Battle of Neighbourhood

REPORT

Introduction

The land of Mumbai is full of opportunities. Located on Maharashtra's coast, Mumbai is India's most-populous city, and its one of the most densely populated urban areas in the world. There is something for everyone here. Mumbai is already full of venues. There are innumerable food joints, restaurants. But if someone wants to open a restaurant, 'where' and 'of what kind' are the obvious questions. And that is what we are here to deal with.

Business Problem

We are here to answer the following:

1. List and visualise the major neighbourhood in Mumbai.
2. What restaurants are there in each of the neighbourhood?
3. Which areas lack what kind of restaurants?
4. What is the best location for a particular kind of restaurant?
5. What kind of restaurants would survive in a particular location?

Target Audience

The target audience is not a big domain. It only suits those who want to invest in opening a restaurant in Mumbai but need assistance regarding the 'where' and 'of what kind' questions. It will help the investors, or business persons in making an informed decision regarding the prospect of opening a restaurant in the city of Mumbai.

Data Description

For this project, we need the following data:

- Mumbai areas Source - https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai Description - Location, longitude and the latitude of areas in Mumbai.

- Different restaurants in the neighbourhood of Mumbai city. Source - Foursquare API Description - By using API, we get the name of the restaurant and the categories.

Using this data will allow exploration and examination to solve the business problem. We will use data science skills such as data scraping, working with API, data cleaning, data wrangling and map visualisation(Folium) and machine learning(k-means clustering).

Methodology:

Data Exploration:

Firstly, we will get the list of neighbourhoods in Mumbai. We have the list available at https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai. We have to do web scraping using Python requests to extract the list of neighbourhood data. This list contains the name, location and the coordinates of each of the neighbourhood in Mumbai.

Data Visualisation:

We will visualise the neighbourhood using the folium package of Python. This map will make sure that the coordinates given in the table are indeed the coordinates of the neighbourhood of Mumbai.

Finding the Venues around Mumbai:

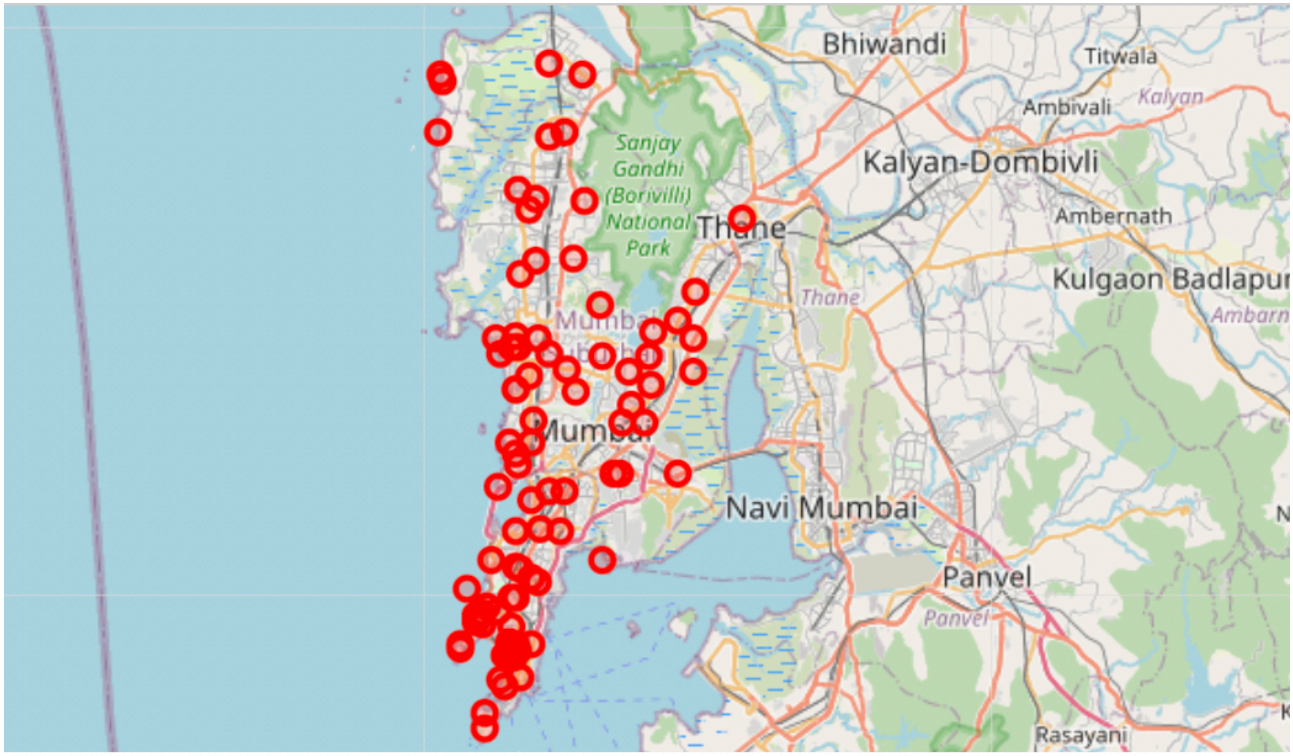
We will now use the foursquare API to get all the venues that are within a radius of 500m. We need to register a Foursquare developer account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venues. With the data, we can examine the number of venues returned for each neighbourhood, as well as the number of a times a particular category of venue is returned for a particular neighbourhood.

Data Clustering:

We will perform clustering on the Venue data by using the k-means clustering algorithm. It is a kind of supervised learning where each of the data points are allotted cluster, overall minimising the intra-cluster distance and maximising the inter-cluster euclidean distance. Here, we will divide the neighbourhood into 3 clusters, depending on the kind of restaurants they have. It will help us in answering the question as to what kind of restaurants can be established in which of the locations.

Result:

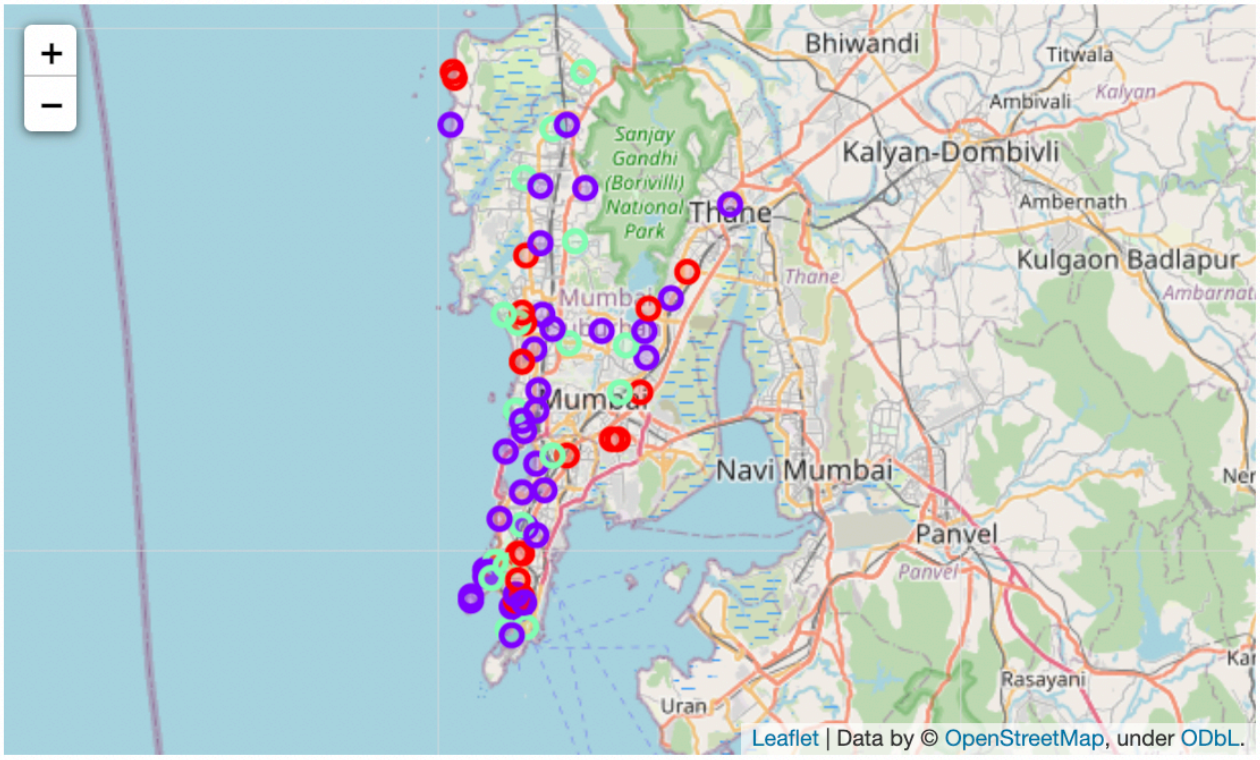
1. Displaying the neighbourhood in Mumbai:



2. The different kinds of restaurants in each of the neighbourhood in Mumbai

	Neighbourhood	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
0	Agripada	Indian Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant	Bengali Restaurant
1	Altamount Road	Indian Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant	Bengali Restaurant
2	Amboli	Halal Restaurant	Asian Restaurant	Chinese Restaurant	Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant
3	Amrut Nagar	Indian Restaurant	Fast Food Restaurant	Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant	Chinese Restaurant
4	Bandstand Promenade	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant
...	...	...	...	...	...	...	...	...
68	Uttan	Indian Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant	Bengali Restaurant
69	Vidyavihar	Restaurant	Fast Food Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant
70	Vile Parle	Indian Restaurant	Japanese Restaurant	Seafood Restaurant	South Indian Restaurant	Goan Restaurant	Afghan Restaurant	American Restaurant
71	Walkeshwar	Fast Food Restaurant	Indian Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant	Asian Restaurant
72	Worli	Indian Restaurant	Fast Food Restaurant	Seafood Restaurant	South Indian Restaurant	Greek Restaurant	Afghan Restaurant	American Restaurant

3. Clustering based on the different kinds of restaurants in the neighbourhood of Mumbai city.



Discussion and Observations:

As we can observe, cluster1 has a dominance of Indian restaurants with it being the most common venue followed by South Indian restaurants. Hence, anyone planning to open an Indian or South Indian restaurant in these regions will not be profitable as there would be intense competition already. Cluster2 has a mix of different restaurants so any restaurant you open in those regions will result in good enough competition, but it would not be the worst decision to open a restaurant here.

Cluster3, as we can see, is dominated by restaurants that are not Indian or south Indian. Hence, it would be profitable if someone opens an Indian restaurant here or even South Indian. The neighbourhood in this cluster has enough of Chinese and Italian and seafood restaurant so it would not be advisable to invest in those kinds of restaurants.

In a nutshell, we can say the following overall: i) Neighbourhoods in cluster1 - Any restaurant but Indian or South Indian. ii) Neighbourhoods in cluster2 - Every restaurant will have competition already. Still, no Indian restaurant as they are still in plenty here. iii) Neighbourhoods in cluster3 - Indian restaurants are not much here. So, they will have good odds of making a profit here. Chinese restaurants are already there in abundance so it would be good to avoid establishing another one here.

Conclusion:

I have completed the process of identifying the business problems, mentioning the data required, preprocessing the data, visualising the results, performing machine learning by k-means, clustering the data into 3 clusters based on their frequency similarities, and then reaching to a definitive solution to the business problems. the aim of the project is to provide recommendations to the relevant investors regarding the best kinds of restaurants that can be opened in a given location, or to find the best location to open a definite kind of a restaurant .