# Coursera Capstone Project

## Final Project

Project- Opening a new Indian or Italian restaurant in Mumbai

By- Rishika Shrivastava

# Introduction-

Mumbai is one of the busiest cities in India. It is also known as the fashion capital of India, home of the great Bollywood industry. With a very dense population in the region, there is a huge demand of eateries and restaurants.

It could either be a café, or an upscale restaurant or street food or any casual dining place among many options. People look for budget friendly, and exotic feast for an outing. These days people go out pretty often, which implies greater income for the owners.

Opening an eatery is not as easy as it sounds and needs a lot of analysis and consideration beforehand. Location is one of the major driving factors which determine whether a restaurant will survive the battle or not.

## Business Problem and audience -

We primarily want to analyse the suburbs of Mumbai to determine which would be the best location for opening a new restaurant. Using various machine learning techniques like grouping and clustering we have to determine from a business perspective, which location would be the best to open a new Italian or Indian Restaurant?

This inspection will be useful for individuals or teams who would like to enter or invest in the culinary business. It would give them a proper insight into the segregations on the basis of cuisine.

# Data and its collection:

For this particular problem, we would need:

- Information on the restaurants in various locations using the predefined dataset
- Latitude and Longitude of the neighbourhoods
- Venue data to perform clustering and get accurate results

**Sources of data and its derivation**:

We get the restaurants information using the Zomato dataset available online (You can get it from Kaggle), which contains the restaurant's name, location, cost, rating among many parameters.

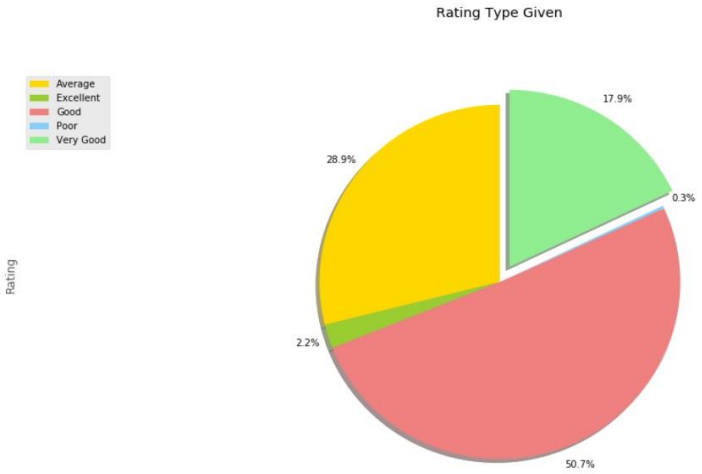
We get the coordinates i.e. The latitude and the longitude of the neighbourhoods using Python's Geocoder.

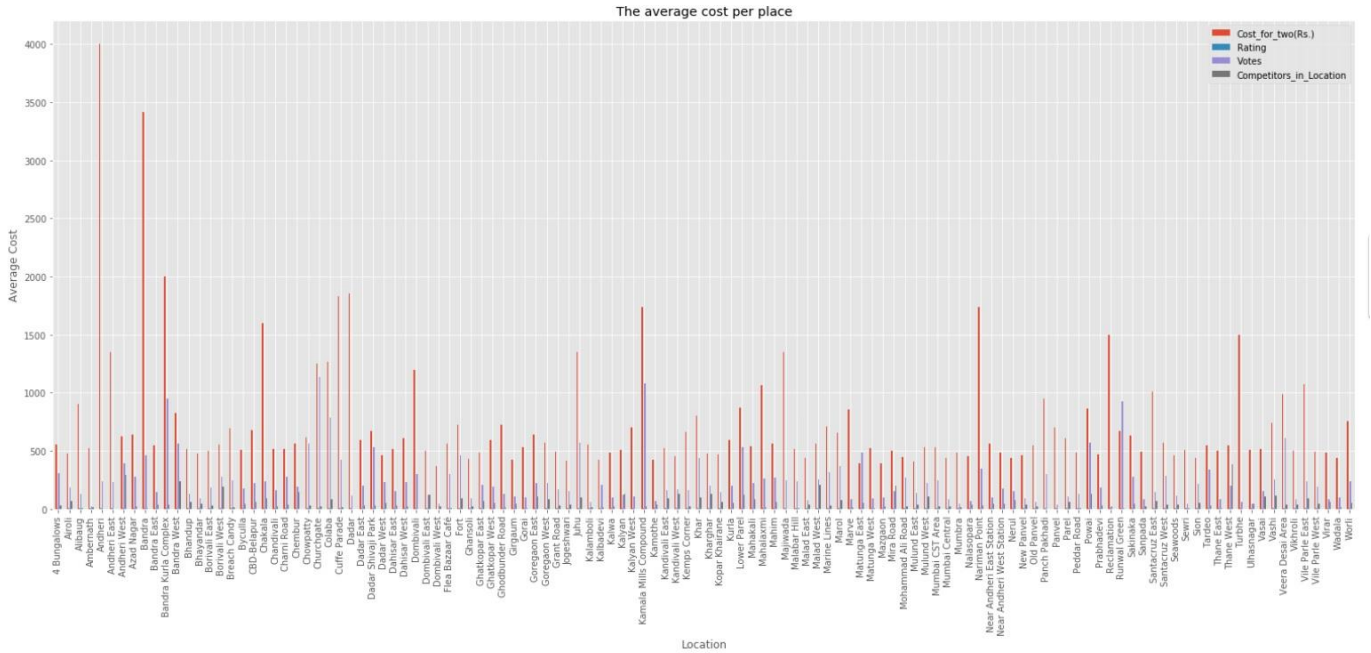
To retrieve venue data we used the Foursquare API, which is a local search-and-discovery mobile app which provides search results for its users. It provides us a variety of venue data, although we are primarily interested in the restaurant categories. We perform data cleaning, data wrangling to get the required dataset to apply the machine learning tools of our choice.

The methodology section is where we will discuss the entire procedure of the analysis and it will be summarized in the next section.

# Methodology:

Initially we begin with our Zomato dataset which contains the data of all the restaurants of Mumbai, India. It contains parameters of which we drop the ones which are not useful to us. We would like to know what the average standard of restaurants in Mumbai. We use the ratings to envision a pie chart. Next we calculate the average of variables like cost, rating, competitors and votes to deduce which could be one of the probable areas to open a business. It gives a rough estimate of all the criteria.



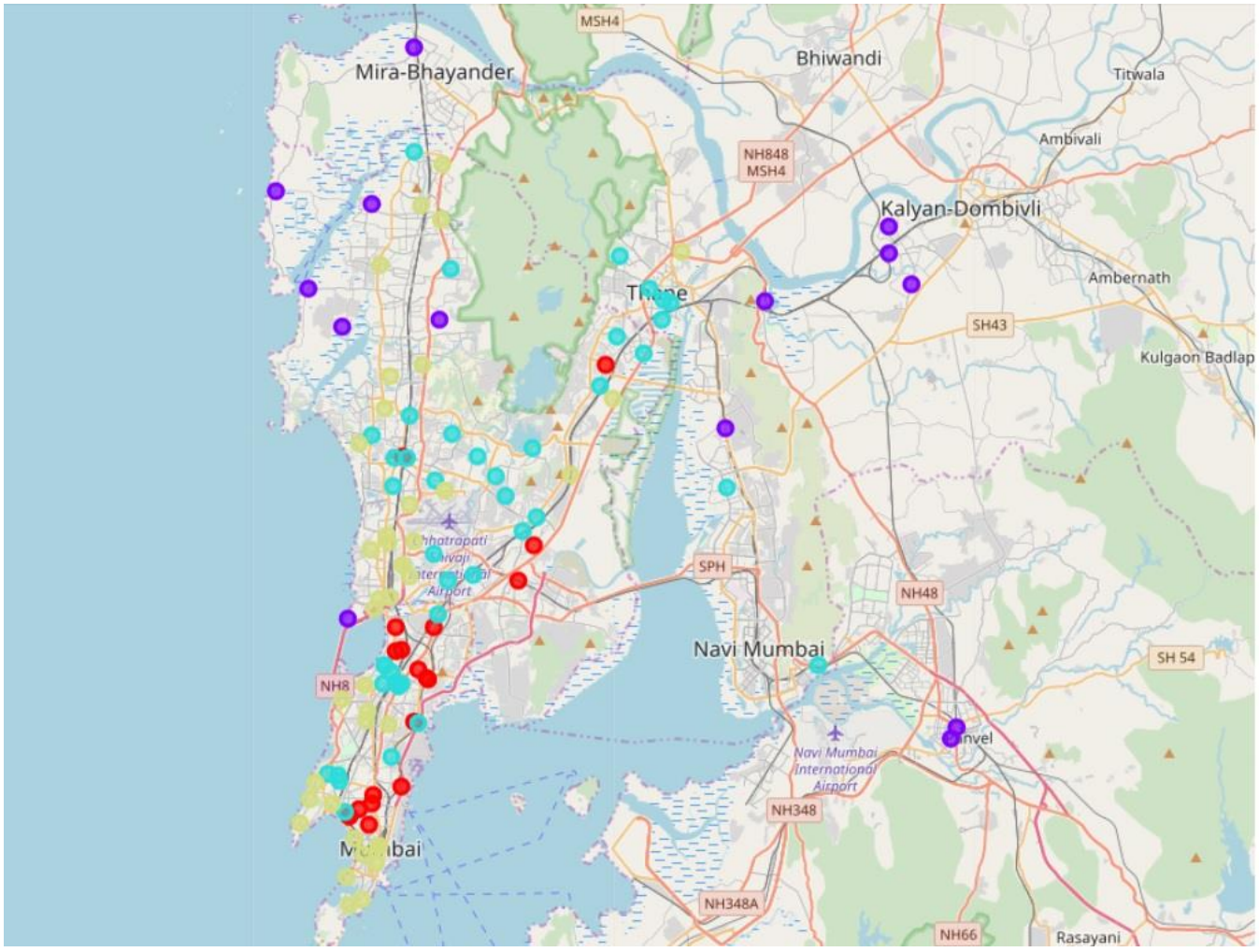
Next we use the Geocoder to get the coordinates of the various neighbourhoods. Foursquare API is used to get the venues within a radius of 2000 meters. WE extract the necessary data like the venue name, location, latitude and longitude and get the unique values from them. Then we check the neighbourhoods individually by calculating the mean of the frequency of their occurrences. This is how we are preparing the data for clustering. Our concerns are Indian and Italian restaurants and we filter the data accordingly.

For the final part we perform K-Means clustering. It was one of the widely used methods for unsupervised ML algorithms. We divide the data into 4 clusters based upon Italian and Indian restaurants as the criteria. The findings help us to clearly assess the best location for a new diner.

# Results:

Upon clustering we come across the following results:

- Cluster 0 contains moderate number of Indian and Italian restaurants.
- Cluster 1 contains the minimum number of restaurants for the same.
- Cluster 2 is highly populated with Indian and Italian restaurants.
- Cluster 3 contains moderate number of Indian and Italian restaurants.




This is a visual representation of our findings, created using Folium.

## Discussion:

As we can clearly see from the map, cluster 2 has the highest concentration of restaurants overall. South Bombay flourishes with the highest number of restaurants. Cluster 1 is low on the numbers and still has scope for expansion and room for more diners. Cluster 0 and 3 have moderate number of eateries and will have more density in comparison to cluster 1. Cluster 2 will have the most intense competition and it would be most difficult to start business there. There are already established ventures there and the business proposition would have to be exceptionally unique. It would be advised to avoid cluster 2, for a new business as chances of loss would be high.

## Recommendations:

Apart from location, the cost, paying capacity of customers, ambience and value for money also matters among many driving factors. Clustering could also be done based on these parameters later onwards.

## Conclusion:

With all procedures fulfilled and performed, i.e. filtering, organizing, arranging and clustering data we have been able to provide the business professionals with results and recommendations.

To answer the question proposed in this project, Cluster 1 would be the most appropriate location to start a new venture. With these results, the stakeholders should be able to capitalize their money in the right region and avoid aiming for areas with potential risk.

## References-


- https://foursquare.com/developers/apps
- https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai