Hotel Recommendation Application

Behzad Sajed Khosrowshahi

March 2019

IBM Data Science Professional Certificate Capstone Project

Recommending Hotels for Tourists

TRAVELLING AROUND THE WORLD IS EASIER AND FASTER THAN BEFORE.

TOURISTS HAVE DIFFERENT PURPOSE TO VISIT A NEW CITY. SOME OF THEM LIKE TO CHECK THE MUSEUMS, SOME OTHER LIKE TO TEST DIFFERENT FOOD, AND THERE ARE TOURISTS THAT LIKE TO SPEND MOST OF THEIR TIME OUTDOOR.

SELECTING A HOTEL IN A CITY THAT YOU DO NOT HAVE ANY IDEA ABOUT IT SHOULD BE VERY TIME CONSUMING AND HARD.

THE PURPOSE OF THIS PROJECT IS TO HELP TOURISTS FOR SELECTING A HOTEL BASED ON THEIR PREFERENCES
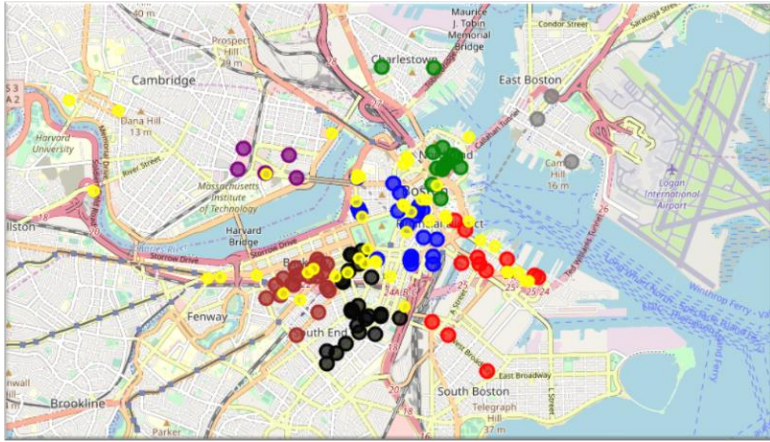
Data Acquisition and Cleaning

- The output of Foursquare API is a JSON object which should be converted into Panda DataFrame

- For Venue dropping all columns except Venue.Name, Venue.Category, Venue.Latitude, Venue.Longitude and Venue.PostalCode

- Writing a function to retrieve the category name from a JSON array.

- Example of Panda Dataframe

	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

Methodology

- Sending a Explore command to Foursquare with a parameter of section for retrieving all venues which have a section attribute
- Sending a Search command to Foursqure with a query parameter = "Hotel" for searching all hotels in the Boston
- Using K-Mean for clustering the venues based on their latitude and longitude.
  - There are 7 clusters
- Add these clusters to the map of the Boston with different seven colors
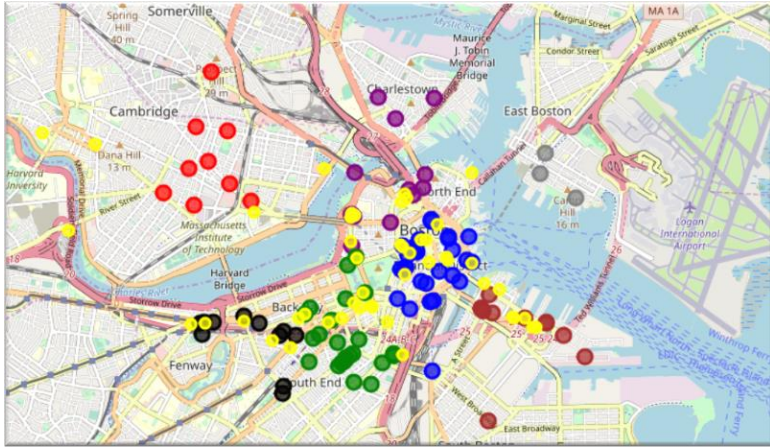- Add hotels to the map with a yellow mark

	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

Results and Discussion

Scenario 1: User Selected 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.

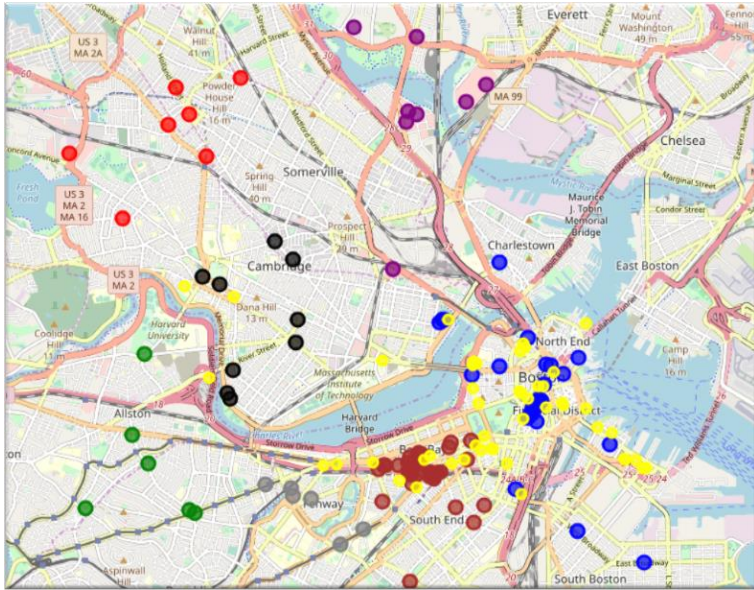|	name	categories	lat	lng	ZipCode
0	Yvonne's	New American Restaurant	42.355664	-71.061466	02108
1	haley.henry	Restaurant	42.357574	-71.059495	02108
2	Trillium Garden On The Greenway	Beer Garden	42.356540	-71.051100	02110
3	The Tip Tap Room	American Restaurant	42.361153	-71.063976	02114
4	O Ya	Sushi Restaurant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

Results and Discussion

- Scenario 2:User Selected Drinks

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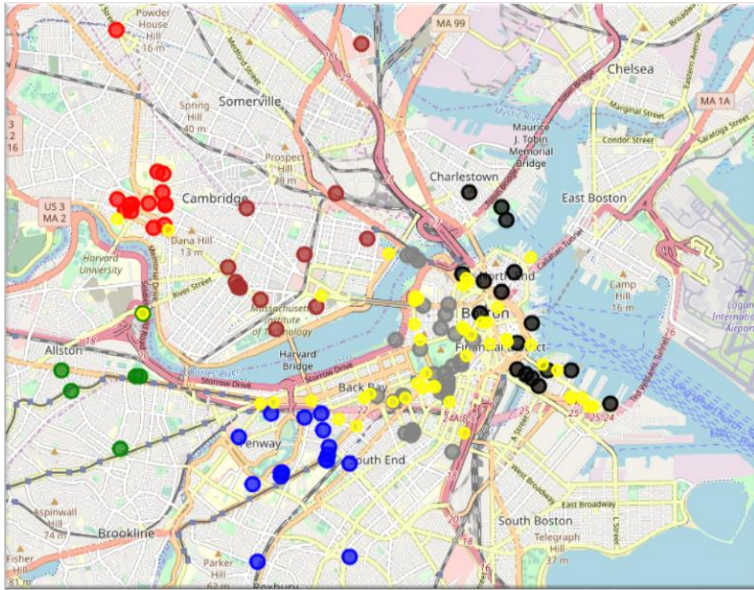
Results and Discussion

- Scenario 3:User Selected Shops

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

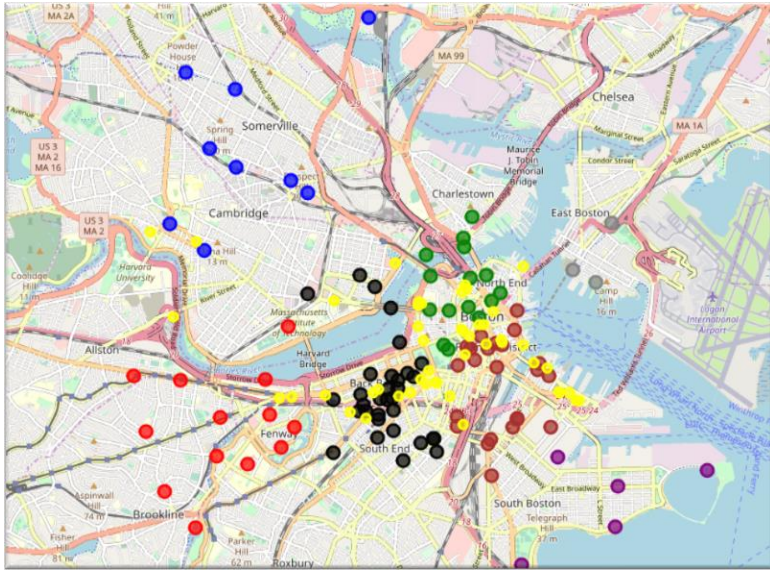|	name	categories	lat	lng	ZipCode
0	Boston Public Market	Market	42.361950	-71.057466	02108
1	Roche Bros Downtown Crossing	Market	42.355363	-71.060117	02110
2	Newbury Comics	Record Shop	42.360643	-71.054413	02109
3	Whole Foods Market	Grocery Store	42.361732	-71.065874	02114
4	Primark	Clothing Store	42.355538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

Results and Discussion

- Scenario 4:User Selected Arts

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

	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

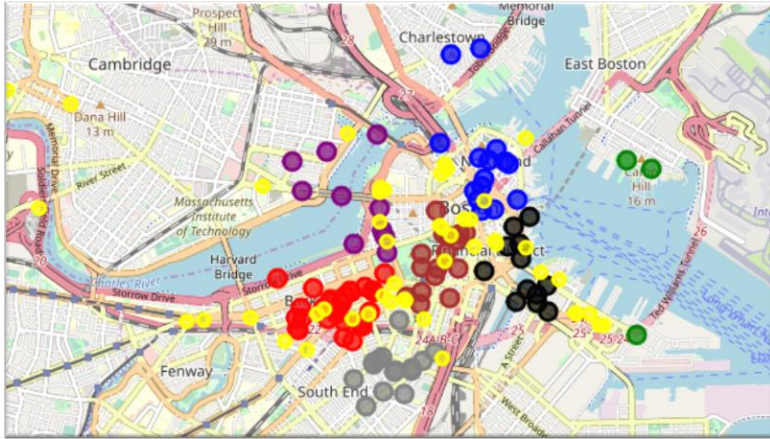|	name	categories	lat	lng	ZipCode
0	Boston Common	Park	42.355487	-71.064882	02111
1	The Rose Kennedy Greenway	Park	42.358227	-71.052055	02110
2	North End Park	Park	42.362488	-71.056477	02113
3	Barry's Bootcamp Boston	Gym / Fitness Center	42.354010	-71.059776	02111
4	Equinox Sports Club Boston	Gym	42.353189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

Results and Discussion

- Scenario 5:User Selected Outdoors

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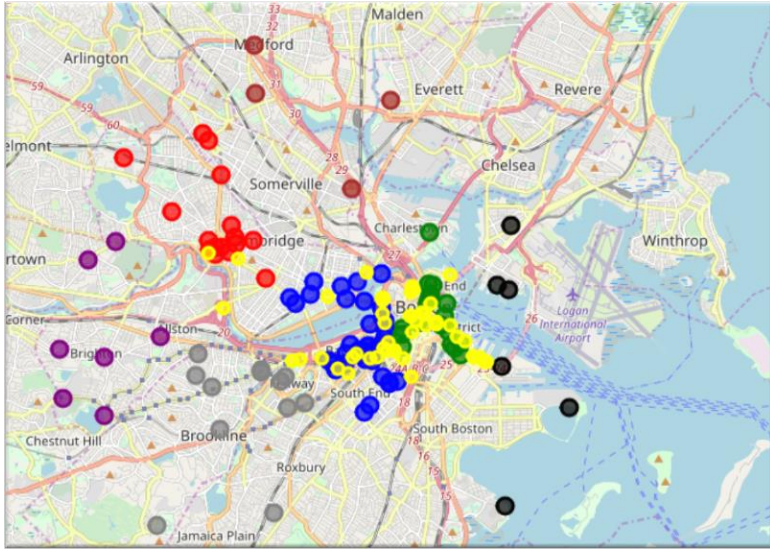
Results and Discussion

- Scenario 6:User Selected 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.

	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

Results and Discussion

- Scenario 7:User Selected 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.

	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

Conclusion

- Analyzing Foursqure Venues Information

- Using K-Mean clustering for hotel recommendation application

- Data Visualization helps user to select a hotel that fits better to his/her preferences.