Coursera Capstone
IBM applied Data science Capstone

Battle of Neighborhoods in Mumbai for opening New Shopping Mall.

By: Saurabh Mahadane

Date: June 6, 2021

Introduction

- Mumbai is the financial capital of India and is one of the most densely populated cities in the world
- The multi-cultural nature of the city brings with it numerous cuisines from all over the globe
- Business owners ,Builders and entrepreneurs might be looking to start a new Shopping Mall in Mumbai
- The task is to identify neighborhoods that have the potential of being good locations for starting a New Shopping Mall

Figure 1: Mumbai view

Data Collection

- The following data was collected for this project:
  - Neighborhood data of Mumbai from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Mumbai
  - Geographical coordinates of Mumbai and all neighborhoods in Mumbai using GeoPy and Geocoder libraries in python
  - Venue data for all neighbourhoods in Mumbai using Foursquare API
- The data was then cleaned to produce the final datasets shown in the upcoming slides

Dataset for Mumbai Neighborhoods

	Neighborhood	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

Figure 2: First 10 rows of Mumbai Neighborhood Dataframe

Methodology

- **Data Visualization**
  - Mumbai neighborhoods data was plotted for providing a better understanding
  - The graph alongside depicts the number of neighborhoods in each location of Mumbai
  - All neighborhoods on the outskirts of the city have been grouped as “Mumbai”
  - Western Suburbs and South Mumbai contain the highest number of neighborhoods

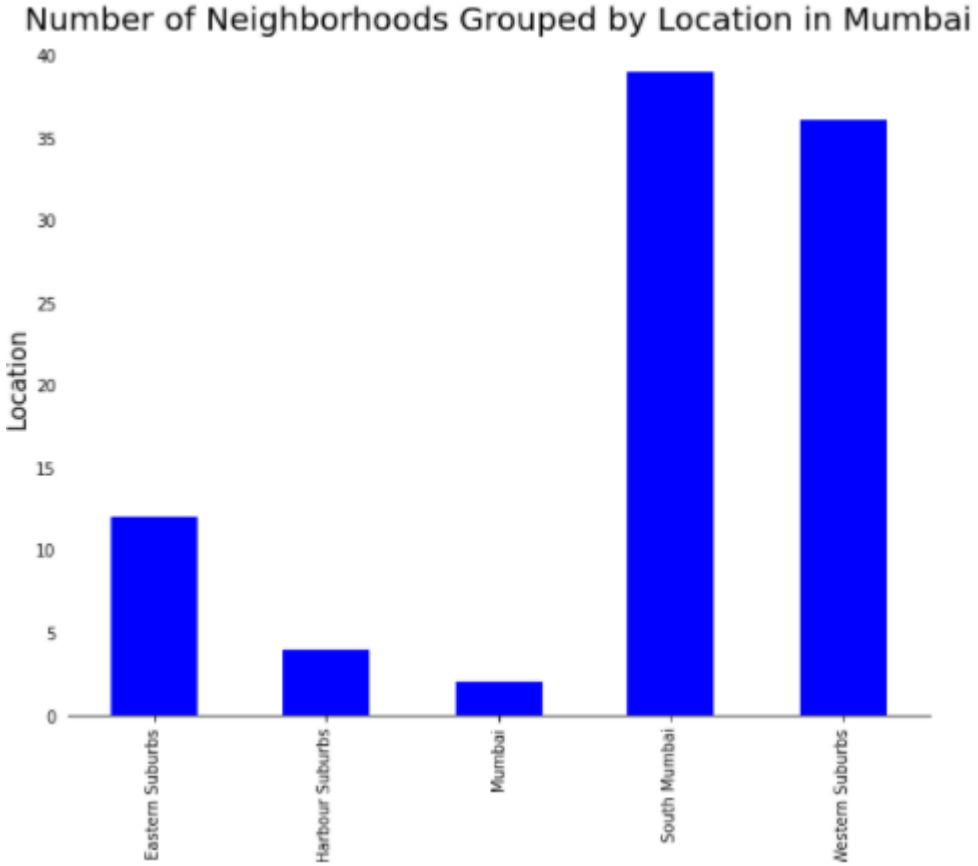

Figure 3: Number of neighbourhoods grouped by location

Methodology Continued

- **Data Visualization**
  - The Folium library in python was used to visualize the spread of all neighborhoods across Mumbai

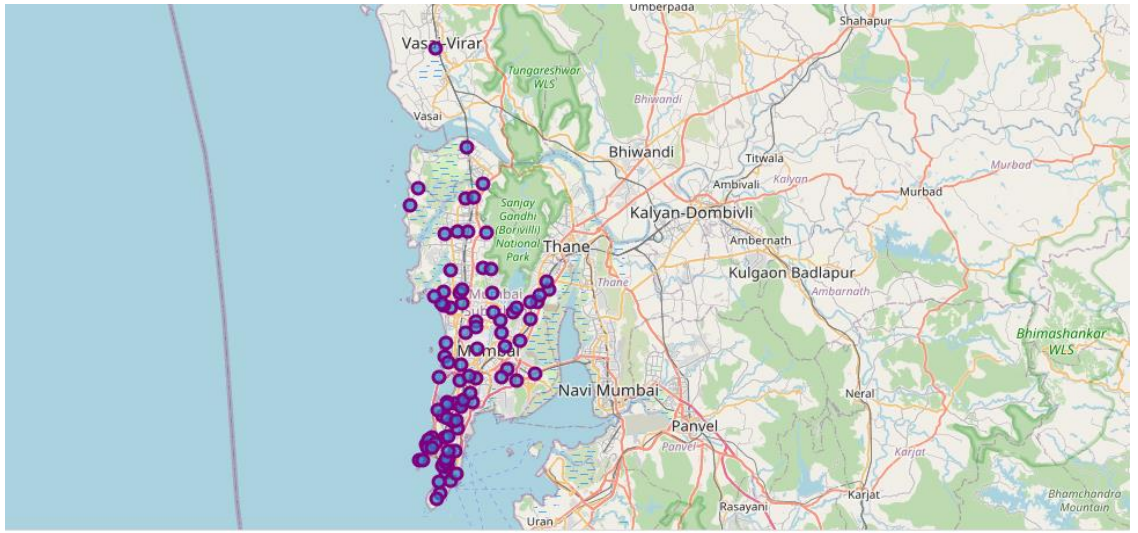

Figure 4: The neighborhood spread across Mumbai

Methodology Continued

- **One-hot Encoding**
  - One-hot Encoding was used to encode venue categories to numeric values with 1 if a venue belongs to a category and 0 if a venue does not belong to a category for all neighborhoods
  - The average is then taken for all venue categories in a neighborhood to produce the dataframe shown

	Neighborhood	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	American Restaurant	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant
0	Aarey Milk Colony	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.022222
1	Agripada	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.014925	0.000000	0.014925	0.0	0.0	0.000000
2	Altamount Road	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.010101
3	Amboli	0.013514	0.0	0.000000	0.000000	0.000000	0.000000	0.013514	0.000000	0.000000	0.000000	0.0	0.0	0.000000
4	Amrut Nagar	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.011494	0.000000	0.000000	0.000000	0.0	0.0	0.011494
5	Asalfa	0.000000	0.0	0.010526	0.010526	0.010526	0.010526	0.010526	0.000000	0.000000	0.000000	0.0	0.0	0.010526
6	Ballard Estate	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.010000	0.0	0.0	0.020000
7	Bandstand Promenade	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.022222
8	Bangur Nagar	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.010000	0.000000	0.0	0.0	0.000000
9	Bhandup	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.045455	0.000000	0.0	0.0	0.000000

Figure 5: One-hot Encoding resulting dataframe

Methodology Continued

- **Unsupervised Learning Model**
  - KMeans clustering was used to cluster neighborhoods in Mumbai based on venue categories
  - The plot shows a maximum Silhouette Score for 4 clusters and thus the n_clusters parameter in KMeans clustering was set to 4

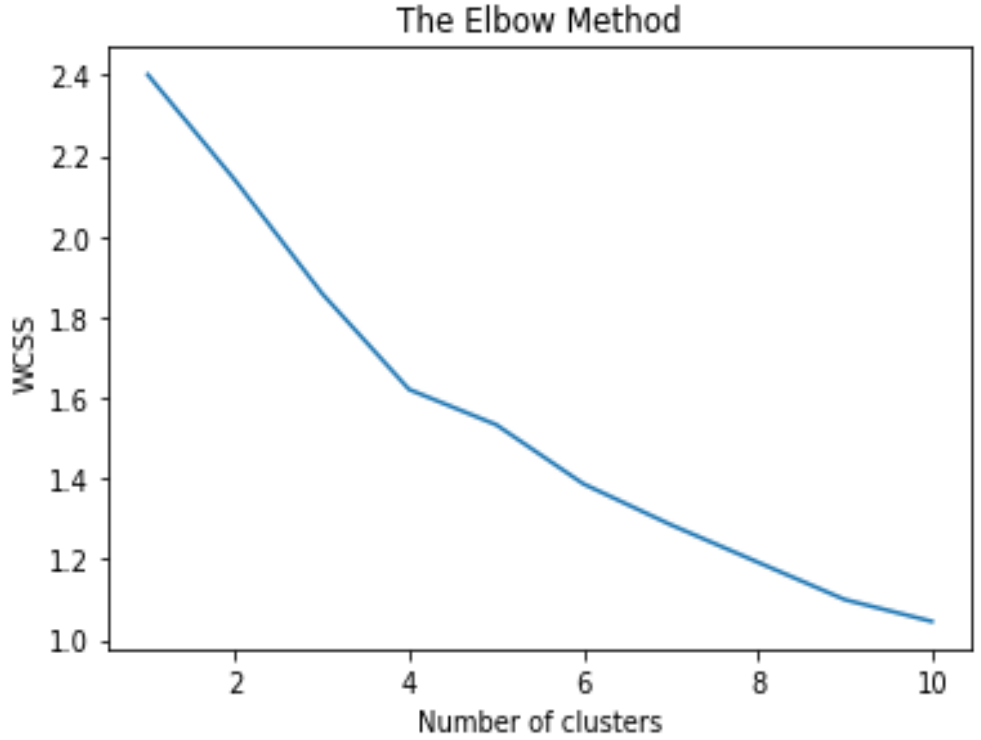

Figure 6: WCSS scores for varying number of clusters in KMeans clustering

Results

- Each neighborhood received a cluster label based on clustering by the KMeans algorithm
- The Cluster Label column along with the Location, Latitude, and Longitude columns were added to the shopping mall venues dataframe found in neighborhood cluster wise.
- This dataframe is shown on for Cluster1 ,Cluster2, Cluster3 and Cluster4.

Results Continued – Cluster 1

- Analyzing each cluster individually

	Neighborhood	Shopping Mall	Cluster Labels	Location	Latitude	Longitude
46	Juhu	0.010000	0	Western Suburbs	19.0149	72.8452
90	Virar	0.012987	0	Western Suburbs	19.0166	72.8585
89	Vile Parle	0.011765	0	Western Suburbs	19.0962	72.8502
3	Amboli	0.013514	0	Western Suburbs	19.1293	72.8464
41	Hiranandani Gardens	0.012987	0	Eastern Suburbs	19.119	72.9068
86	Versova	0.010000	0	Western Suburbs	19.1377	72.8135
50	Kemps Corner	0.010000	0	South Mumbai	18.9647	72.8054
40	Hindu colony	0.010309	0	South Mumbai	19.0197	72.8474
51	Khar Danda	0.010000	0	Western Suburbs	19.0843	72.8269
68	Naigaon	0.011111	0	Western Suburbs	19.0119	72.8453
64	Matunga	0.010989	0	South Mumbai	19.0272	72.8559

Figure 7: Cluster 1

Results Continued – Cluster 2

	Neighborhood	Shopping Mall	Cluster Labels	Location	Latitude	Longitude
49	Kanjurmarg	0.045455	1	Eastern Suburbs	19.1314	72.9357
75	Pant Nagar	0.048780	1	Eastern Suburbs	19.0863	72.915
83	Thakur village	0.057692	1	Western Suburbs	19.2102	72.8754
87	Vidyavihar	0.050505	1	Eastern Suburbs	19.08	72.8973
88	Vikhroli	0.044444	1	Eastern Suburbs	19.1111	72.9278
71	Navy Nagar	0.041667	1	South Mumbai	18.906	72.8155
23	Cotton Green	0.046154	1	South Mumbai	18.9862	72.8412
0	Aarey Milk Colony	0.066667	1	Western Suburbs	19.1703	72.8711
13	C.G.S. colony	0.052632	1	South Mumbai	19.1389	72.9382
9	Bhandup	0.045455	1	Eastern Suburbs	19.1456	72.9486

Figure 8: Cluster 2

Results Continued – Cluster 3

	Neighborhood	Shopping Mall	Cluster Labels	Location	Latitude	Longitude
19	Chembur	0.0	2	Harbour Suburbs	19.054	72.8997
10	Bhayandar	0.0	2	Western Suburbs	19.3074	72.8518
67	Nahur	0.0	2	Eastern Suburbs	19.1537	72.9467
66	Mumbai Central	0.0	2	South Mumbai	18.9697	72.8151
65	Mira Road	0.0	2	Western Suburbs	19.2656	72.8711
11	Bhuleshwar	0.0	2	South Mumbai	18.9512	72.83
63	Marol	0.0	2	Western Suburbs	19.1192	72.8827
62	Marine Lines	0.0	2	South Mumbai	18.9434	72.8232
60	Mankhurd	0.0	2	Harbour Suburbs	19.0485	72.9322
59	Malabar Hill	0.0	2	South Mumbai	18.95	72.795
58	Mahul	0.0	2	Harbour Suburbs	19.0454	72.8932
57	Mahim	0.0	2	South Mumbai	19.0407	72.8431

Figure 9: Cluster 3

Results Continued – Cluster 4

	Neighborhood	Shopping Mall	Cluster Labels	Location	Latitude	Longitude
5	Asalfa	0.031579	3	Eastern Suburbs	19.0953	72.8926
24	Cuffe Parade	0.020619	3	South Mumbai	18.913	72.8205
54	Lower Parel	0.020000	3	South Mumbai	18.9981	72.8281
28	Dagdi Chawl	0.022472	3	South Mumbai	18.9771	72.8291
30	Dava Bazaar	0.022727	3	South Mumbai	19.1314	72.927
77	Poisar	0.022989	3	Western Suburbs	19.2116	72.8527
76	Parel	0.030000	3	South Mumbai	18.9957	72.84
8	Bangur Nagar	0.030000	3	Western Suburbs	19.1674	72.8323
72	Nehru Nagar	0.020000	3	Eastern Suburbs	19.0005	72.8228
26	Currey Road	0.021739	3	South Mumbai	18.9952	72.8346
92	Worli	0.021739	3	South Mumbai	19.0074	72.8169

Figure 10: Cluster 4

Results Continued

- Visualization of neighborhood clusters was done using Folium in python
- Different clusters correspond to different colors

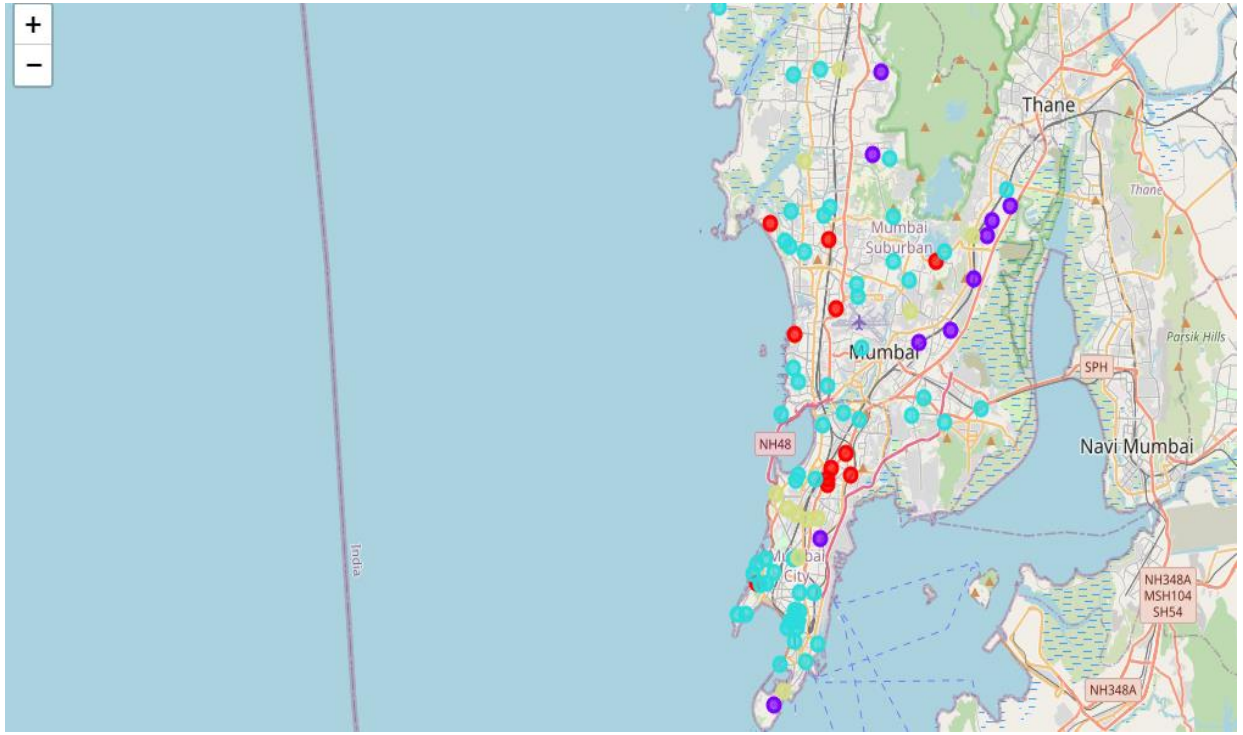

Figure 11: Visualizing the clustering of neighborhoods in Mumbai

Discussion

By analysing the four clusters obtained we can see that

- Clusters 3 have negligible of shopping Malls in comparison with other clusters.
- Clusters 1 have minimum no of Shopping malls in comparison to Cluster 2&4
- Recommendation: To find the least competition among the other Mall properties as per analysis Cluster 1 areas/locations would be advisable as Best recommendation locations.
- Thus, the most optimal neighborhoods for opening shopping mall are in cluster 1

Conclusion

- In this project, the neighbourhoods in Mumbai, India have been successfully analysed for determining which would be the best neighbourhoods for opening a New Shopping mall.
- Based on the analysis carried out, neighbourhoods in cluster 1 are recommended as locations for the New Shopping mall.
- The stakeholders and investors can further tune this by considering various other factors like transport, legal requirements, and costs associated which were out of the scope for this project and thus were not considered.