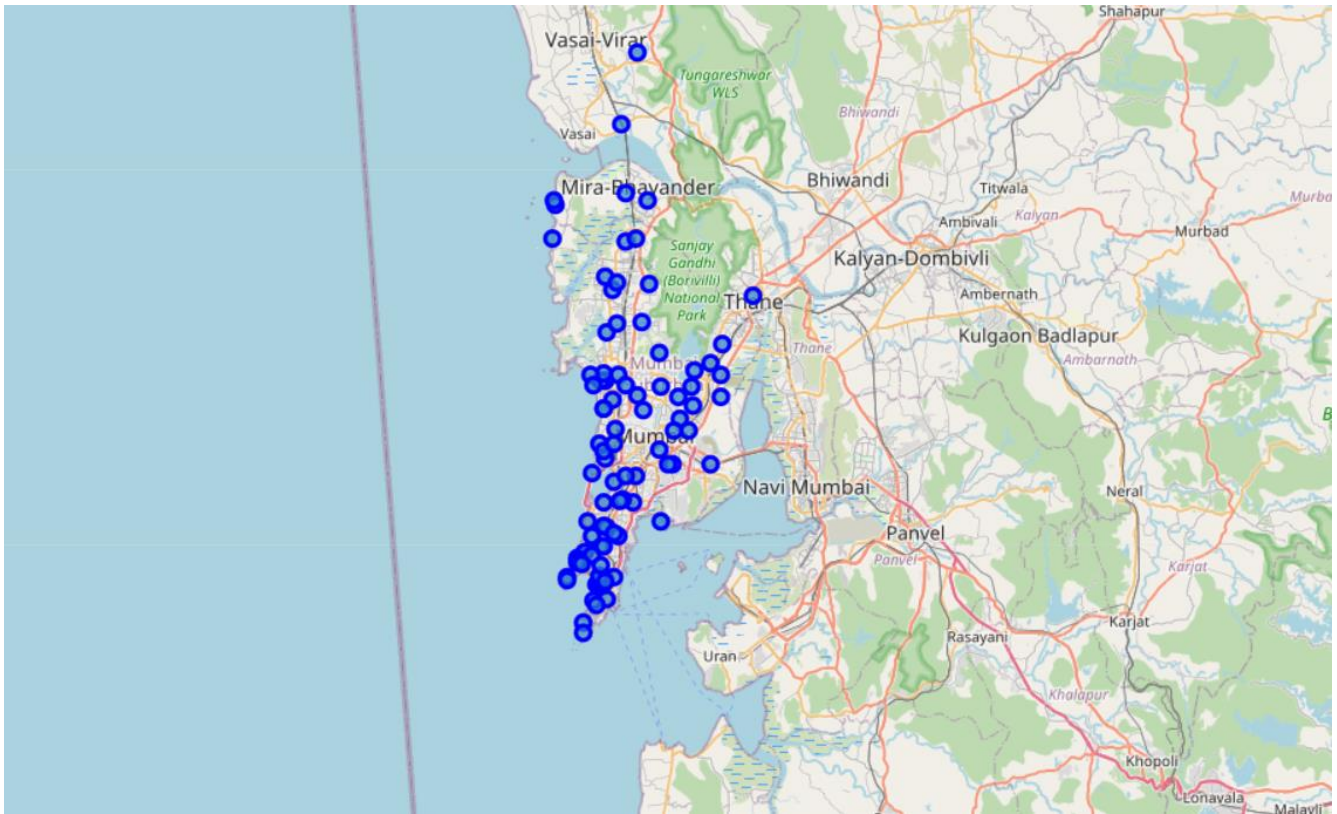# IBM Data Science Capstone Project

The Battle of Neighborhoods – Mumbai

By : Lucky Rathod

January 2021

Neighborhoods Visualization of Mumbai City

1.INTRODUCTION : Business Problem

This project deals with discussing the neighborhoods of **Mumbai**, The Financial Hub of India.This project would specifically help Business people planning to start Restaurants, Hotels, Gym/FitnessCamp/Yoga studio etc.in Mumbai, Maharashtra, India.

Mumbai, formerly known as Bombay, **Capital of Maharashtra state, India**. It is the country's financial and commercial centre and its principal port on the Arabian Sea. Located on Maharashtra's coast, Mumbai is India's most-populous city and it is one of the largest and most densely populated urban areas in the world. Mumbai is home for 1.84 Cr people. The total area of Mumbai is 603.4 sq.km. Average temperature of Mumbai is 26.9°C.

Mumbai - City that never Sleeps

The Foursquare API is used to access the venues in the neighborhoods. Since, it returns less venues in the neighborhoods, we would be analysing areas for which countable number of venues are obtained. Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task. The optimal number of clusters can be obtained using silhouette score.

Folium visualization library can be used to visualize the clusters superimposed on the map of Mumbai city. These clusters can be analyzed to help small scale business owners select a suitable location for their need such as Hotels, Shopping Malls, Restaurants or even specifically Indian restaurants or Coffee shops or Gym/Fitness camp/Yoga studio.

The major **Target Audience** would be small-scale business owners and stake holders planning to start their business at a location in Mumbai. This project would help them find the optimal location based on the category of their business such as,

- What is the best location to start a new hotel in Mumbai with restaurants around?
- Which area is best suitable for opening a Gym/Yoga/Fitness Camp in Mumbai?

2.DATA Requirements

Mumbai has multiple neighborhoods. Lets use the following dataset which has the list of Neighborhoods in Mumbai along with their Latitude and Longitude:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai#Mumbai_neighbourhood_coordintes

This Wikipedia page contains a table named "Mumbai neighbourhood coordinates". We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages.

We will use **Pandas read_html** library to Scrape data from Wikipedia page.

	Neighbourhood	Location	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.129300	72.843400
1	Chakala, Andheri	Western Suburbs	19.111388	72.860833
2	D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373
3	Four Bungalows	Andheri,Western Suburbs	19.124714	72.827210
4	Lokhandwala	Andheri,Western Suburbs	19.130815	72.829270
5	Marol	Andheri,Western Suburbs	19.119219	72.882743
6	Sahar	Andheri,Western Suburbs	19.098889	72.867222
7	Seven Bungalows	Andheri,Western Suburbs	19.129052	72.817018
8	Versova	Andheri,Western Suburbs	19.120000	72.820000
9	Mira Road	Mira-Bhayandar,Western Suburbs	19.284167	72.871111
10	Bhayandar	Mira-Bhayandar,Western Suburbs	19.290000	72.850000

Fig 1 : Mumbai Neighborhood Data with Latitude and Longitude

Next the details of venues in each neighborhood namely **Venue, Venue Latitude, Venue Longitude, Venue Category** data needs to be obtained. Here, Foursquare API is used to obtain this data.

https://foursquare.com/

A total of 1365 venues data have been obtained from Foursquare. The resultant venues dataset, (shown in Fig 2) is used for the analysis process.

```
venues_df.head()
```

(1365, 8)

	Neighbourhood	Location	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Amboli	Andheri,Western Suburbs	19.1293	72.8434	Cafe Arfa	19.128930	72.847140	Indian Restaurant
1	Amboli	Andheri,Western Suburbs	19.1293	72.8434	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
2	Amboli	Andheri,Western Suburbs	19.1293	72.8434	Subway	19.127860	72.844461	Sandwich Place
3	Amboli	Andheri,Western Suburbs	19.1293	72.8434	Cafe Coffee Day	19.127748	72.844663	Coffee Shop
4	Amboli	Andheri,Western Suburbs	19.1293	72.8434	Spices & Chillies	19.127765	72.844131	Asian Restaurant

Fig2 : Mumbai Venues Dataset

3.METHODOLOGY

Now, we have the neighborhoods data of Mumbai **(93 neighborhoods).** We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of **1365 venues** have been obtained in the whole city and **181 unique categories**. But as seen we have multiple neighborhoods with less than 10 venues returned. In order to create a good analysis let's consider only the neighborhoods with more than 10 venues.
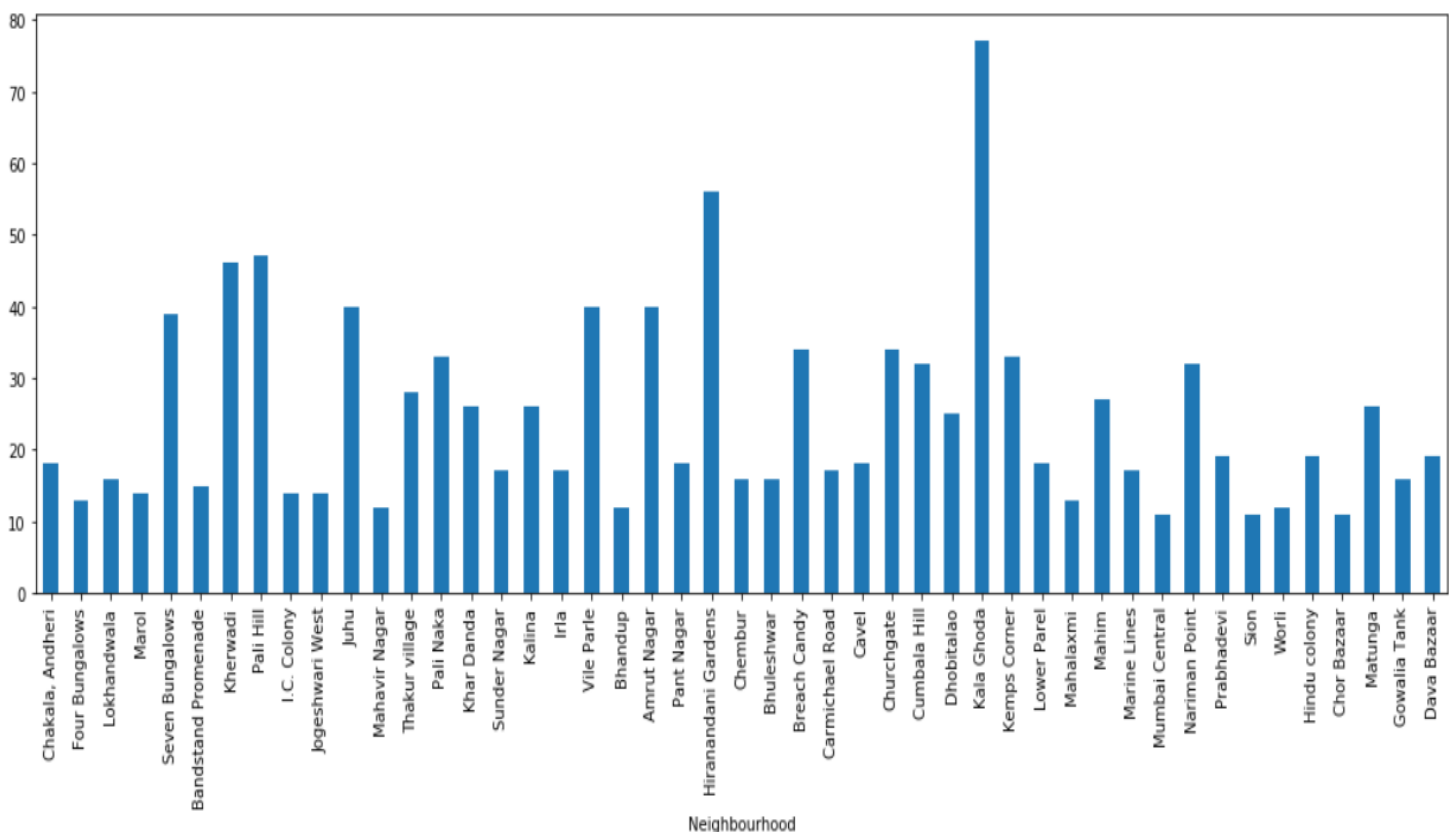
We can perform one hot encoding on the obtained data set and use it find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. Here **K - Nearest Neighbor** clustering technique have been used. To find the optimal number of **clusters silhouette score** metric technique is used.

The clusters obtained can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest business people, suitable locations based on the category.

4.ANALYSIS

Looking into the dataset we found that there were many neighborhoods with less than 10 venues which can be remove before performing the analysis to obtain better results. The following plot shows only the neighborhoods from which 10 or more than 10 venues were obtained. The resultant dataset consists of **47 neighborhoods.**

Fig3: Filtered Neighborhood Dataset

Next, we will perform one hot encoding on the filtered data to obtain the venue categories in each neighborhood. Then group the data by neighborhood and take the mean value of the frequency of occurrence of each category.Sample Output is shown below

	Neighbourhoods	Afghan Restaurant	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank
0	Agripada	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.200	0.0
1	Altamount Road	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.125	0.0
2	Amboli	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.125000	0.0	0.0	0.0	0.000	0.0
3	Amrut Nagar	0.025	0.025	0.0	0.0	0.0	0.000000	0.0	0.0	0.050000	0.0	0.0	0.0	0.000	0.0
4	Asalfa	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000	0.0
5	Ballard Estate	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000	0.0
6	Bandstand Promenade	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000	0.0
7	Bangur Nagar	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000	0.0
8	Bhandup	0.000	0.000	0.0	0.0	0.0	0.083333	0.0	0.0	0.083333	0.0	0.0	0.0	0.000	0.0
9	Bhayandar	0.000	0.000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000	0.0

Fig4 : Mean of frequrncy of occurrence of each category

The Above dataset is used to obtain the top 10 most common venues in each neighborhood i.e. the 10 venues with the highest mean of frequency of occurrence.A sample for first 5 Neighborhood is show in figure below

	Neighbourhood	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	Agripada	Restaurant	Gym	Coffee Shop	Indian Restaurant	Bakery	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space
1	Altamount Road	Café	Bakery	Fast Food Restaurant	Coffee Shop	Sandwich Place	Theater	Indian Restaurant	Yoga Studio	Farmers Market	Falafel Restaurant
2	Amboli	Halal Restaurant	Indian Restaurant	Coffee Shop	Chinese Restaurant	Sandwich Place	Fast Food Restaurant	Asian Restaurant	Park	Gourmet Shop	Greek Restaurant
3	Amrut Nagar	Indian Restaurant	Café	Clothing Store	Fast Food Restaurant	Electronics Store	Asian Restaurant	Afghan Restaurant	Bookstore	Sandwich Place	Bowling Alley
4	Asalfa	Food Truck	Campground	Yoga Studio	Dog Run	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Electronics Store

Fig5 : Ten most common venues in each Neighborhood

The resultant dataset can be used for the clustering algorithm. Here, the K-Nearest Neighbor (KNN) clustering algorithm is used. It is an unsupervised machine learning technique that clusters the given data into K number of clusters. For optimal result we need to select the best value for K. Here, the silhouette score is used to find the best value for K. A range of values from 2 to 10 was considered, KNN clustering was performed on the dataset and the silhouette score was calculated and plotted on a line plot as shown in **Fig6.** From the plot we can see that a K value of 6 provides the best score. This K value is used for the K-Means Clustering Technique. The K-Means labels obtained were included in the top neighborhoods dataset for examining the characteristics of each cluster.
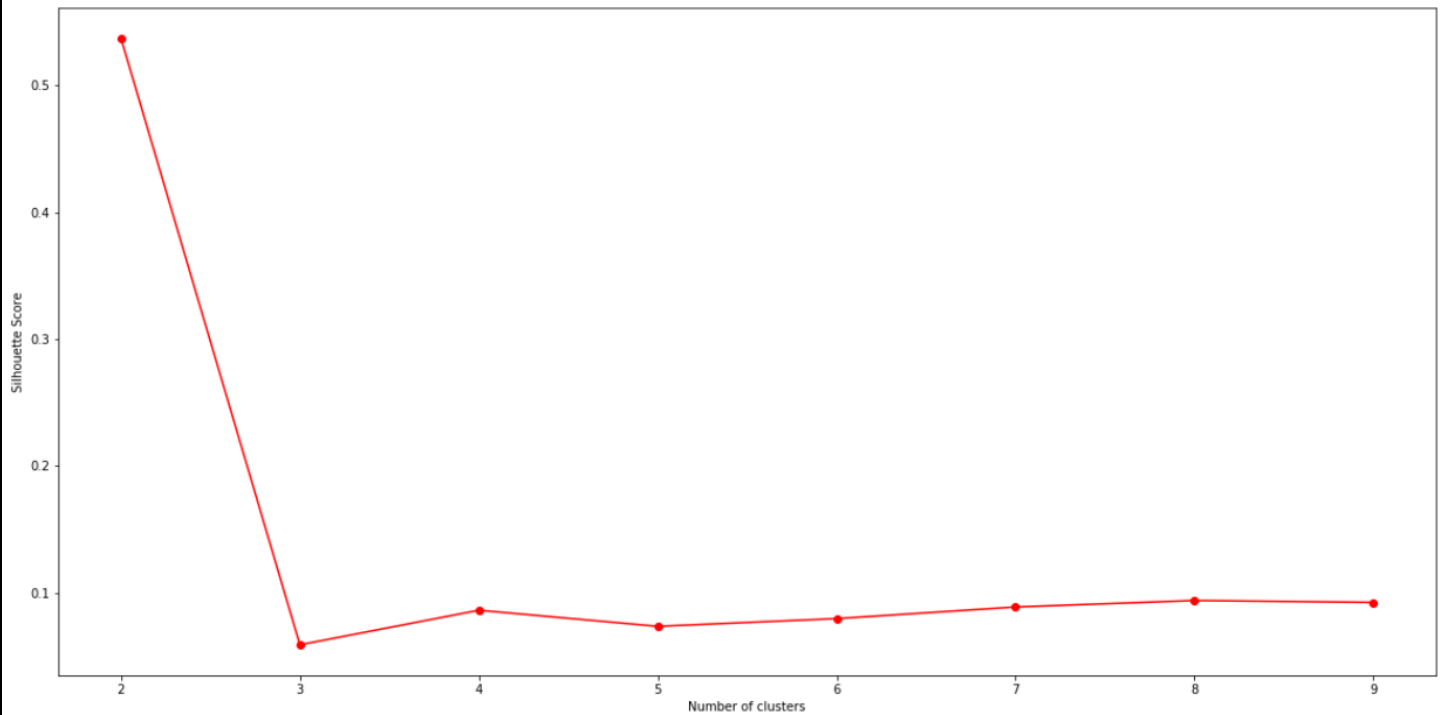
Fig6 : Silhouette Score for different Number of Clusters

The K-Means labels obtained were included in the top neighborhoods dataset for examining the characteristics of each cluster

5.RESULTS

Let's examine the 6 clusters and find the discriminating venue categories that distinguish each cluster. For this purpose, lets also look into the five most common venue category in each cluster.

5.1 Cluster 1 (Green Dots in Map)

	Neighbourhood	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
11	Uttan	Mira-Bhayandar,Western Suburbs	19.280000	72.785000	0.0	Beach	Playground	Indian Restaurant	Bus Station	Resort	Yoga Studio	Dog Run	Fast Food Restaurant
16	Gorai	Borivali (West),Western Suburbs	19.250057	72.782021	0.0	Resort	Seafood Restaurant	Aquarium	Indian Restaurant	Beach	Yoga Studio	Fast Food Restaurant	Farmers Market
64	Dongri	South Mumbai	19.283333	72.783333	0.0	Beach	Playground	Indian Restaurant	Bus Station	Resort	Yoga Studio	Dog Run	Fast Food Restaurant

Represented by Green dots in Map

Count of Each Venue in Cluster 1 :

```
Farmers Market        3
Fast Food Restaurant  3
Yoga Studio           3
Beach                 3
Falafel Restaurant    3
Resort                3
Indian Restaurant     3
Bus Station           2
Dog Run               2
Playground            2
dtype: int64
```

5.2 Cluster 2 (Yellow Dot in Map)

Neighbourhood	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	9th Most Common Venue	10th Most Common Venue
Cuffe Parade	South Mumbai	18.91	72.81	1.0	Garden	Yoga Studio	Dog Run	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Electronics Store	Dumpling Restaurant

Count of Each Venue in Cluster 2 :

```
Dumpling Restaurant      1
Farmers Market           1
Dog Run                  1
Falafel Restaurant       1
Fast Food Restaurant     1
Yoga Studio              1
Garden                   1
Electronics Store        1
Field                    1
Event Space              1
dtype: int64
```

5.3 Cluster 3 (Blue Dots in Map)

Neighbourhood	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
Amboli	Andheri,Western Suburbs	19.129300	72.843400	2.0	Halal Restaurant	Indian Restaurant	Coffee Shop	Chinese Restaurant	Sandwich Place	Fast Food Restaurant	Asian Restaurant
Chakala, Andheri	Western Suburbs	19.111388	72.860833	2.0	Hotel	Multiplex	Café	Restaurant	Fast Food Restaurant	Liquor Store	Ice Cream Shop
D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373	2.0	Gym / Fitness Center	Snack Place	Vegetarian / Vegan Restaurant	Pizza Place	Cocktail Bar	Playground	Indian Restaurant
Four Bungalows	Andheri,Western Suburbs	19.124714	72.827210	2.0	Bakery	Chinese Restaurant	Market	Electronics Store	Juice Bar	Residential Building (Apartment / Condo)	B
Lokhandwala	Andheri,Western Suburbs	19.130815	72.829270	2.0	Coffee Shop	Shopping Mall	Women's Store	Indian Restaurant	Liquor Store	Lounge	Market

Count of Each Venue in Cluster 3 :

```
Indian Restaurant       46
Fast Food Restaurant    39
Café                    35
Farmers Market          24
Bakery                  22
Coffee Shop             22
Chinese Restaurant      22
Falafel Restaurant      21
Bar                     21
Pizza Place             19
dtype: int64
```

5.4 Cluster 4 (Red dots in Map)

	Neighbourhood	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	9th Common Venue
10	Bhayandar	Bhayandar,Western Suburbs	19.29	72.85	3.0	Shipping Store	Yoga Studio	Dog Run	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Electronics

Count of Each Venue in Cluster 4 :

```
Dumpling Restaurant      1
Farmers Market           1
Dog Run                  1
Event Space              1
Falafel Restaurant       1
Fast Food Restaurant     1
Yoga Studio              1
Electronics Store        1
Field                    1
Shipping Store           1
dtype: int64
```

5.5 Cluster 5 (Pink dots in Map)

	Neighbourhood	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	
45	Indian Institute of Technology Bombay campus	Powai,Eastern Suburbs	19.133636	72.915358	4.0	Indian Restaurant	Concert Hall	Café	Coffee Shop	Event Space	Field	Fast Food Restaurant	Farmers Market	R
62	Currey Road	South Mumbai	18.994000	72.833000	4.0	Coffee Shop	Maharashtrian Restaurant	Plaza	Indian Restaurant	Yoga Studio	Dog Run	Fast Food Restaurant	Farmers Market	R
82	Hindu colony	Dadar,South Mumbai	19.019700	72.847900	4.0	Indian Restaurant	Coffee Shop	Café	Chinese Restaurant	Fast Food Restaurant	Farmers Market	Vegetarian / Vegan Restaurant	Lounge	

Count of Each Venue in Cluster 5 :

```
Coffee Shop              3
Fast Food Restaurant     3
Indian Restaurant        3
Farmers Market           3
Café                     2
Falafel Restaurant       2
Event Space              2
Electronics Store        1
Dog Run                  1
Plaza                    1
dtype: int64
```

5.5 Cluster 6 (Green dots in Map)

	Neighbourhood	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 Common V
39	Pant Nagar	Ghatkopar,Eastern Suburbs	19.080000	72.910000	5.0	Indian Restaurant	Ice Cream Shop	Park	Convenience Store	Snack Place	Food Truck	Spa	Mu
40	Kanjurmarg	Eastern Suburbs	19.130000	72.940000	5.0	Bakery	Indian Restaurant	Multiplex	Yoga Studio	Field	Fast Food Restaurant	Farmers Market	Resta
42	Nahur	Mulund,Eastern Suburbs	19.157000	72.941000	5.0	Indian Restaurant	Ice Cream Shop	Bus Station	Restaurant	Yoga Studio	Dog Run	Fast Food Restaurant	Fa N
54	Bhuleshwar	South Mumbai	18.950000	72.830000	5.0	Indian Restaurant	Food	Ice Cream Shop	Fast Food Restaurant	Market	Restaurant	Jewelry Store	Cl
77	Walkeshwar	South Mumbai	18.947596	72.795957	5.0	Convenience Store	Coffee Shop	Ice Cream Shop	Indian Restaurant	Lighthouse	Donut Shop	Field	Fast Resta
80	Dagdi Chawl	Byculla,South Mumbai	18.977129	72.829131	5.0	Indian Restaurant	Restaurant	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Electi
84	Chira Bazaar	Kalbadevi,South Mumbai	18.948140	72.825462	5.0	Indian Restaurant	Café	Outdoors & Recreation	Train Station	Arts & Crafts Store	Yoga Studio	Field	Fast Resta
86	Chor Bazaar	Kamathipura,South Mumbai	18.960321	72.827176	5.0	Indian Restaurant	Market	Antique Shop	Dessert Shop	Asian Restaurant	Ice Cream Shop	BBQ Joint	Resta
90	Dava Bazaar	South Mumbai	18.946882	72.831362	5.0	Indian Restaurant	Multiplex	Ice Cream Shop	Middle Eastern Restaurant	Electronics Store	Restaurant	Café	Je

Count of Each Venue in Cluster 6 :

```
Indian Restaurant        9
Fast Food Restaurant     7
Ice Cream Shop           6
Restaurant               5
Yoga Studio              5
Falafel Restaurant       5
Farmers Market           5
Electronics Store        3
Multiplex                3
Field                    3
dtype: int64
```

6.DISCUSSION

Now that we have the clusters and the top venue categories let's visualize the top 5 venue category in each Cluster for comparison.
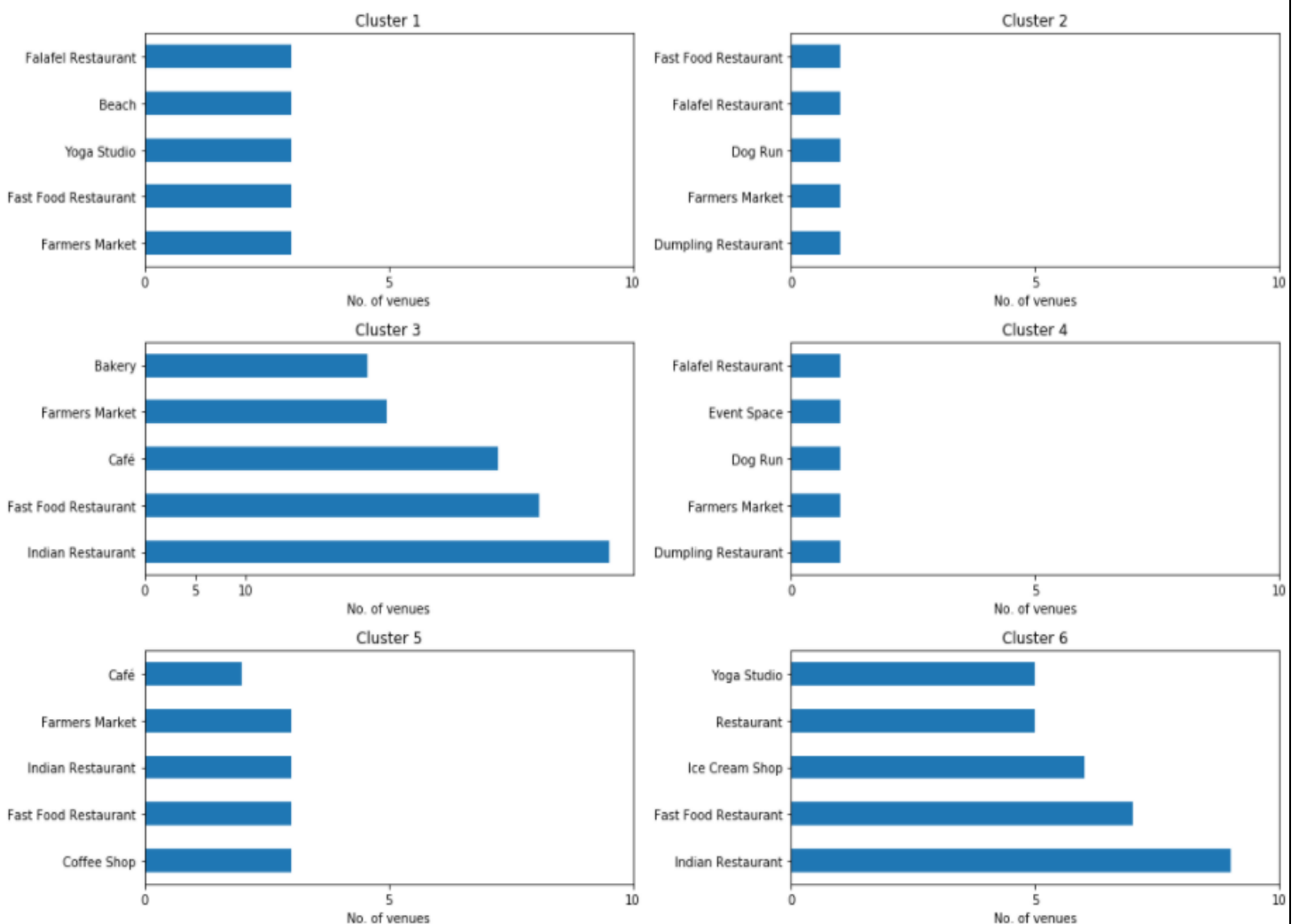
This plot can be used to suggest valuable information to Business persons. Let's discuss a few examples considering they would like to start the following category of business.

1. Hotel

The neighborhoods in cluster 3 and Cluster 6 has the greatest number of Restaurants, hence opening one here is the best choice.We can also open one at the neighborhoods in cluster 2 or 4 because there not many hotels.

2. Gym/Yoga/Fitness Camp

The neighborhoods in cluster 1 and cluster 6 has many food restraurant.Also there are many yoga studios in this neighborhood it is not a great idea.So it will be best to open a yoga/gym/fitness camp in Cluster 5 because it contains cafe and restraunts but there are not many gym/yoga/fitness camp.

Similarly, based on the requirement suggestions can be provided about the neighborhood that would be best suitable for the business.Fig7 shows a map of Mumbai with neighborhood clusters superimposed on top of it.This map can be used to suggest vast location to start a new businee based on the category
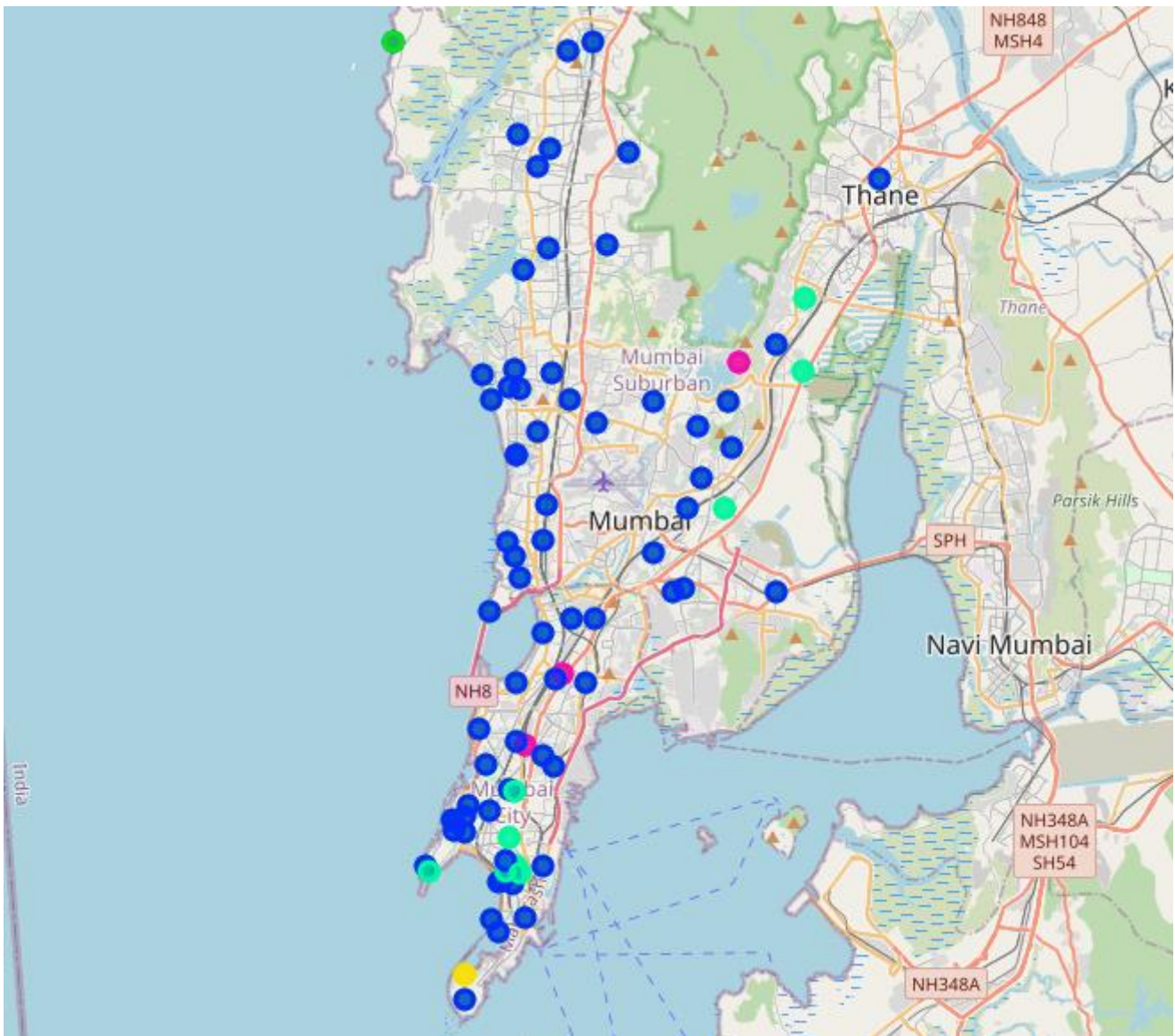

Fig7 : Map of Mumbai with clusters superimposed on Top

For example, the highlighted location shown in Fig 8 consists of Cluster 3 and Cluster 6, whose neighborhoods have many Restaurants but less Hotels. Thus, this would be a suitable location for building a hotel.
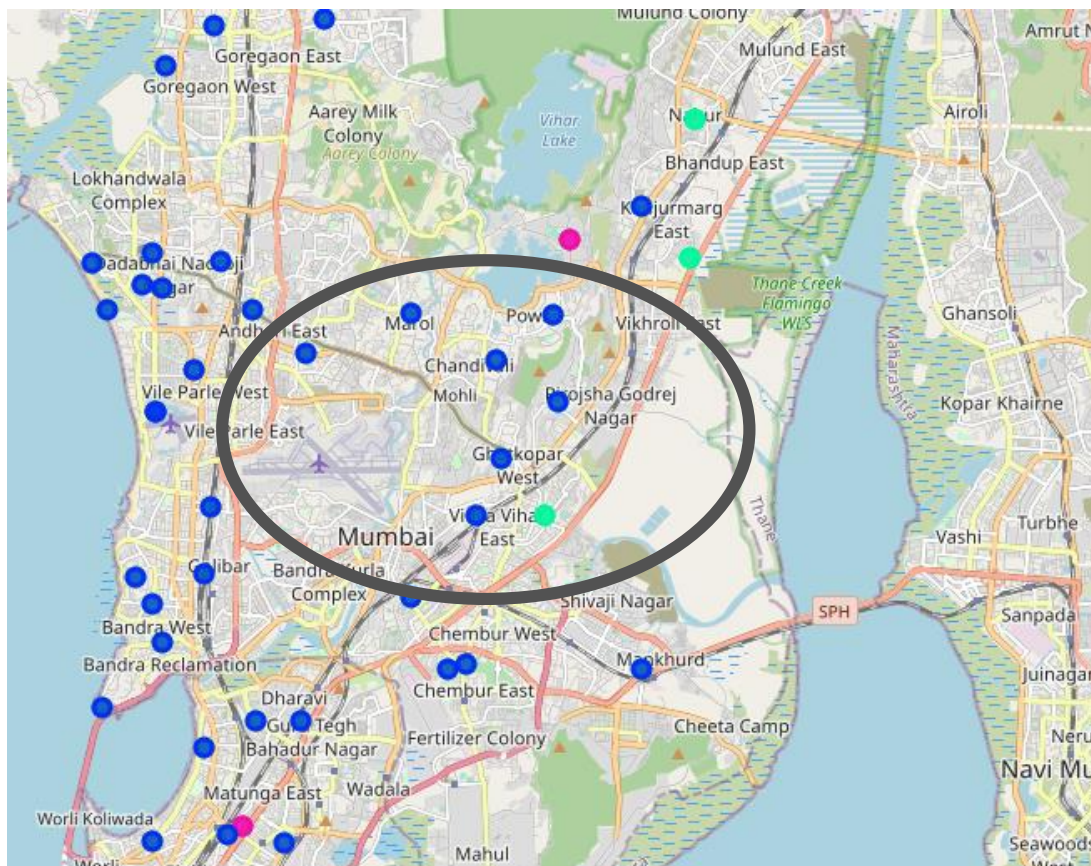

Fig8 : Location Suitable for New Hotel

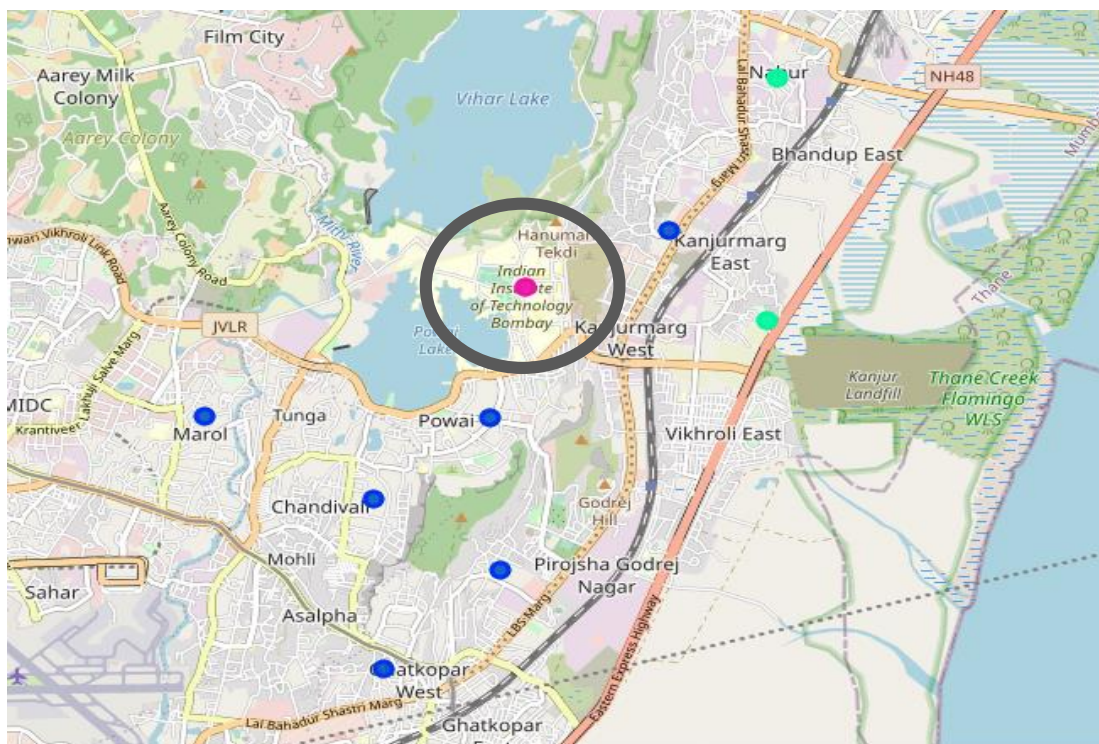In Below Fig9 , highlighted location is best suitable for building gym/yoga studio/fitness camp

Fig9 : Location Suitable for New Gym/Yoga studio/Fitness camp

7.Conclusion

Purpose of this project was to analyze the neighborhoods of Mumbai and create a clustering model to suggest personals places to start a new business based on the category. The neighborhoods data was obtained from an online source and the Foursquare API was used to find the major venues in each neighborhood. But we found that many neighborhoods had less than 10 venues returned. In order to build a good Data Science model, we filtered out these locations. The remaining locations were used to create a clustering model. The best number of clusters i.e. 6 was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster.

A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. Both these can be used by stakeholders to decide the location for the particular type of business. A major drawback of this project was that the Foursquare API returned only few venues in each neighborhood. As a future improvement, better data sources can be used to obtain more venues in each neighborhood. This way the neighborhoods that were filtered out can be included in the clustering analysis to create a better decision model.