2019

Hotel Recommendation Application

Behzad Sajed Khosrowshahi

IBM Data Science Professional Certificate

Capstone Project

3/5/2019

Introduction

Background

These days travelling to different places in the world is faster and easier than before. There are different reasons for a tourist to visit a new city. There are tourists that like to visit different countries to experience different cultures which includes arts, food. There are travelers who love nature and would like to spend most of their time in the nature. Also, couple of friends would like to have parties and nightlife in a city for an occasion. And at the end some cities are perfect for shopping.

As we can see there are different purposes for each tourist to visit a new city. After arriving to a new city, the most important things that tourists need to arrange before arriving is accommodation. It could be time consuming to check all hotels in a city within your budget and find the best one that addresses all your preferences. If we have a system that help us to choose the accommodation, it would be great.

Problem

Based on City's data and also the user's preference, we should have a system to recommend us the best hotels that fit to our requirements. I called this project **Hotel Recommendation Application.** This project is going to analyze city's venue data and based on user's preference, it shows the list of hotels that user may select them.

Interest

This application is going very helpful for tourists that want to visit a new place and they do not have any idea where they should book their hotel.

For an example, if a visitor like to spend more time for visiting arts, so maybe it is better to book a hotel in an area that it is closer to most of the arts venues. Another visitor may prefer a hotel close to area that has more restaurant. So, the application can recommend a different list of hotels.

Data Acquisition and Cleaning

Data Sources

For this application I need these data:

- Visitor Preference: Visitor can an option to choose one these preferences.
    - Food which includes Restaurant, Ice Cream shop, and any venue is related to Food
    - Drinks that can be Bar, Brewery, etc.
    - Shops which meaning shopping center, store.
    - Arts that has museum, art gallery, live music.
    - Outdoors for park, beach, hiking.
    - Sights which Historical Site, Library are part of it.
    - Top Picks which the interesting points in the city that most people visit it.
- City's Venues Information which I used the Foursquare API to collect this data.
- City's Hotel Information that again I used the Foursquare API.
- City's Geographical Information: I used the Geopy library to retrieve Boston Latitude and Longitude.
- Map of the Boston that Folium library helps us.

Data Cleaning

The Data that downloaded from Foursquare is a JSON object. These JSON text needs to be converted to Panda's dataframe for data analyzing. After analyzing the first response from Foursquare, I figured out I need to drop couple of columns. Also, I needed to write a function to retrieve data from any array in a JSON text. And at the end I needed to rename couple of columns.

Feature Selection

I used two Foursquare commands to extract data from their databased. The first one is **explore**. This is the command I used

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&near={}&section={}&radius={}&limit={}

cliend_id, and clien_secret are my credential for Foursquare. And the *v* parameter is the version which I should send to Foursquare. The *near* parameter is the city which in this project is Boston, MA. And the *section* parameter is user's preference (e.g. Food, Arts, etc). The last two parameters are *Radius* for the

maximum distance between venue and the Boston's latitude and longitude and then *Limit* which sends the maximum number of venues that this command replies.

This Foursquare explore command replies with a dataset with lots of features. For more information, you can check https://foursquare.com/dev/docs/venues/explore. For this project I just needed to keep these features:

- Venue Name Name of the Venue
- Venue Category which can be the category of the Venue (e.g. Food, Arts, Outdoors, etc)
- Venue's latitude and longitude for the venue latitude and longitude.
- Venue's Zip Code that has the Zip code of the venue

And I dropped other features.

This is an example of the outcome of the data.

	name	categories	lat	lng	ZipCode
0	Sam LaGrassa's	Sandwich Place	42.356870	-71.059960	02108
1	Yvonne's	New American Restaurant	42.355664	-71.061466	02108
2	haley.henry	Restaurant	42.357574	-71.059495	02108
3	sweetgreen	Salad Place	42.357704	-71.058713	02108
4	Grotto	Italian Restaurant	42.359915	-71.062807	02114

The second Foursquare command that I used is **search** command. This is the template of the command:

https://api.foursquare.com/v2/venues/search?&client_id={}&client_secret={}&v={}&near={}&query={}&radius={}&limit={}

This command is kind of the same as Explore command but I just need to replace the *section* parameter to *query* for adding "Hotel" to this command.

The result of the search command is kind of the same but has different field. Again, I needed to drop all unused columns and rename some columns. For extracting the category of a search command's result, I needed to write a new function which is called *get_search_category_type.*

This is an example result of search command.

	Name	categories	Latitude	Longitude	ZipCode
0	Kimpton Nine Zero Hotel	Hotel	42.357388	-71.060696	02108
1	Godfrey Hotel	Hotel	42.354748	-71.061613	02111
2	Millennium Bostonian Hotel Boston	Hotel	42.361006	-71.056066	02109
3	XV Beacon Hotel	Hotel	42.358217	-71.061900	02108
4	Club Quarters Hotel in Boston	Hotel	42.356241	-71.057579	02110

Methodology

In this project there are four phases which should be run in sequence. First, I am sending the explore command to Foursquare based on user's preference and then convert the result into a panda dataframe.

And then, send the Foursquare search command to get info of all hotels nearby the city that user is going to visit.

After that the application is going to create clusters based on this nearby venues' latitude and longitude. The method that clusters will be created is K-Mean. The number of clusters is seven in this application.

After clusters are defined, then we add them to the map. I used seven different colors to distinguish them. These colors are **Red, Blue, Black, Purple, Green, Gray and Brown.**

And the final step is adding hotels' location to the map of the city to help visitor to choose a hotel nearby her/his preference. The color of the hotels in the map is **yellow.**

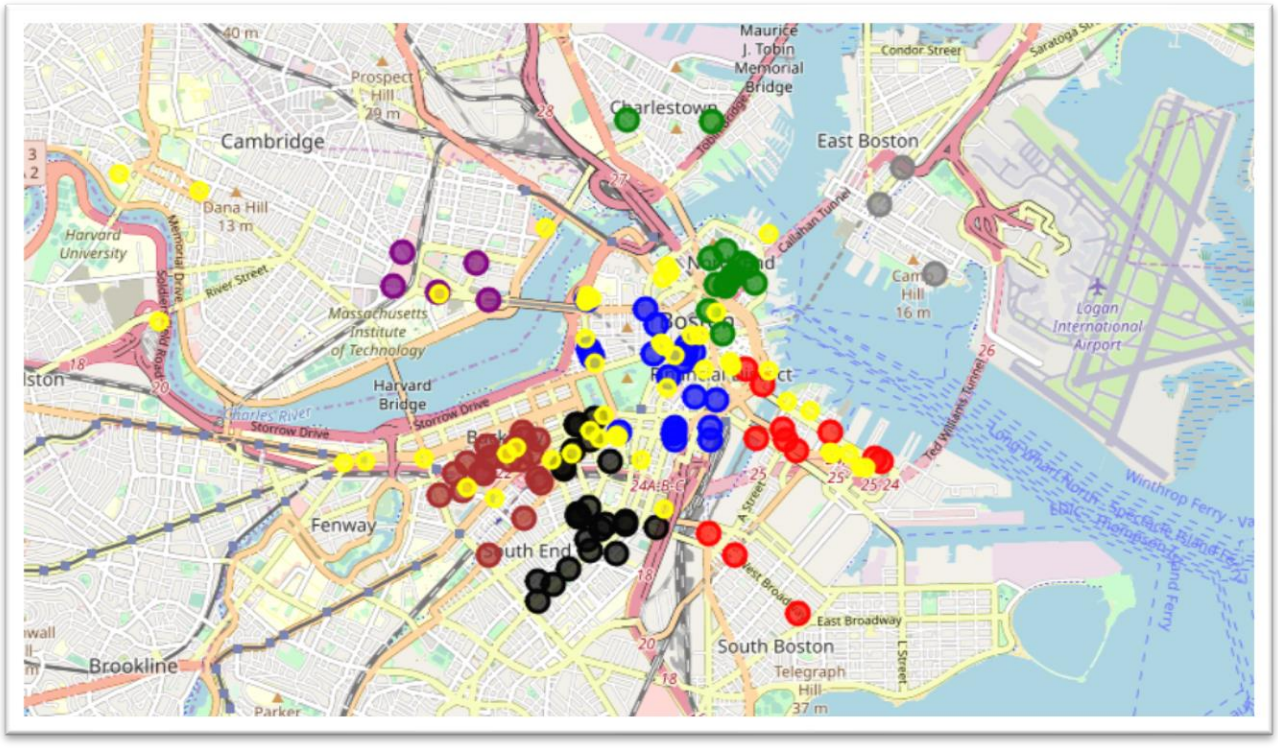
Results and Discussions

I ran this application for different seven scenarios based on seven different preferences that user can choose.

These are the results for each scenario.

Scenario 1: User Selected Food

	name	categories	lat	lng	ZipCode
0	Yvonne's	New American Restaurant	42.355664	-71.061466	02108
1	haley.henry	Restaurant	42.357574	-71.059495	02108
2	Trillium Garden On The Greenway	Beer Garden	42.356540	-71.051100	02110
3	The Tip Tap Room	American Restaurant	42.361153	-71.063976	02114
4	O Ya	Sushi Restaurant	42.351502	-71.056763	02111



Scenario 1: User Preference = Food

In this scenario, it seems if we select a hotel close in black cluster, we have more options for food in Brown, Blue and Black clusters.

Scenario 2: User Selected Drinks

	name	categories	lat	lng	ZipCode
0	Yvonne's	New American Restaurant	42.355664	-71.061466	02108
1	haley.henry	Restaurant	42.357574	-71.059495	02108
2	Trillium Garden On The Greenway	Beer Garden	42.356540	-71.051100	02110
3	The Tip Tap Room	American Restaurant	42.361153	-71.063976	02114
4	O Ya	Sushi Restaurant	42.351502	-71.056763	02111

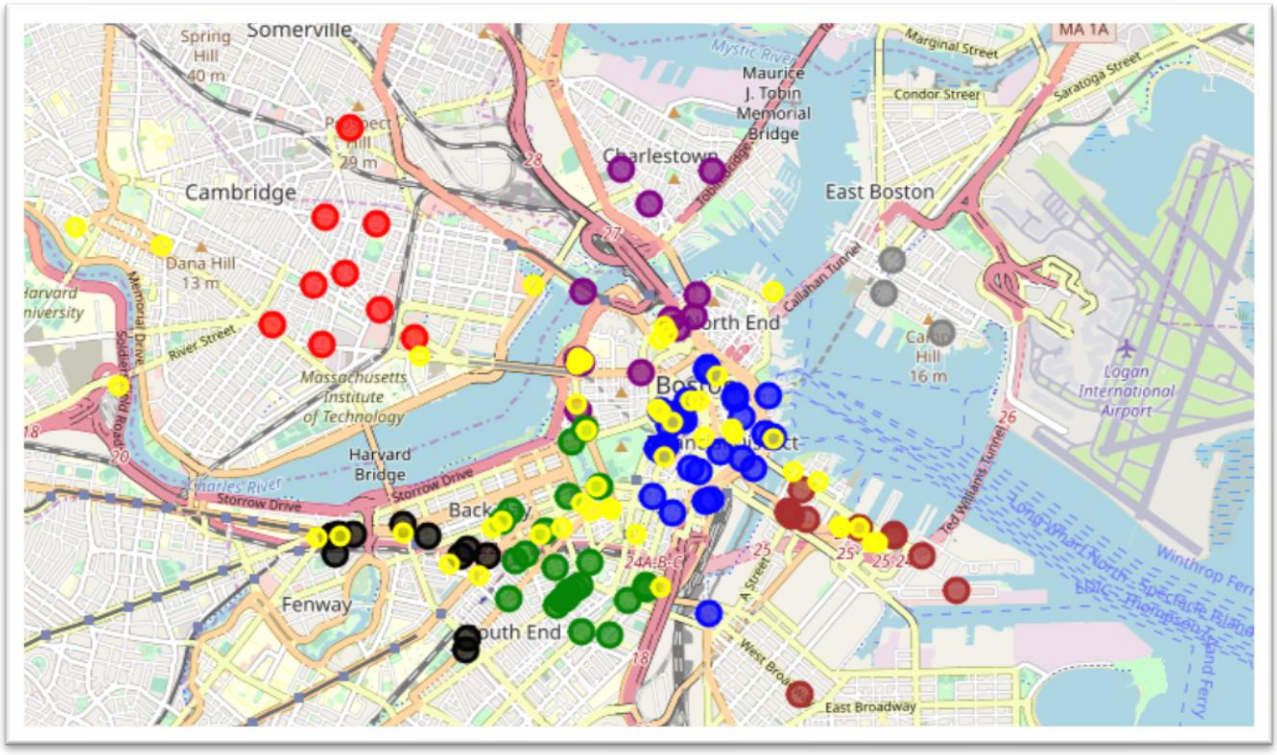


Figure 2: User Preference = Drinks

In this scenario, it seems we should select hotels in border of green and blue clusters which give us more picks.

Scenario 3: User Selected Shops

	name	categories	lat	lng	ZipCode
0	Boston Public Market	Market	42.361950	-71.057466	02108
1	Roche Bros Downtown Crossing	Market	42.355363	-71.060117	02110
2	Newbury Comics	Record Shop	42.360643	-71.054413	02109
3	Whole Foods Market	Grocery Store	42.361732	-71.065874	02114
4	Primark	Clothing Store	42.355538	-71.060095	02110

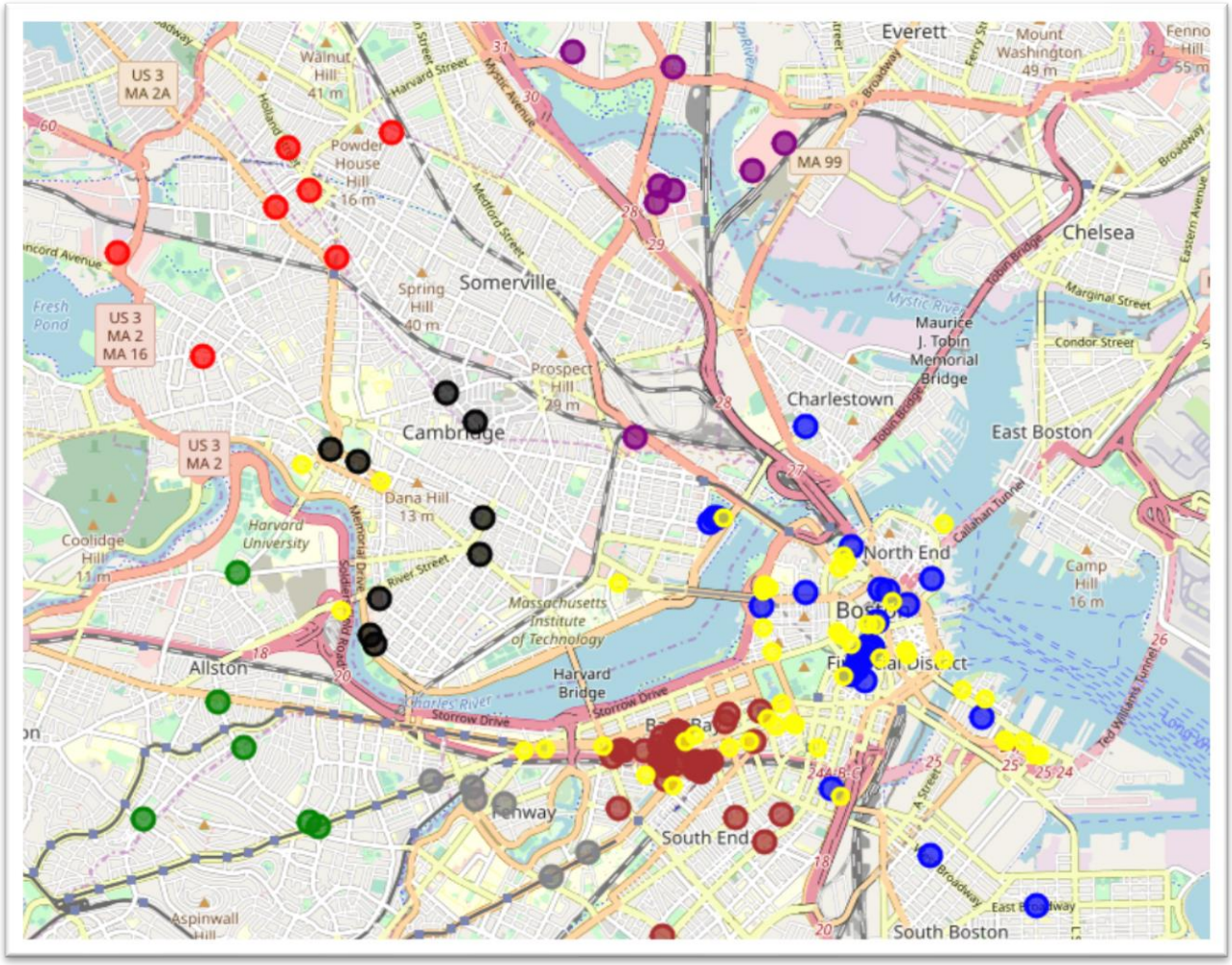


Figure 3: User Preference = Shops

In this scenario, for sure we should select a hotel in blue or brown clusters.

Scenario 4: User Selected Arts

	name	categories	lat	lng	ZipCode
0	Boston Opera House	Opera House	42.354014	-71.062602	02111
1	Improv Asylum Theatre	Comedy Club	42.362931	-71.055404	02113
2	Make Way For Ducklings	Outdoor Sculpture	42.355569	-71.069764	02108
3	Cutler Majestic Theatre	Theater	42.351545	-71.064850	02116
4	Museum of African American History	History Museum	42.360058	-71.065287	02114

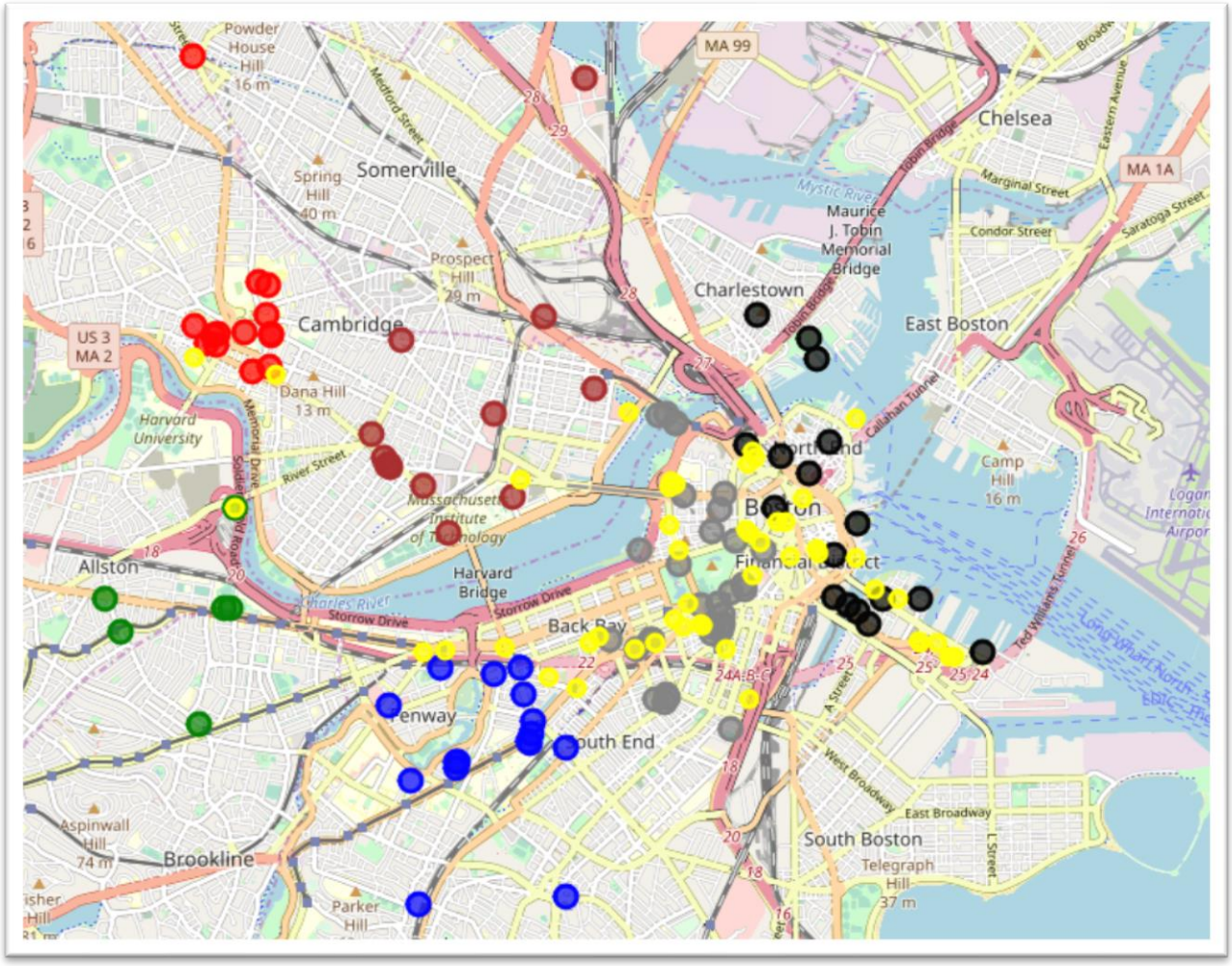


Figure 4: User Preference = Arts

In this scenario, it seems selecting a hotel in black or gray cluster helps us to spend more time for visiting Arts venues.

Scenario 5: User Selected Outdoors

	name	categories	lat	lng	ZipCode
0	Boston Common	Park	42.355487	-71.064882	02111
1	The Rose Kennedy Greenway	Park	42.358227	-71.052055	02110
2	North End Park	Park	42.362488	-71.056477	02113
3	Barry's Bootcamp Boston	Gym / Fitness Center	42.354010	-71.059776	02111
4	Equinox Sports Club Boston	Gym	42.353189	-71.063053	02111

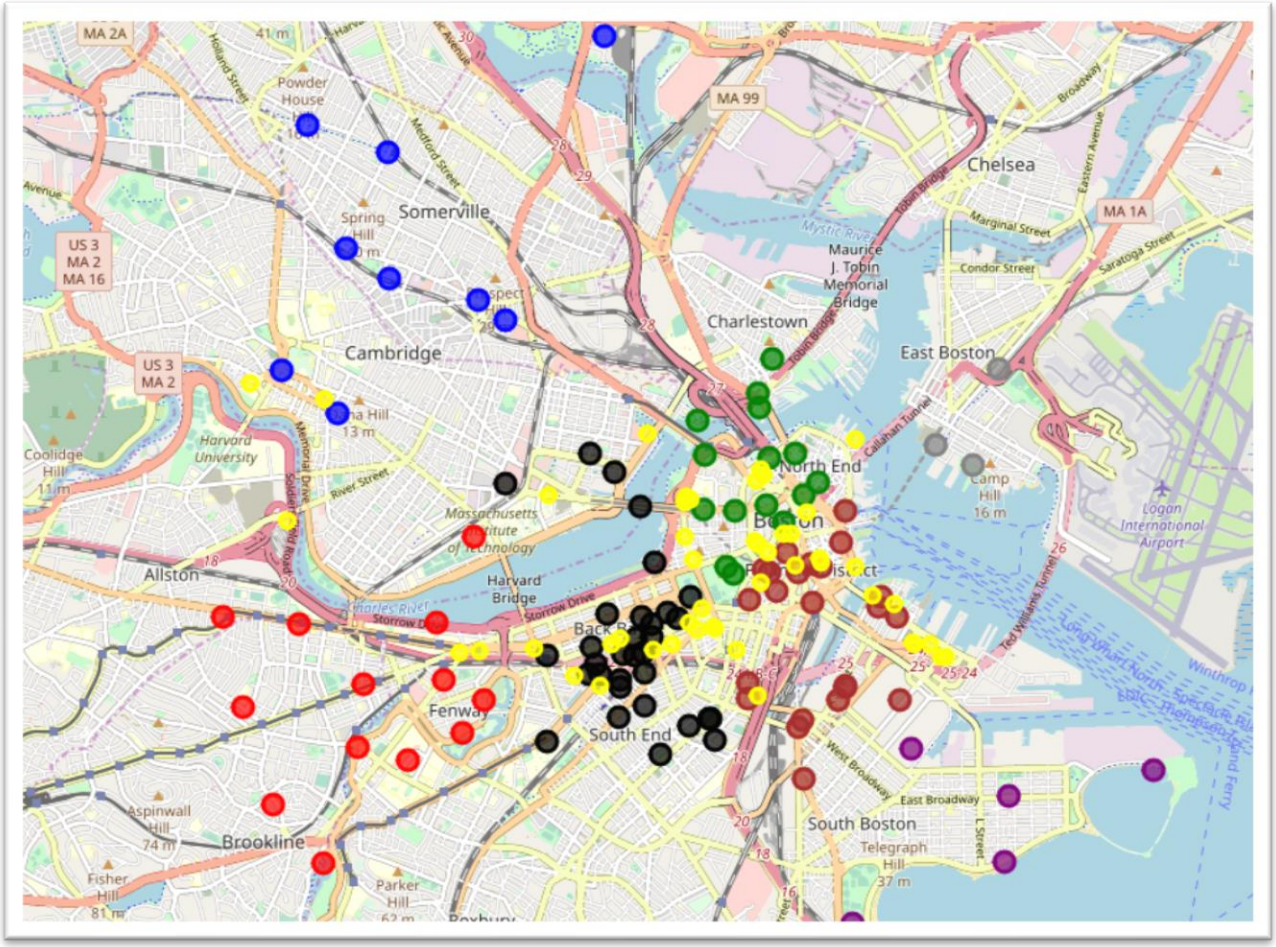


Figure 5: User Preference = Outdoors

In this scenario, hotels should be selected in the brown cluster gives us to be close to venues in black and green clusters.

Scenario 6: User Selected Sights

	name	categories	lat	lng	ZipCode
0	Sam LaGrassa's	Sandwich Place	42.356870	-71.059960	02108
1	Boston Athenaeum	Library	42.357481	-71.061838	02108
2	The Freedom Trail	Historic Site	42.357314	-71.061038	02111
3	Yvonne's	New American Restaurant	42.355664	-71.061466	02108
4	haley.henry	Restaurant	42.357574	-71.059495	02108

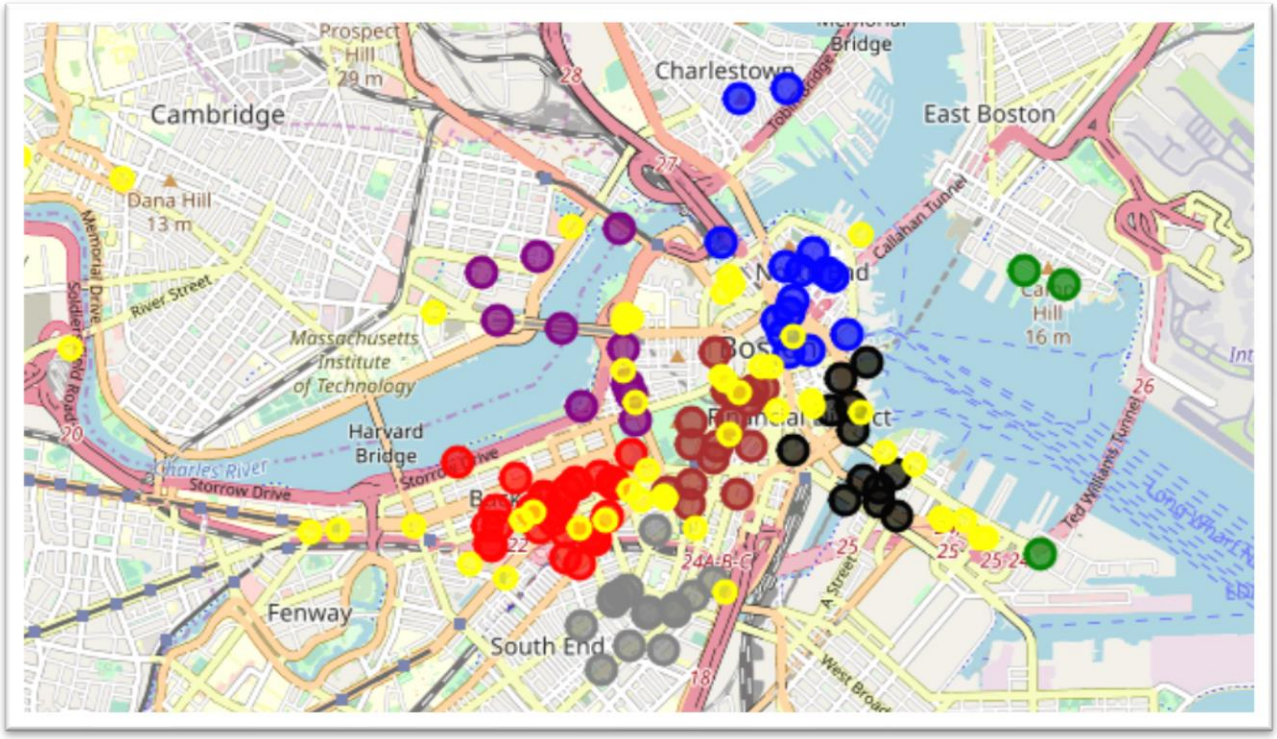


Figure 6: User Preference = Sights

In this scenario, selecting a hotel in the border in brown cluster helps us to have short distance to most of the sights in Boston.

Scenario 7: User Selected Top Picks

	name	categories	lat	lng	ZipCode
0	The Rose Kennedy Greenway	Park	42.358227	-71.052055	02110
1	Flour Bakery + Cafe	Bakery	42.373117	-71.122349	02138
2	Sam LaGrassa's	Sandwich Place	42.356870	-71.059960	02108
3	Museum of Fine Arts	Art Museum	42.339110	-71.094012	02115
4	Cutty's	Sandwich Place	42.333246	-71.119255	02445

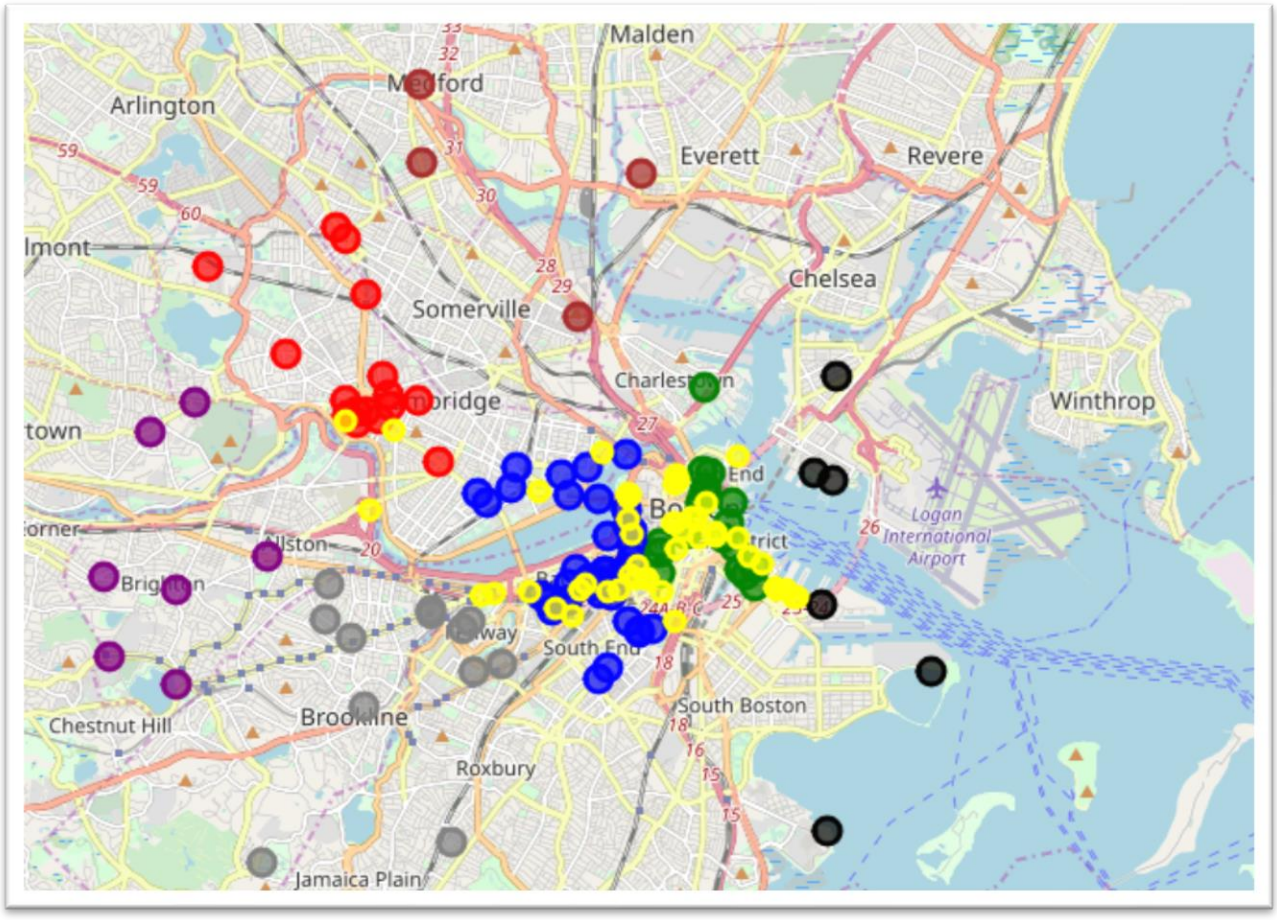


Figure 7: User Preference = Top Picks

In this scenario, selecting a hotel in green or blue clusters helps us to visit places that most of tourists visit when they are in Boston.

Conclusion

In this project, I use the venue information in the city of Boston to recommend tourists for choosing a hotel that can help to spend most of the time in the city to do what they prefer. These preferences could be visiting different restaurant, museums, parks and/or those interesting points in the city that most the tourists for this city check it.