

DATA SCIENCE CAPSTONE PROJECT

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 and entrepreneurs might be looking to start a new restaurant in Mumbai
- The task is to identify neighborhoods that have the potential of being good locations for starting a new restaurant



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 neighborhoods 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	Western Suburbs	19.1293	72.8464
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623
2	D.N. Nagar	Western Suburbs	19.1241	72.8325
3	Four Bungalows	Western Suburbs	19.1263	72.8243
4	Lokhandwala	Western Suburbs	19.1432	72.8249
5	Marol	Western Suburbs	19.1192	72.8827
6	Sahar	Western Suburbs	19.1027	72.8626
7	Seven Bungalows	Western Suburbs	19.1315	72.817
8	Versova	Western Suburbs	19.1377	72.8135
9	Mira Road	Western Suburbs	19.2657	72.8711

Figure 2: First 10 rows of Mumbai Neighborhood Dataframe

Dataset for Top 10 Most Common Venues

	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	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure 3: First 10 rows of top 10 most common venues for all neighborhoods

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

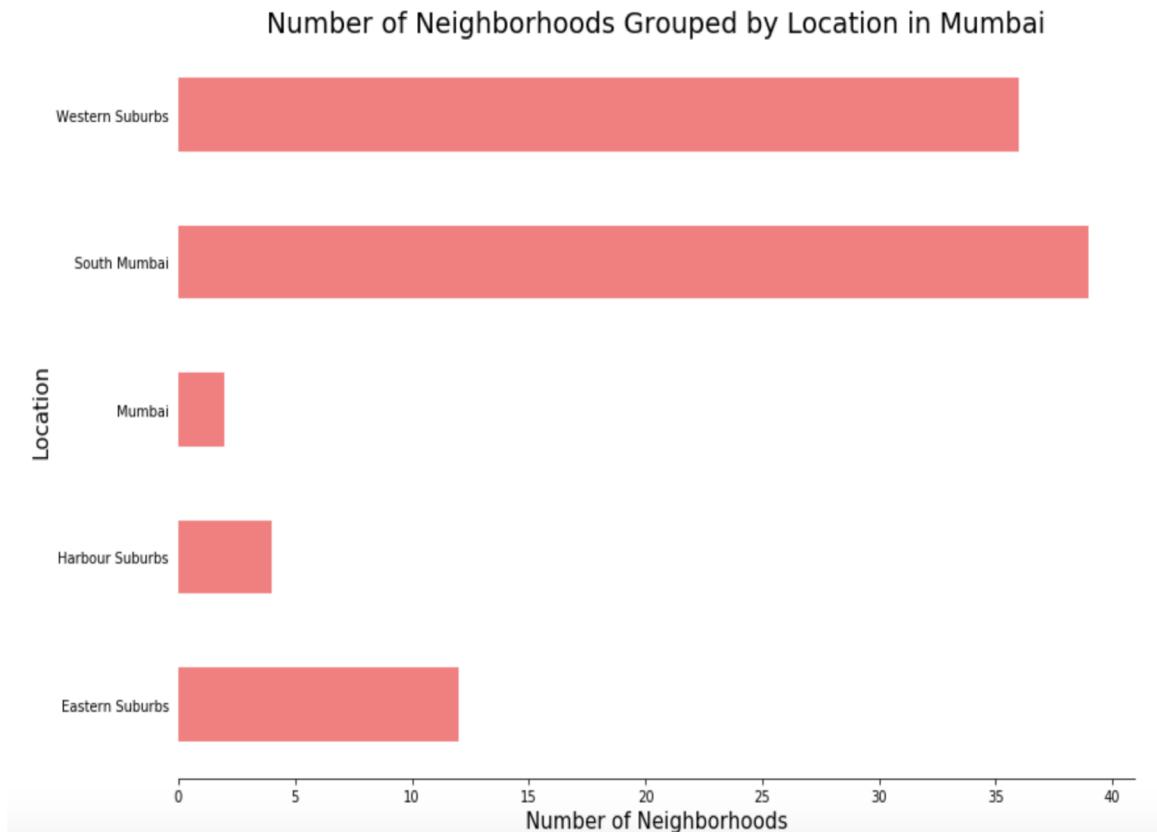


Figure 4: Number of neighborhoods grouped by location

Methodology Continued

- **Data Visualization**

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

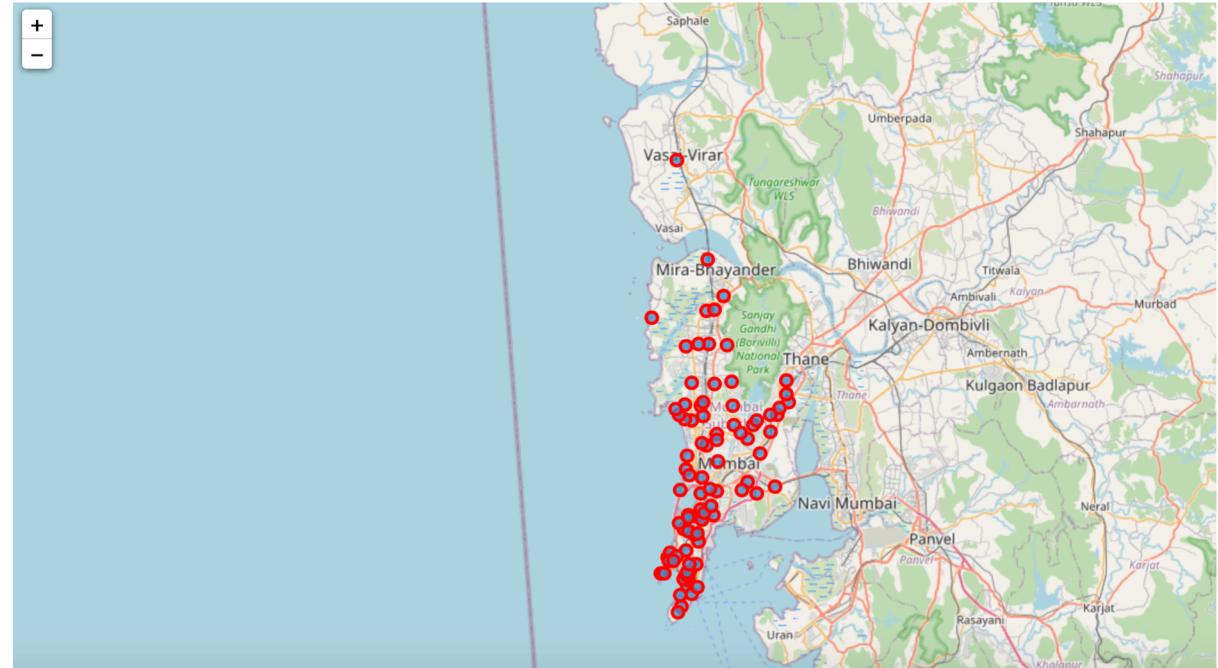


Figure 5: The neighborhood spread across Mumbai

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	ATM	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	... Trail	Train	Train Station	Vegetarian / Vegan Restaurant	Whisky Bar	Wine Bar	Wine Shop	Women's Store	Yoga Studio	Zoo	
0	Amboli	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0
1	Chakala, Andheri	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.047619	0.0	0.0	0.000000	0.000000	0.0	0.0
2	D.N. Nagar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.043478	0.0	0.0	0.000000	0.021739	0.0	0.0
3	Four Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.030303	0.0	0.0	0.000000	0.015152	0.0	0.0
4	Lokhandwala	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.010753	0.0	0.0	0.000000	0.010753	0.0	0.0
5	Marol	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0
6	Sahar	0.0	0.0	0.033333	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0
7	Seven Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.014925	...	0.0	0.0	0.0	0.029851	0.0	0.0	0.000000	0.000000	0.0	0.0
8	Versova	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.025000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.025	0.000000	0.0	0.0
9	Mira Road	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.066667	0.0	0.0

10 rows x 221 columns

Figure 6: 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 5 clusters and thus the n_clusters parameter in KMeans clustering was set to 5



Figure 7: Silhouette 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 top 10 most common venues dataframe to provide the final results
- This dataframe is shown on the next slide

Results Continued

	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	10th Most Common Venue
0	Amboli	Western Suburbs	19.1293	72.8464	1	Indian Restaurant	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	1	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	0	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Western Suburbs	19.1263	72.8243	0	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Western Suburbs	19.1432	72.8249	0	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Western Suburbs	19.1192	72.8827	1	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Western Suburbs	19.1027	72.8626	0	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Western Suburbs	19.1315	72.817	0	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Western Suburbs	19.1377	72.8135	0	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Western Suburbs	19.2657	72.8711	1	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure 8: First 10 rows of clustering neighborhoods in Mumbai dataframe

Results Continued – Cluster 1

- Analyzing each cluster individually

	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
2	D.N. Nagar	Western Suburbs	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Western Suburbs	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Western Suburbs	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
6	Sahar	Western Suburbs	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Western Suburbs	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Western Suburbs	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
10	Bhayandar	Western Suburbs	Bakery	Pizza Place	Lake	Electronics Store	Ice Cream Shop	Diner	Indian Restaurant	Food Truck	Restaurant	Factory
11	Uttan	Western Suburbs	ATM	Café	Food Truck	Bakery	River	Restaurant	Metro Station	Design Studio	Fish Market	Fish & Chips Shop
12	Bandstand Promenade	Western Suburbs	Coffee Shop	Café	Tea Room	Scenic Lookout	Deli / Bodega	Performing Arts Venue	Food Truck	Indian Restaurant	Fast Food Restaurant	Lounge
13	Kherwadi	Western Suburbs	Café	Indian Restaurant	Hookah Bar	Restaurant	Italian Restaurant	Pizza Place	Seafood Restaurant	Bar	Chinese Restaurant	South American Restaurant

Figure 9: Cluster 1

Results Continued – Cluster 2

	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	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Western Suburbs	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
5	Marol	Western Suburbs	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
9	Mira Road	Western Suburbs	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar
15	I.C. Colony	Western Suburbs	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Bar	Dessert Shop	Pizza Place	Juice Bar	Basketball Court	Soccer Field
18	Aarey Milk Colony	Western Suburbs	Fast Food Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Shopping Mall	Breakfast Spot	Snack Place	Bookstore	Coffee Shop	Donut Shop	Bar
19	Bangur Nagar	Western Suburbs	Coffee Shop	Indian Restaurant	Clothing Store	Multiplex	Fast Food Restaurant	Pizza Place	Electronics Store	Smoke Shop	Middle Eastern Restaurant	Pub
20	Jogeshwari West	Western Suburbs	Indian Restaurant	Men's Store	Asian Restaurant	Fried Chicken Joint	Mughlai Restaurant	Café	Chinese Restaurant	Smoke Shop	Flea Market	Ice Cream Shop
21	Juhu	Western Suburbs	Indian Restaurant	Coffee Shop	Flower Shop	Shopping Mall	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Movie Theater	Café	Lounge	Breakfast Spot
23	Poisar	Western Suburbs	Indian Restaurant	Train Station	Dessert Shop	Sandwich Place	Food	Electronics Store	Fast Food Restaurant	Men's Store	Mexican Restaurant	Snack Place

Figure 10: Cluster 2

Results Continued – Cluster 3, 4, and 5

	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
36	Bhandup	Eastern Suburbs	Train Station	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Zoo	Donut Shop	Flea Market	Fish Market	Fish & Chips Shop	Field
40	Kanjurmarg	Eastern Suburbs	Train Station	Gift Shop	Gym	Multiplex	Asian Restaurant	Chinese Restaurant	Cupcake Shop	Hotel	Field	Fast Food Restaurant
90	Dava Bazaar	South Mumbai	Train Station	Indian Restaurant	Cupcake Shop	Hotel	Fish Market	Café	Fast Food Restaurant	Coffee Shop	Asian Restaurant	Smoke Shop

Figure 11: Cluster 3

	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
50	Mankhurd	Harbour Suburbs	Train Station	Coffee Shop	Sports Bar	Bus Station	Zoo	Electronics Store	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant

Figure 12: Cluster 4

	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	Fast Food Restaurant	Dive Bar	Fish Market	Fish & Chips Shop	Field	Farmers Market	Farm	Falafel Restaurant

Figure 13: Cluster 5

Results Continued

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

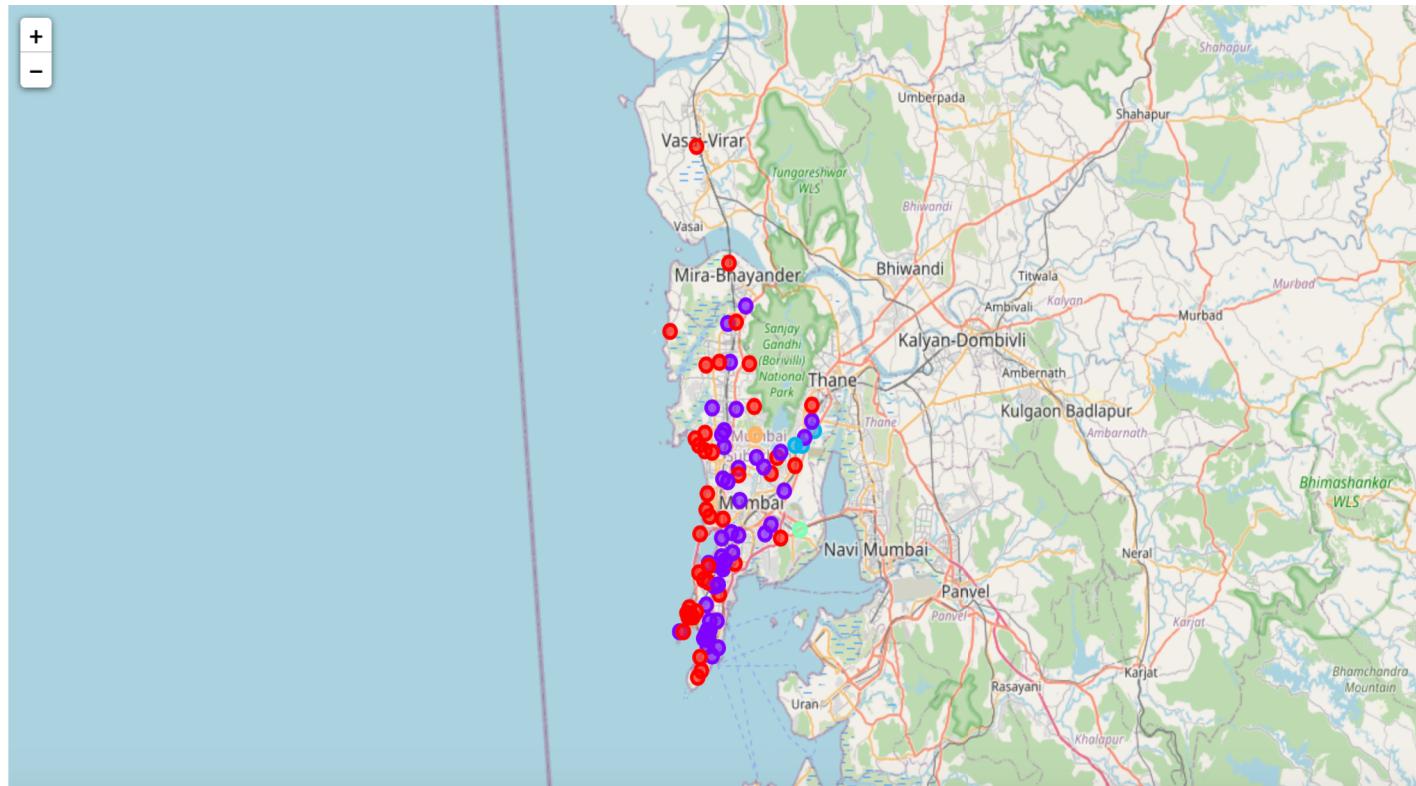


Figure 14: Visualizing the clustering of neighborhoods in Mumbai

Discussion

- By analyzing the clusters, it is evident that neighborhoods in cluster 3, 4, and 5 are not well suited for a new restaurant since they contain a high degree of other venues like train station, bus station, fish market, gym and smoke shop
- Comparing clusters 1 and 2, we find that neighborhoods in cluster 1 contain a much higher degree of food joints
- Neighborhoods in cluster 2 contain many other venues like soccer field, flea market, smoke shop, gym, train station, dance studio and music store
- Thus, the most optimal neighborhoods for starting a restaurant are in cluster 1

Discussion Continued

- Plotting final results using Folium library in python

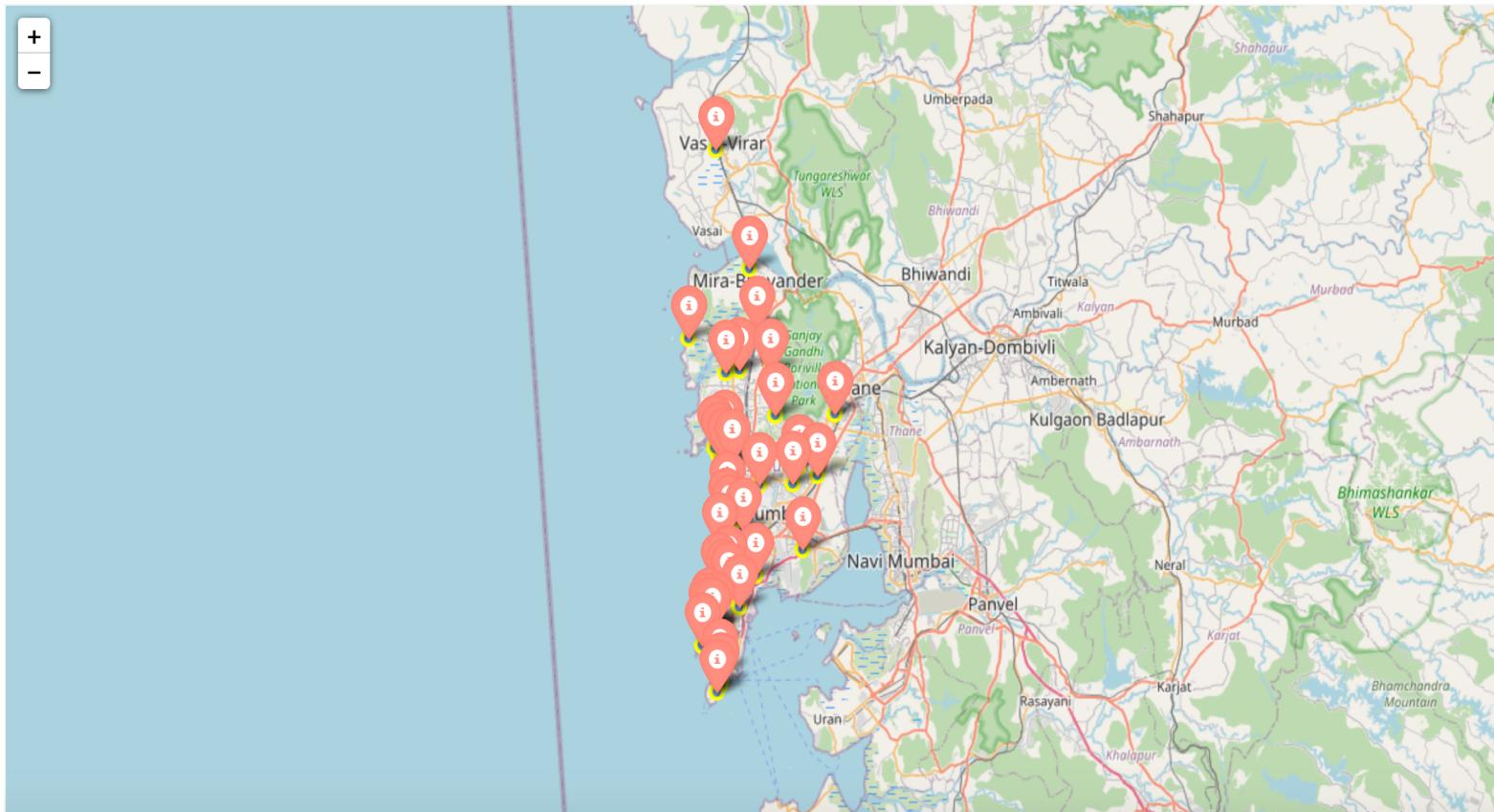


Figure 15: Neighborhoods in Mumbai that are most suited
for starting a new restaurant (Cluster 1)

Conclusion

- 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