# Coursera Capstone Project

## The Battle of Neighborhoods

## Opening a Restaurant in Mumbai, India

By: Bhaskar Sahoo



Fig.1: City of Mumbai (Photo source: https://www.maakan.com/ )

# Introduction

Restaurants are places for spending quality family time outside home and often provide escape from grinding household chores. The number of working people have increased in recent years and so has the demand for good restaurants. Eating out is fun and exploring different cuisines is exciting. From an economic point of view, restaurant business is a major service industry providing employment to many people at various levels. Moreover, this business support other sectors such as tourism, where it caters to the taste of tourists in a place outside their home country. Restaurant business one of the most profitable ventures in the service industry.

Mumbai is second most populous city in India and the seventh most populous city in the world with a population of almost 20 million (Wikipedia). It is the financial and commercial capital of the India and have hundreds of thousands of people from all over the country. As a metropolis it provides immense business potential and restaurant business is one among them. However,

opening a new restaurant requires lot of strategies and planning. Particularly, the location of the restaurant is the most important decision as it will determine the success/failure of the business venture. Demography around the proposed location and land use (e.g. residential, industrial or official buildings, etc.) play important roles as well.

## Business Problem

Location and land use of the proposed location for a restaurant are instrumental in the success of a new restaurant. The business owner needs to review and select the location where there is a demand for the type of cuisine it caters. This would ensure profitable returns and subsequent success of the business. The objective of this project to analyze and select the best locations in the neighbourhood of Mumbai to open a new restaurant. Machine learning and data science methodologies will be used to this analysis in selecting a suitable restaurant location.

## Target Audience

This project will particularly be interest to aspiring small business owners who are looking for opportunities in the culinary industry in Mumbai. Also, being a multi-cultural hub for people from various culinary choices, this would interest workers in the food industry (e.g. chefs and food experts). Moreover, it will be useful for the local populace to whom the business would cater their needs.

## Data

The data required for this project were collected from Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Mumbai) using web scarping techniques. The following information were required for the analysis.

- List of all the neighbourhoods in Mumbai.
- Geolocations of the neighbourhoods (Latitude and Longitude)
- Venue data (for Clustering)

## Methodology

For the purpose of opening a new restaurant, a list of neighbourhoods in Mumbai is required. Since the suburbs in the city are one of the most crowded places in the world, a wider geographical coverage would help in deciding the best business location. The list was obtained from Wikipedia (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Mumbai), which is a reliable source of geo-spatial information. Python "webscraping" method and "beautifulsoup" package were used to extract the neighbourhood data in a tabular form. Special care was taken to choose the right listing of neighbourhoods in the Wikipedia page as some webpages contained formats incompatible with the beautifulsoup package.

For geospatial analysis, locations of the neighbourhoods must be expressed by geographical coordinate of latitudes and longitudes. The Foursquare API is a social location service which associates geographical names with their accurate locational information. Python Geocoder package was used to convert geographical addresses into their geographical coordinates, which is later used by the Foursquare API. The extracted data were populated in a Pandas DataFrame, which was displayed in a map using Folium package. A visual inspection of the map provides a check for appropriate plotting of the geographic coordinate data in Mumbai. The map is shown in Fig. 2.
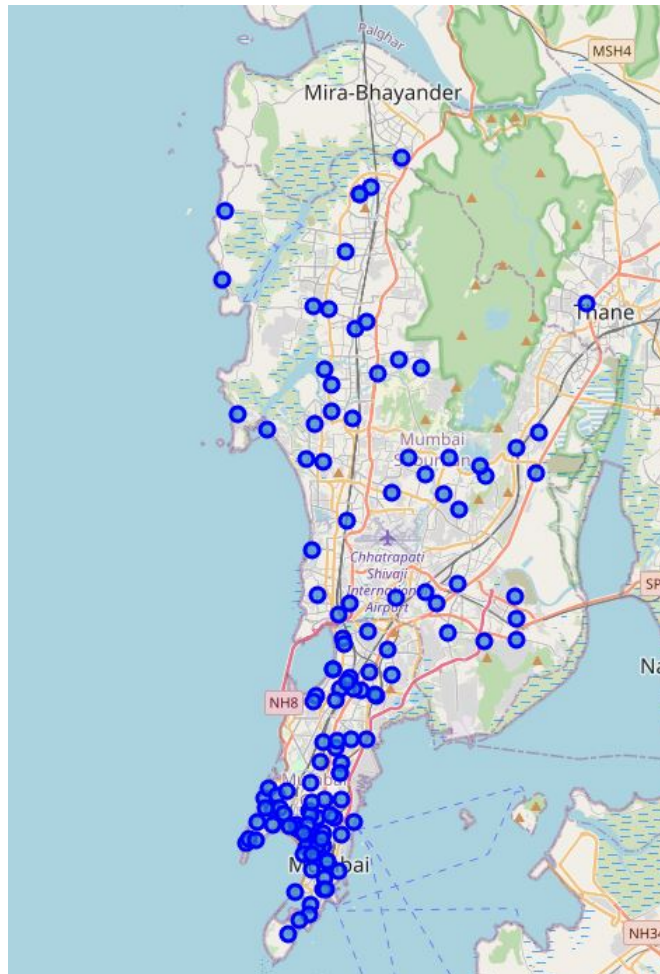


Fig. 2: Map of the neighbourhoods in Mumbai, India (generated using Python Folium package)

The region of interest (ROI) considered in this report included an area within 1 km radius of the location of Mumbai. Top 100 venues within the ROI were obtained using Foursquare API. Registration for a Foursquare Developer account was required for using the service via a Foursquare ID and Foursquare secret key. Next, we made API calls to Foursquare using the geographical coordinates of the neighbourhoods in a Python loop. Venue data were returned by Foursquare in a JSON file, whence information on name, category, latitude and longitude was

obtained. This information may be used to explore the number and unique categories of venues in each neighbourhood. The neighbourhoods were grouped by venues and their categories. This tabular data gives information on the number of venues contained within a neighbourhood. Further, these data are filtered for the type "Restaurant" for final clustering. In the final step, a k-means clustering algorithm was used to cluster the data into 3 clusters. This clustering method is the one of the most popular unsupervised machine-learning techniques, where k centroids are identified. Every data point is assigned to the nearest cluster while keeping the centroid size a minimum. This clustering helped identify neighbourhoods with the highest and lowest concentrations of restaurants. Viability of a new restaurant business can be examined from this clustering.

# Results

Results from the k-means clustering divided the neighbourhoods in Mumbai into 3 clusters based on the frequency of occurrence of existing restaurant businesses (Fig. 3). The results can be summarized as follows:
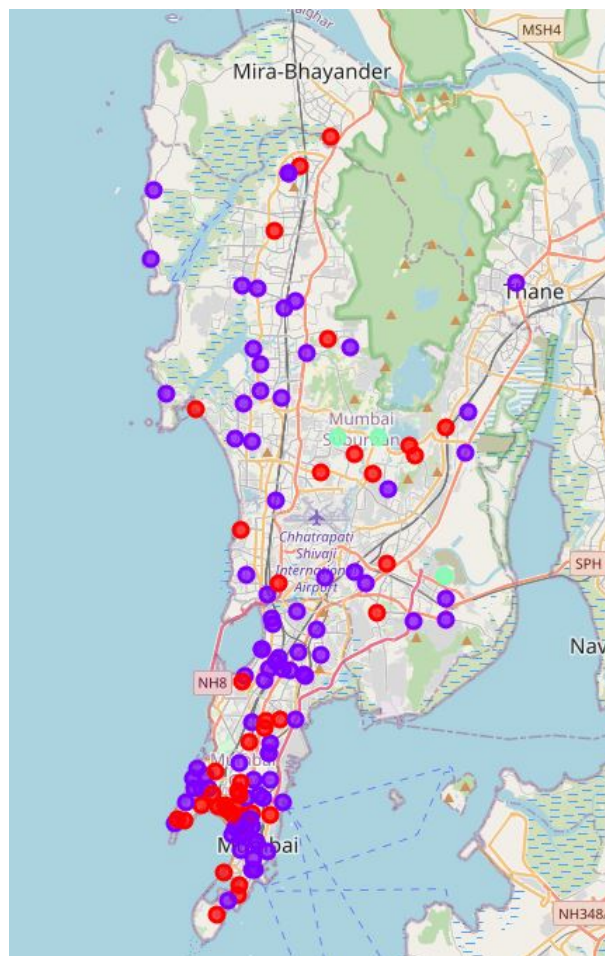


Fig. 3: k-means clustering of restaurants in Mumbai neighbourhoods.

Cluster 1 (Red): Neighbourhoods have the 2nd highest concentration of restaurants.

Cluster 2 (Purple): Neighbourhoods have the lowest concentration of restaurants.

Cluster 3 (Green): Neighbourhoods with the highest concentration of restaurants

# Discussion

It is observed from Fig. 3 that there are fewer restaurants in West- and South- Mumbai (Cluster 2, purple). While Cluster 3 (green) in the central and south-eastern parts of Mumbai has the highest concentration of restaurants, Cluster 1 (red) in predominantly south and central Mumbai has the second highest concentration of restaurants. Therefore, it is prudent to open a new restaurant in the Cluster 2 neighbourhoods. For example, neighbourhoods of Dadar and Matunga with many Cluster 2 venues could be ideal venues for new restaurants. This would ensure minimum competition from similar establishments in the area.

This report is based on clustering of only the existing concentration of restaurants in the neighbourhoods of Mumbai. However, more data on the demography may be required for a more appropriate decision as people from different regions of the country/abroad, who live in the city and its suburbs have different culinary preferences. Moreover, the type of restaurant must be considered while selecting the venues. For example, fast food restaurants would be more popular near office building and educational institutions.

# Conclusion

This report depicts the use of machine learning algorithm for the identification of most suitable venues for restaurant business. A k-means algorithm using 3 clusters was used to decide best restaurant locations based on the number of such businesses already existing in the neighbourhoods. For this objective, we have defined the business problem, gathered required data from relevant sources, performed data analysis using machine learning techniques and provided recommendations based on the results obtained. This is aimed at helping aspiring small business owners to take informed decisions in opening new restaurants in the neighbourhood of Mumbai.

# References

https://www.maakan.com/

https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Mumbai