

Capstone Project – The Battle of Neighbourhood's

Author: Chandan Kumar Reddy

Data: 30-12-2018

Problem Statement

Mumbai is a financial capital of India. So people come from various parts of the country and across the globe. So within growing population in Mumbai there will be a huge demand for any business to setup. So the problem is to find which business is suitable for which neighbourhood and which neighbourhoods are similarly related each other in terms of taste, life, habits etc. Knowing the people's demand across the neighbourhoods we can able to suggest which business would be better to set up, so that they decrease the risks and earns profits only at the certain locations where it is people are liking.

In this Project we will cluster the Neighbourhoods, group them and find similarities and relationship between the neighbourhoods.

Target audience: -

Potential clients looks to start a business Mumbai but not aware which business to setup due to lack of knowledge and volatile market conditions.

Stakeholders: -

All small and medium scale businesses.

Data Section :

Following sources of data are used while executing the Capstone Project: -

Data title: - Website

Postal Codes of Mumbai, India from scraping the website mapsofindia.

Type of data: -

Dataset in form of html file converted into data frame and saved in CS format

Duration: -

December 2018

Description of the dataset: -

Data is driven from scraping the website maps of India, Where we got Postal codes and Neighbourhoods of Mumbai city, India. The web site is last updated date was 30th November 2015.

Source: -

<https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/>

Data title: - Here Maps Geocoding API

Type of data: -

JSON

Duration: - N/A

Description of the data: -

Location coordinates obtained by Here Rest API calls.

Location Information obtained from the data frame is used to obtain the location coordinates from Here Maps API.

A separate Python script has been developed to extract the unique street names, district names from Here Rest API's using geocoder.

Source: -

Here

Data title: -

Foursquare location data

Type of data: -

JSON

Duration: -

N/A

Description of the data: -

Location coordinates obtained by Foursquare API calls.

To determine the proximity of various amenities as per the client's requirement, Foursquare location data is used.

Source: -

<https://foursquare.com/>

Methodology

In this Project, we will convert addresses into their equivalent latitude and longitude values. Also, you will use the Foursquare API to explore neighbourhoods in Mumbai. We will use the explore function to get the most common venue categories in each neighbourhood, and then use this feature to group the neighbourhoods into clusters. We will use the k-means clustering algorithm to complete this task. Finally, we will use the Folium library to visualize the neighbourhoods in Mumbai and their emerging clusters.

First we will import all required libraries and install packages required.

1. Download and Explore Dataset
2. Load and explore the data
3. Transform the data into a pandas Dataframe
4. Using Geopy library to get the latitude and longitude values of Mumbai City.
5. Create a map of Mumbai with neighbourhoods superimposed on top.
6. Defining Foursquare Credentials and Version.
7. Exploring the first neighbourhood in our dataframe.
8. Get the top 10 venues that are in Mumbai within a radius of 500 meters.
9. Send the GET request and examine the results.
10. Explore Neighbourhoods in Mumbai.
11. Checking the size of the resulting dataframe.
12. Finding out how many unique categories can be curated from all the returned venues.
13. Analysing Each Neighbourhood.
14. Group the rows by neighbourhood and by taking the mean of the frequency of occurrence of each category.
15. Create each neighbourhood along with the top 5 most common venues and put it into a pandas dataframe.
16. Cluster Neighbourhoods and run k-means to cluster the neighbourhood into 5 clusters.
17. Visualize the resulting clusters
18. Finally Examine Clusters.

Results

Data driven from by scraping the website and copied to data frame.

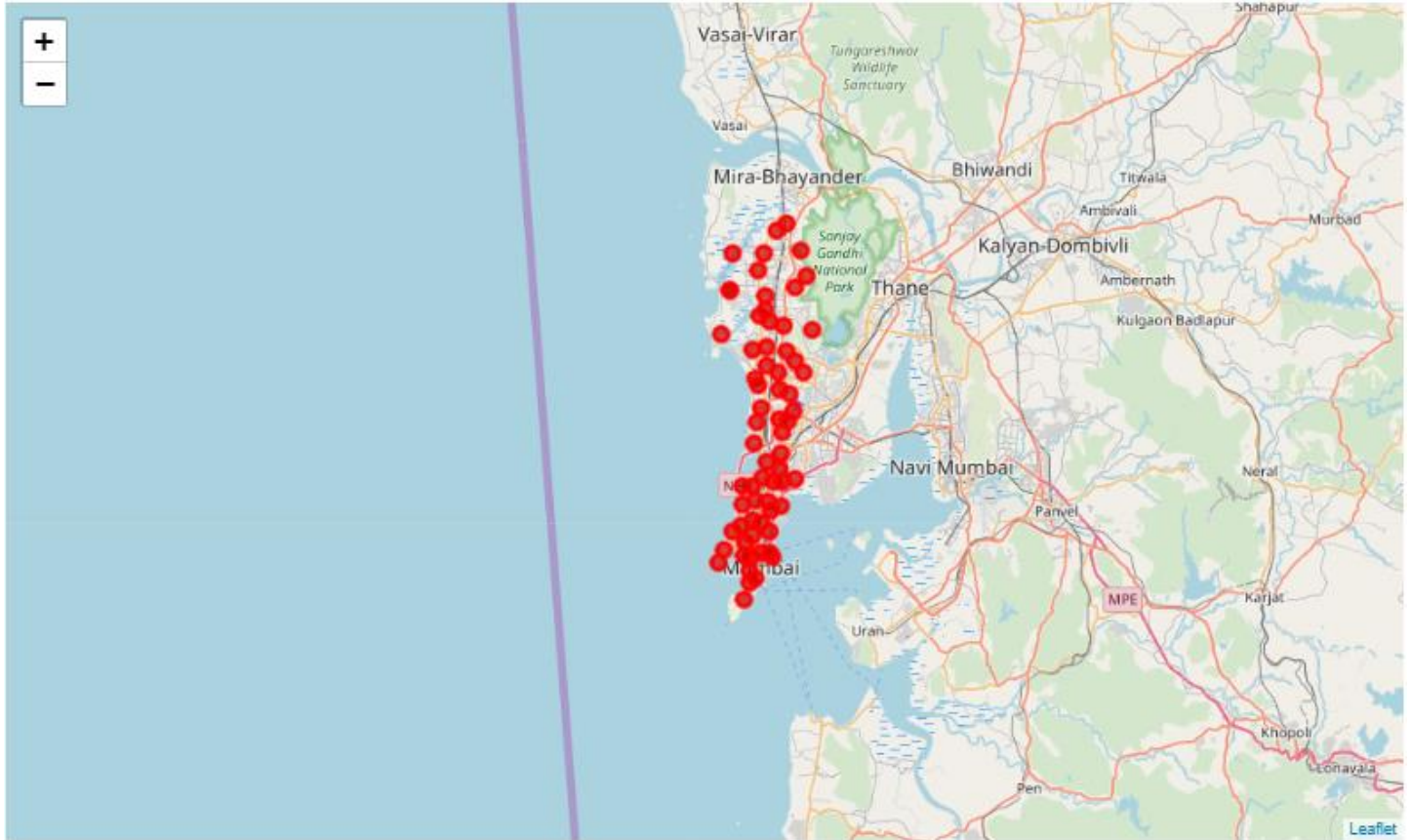
	Location	Pincode	State	District
0	Location	Pincode	State	District
1	A I staff colony	400029	Maharashtra	Mumbai
2	Aareymilk Colony	400065	Maharashtra	Mumbai
3	Agripada	400011	Maharashtra	Mumbai
4	Airport	400099	Maharashtra	Mumbai

	Pincode	Location
0	400001	Bazargate, M.P.t., Mumbai., Stock Exchange, Ta...
1	400002	Kalbadevi, Ramwadi, S. c. court, Shroff Mahaja...
2	400003	B P t colony, B.P.lane, Mandvi, Masjid, Noor B...
3	400004	Ambewadi, Charni Road, Chaupati, Girgaon, Madh...
4	400005	Asvini, Colaba, Holiday Camp, V.W.t.c.

Using Here API fetching all the co-ordinates of the neighbourhoods.

	Pincode	Location	Latitude	Longitude
0	400001	Bazargate, M.P.t., Mumbai., Stock Exchange, Ta...	18.9479	72.8466
1	400002	Kalbadevi, Ramwadi, S. c. court, Shroff Mahaja...	18.9479	72.8262
2	400003	B P t colony, B.P.lane, Mandvi, Masjid, Noor B...	18.9522	72.8351
3	400004	Ambewadi, Charni Road, Chaupati, Girgaon, Madh...	18.9496	72.8191
4	400005	Asvini, Colaba, Holiday Camp, V.W.t.c.	18.9086	72.8183

Visualizing the data



Getting the top five venues frequency from all the Neighbourhoods.

----A I staff colony, Santacruz P&t colony----

```

venue freq
0 Chinese Restaurant 0.50
1 Burger Joint 0.25
2 Asian Restaurant 0.25
3 Middle Eastern Restaurant 0.00
4 Mobile Phone Shop 0.00

```

----Aareymilk Colony, Nagari Niwara----

```

venue freq
0 Hotel 0.14
1 Resort 0.14
2 Gym / Fitness Center 0.14
3 Café 0.14
4 Golf Course 0.14

```

----Agripada, Chinchpokli, Haines Road, Jacob Circle----

```

venue freq
0 Tea Room 0.17
1 Gym 0.17
2 Indian Restaurant 0.17
3 Coffee Shop 0.17
4 Restaurant 0.17

```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A I staff colony, Santacruz P&t colony	Chinese Restaurant	Burger Joint	Asian Restaurant	Yoga Studio	Donut Shop	Field	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant
1	Aareymilk Colony, Nagari Niwara	Gym / Fitness Center	Farm	Golf Course	Hotel	Restaurant	Café	Resort	Diner	Farmers Market	Falafel Restaurant
2	Agripada, Chinchpokli, Haines Road, Jacob Circle	Gym	Indian Restaurant	Coffee Shop	Restaurant	Tea Room	Racetrack	Farmers Market	Farm	Falafel Restaurant	Electronics Store
3	Airport, International Airport, Sahar P & t co...	Airport	Business Service	Yoga Studio	Donut Shop	Field	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Electronics Store
4	Ambewadi, Charni Road, Chaupati, Girgaon, Madh...	Harbor / Marina	Restaurant	Gastropub	Indian Restaurant	Food Truck	Juice Bar	Pizza Place	Gym	Train Station	Aquarium

Applying K-means to for clustering the data

```
# set number of clusters
kclusters = 10

mumbai_grouped_clustering = mumbai_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(mumbai_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_

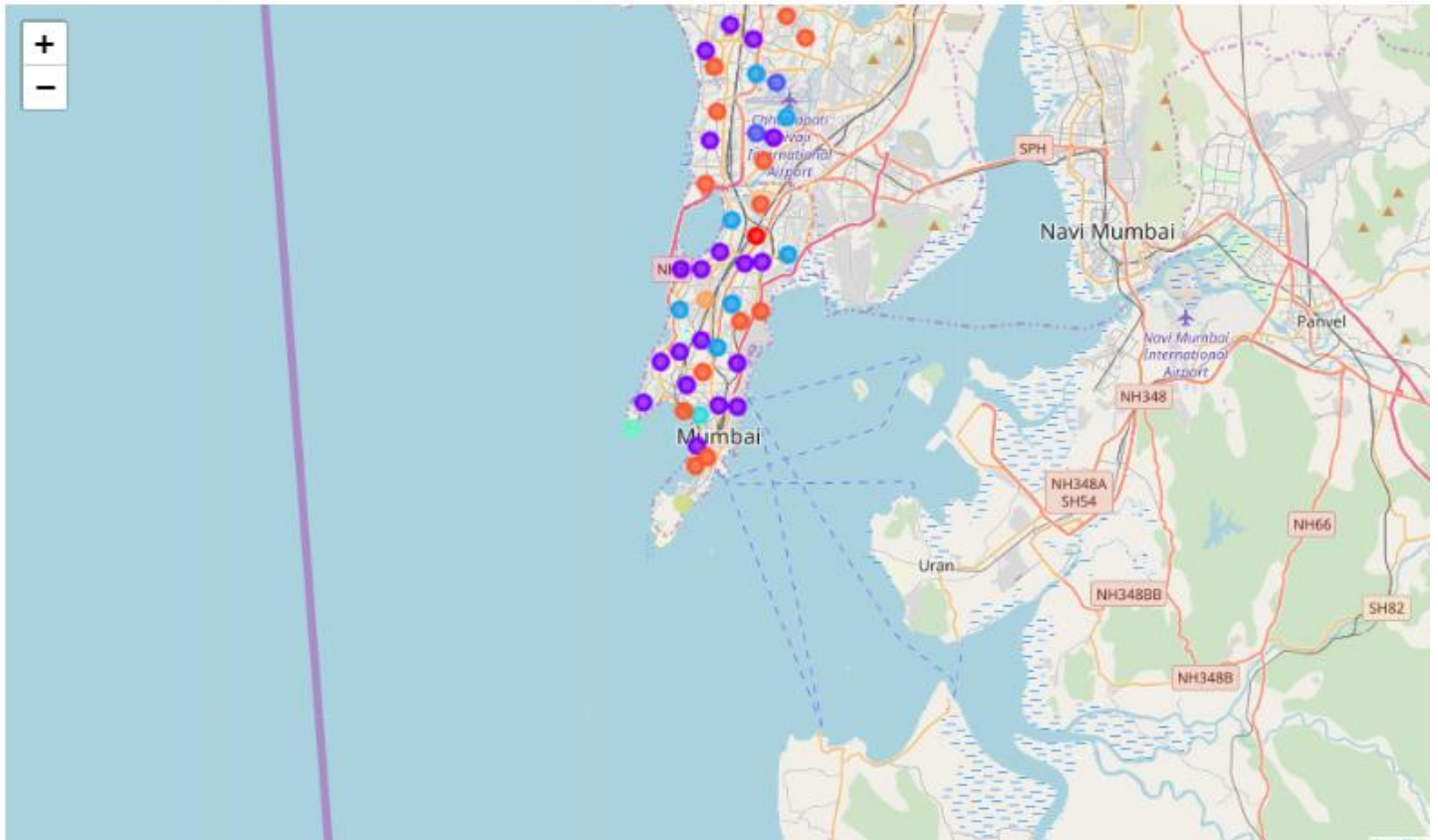
array([4, 1, 9, 7, 1, 1, 9, 1, 1, 1, 3, 8, 1, 9, 3, 9, 3, 0, 1, 9, 1, 1, 3,
       1, 3, 1, 1, 9, 9, 1, 5, 3, 1, 9, 9, 1, 3, 9, 2, 9, 3, 1, 9, 1, 9, 6,
       9, 1, 1, 1, 1, 9, 1, 1, 9, 4, 1, 1, 2, 1, 3, 9, 1], dtype=int32)
```

	Pincode	Location	Latitude	Longitude	Cluster Labels
1	400002	Kalbadevi, Ramwadi, S. c. court, Shroff Mahaja...	18.9479	72.8262	1
2	400003	B P t colony, B.P.lane, Mandvi, Masjid, Noor B...	18.9522	72.8351	3
3	400004	Ambewadi, Charni Road, Chaupati, Girgaon, Madh...	18.9496	72.8191	3
4	400005	Asvini, Colaba, Holiday Camp, V.W.t.c.	18.9086	72.8183	3
5	400006	Malabar Hill	18.9539	72.7995	3

Combining all the data into one dataframe.

	Pincode	Location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	400002	Kalbadevi, Ramwadi, S. c. court, Shroff Mahaja...	18.9479	72.8262	4	Indian Restaurant	Café	Chinese Restaurant	Bar	Cheese Shop	Snack Place	Train Station	Department Store
2	400003	B P t colony, B.P.lane, Mandvi, Masjid, Noor B...	18.9522	72.8351	1	Indian Restaurant	Convenience Store	Electronics Store	Restaurant	Rest Area	Café	Dessert Shop	Middle Eastern Restaurant
3	400004	Ambewadi, Charni Road, Chaupati, Girgaon, Madh...	18.9496	72.8191	9	Harbor / Marina	Restaurant	Gastropub	Indian Restaurant	Food Truck	Juice Bar	Pizza Place	Gym
4	400005	Asvini, Colaba, Holiday Camp, V.W.t.c.	18.9086	72.8183	7	Park	Chinese Restaurant	Basketball Court	Restaurant	Department Store	Asian Restaurant	Ice Cream Shop	Fast Food Restaurant
5	400006	Malabar Hill	18.9539	72.7995	1	Gym	Convenience Store	Restaurant	Dessert Shop	Park	Comfort Food Restaurant	Fast Food Restaurant	Farmers Market

Visualize the clustered data



5. Discussion

Based on analyses and visualization we can check the top 5 venues for each neighbourhood and can see that in which venue category's people are tending to spend their free time.

By analysing these Neighbourhoods, Business can know which will be probably best place to setup their business.

Conclusion

The decision of a buyer is influenced by current demand in the market. So, based upon the findings summarized in the results and discussion sections, following conclusions can be made: -

1. While making recommendations to a prospective client, it is imperative to know his/ her immediate needs and requirements besides the budget. This would help to catch his/ her attention.
2. Knowledge about the most recent market prices can be very helpful for the client and can help him take a decision.