

Exploring Neighborhoods of Mumbai for Starting Café

Applied Data Science

Capstone Project

By: Abhijeet Mukherjee

Date: March 22, 2021

Introduction

- ▶ Mumbai is one of the busiest metropolitan in India, being the financial capital and one of the densely populated City. Due to multiple iconic spots, Mumbai is a major tourist attraction every year. I have been working in Mumbai for a while now and have unique experience with the neighborhoods. It is one of the major IT hub and everyday thousands of professionals prefer business meetups in Café. This is a everyday dilemma faced by IT professional like myself to actually find a place to carry out small tasks like sending mails, making calls without background disturbance or grabbing a coffee between a hectic day. Thus, the aim of this project is to study the neighborhoods in Mumbai to determine possible locations for starting a Café . This project can be useful for business owners and entrepreneurs who are looking to invest in a Café in Mumbai. The main objective of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

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 neighborhoods in Mumbai using Foursquare API

Data Collection

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

Dataset for Mumbai Neighborhoods

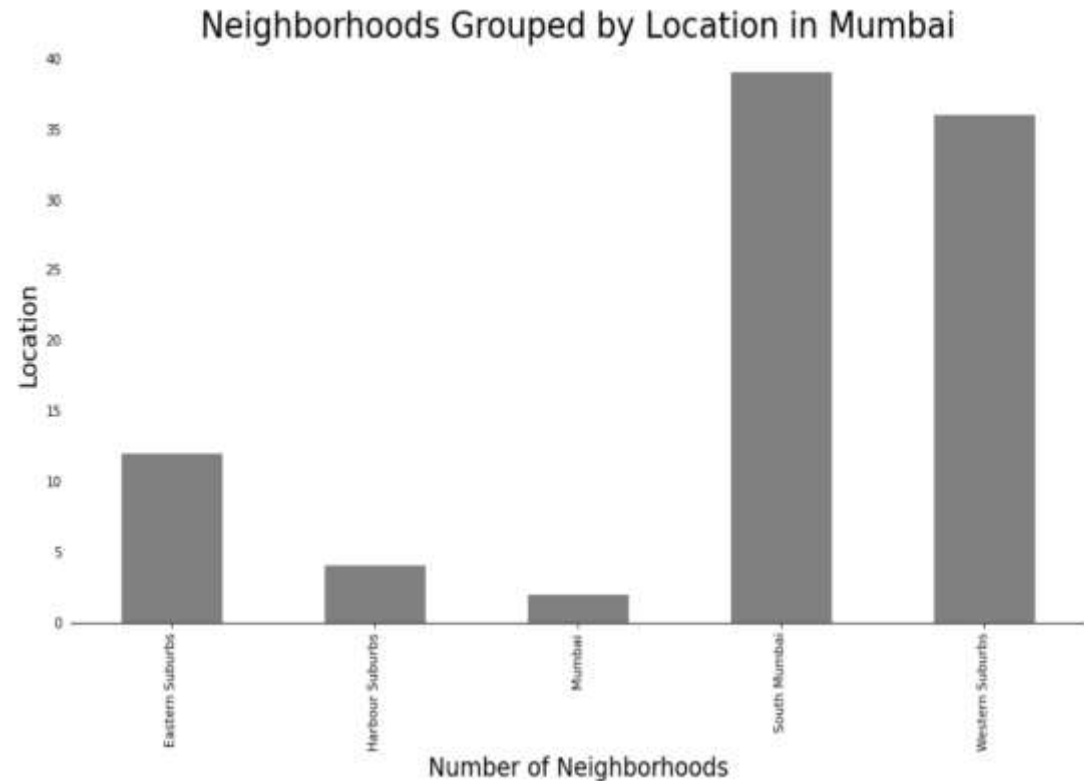
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Indian Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Bar	Asian Restaurant	Bakery	Camera Store	Falafel Restaurant	Snack Place
1	Chakala, Andheri	Hotel	Café	Indian Restaurant	Fast Food Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Restaurant	Hotel Bar	Asian Restaurant	Bar
2	D.N. Nagar	Bar	Gym / Fitness Center	Pub	Pizza Place	Indian Restaurant	Snack Place	Women's Store	Vegetarian / Vegan Restaurant	Lounge	Market
3	Four Bungalows	Indian Restaurant	Chinese Restaurant	Café	Pub	Ice Cream Shop	Bar	Seafood Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Gym / Fitness Center
4	Lokhandwala	Indian Restaurant	Chinese Restaurant	Café	Pub	Gym / Fitness Center	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Bar	Lounge
...
88	Parel	Indian Restaurant	Coffee Shop	Chinese Restaurant	Pharmacy	Clothing Store	Playground	Plaza	Multicuisine Indian Restaurant	Restaurant	Roof Deck
89	Gowalia Tank	Indian Restaurant	Café	Bakery	Coffee Shop	Fast Food Restaurant	Snack Place	Chinese Restaurant	Pizza Place	Electronics Store	Salon / Barbershop
90	Dava Bazaar	Train Station	Café	Clothing Store	Food Truck	Beer Garden	Asian Restaurant	Cupcake Shop	Coffee Shop	Fish Market	Falafel Restaurant
91	Dharavi	Indian Restaurant	Fast Food Restaurant	Music Venue	Shoe Store	Seafood Restaurant	Sandwich Place	Lake	Food & Drink Shop	Café	Garden
92	Thane	Pizza Place	Performing Arts Venue	Dessert Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Electronics Store	Donut Shop

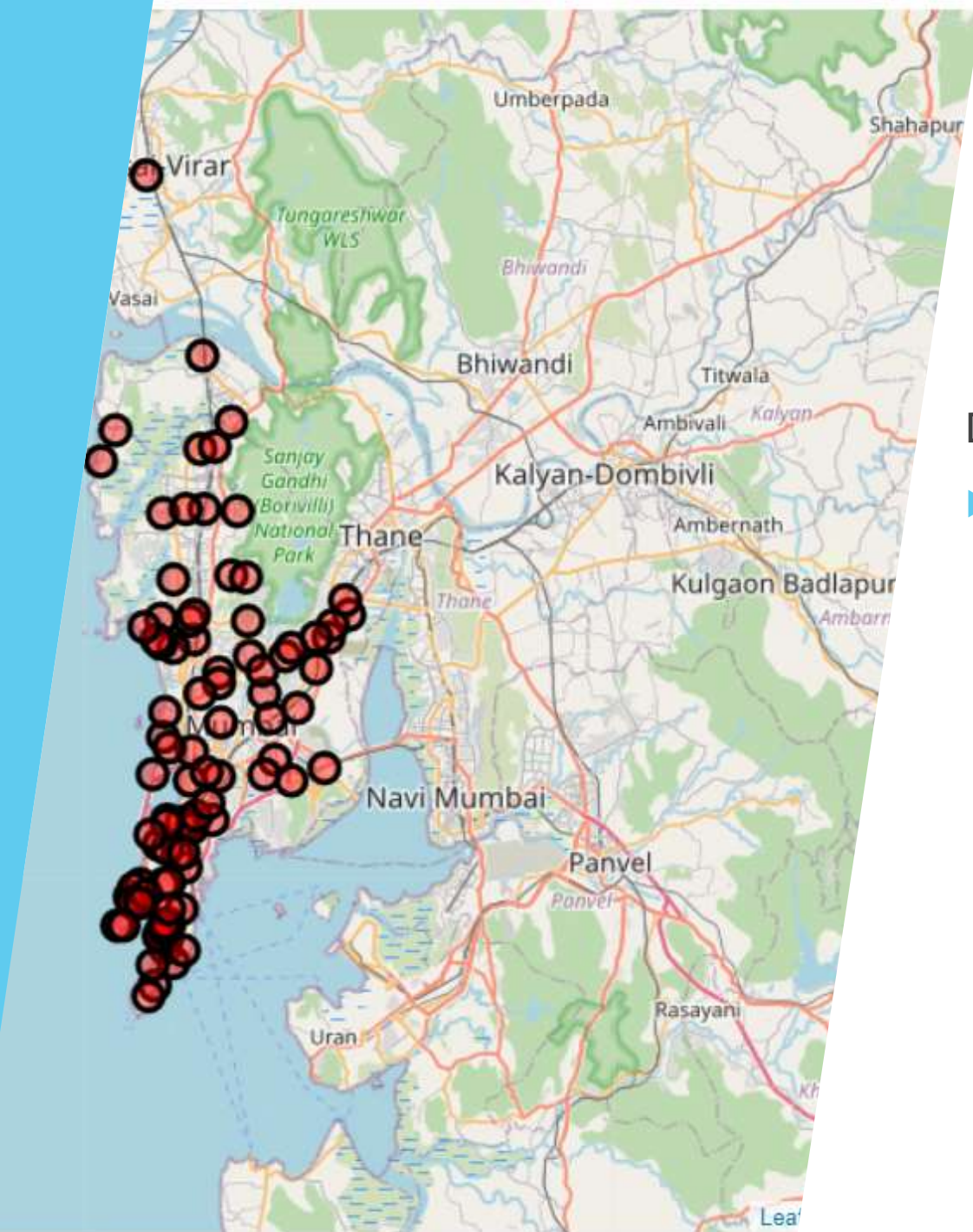
Dataset for Top 10 Most Common Venues

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





Continued...

Data Visualization

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

► One-hot Encoding

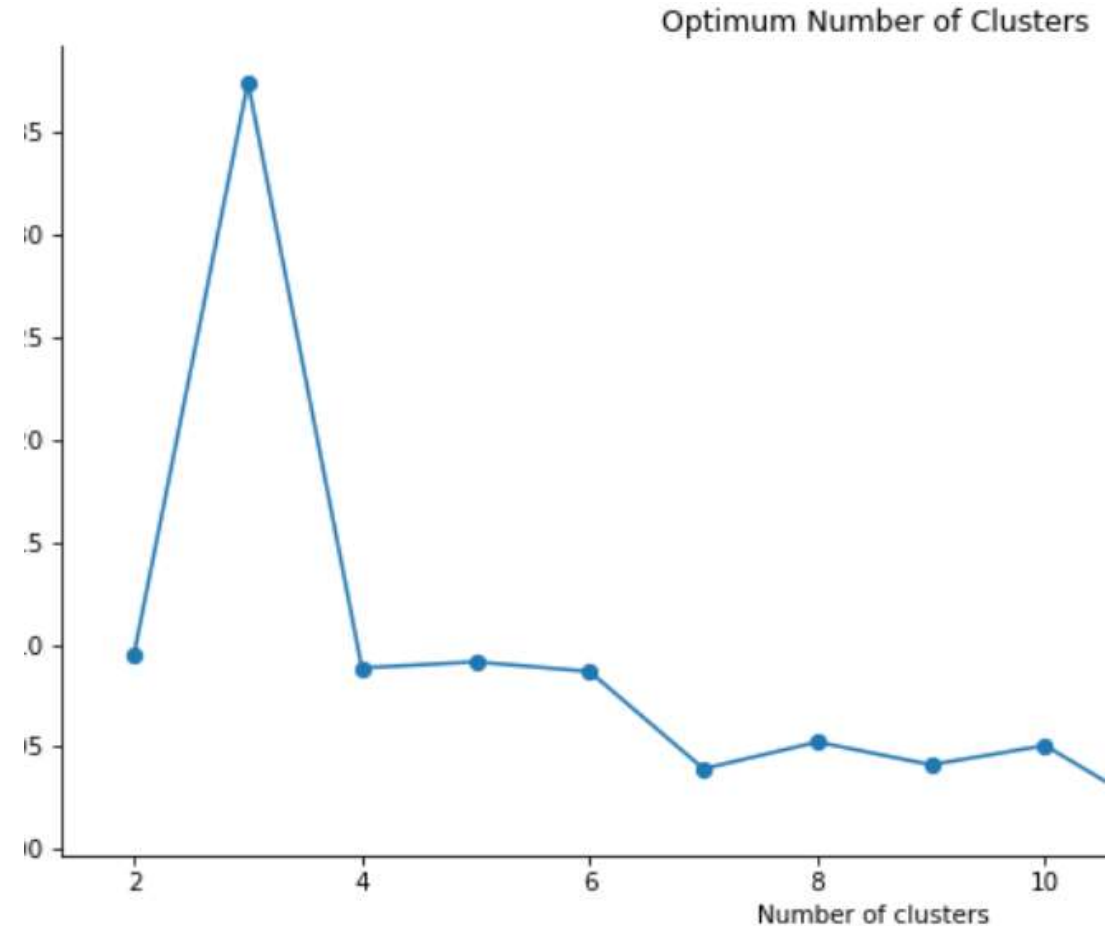
One-hot Encoding was used to encode venue categories to numeric values with 1 if a venue belongs to a category and 0 if a venue does not belong to a category for all neighborhoods • The average is then taken for all venue categories in a neighborhood to produce the data frame shown

	Neighborhood	Accessories Store	Airport	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	...	Train	Train Station	Travel & Transport	Vegetarian / Vegan Restaurant	Whisky Bar
0	Amboli	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.000000	0.0
1	Chakala, Andheri	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.048780	0.0
2	D.N. Nagar	0.021739	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.043478	0.0
3	Four Bungalows	0.014925	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.029851	0.0
4	Lokhandwala	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.010870	0.0
5	Marol	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.000000	0.0
6	Sahar	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.000000	0.0
7	Seven Bungalows	0.015873	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.031746	0.0
8	Versova	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.025	...	0.0	0.0	0.0	0.000000	0.0
9	Mira Road	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	...	0.0	0.0	0.0	0.062500	0.0

► Unsupervised Learning Model

KMeans clustering was used to cluster neighborhoods in Mumbai based on venue categories

The plot shows a maximum Silhouette Score for 5 clusters and thus the n_clusters parameter in KMeans clustering was set to 5



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 top 10 most common venues dataframe to provide the final results
- ▶ This dataframe is shown on the next slide

	Neighborhood	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
0	Amboli	Western Suburbs	19.1293	72.8464	1	Indian Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Bar	Asian Restaurant	Bakery	Camera Store	Falafel Restaurant
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	1	Hotel	Café	Indian Restaurant	Fast Food Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Restaurant	Hotel Bar	Asian Restaurant
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	1	Bar	Gym / Fitness Center	Pub	Pizza Place	Indian Restaurant	Snack Place	Women's Store	Vegetarian / Vegan Restaurant	Lounge
3	Four Bungalows	Western Suburbs	19.1264	72.8242	1	Indian Restaurant	Chinese Restaurant	Café	Pub	Ice Cream Shop	Bar	Seafood Restaurant	Coffee Shop	Vegetarian / Veg Restaurant
4	Lokhandwala	Western Suburbs	19.1432	72.8249	1	Indian Restaurant	Chinese Restaurant	Café	Pub	Gym / Fitness Center	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Bar
...
88	Parel	South Mumbai	18.9957	72.84	1	Indian Restaurant	Coffee Shop	Chinese Restaurant	Pharmacy	Clothing Store	Playground	Plaza	Multicuisine Indian Restaurant	Restaurant
89	Gowalia Tank	South Mumbai	18.9645	72.8112	1	Indian Restaurant	Café	Bakery	Coffee Shop	Fast Food Restaurant	Snack Place	Chinese Restaurant	Pizza Place	Electronics Store
90	Dava Bazaar	South Mumbai	19.1314	72.927	0	Train Station	Café	Clothing Store	Food Truck	Beer Garden	Asian Restaurant	Cupcake Shop	Coffee Shop	Fish Market
91	Dharavi	Mumbai	19.0467	72.8546	1	Indian Restaurant	Fast Food Restaurant	Music Venue	Shoe Store	Seafood Restaurant	Sandwich Place	Lake	Food & Drink Shop	Café
92	Thane	Mumbai	19.1409	72.8826	2	Pizza Place	Performing Arts Venue	Dessert Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Electronics Store

Clustered Neighbourhood

CLUSTER 1

```
In [66]: mum_merged.loc[mum_merged['Cluster Labels'] == 0, mum_merged.columns[[0] + [1] + list(range(5, mum_merged.shape[1]))]]
```

Out[66]:

	Neighborhood	Location	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
32	Nalasopara	Western Suburbs	Multiplex	Pizza Place	Fast Food Restaurant	Ice Cream Shop	Diner	Department Store	Bus Station	Farmers Market	Falafel Restaurant	Factory
36	Bhandup	Eastern Suburbs	Train Station	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Zoo	Dhaba	Field	Farmers Market	Falafel Restaurant	Factory
40	Kanjurmarg	Eastern Suburbs	Train Station	Chinese Restaurant	Asian Restaurant	Gym	Food Truck	Cupcake Shop	Multiplex	Gift Shop	Diner	Dive Bar
50	Mankhurd	Harbour Suburbs	Bus Station	Coffee Shop	Sports Bar	Train Station	Zoo	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space
59	Cotton Green	South Mumbai	Plaza	Train Station	Pizza Place	Multiplex	Ice Cream Shop	Fast Food Restaurant	Falafel Restaurant	Factory	Event Space	Electronics Store
90	Dava Bazaar	South Mumbai	Train Station	Café	Clothing Store	Food Truck	Beer Garden	Asian Restaurant	Cupcake Shop	Coffee Shop	Fish Market	Falafel Restaurant

Cluster 1

CLUSTER 2

```
[67]: mum_merged.loc[mum_merged['Cluster Labels'] == 1, mum_merged.columns[[0] + [1] + list(range(5, mum_merged.shape[1]))]]
```

```
1[67]:
```

	Neighborhood	Location	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	Amboli	Western Suburbs	Indian Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Bar	Asian Restaurant	Bakery	Camera Store	Falafel Restaurant	Snack Place
1	Chakala, Andheri	Western Suburbs	Hotel	Café	Indian Restaurant	Fast Food Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Restaurant	Hotel Bar	Asian Restaurant	Bar
2	D.N. Nagar	Western Suburbs	Bar	Gym / Fitness Center	Pub	Pizza Place	Indian Restaurant	Snack Place	Women's Store	Vegetarian / Vegan Restaurant	Lounge	Market
3	Four Bungalows	Western Suburbs	Indian Restaurant	Chinese Restaurant	Café	Pub	Ice Cream Shop	Bar	Seafood Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Gym / Fitness Center
4	Lokhandwala	Western Suburbs	Indian Restaurant	Chinese Restaurant	Café	Pub	Gym / Fitness Center	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Bar	Lounge
...
86	Chor Bazaar	South Mumbai	Indian Restaurant	Dessert Shop	BBQ Joint	Bus Station	Ice Cream Shop	Rest Area	Restaurant	Food	Middle Eastern Restaurant	Chinese Restaurant
87	Matunga	South Mumbai	Indian Restaurant	Fast Food Restaurant	Café	Snack Place	Ice Cream Shop	Chinese Restaurant	South Indian Restaurant	Vegetarian / Vegan Restaurant	Train Station	Coffee Shop
88	Parel	South Mumbai	Indian Restaurant	Coffee Shop	Chinese Restaurant	Pharmacy	Clothing Store	Playground	Plaza	Multicuisine Indian Restaurant	Restaurant	Roof Deck
89	Gowalia Tank	South Mumbai	Indian Restaurant	Café	Bakery	Coffee Shop	Fast Food Restaurant	Snack Place	Chinese Restaurant	Pizza Place	Electronics Store	Salon / Barbershop
91	Dharavi	Mumbai	Indian Restaurant	Fast Food Restaurant	Music Venue	Shoe Store	Seafood Restaurant	Sandwich Place	Lake	Food & Drink Shop	Café	Garden

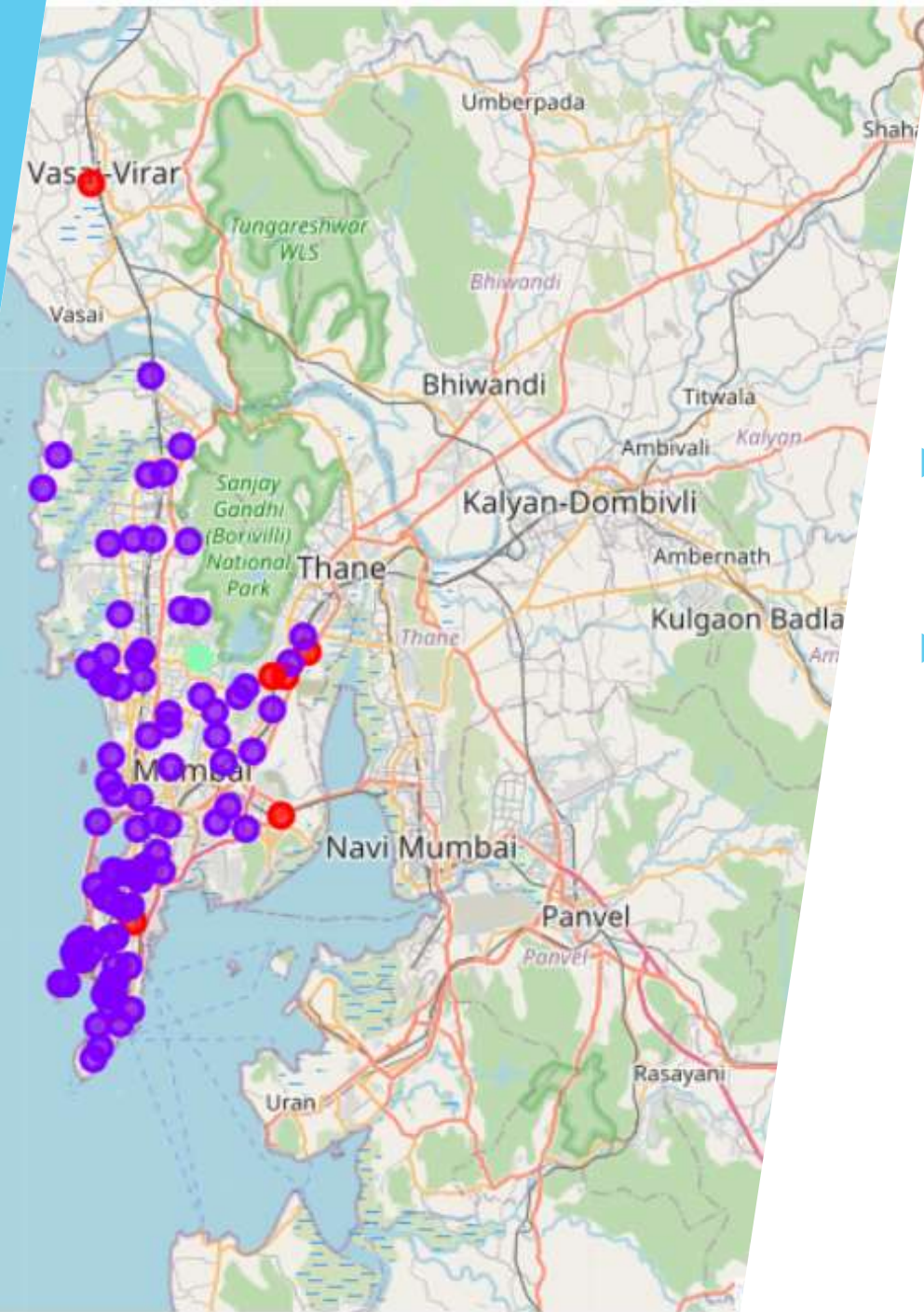
Cluster 2

CLUSTER 3

```
mum_merged.loc[mum_merged['Cluster Labels'] == 2, mum_merged.columns[[0] + [1] + list(range(5, mum_merged.shape[1]))]]
```

	Neighborhood	Location	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
92	Thane	Mumbai	Pizza Place	Performing Arts Venue	Dessert Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Factory	Event Space	Electronics Store	Donut Shop

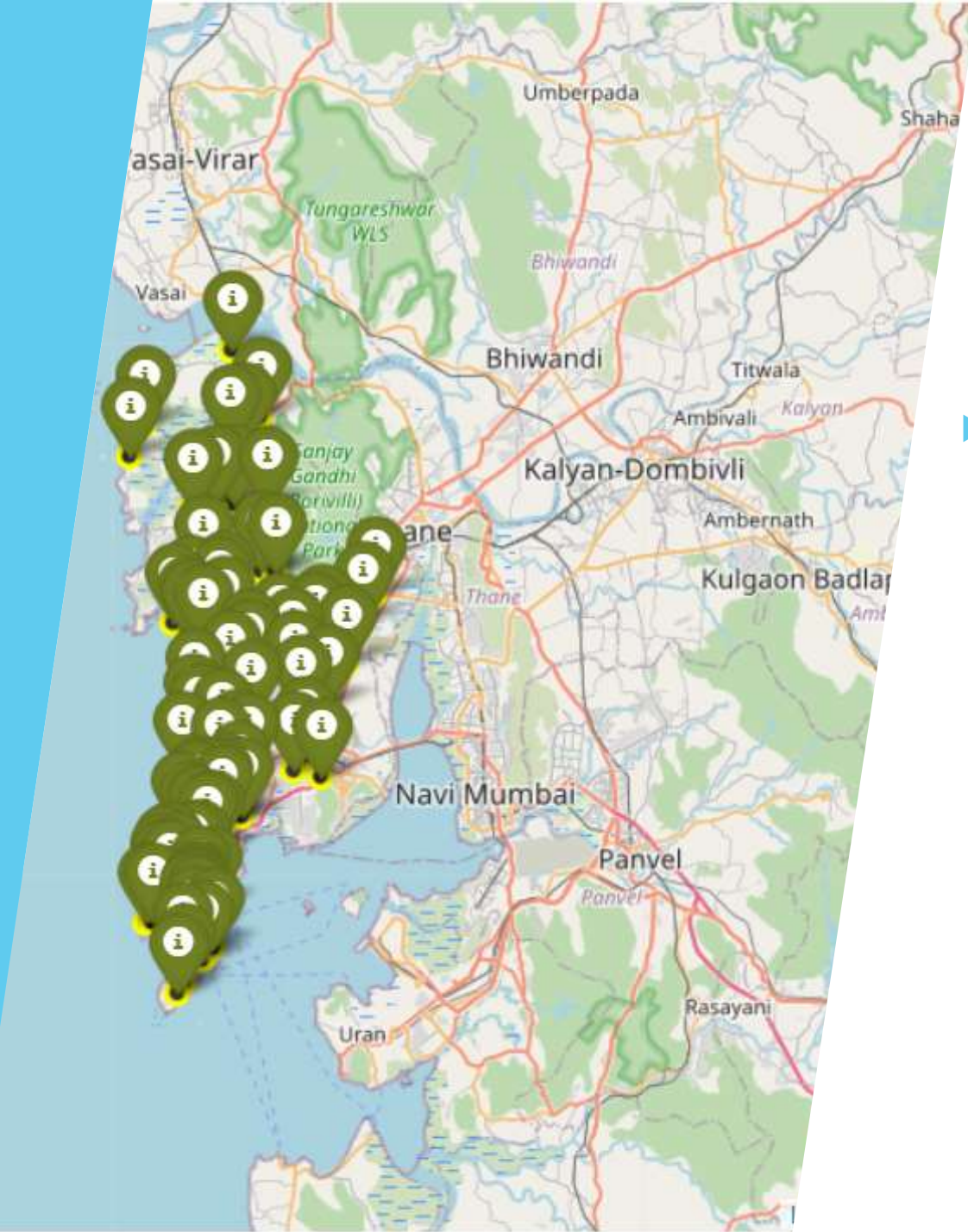
Cluster 3



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

- ▶ By analyzing the three clusters obtained we can see that some of the clusters are more Cafe, whereas, other clusters are less suited. Neighborhoods in clusters 1 and 3 contain a small percentage of restaurants, hotels, cafe and pubs in their top 10 common venues. These clusters contain a higher degree of other venues like train station, bus station, fish market, gym, performing arts venue and smoke shop, to name a few. Thus, they are not well suited for opening a new cafe. On the other hand, neighborhoods in cluster 2 contain a much higher degree of restaurants, hotels, multiplex, cafes, bars and other food joints. Thus, the neighborhoods in this clusters would be well suited for opening a new restaurant.
- ▶ Most neighborhoods in cluster 2 seem to have Cafe as their top most common venue; however, on careful analysis we can see that neighborhoods in cluster 2 also contain other venues like soccer field, flea market, smoke shop, gym, train station, dance studio, music store, cosmetics shop and so on. Thus, it is recommended that the new Cafe can be opened in the neighborhoods belonging to cluster 2.

Discussion



- Plotting final results using Folium library in python

- ▶ In this project, the neighborhoods in Mumbai, India have been successfully analyzed for determining which would be the best neighborhoods for opening a new restaurant
- ▶ Based on the analysis carried out, neighborhoods in cluster 1 are recommended as locations for the new restaurant
- ▶ 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

Conclusion