Analyzing the Neighborhoods in Mumbai for Starting a Restaurant

Applied Data Science Capstone Project

By: Jean Charles

Date: July 13, 2024

Table of Contents

Introduction

Mumbai, the financial capital of India, is among the most densely populated cities globally. Located on India's west coast, it attracts significant tourism from around the world each year. Having grown up in Mumbai, I have a deep affection for the city. It stands as a major global hub, characterized by its remarkable diversity, with residents from various ethnic backgrounds. This multicultural fabric has introduced a wide range of international cuisines to the city.

Given India's general passion for food, and my personal enthusiasm for exploring different cuisines and flavors, this project aims to analyse Mumbai's neighbourhoods to identify potential locations for opening a restaurant. The findings will be valuable for business owners and entrepreneurs seeking to invest in the restaurant industry in Mumbai. The primary objective of this project is to meticulously analyse relevant data to provide informed recommendations for stakeholders.

Data Collection

The following data is required for the project:

1. **Neighbourhood Data of Mumbai**

2. **Geographical Coordinates of Mumbai and All Neighbourhoods in Mumbai**

3. **Venue Data for Neighbourhoods in Mumbai**Neighbourhoods Data

The data for the neighbourhoods in Mumbai was scraped from [this Wikipedia page] (https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Mumbai). The data is read into a Pandas Data Frame using the `read HTML () ` method. The main reason for this approach is that the Wikipedia page provides a comprehensive and detailed table that can be easily scraped using Pandas' `read_html()` method. Figure 1 displays the top 10 rows of the DataFrame.

	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 1: Top 10 rows of Mumbai neighborhoods data scraped from Wikipedia.

Geographical Coordinates

The geographical coordinates for Mumbai were retrieved using the GeoPy library in Python. These coordinates are essential for generating a map of Mumbai using the Folium library in Python. Figure 2 presents the code used to obtain the geographical coordinates of Mumbai.

```
address = 'Mumbai, IN'
geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Mumbai are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinates of Mumbai are 19.0759899, 72.8773928.
```

Figure 2: Obtaining geographical coordinates of Mumbai.

The geocoder library in Python was utilized to acquire latitude and longitude data for various neighbourhoods in Mumbai. These coordinates were compared with those provided on Wikipedia within our Data Frame. If the absolute difference exceeded 0.001, the coordinates from geocoder were used to refine our dataset. Subsequently, these refined coordinates were employed for plotting neighbourhoods using the Folium library in Python. Figure 3 displays the comparison between the coordinates obtained from Wikipedia ('Latitude', 'Longitude') and those obtained from geocoder ('Latitude1', 'Longitude1'). It also presents the absolute differences between the two latitude columns ('Latdiff') and the two longitude columns ('Longdiff'). Only the top 10 rows are shown.

	Neighborhood	Location	Latitude	Longitude	Latitude1	Longitude1	Latdiff	Longdiff
0	Amboli	Western Suburbs	19.1293	72.8464	19.1291	72.8464	0.00024	0.00304
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	19.1084	72.8623	0.003028	0.001497
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	19.1251	72.8325	0.000965	0.001107
3	Four Bungalows	Western Suburbs	19.1263	72.8243	19.1263	72.8243	0.001606	0.00288
4	Lokhandwala	Western Suburbs	19.1432	72.8249	19.1432	72.8249	0.012345	0.0044
5	Marol	Western Suburbs	19.1192	72.8827	19.1191	72.8828	0.000169	6.7e-05
6	Sahar	Western Suburbs	19.1027	72.8626	19.1027	72.8626	0.00376476	0.00464166
7	Seven Bungalows	Western Suburbs	19.1315	72.817	19.1315	72.8165	0.00240802	0.000558001
8	Versova	Western Suburbs	19.1377	72.8135	19.1377	72.8135	0.01769	0.00652
9	Mira Road	Western Suburbs	19.2657	72.8711	19.2657	72.8707	0.0184624	0.000418149

Figure 3: Absolute difference between latitude and longitude values obtained from Wikipedia and Geocoder.

Figure 4 displays the top 10 rows of the finalized Mumbai neighbourhoods DataFrame. This DataFrame includes refined latitude and longitude values obtained after comparing and replacing coordinates from the geocoder library where necessary. Additionally, unnecessary columns have been dropped for clarity and focus.

	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 4: Final Mumbai neighborhoods dataframe.

Venue Data

Venue data was collected using the Foursquare API, providing recommendations for all neighbourhoods in Mumbai. This dataset serves to analyse popular venues across neighbourhoods and construct an unsupervised learning model for clustering. Each neighbourhood's venue recommendations were retrieved with a limit of 200 venues per neighbourhood and within a radius of 1 km from the neighbourhood's geographical coordinates. Figure 5 displays the top 10 rows of cleaned data obtained from the

Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.1293	72.84644	Cafe Arfa	19.128930	72.847140	Indian Restaurant
1	Amboli	19.1293	72.84644	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
2	Amboli	19.1293	72.84644	Shawarma Factory	19.124591	72.840398	Falafel Restaurant
3	Amboli	19.1293	72.84644	Jaffer Bhai's Delhi Darbar	19.137714	72.845909	Mughlai Restaurant
4	Amboli	19.1293	72.84644	Narayan Sandwich	19.121398	72.850270	Sandwich Place
5	Amboli	19.1293	72.84644	Persia Darbar	19.136952	72.846822	Indian Restaurant
6	Amboli	19.1293	72.84644	Domino's Pizza	19.131000	72.848000	Pizza Place
7	Amboli	19.1293	72.84644	Garden Court	19.127188	72.837478	Indian Restaurant
8	Amboli	19.1293	72.84644	Subway	19.127860	72.844461	Sandwich Place
9	Amboli	19.1293	72.84644	Sarvodaya Veg. Restaurant	19.123760	72.850893	Indian Restaurant

Figure 5: Data obtained from Foursquare API after cleaning.

Methodology

This section provides details for the methodology used in the project.

Data Visualization

Figure 6 presents a bar plot illustrating the distribution of neighbourhoods across different locations in Mumbai. This visualization provides a basic understanding of the number of neighbourhoods in each area of the city.

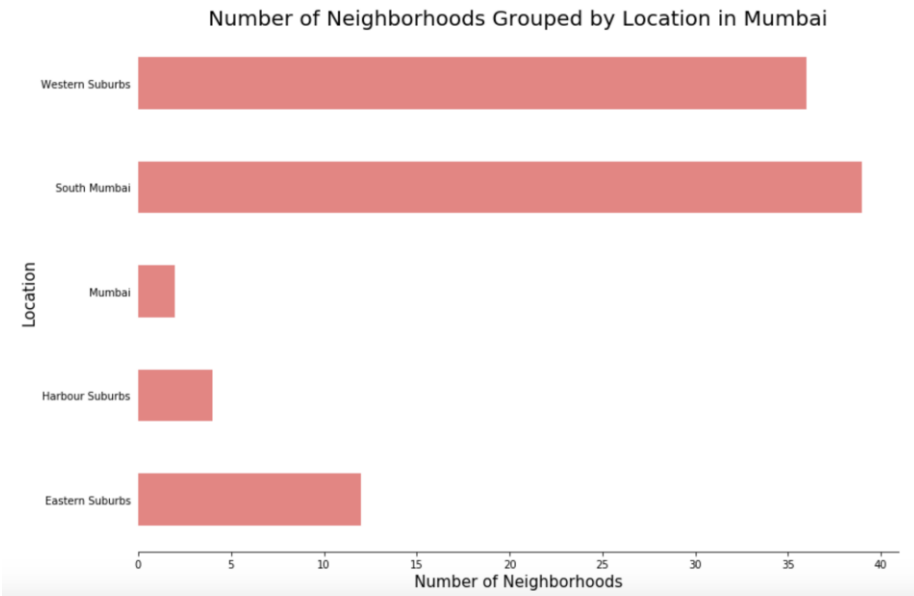

Figure 6: Number of neighbourhoods grouped by location.

Figure 7 displays a map generated using Folium, illustrating the distribution of neighbourhoods across Mumbai. The map highlights how neighbourhoods are spread throughout the city, including those located at the outskirts categorized simply as "Mumbai."

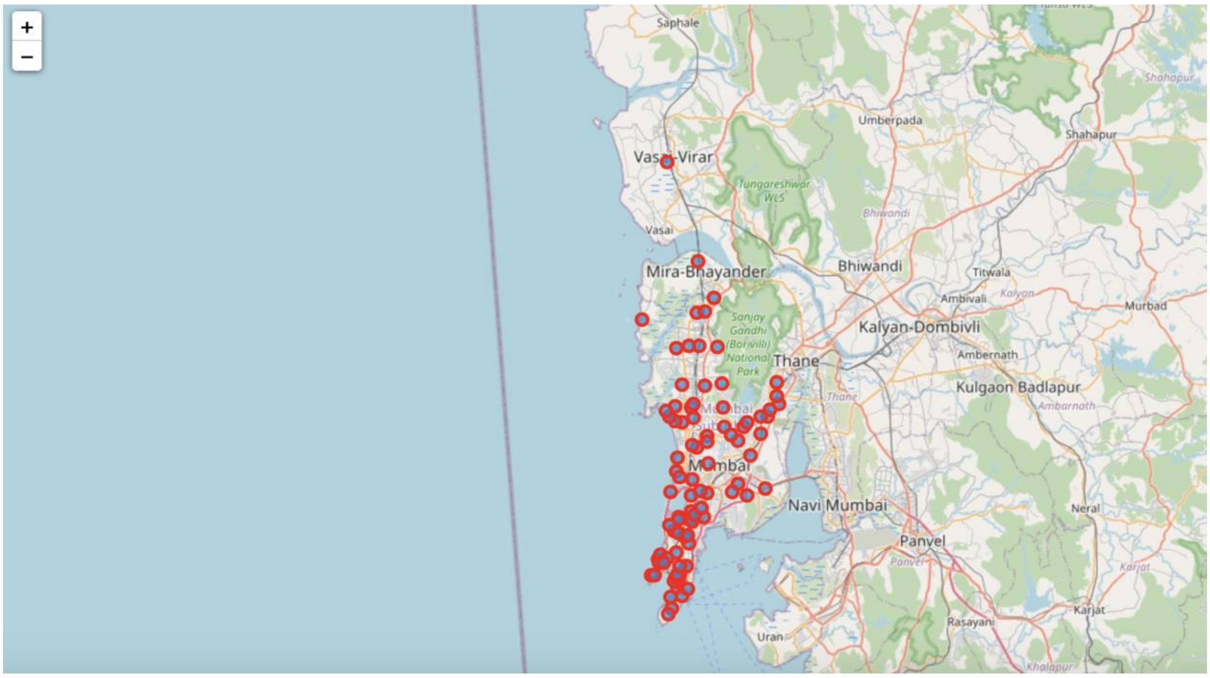

Figure 7: Depicting the neighbourhood spread across Mumbai.

Feature Extraction

Feature extraction involved transforming the Foursquare API data (referenced in Figure 5) to prepare for building the unsupervised learning model. This process required converting the "Venue Category" column into a numeric format suitable for modelling. One-hot Encoding was employed for this purpose, where each unique venue category was converted into a separate column. A value of 1 was assigned to a neighbourhood's row in a specific category column if it included venues from that category. column and if a neighbourhood venue does not belong to the category, the value would be 0. This process was repeated

A sparse matrix was generated from all venues across neighbourhoods using One-hot Encoding, resulting in a Data Frame where each neighbourhood's row contains columns representing unique venue categories with values of 1 or 0 indicating presence or absence of venues in those categories. Subsequently, the Data Frame was grouped by neighbourhood name, and the average value across all categories was computed. Figure 8 presents the top 10 rows of this averaged DataFrame.

	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	0.000	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	0.000	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	0.000	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	0.000	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	0.000	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	0.000	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	0.000	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	0.000	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	0.000	0.066667	0.0	0.0

10 rows × 221 columns

Figure 8: One-hot Encoding resulting dataframe.

Feature Extraction

To prepare for the unsupervised learning model, a sparse matrix was created using One-hot Encoding across all neighbourhoods and venue categories from the Foursquare API data (as depicted in Figure 5). Each neighbourhood's row in this matrix indicates the presence (1) or absence (0) of venues in various categories. After grouping by neighbourhood name, the average presence across categories was computed, resulting in a DataFrame (Figure 8) where most values are 0 due to the diversity of venue categories and varying presence in neighbourhoods.

Top 10 Common Venues DataFrame

Additionally, a DataFrame was constructed listing the top 10 most frequent venue categories across all neighbourhoods. Although not directly part of Feature Extraction, this DataFrame provides insight into common venue types and will be used later to integrate findings from the unsupervised learning model. Figure 9 displays the top 10 rows of this DataFrame.

	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 9: Top 10 most common venues for neighborhoods.

Unsupervised Learning

K-means, an unsupervised learning technique, was employed to cluster neighbourhoods based on nearby venue categories. An essential step in this process was determining the optimal number of clusters. This was achieved using the Silhouette score, calculated across a range of clusters from 2 to 15. Figure 10 presents the resulting number of clusters identified and their corresponding Silhouette scores, aiding in the selection of the optimal clustering configuration.

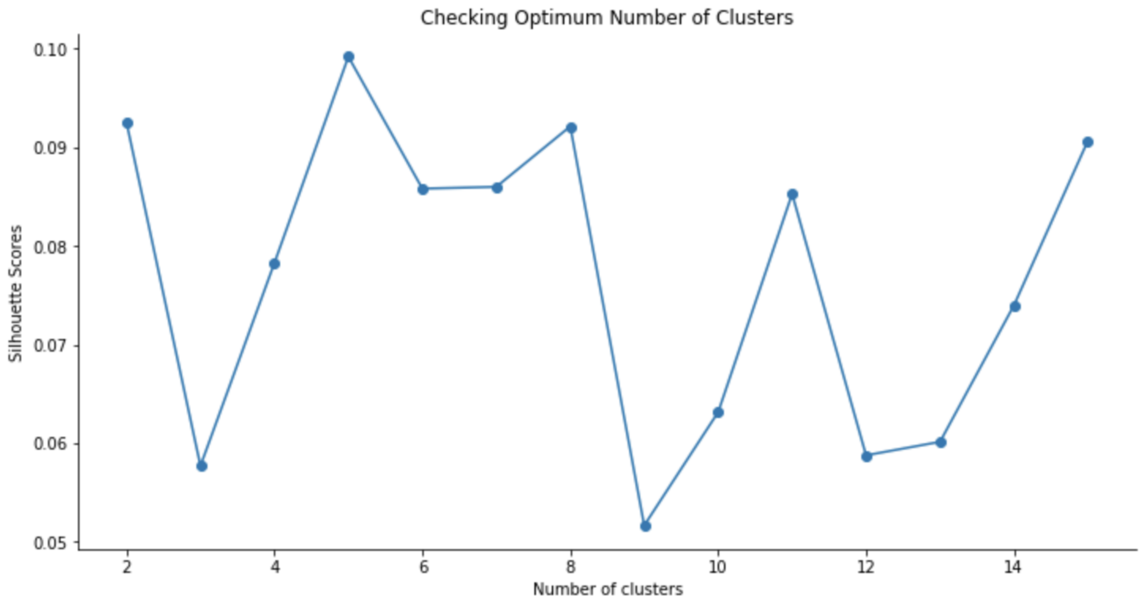

Figure 10: Silhouette scores for different number of clusters.

While the Silhouette scores indicate moderate clustering quality across various k-values, it is apparent that higher scores are not achieved consistently. Nonetheless, the data will be clustered using the k-means algorithm. Based on the analysis shown in Figure 10, 5 clusters will be utilized as it yields the highest Silhouette score within the range evaluated. This choice aims to achieve the most effective clustering of neighbourhoods based on venue categories.

Results

Using the k-means clustering algorithm, neighborhoods in Mumbai were grouped into clusters based on their nearby venue categories. Each neighborhood was assigned a cluster label representative of its cluster assignment. These cluster labels were integrated into the DataFrame from Figure 9, along with the Location, Latitude, and Longitude columns, providing a comprehensive summary of the clustering results

Figure 11: Clustering neighbourhoods in Mumbai.

To delve deeper into specific clusters, neighbourhoods within each cluster were extracted based on their assigned cluster labels. Below are details for each cluster, showcasing neighbourhoods within clusters that contain a significant number of neighbourhoods. Each cluster's details are presented with a limit of 10 rows for clusters with extensive neighbourhood representation.

	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 12: Cluster 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
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 13: 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
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 14: 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 15: 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 16: Cluster 5.

Figure 17 depicts a map of Mumbai where neighbourhoods are plotted with markers of different colours to distinguish between clusters identified through the k-means clustering model. This visualization provides an overview of how neighbourhoods are spatially distributed across Mumbai, grouped according to their similarities in nearby venue categories as determined by the clustering analysis.

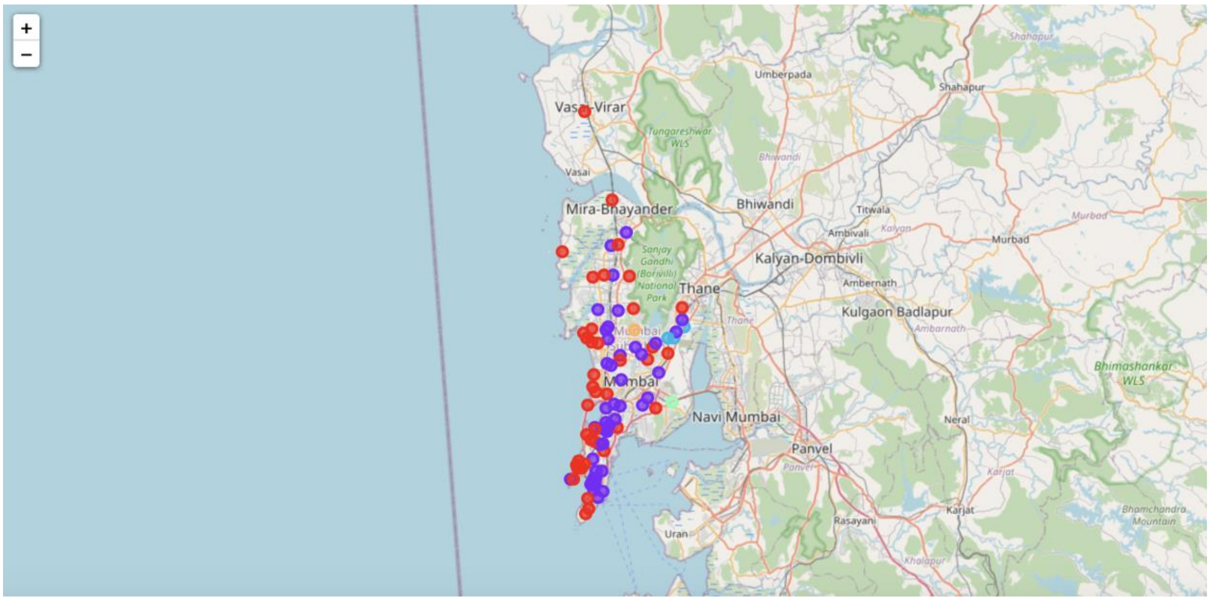

Figure 17: Visualizing the clustering of neighbourhoods in Mumbai.

Discussion

Upon analysing the five clusters generated from the k-means clustering:

- **Clusters 3, 4, and 5**: These clusters exhibit a low percentage of restaurants, hotels, cafes, and pubs among their top 10 common venues. Instead, they are characterized by venues such as train stations, bus stations, fish markets, gyms, performing arts venues, and smoke shops. These neighbourhoods are less suitable for opening a new restaurant.
- **Clusters 1 and 2**: These clusters feature a higher concentration of restaurants, hotels, multiplexes, cafes, bars, and other food joints in their top 10 common venues. They are therefore more conducive to starting a new restaurant.

Comparing clusters 1 and 2:

- **Cluster 1**: Neighbourhoods here show a diverse range of food joints including restaurants, tea rooms, bakeries, cafes, steakhouses, and pubs. They also feature a variety of cuisines such as Japanese, Indian, Chinese, Italian, and seafood restaurants.
- **Cluster 2**: While Indian Restaurants dominate as the most common venue in cluster 2, further inspection reveals a mix of other venues like soccer fields, flea markets, smoke shops, gyms, train stations, dance studios, music stores, and cosmetics shops.

Recommendation:

Given the higher diversity and concentration of food-related venues in cluster 1, it is recommended to open the new restaurant in neighbourhoods belonging to this cluster.

Mapping Recommended Neighbourhoods

Figure 18 displays a map highlighting neighbourhoods from cluster 1, ideal for opening a new restaurant based on their favourable venue profile.

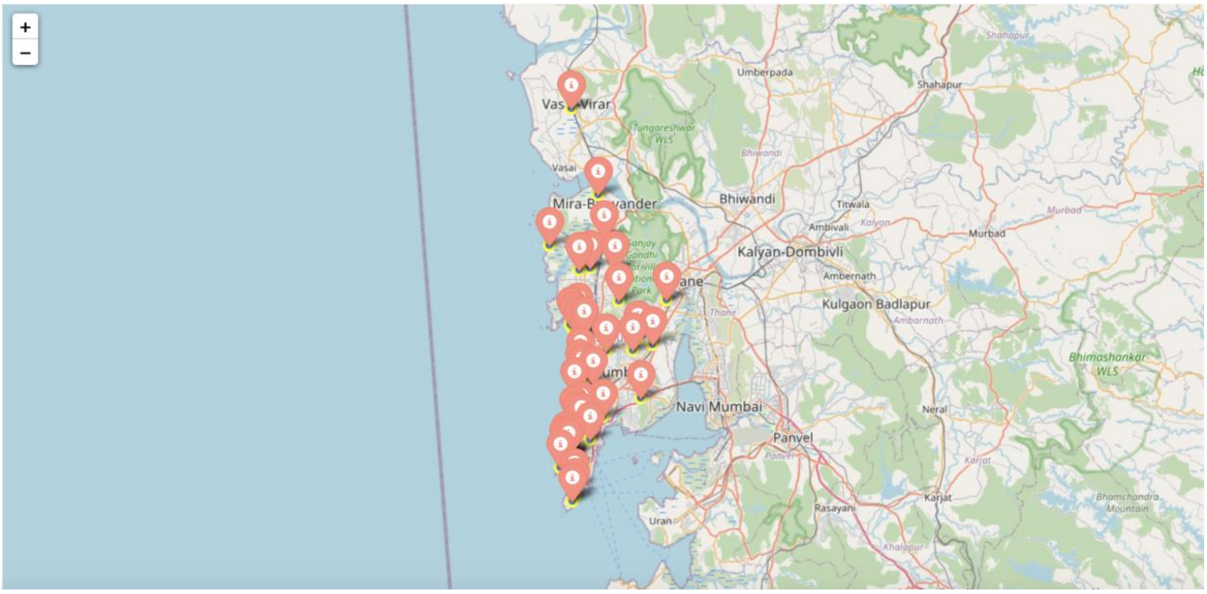

Figure 18: Neighbourhoods most suited for starting a new restaurant.

Conclusion

The neighbourhoods in Mumbai, India have been systematically analysed to identify optimal locations for opening a new restaurant. Based on the analysis, neighbourhoods in cluster 1 have been recommended as the most suitable locations. Figure 18 illustrates these recommended neighbourhoods on a map.

Considerations for Stakeholders and Investors: While cluster analysis provides a solid foundation, stakeholders and investors are encouraged to further refine their decision-making process by considering additional factors such as transportation accessibility, legal requirements, and associated costs. These factors, although outside the scope of this project, play crucial roles in the success of restaurant ventures.

By leveraging this analytical approach, stakeholders can make informed decisions to maximize the potential success of their restaurant investments in Mumbai.

Final Comments

Note 1: In order to view the code for this project, kindly refer to the notebook on the github repository at: