**CAPSTONE PROJECT: BATTLE OF NEIGHBOURHOODS**

**1.INTRODUCTION/BUSINESS PROBLEM:**

**1.1 BACKGROUND:**

Food, irrespective of the place, is loved by people of all races and religion. It is something considerably basic needed for survival and hence starting a business-like restaurant or investing in one which satisfies that need can be lucrative if done correct.

**1.2 PROBLEM:**

Ideally, for a restaurant to be a successful one, it must grab a lot of people’s attention, entice them to enter and assuming that the service is good eventually it would become a successful one. Hence the location is the crux of the business. So before opening a restaurant there are few factors to be considered like the location, the cuisine type, the locality, and the local competition, permits, expenses like construction, salary etc. Few things that can be solved using data science are the location, the cuisine type and in which borough to start.

For the sake of the capstone project, I have selected a US state which is famous for the prestigious Universities located here. I have selected the Massachusetts state, where millions of students from almost all-around the world come here and pursue higher studies and jobs. Since it is a global hub for students, we can expect people from different countries having various preferences in the cuisine type to be living there.

**1.3 OBJECTIVE**

The goal of the project will be to identify the city, the neighborhood suitable for the restaurant and the type of restaurant

**2. DATA:**

**2.1 DATA SOURCES**

1. A csv file containing the zip, zip code name, city, state, county name of Massachusetts can be downloaded from the following link.

<https://www.downloadexcelfiles.com/us_en/download-list-us-zip-codes-massachusetts-state#.XukY8WgzZPY>

2. For the venue details we can use the foursquare api

<https://foursquare.com/developers/apps>

3. For the latitude and longitude data we use a library name pygeocode,

Input: a list of pin code (5 digits)

Output: A data frame with the following columns: - postal code, country code, place name, state name, accuracy, county code, latitude and longitude, community name and community code

4. After targeting which city we are going to focus on, a csv file containing the neighborhood details was created manually using the help of the website listed below:

<https://www.unitedstateszipcodes.org/>

Both the csv files are attached in the GitHub repository

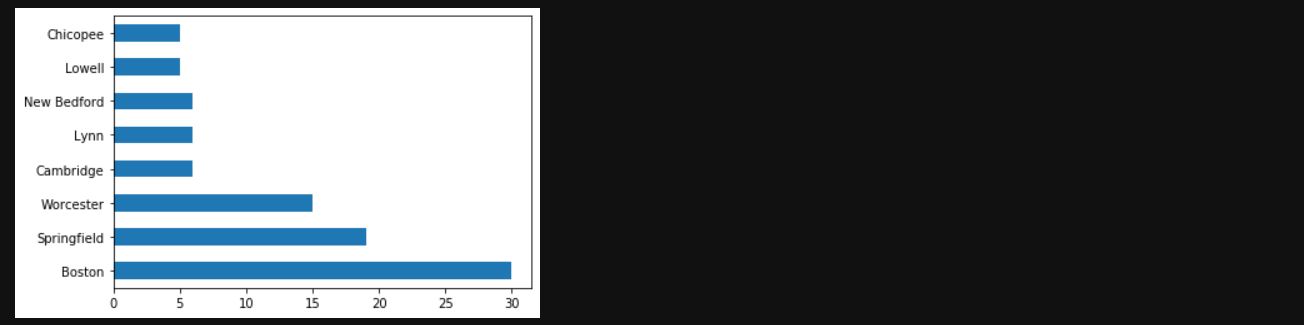
**2.2 DATA PREPROCESSING**

* Firstly, the data from the US\_zip.csv file is loaded into the notebook
* To get the location co-ordinates for each zip-codes, we use pygeocode library
* The zip codes must be passed as a 5-digit character for the library to process them and fetch the co-ordinates. So, we perform string operations in it.
* The final output is data frame with the following columns: postal code, country code, place name, state name, accuracy, county code, latitude and longitude, community name and community code.

**3. EXPLORATORY DATA ANALYSIS**

**3.1 SELECTION OF CITY**

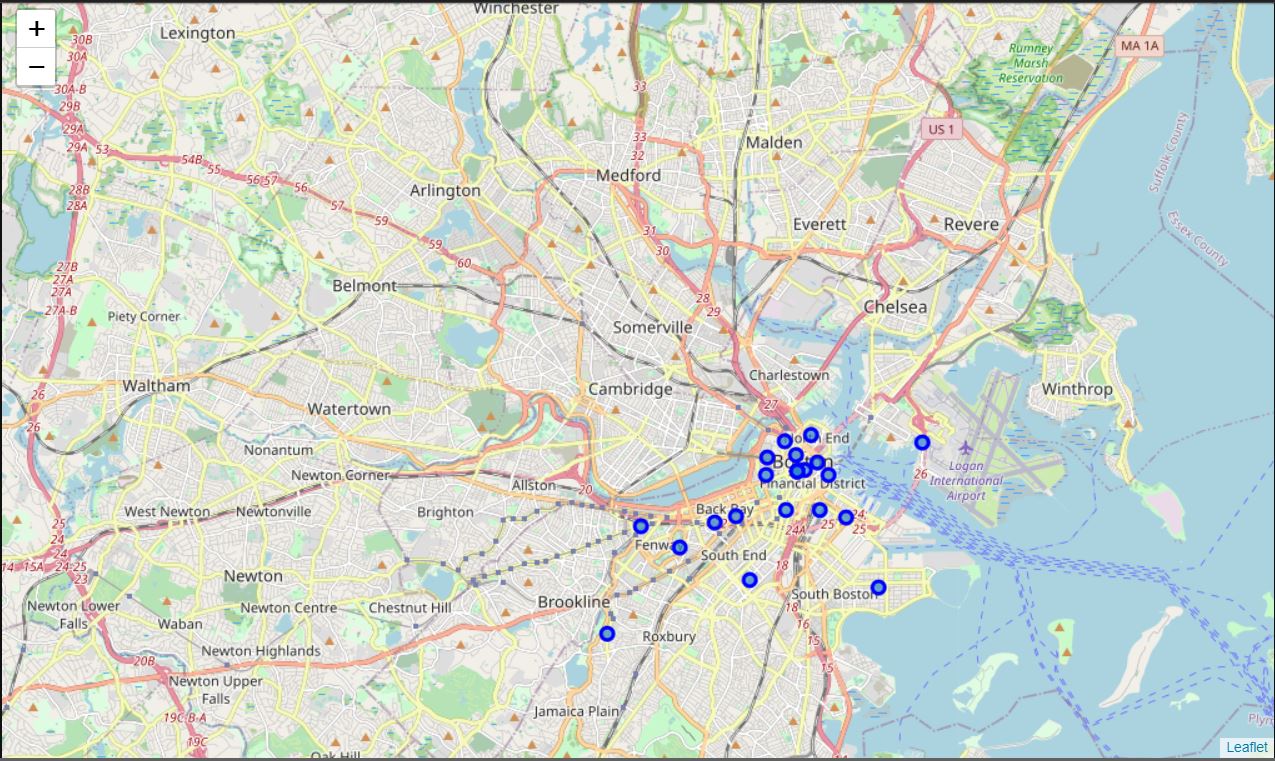
This section address to our first objective where to start a restaurant in Massachusetts. Normally it is better to start in a big city that has highest population, because when it comes to restaurants the more people the better outcome. We can find the largest city by plotting the number of neighborhoods each city has. A bar chart displaying the top 8 cities in Massachusetts with highest neighborhood counts



**Figure 3.1 Neighborhood Count**

From the chart above, it is obvious that Boston is the largest city among them with almost 30 neighborhoods. So, I’ll be concentrating on the city in this project and if a investor or client is looking to open a restaurant in Massachusetts, Boston is one fine place to begin, provided that MIT, Harvard, NEU and Boston University are located there. Hence you can find a lot of people from various places of the world living there.

So, from the data frame, another data frame that contains the details of Boston city is created. A csv file that contains the neighborhood details with their zip codes was manually created since it was not readily available in the internet. The data frame that was newly created is updated with the neighborhood details.



**Figure 3.2 Neighborhoods In Boston**

The above map helps us to visualize the various neighborhoods in Boston city.

**3.2 ANALYSING EACH NEIGHBORHOOD**

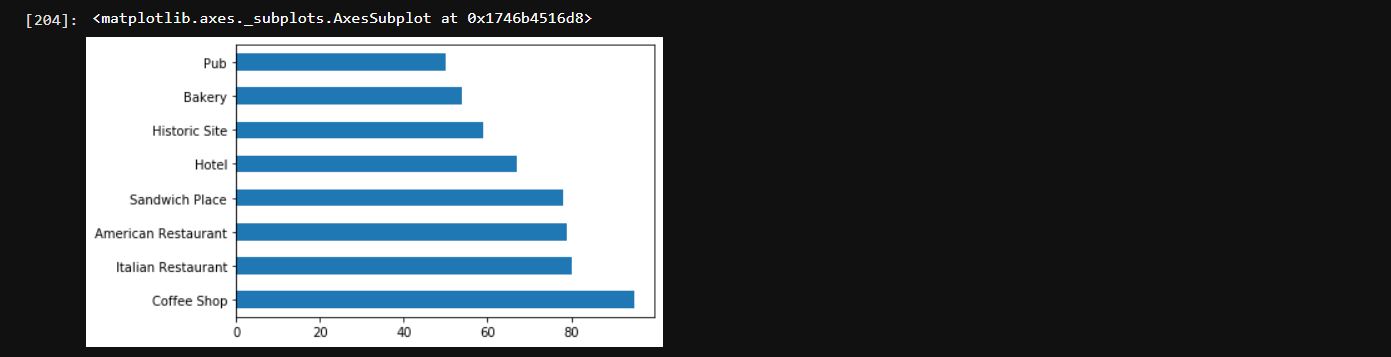
Once we have a narrowed it down to 29 neighborhoods, we must get the details of the venues in those neighborhoods using the Foursquare API. Using this API, we get a data frame that contains the neighborhood name, venue name, the category of the venue, location details and zip codes. Using the data, we can find out the famous venues in each neighborhood.



**Figure 3.3 Data frame Containing The Famous Venues In Each Neighborhood**

**3.3 SELECTING THE TYPE OF RESTAURANT**

Then we can search for what kind of restaurant is most common in Boston by plotting their numbers.



**Figure 3.4 Venue Counts**

Thus, from the plot we can understand there are more coffee shops in Boston than any other type of restaurants with around 100 establishments in the city. Hence it is recommended that the client better plan to open a café.

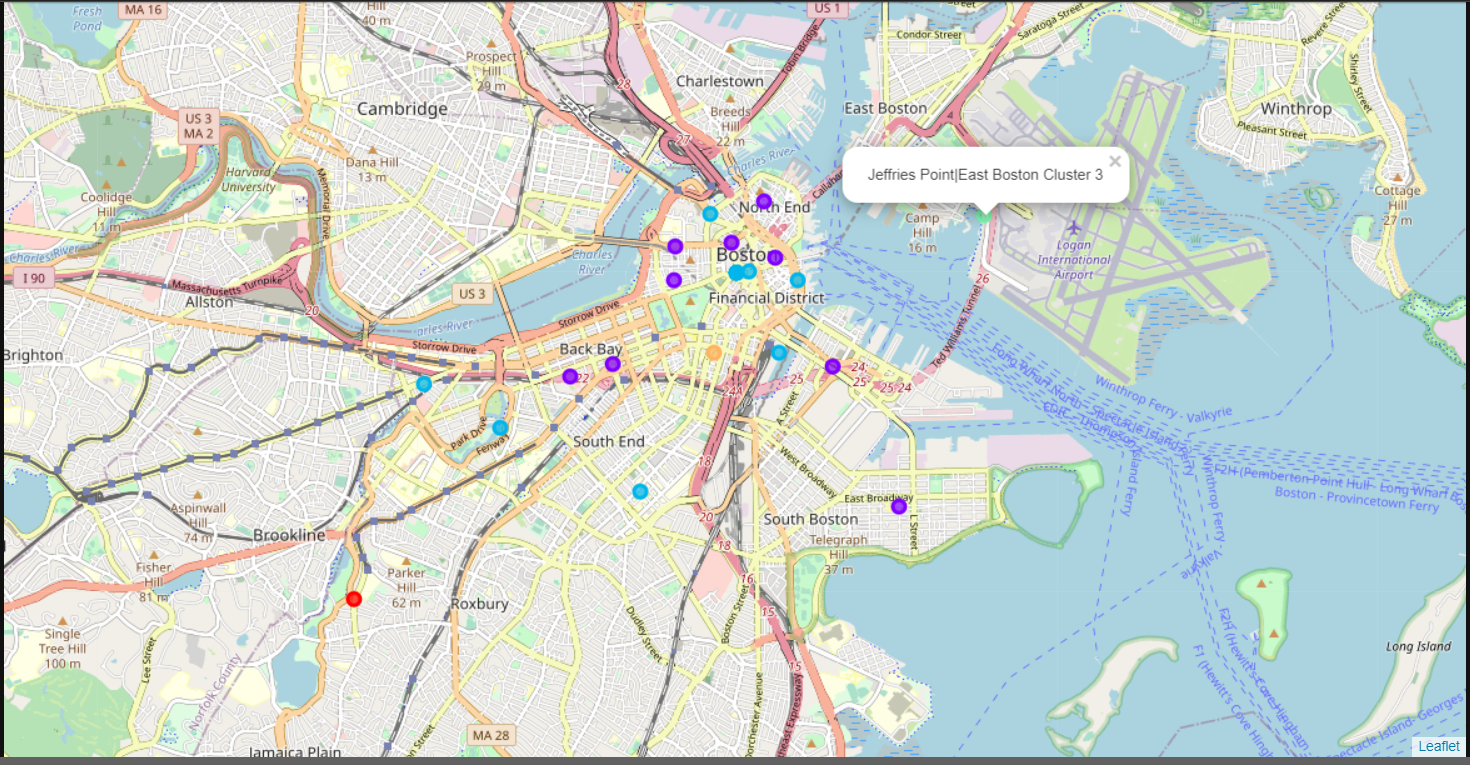
**4. CLUSTER GENERATION**

To cluster the neighborhood based on the categories of venues we use K-means clustering to produce 5 clusters. The resulting data frame assigns each neighborhood a cluster along with their most famous neighborhoods



**Figure 4.1 Clustered Neighborhoods**

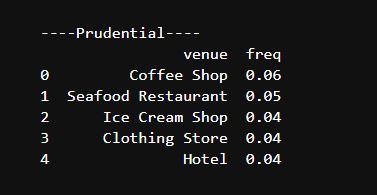
Later these clusters are plotted which could help us in determining the neighborhood for establishing a coffee shop.



**Figure 4.2 Neighborhood Clusters Plotted Over Boston Map**

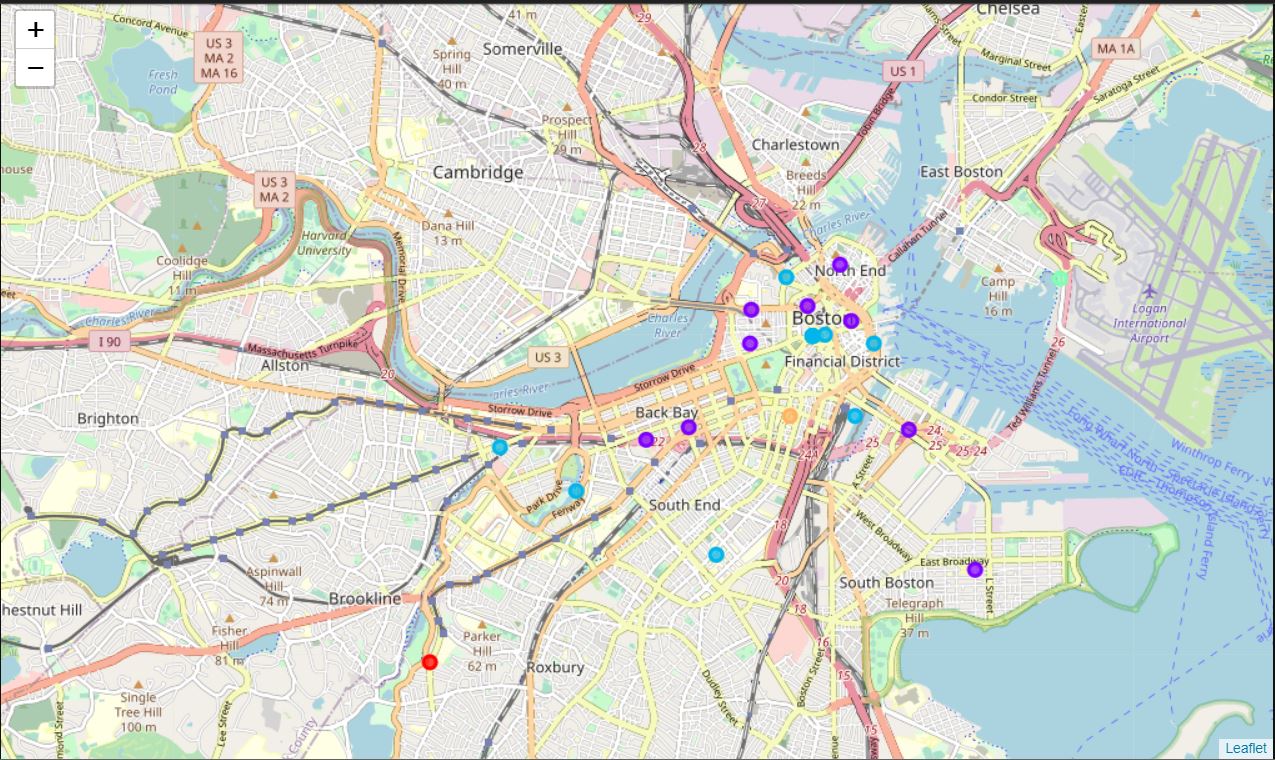
**5. RESULT:**

Now all we must do is to pinpoint the exact location or neighborhood that would be suitable for our café. By finding out the common venues in each neighborhood we can solve this problem.

**Figure 5.1 Common Venues In Prudential Figure 5.2 Common Venues In Back Bay**

Back bay and Prudential both neighborhoods seem to be a perfect place to start a café.

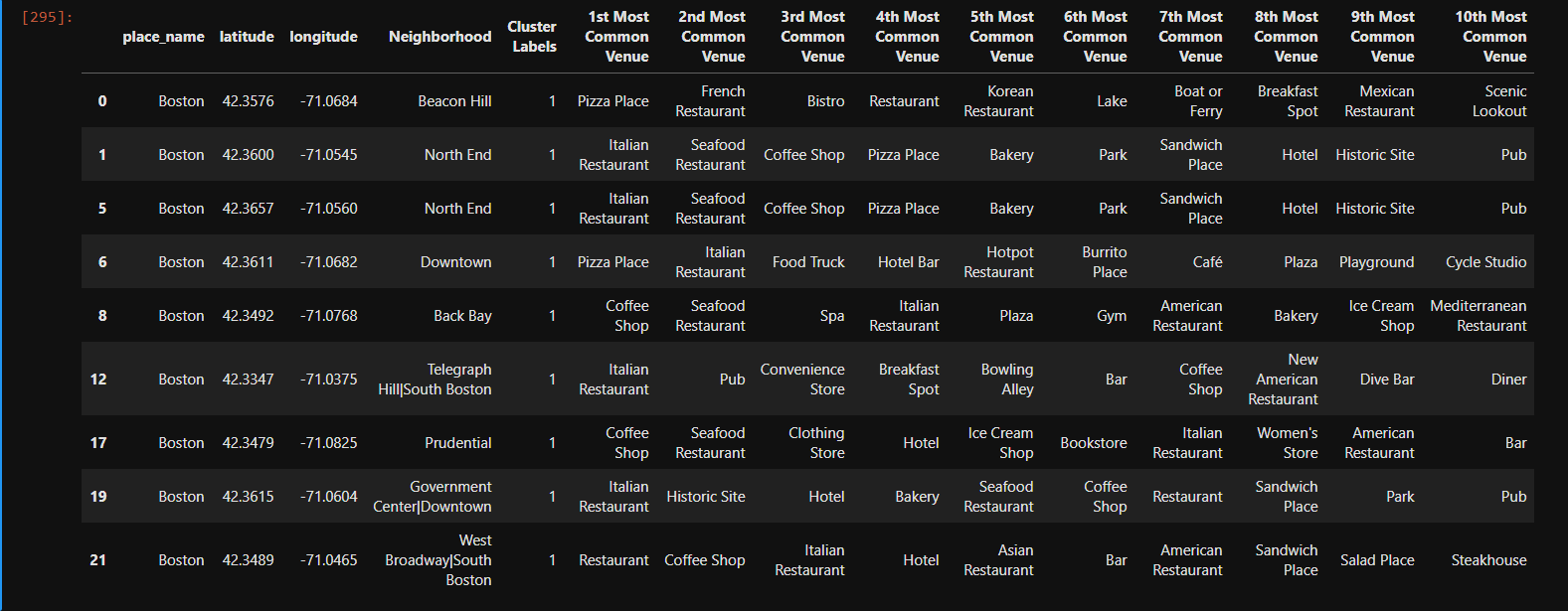


**Figure 5.3 Favorable Neigborhoods**

Both the neighborhoods belong to the same cluster where coffee shops are quite famous and are near to each other. Hence opening a café in either of these neighborhoods in Boston is recommended.

**6. DISCUSSION:**

It is quite possible that both the neighborhoods may be saturated with too many coffee shops, in that case opening another shop is not such a good idea. But based on observation, coffee shops make it to the top 6 most common venues in all those neighborhoods of that cluster (cluster=1). Even if not in Back bay or Prudential neighborhood, it is fine to open a coffee shop in any of those neighborhoods that fall under that cluster.



**Figure 6.1 Cluster 1 Common Venues**

It is also noted that Italian restaurants are also quite perform well in this cluster so Italian restaurants can be considered as second option for this cluster.

**7. CONCLUSION**

This study assumes that more the number of a particular type of restaurant the more popular it is. But there are lot of other factors that is involved in the success of a restaurant. Factors like customer critics, theme of the restaurant, ease of obtaining permits, cost to run a restaurant have not been considered for this study. This study only recommends the type of restaurant and the neighborhoods where it runs well.