

Battle of Neighbourhoods – Mumbai

Final Report

Introduction:

In India, many working professionals who hail from different parts of the country, have to shift to metro cities for work and this project aims to help these people in exploring the neighbourhoods so that they can make wise choice while relocating. Many factors should be looked at when exploring neighbourhoods such as Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, ease of accessing to Cafe, etc.

This Project aim to create an analysis of features for a people migrating to Mumbai to search a best neighbourhood as a comparative analysis between neighbourhoods.

Problem Which Tried to Solve:

The main objective is to separate neighbourhoods by their similarities based on nearby venues, for the people who are shifting there.

The Location:

Mumbai is a popular corporate destination in India and lots of new people relaocate in this city every day. As a result, it is one of the most diverse and multicultural areas in the country, being home to various religious groups and places of worship.

Foursquare API:

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Clustering Approach:

We will have to identify similar neighbourhoods in the city of Mumbai and separate different clusters based on the features discussed above. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Libraries Which are Used to Developed the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

Data Description:

Data Link: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

Will use Mumbai Areas table which can be scraped from above link. Dataset consists of latitude and longitude, Area and Location.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

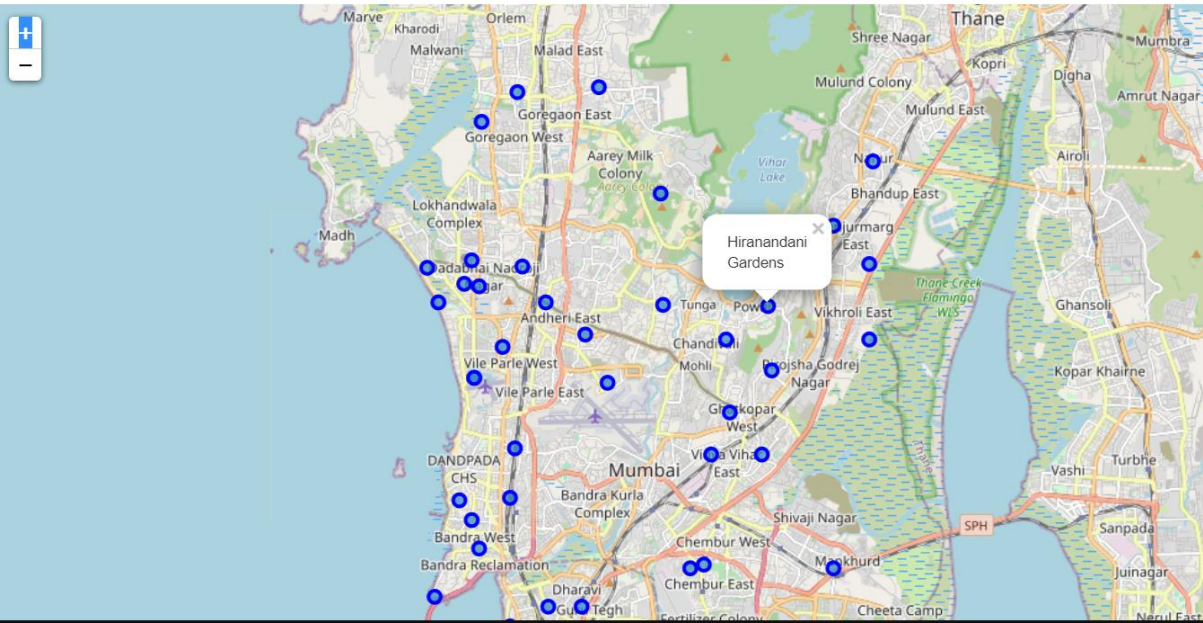
Methodology

K-means clustering approach was used to separate neighbourhoods that are different from each other in terms of nearby facilities.

Results

Five clusters amongst 94 neighbourhoods in Mumbai were generated with one of the cluster having only 4 neighbourhoods which indicated that nearby venues in these neighbourhoods are different compared to other clusters.

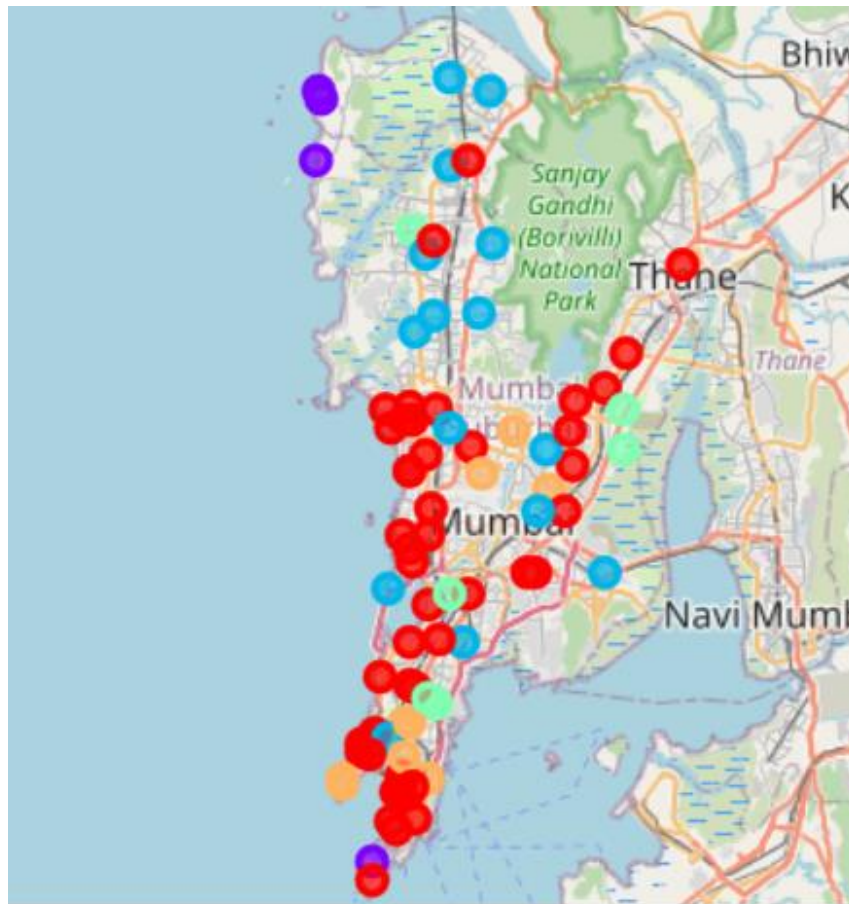
Map of neighbourhoods of Mumbai



Most common nearby venues for each neighbourhood

	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	Agripada	Indian Restaurant	Coffee Shop	Japanese Restaurant	Racetrack	Restaurant	Bank	Bakery	Electronics Store	Flea Market	Fish Market
1	Altamount Road	Café	Bakery	Sandwich Place	Coffee Shop	Chinese Restaurant	Indian Restaurant	Snack Place	Park	Pizza Place	Electronics Store
2	Amboli	Coffee Shop	Pizza Place	Asian Restaurant	Indian Restaurant	Bar	Italian Restaurant	Fast Food Restaurant	Electronics Store	Chinese Restaurant	Sandwich Place
3	Amrut Nagar	Indian Restaurant	Café	Clothing Store	Diner	Pizza Place	Chinese Restaurant	Lounge	Fast Food Restaurant	Gourmet Shop	Food & Drink Shop
4	Asalfa	Convenience Store	Indian Restaurant	Bus Station	Donut Shop	Yoga Studio	Event Space	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop

Clusters of Neighbourhoods



Discussion

In this project, I analyzed 93 neighbourhoods of Mumbai city to find similar neighbourhoods and separate dissimilar neighbourhoods. I extracted data of neighbourhoods of Mumbai from Wikipedia page. After cleaning and preprocessing the data, I used Foursquare API to generate nearby venues data for each of neighbourhoods. Then proceeded with plotting top 10 most common venues in each neighbourhood. This data was fed into k-means clustering model for segmentation

Conclusion

In this project, using k-means cluster algorithm I separated the neighbourhood into five different clusters which have very-similar neighbourhoods around them. So a person shifting to Mumbai and searching for a place can get an idea looking at the clusters about which neighbourhoods offer similar nearby facilities and then he can decide based on budget and other details where he wants to reside.