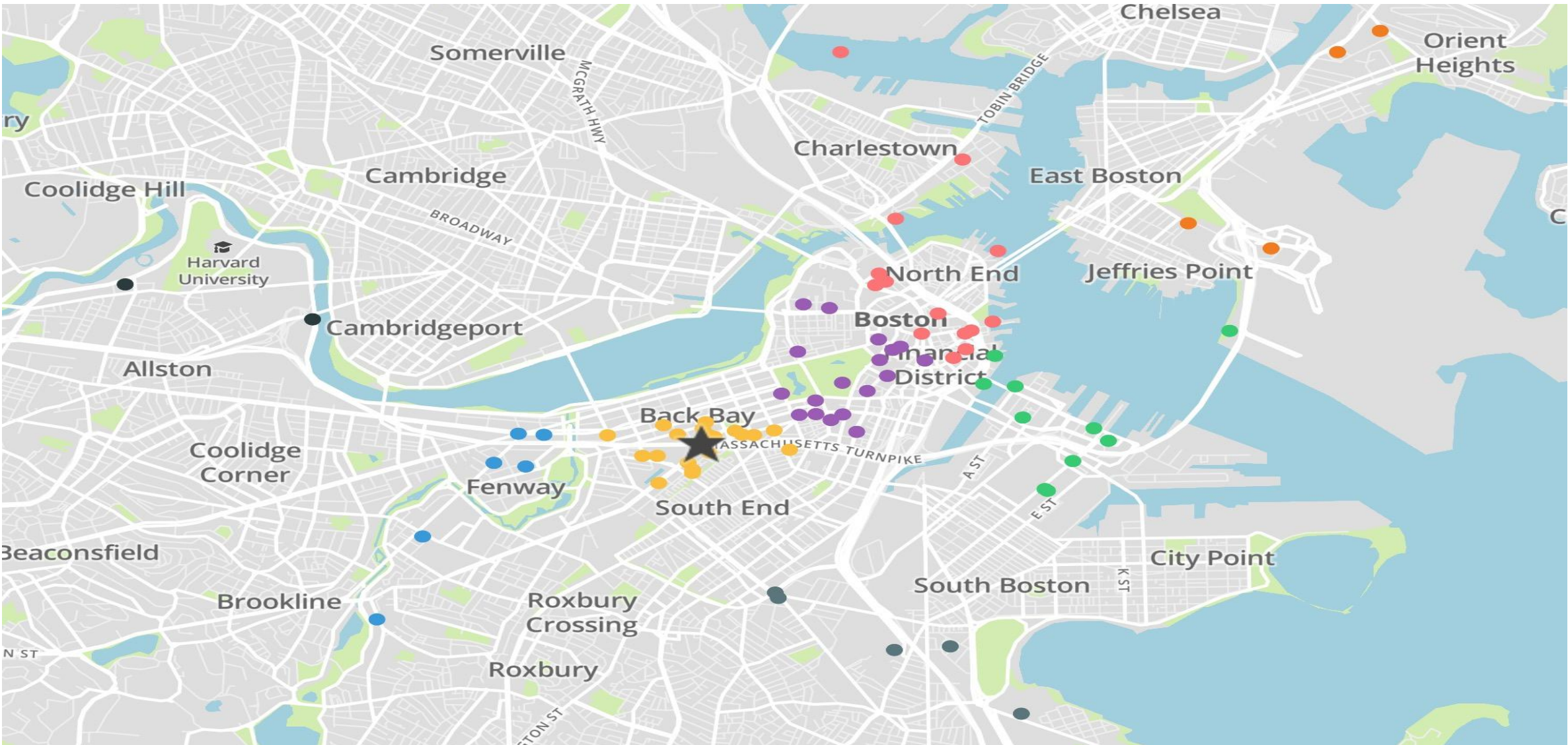
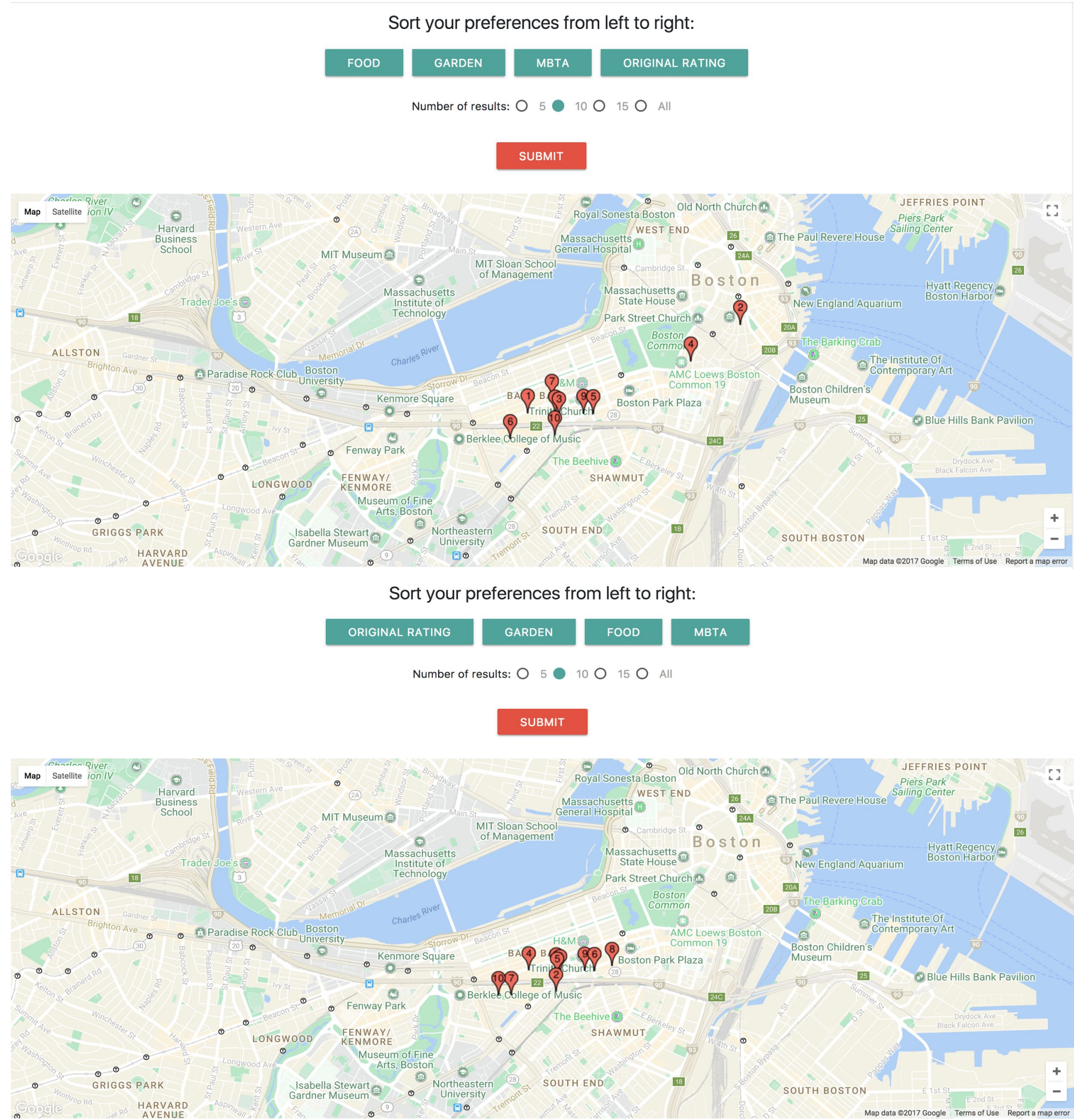


Figure 1. Cluster with K-means(10 clusters)

- After discarding crime factor , we recalculate the score of each hotel by original score, foods, mbta stops, and gardens. We still use the average score of the four normalized scores.
- After analyze the error and Silhouette Coefficient score, we find that the best number of cluster should be 10.
- Finally we use the coordinates of the hotels and the new custom score(without crime factor) for K-means algorithm.
- Instead of using matplotlib package in python, we use an another powerful plot package called plotly. By using this, we successfully plot each renting point onto a real map.
- As shown above, for different clusters, we mark with different color. We select the middle of the cluster with highest average score as the best location to build a new hotel, which is the black star on the map.

## Visualization



- We made an interactive client-server web application to visualize our project. We use Django as backend server and plain javascript as frontend.
- Our application can let users decide the order of the four factors including origin hotel score, food, mbta stops and gardens by their own preference and shows the best several hotels that our algorithms recommended. Users can also choose the number of results to show on the map.
- To provide user a better experience, we also apply other javascript libraries such as Bootstrap, Sortable.js, Google Map Javascript API and Materialize.

## Future Work

- Since food establishment licenses dataset contains many small cafeterias and our project targets hotel customers, we can filter those cafeterias and make the food dataset more target specified.
- We can include more factors such as nearby attractions in the score system to make the score reflect the hotel's quality more accurately.
- When we visualize the result hotel, we only display the hotel name and the a pinned location on the map. we can show more information about the information including photos and external links to book the hotel.