# IBM Data Science Capstone Project

*Restaurant Recommendation in Mumbai, India*

By: Kaushik P. Metha

**Table of contents:**

# 1. Introduction:

This report is a part of the final course of IBM Data Science Specialization hosted by Coursera.

If you are looking for a culinary escapade, Mumbai is, by all means the gastronomic hub of India. The restaurant industry offers a wide variety of cuisines like 'Indian', 'Mexican', 'Italian' and many more. There's something for everyone. The restaurants in Mumbai are considerably heterogeneous to provide pocket-friendly casual meals or a lavish dining experience. Mumbai's Khau Gallis offer some of the most finger-licking and delicious street food.

The target audience for this project are the people of Mumbai, tourists etc. Now, let's dig right into the mouth-watering delights Mumbai has to offer.

# 2. Problem Definition:

Mumbai is a massive city with a very cosmopolitan feel and this is the reason it has a diverse range of cuisines. There are plenty of restaurants where you can have pocket-friendly meal or a lavish dining experience.

One of the main problems is that people are unaware of famous restaurants and places available in a specific locality. This not only applies to the tourists but also to the local people. Today, there are a large number of restaurants established in Mumbai. With the increase in number of restaurants people often get confused about the best-suited restaurant according to their preferences. In addition to that, people often face a hard time to find out the best place near their locality. The main aim of this project is to recommend restaurants based on the location of the user.

## 3. Data:

The following data will be used to solve the problem
1) List of all the neighbourhoods in Mumbai.
2) Latitude and Longitude of those neighbourhoods.
3) Data related to the restaurants in those neighbourhoods.

The list of all the neighbourhoods in Mumbai and their location (Latitude and Longitude) is extracted from Wikipedia. Link to the Wikipedia page: *https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai*. BeautifulSoup (a web scraping package in python) is used to extract the neighbourhood data.

The data related to the restaurants in a neighbourhood is obtained by using the Foursquare API. The name of the restaurant, category, the location of the restaurant etc can be extracted using the Foursquare API.

The list of all the restaurants in Mumbai is obtained by web scraping Zomato. Link: https://www.zomato.com/mumbai/restaurants. Name of the restaurant, rating, number of votes, cost for 2 is extracted from Zomato.

# 4. Methodology:

The first step of the project was to collect the list of neighbourhoods in the city of Mumbai. This list is available on Wikipedia (*https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai*). The neighbourhood, suburb and the coordinates of the neighbourhood were extracted using the BeautifulSoup library. The coordinates of few neighbourhoods were incorrect on the Wikipedia page. After correcting those coordinates and performing some data cleaning and preparation techniques, the following dataset of Mumbai neighbourhoods was obtained

|   | Area | Location | Latitude | Longitude |
|---|------|----------|----------|-----------|
| 0 | Amboli | Western Suburbs | 19.129300 | 72.843400 |
| 1 | Chakala, Andheri | Western Suburbs | 19.111388 | 72.860833 |
| 2 | D.N. Nagar | Western Suburbs | 19.124085 | 72.831373 |
| 3 | Four Bungalows | Western Suburbs | 19.124714 | 72.827210 |
| 4 | Lokhandwala | Western Suburbs | 19.130815 | 72.829270 |

The next step was to obtain a list of all the restaurants in Mumbai. Restaurant related data like name of the restaurant, rating, number of votes, price etc was scraped from www.zomato/mumbai/restaurants. The data of around 20000 restaurants was scraped from the above website. After running some data cleaning and data preparation techniques the following dataset was obtained.

|   | Name | Cost for 2 | Dining_rating | Dining_votes | Delivery_rating | Delivery_votes |
|---|------|-----------|---------------|--------------|-----------------|----------------|
| 0 | Yasin's Food Inn | 700.0 | 4.1 | 2182.0 | 3.8 | 99400.0 |
| 1 | McDonald's | 400.0 | 4.0 | 948.0 | 4.1 | 11400.0 |
| 2 | Joey's Pizza | 800.0 | 4.8 | 8922.0 | 4.2 | 4081.0 |
| 3 | Guru Kripa | 300.0 | 4.6 | 3671.0 | 4.2 | 21500.0 |
| 4 | Cafe Safar | 450.0 | 4.1 | 2174.0 | 3.9 | 89100.0 |

The coordinates of Mumbai were obtained and the neighbourhoods of Mumbai were plotted on a map with the help of 'folium' library
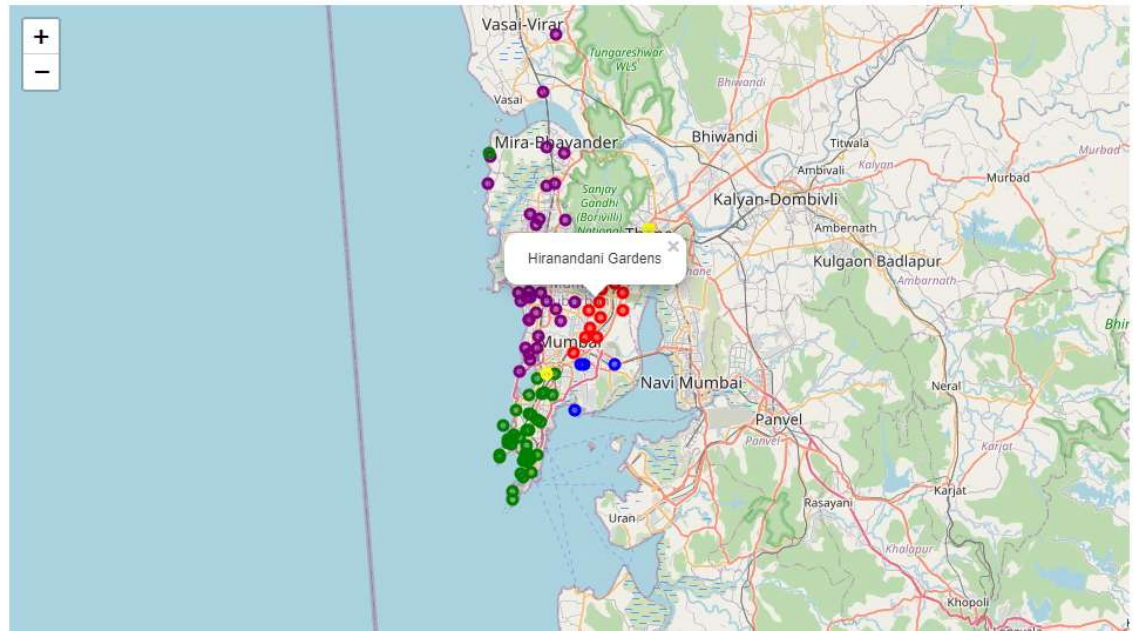


Mumbai neighbourhoods

Next, I used FourSquare API to explore the nearby restaurants of the first Neighbourhood of the dataset. The API requests returned the data of 100 nearby venues within a radius of 3000meters. Venue Name, Category, Address, Coordinates, Distance from current neighbourhood was stored in a data frame. Then the data of restaurants was extracted from the data frame ( Categories that have the word 'Restaurant' in it ). After some data cleaning, the following dataset was obtained (Only of first neighbourhood)

| | Name | Category | Address | Street | Distance | Latitude | Longitude | Id |
|---|---|---|---|---|---|---|---|---|
| 0 | Shawarma Factory | Falafel Restaurant | Dadabhai Road | Off JP Road, Near Navrang Cinema | 611 | 19.124591 | 72.840398 | 5174e2be498e39cf0d1c20cb |
| 1 | Garden Court | Indian Restaurant | Opp. Andheri Sports Complex, Andheri | JP Rd. | 665 | 19.127188 | 72.837478 | 4b46128bf964a520751526e3 |
| 2 | Shetty's Corner | Chinese Restaurant | Bhavan's College | NaN | 765 | 19.124845 | 72.837858 | 4df24b2cd1649c8a28de604e |
| 3 | Persia Darbar | Indian Restaurant | Jogeshwari West | S.V. Road | 924 | 19.136952 | 72.846822 | 5247116c11d2982261defb5d |
| 4 | Jaffer Bhai's Delhi Darbar | Mughlai Restaurant | Jogeshwari West | NaN | 973 | 19.137714 | 72.845909 | 55fc3615498e141bd45da525 |

A generalised code was written which returned the dataset of nearby restaurants of the selected neighbourhood. A dropdown menu widget is used to select the neighbourhood. The dataset of nearby restaurants was merged with the Zomato dataset. This newly merged dataset contains Name of the Restaurant, Category, Address, Distance, Coordinates, Ratings, Votes etc

| | Name | Category | Address1 | Address2 | Distance(m) | Latitude | Longitude | Id | Cost for 2 | Dining_rating | Dining_votes | Del |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Sharda Bhavan | Indian Restaurant | Lakhamsi Nappu Rd., Matunga | NaN | 1170 | 19.026783 | 72.850841 | 4b87b1f8f964a52084c731e3 | 200.000000 | 4.1 | 441.0 | |
| 1 | Terttulia | Italian Restaurant | Ground Floor, Hotel Parkway | Near Natural Ice Cream, Ranade Road Extension | 1301 | 19.025638 | 72.835416 | 50d801d6e4b09e2e64f6f437 | 388.557387 | 4.4 | 2191.0 | |
| 2 | Jai Hind Lunch Home | Seafood Restaurant | Tulsipipe Road | Dadar West | 2380 | 19.002183 | 72.829512 | 4bfe4e1855539c7488d2bcf3 | 850.000000 | 4.4 | 1180.0 | |
| 3 | Guru Kripa | Indian Restaurant | Guru Kripa Building, Near SIES College | Rd No 24 | 3307 | 19.042955 | 72.861796 | 4b0587d8f964a52008a422e3 | 300.000000 | 4.6 | 3671.0 | |
| 4 | Ming Yang | Chinese Restaurant | Taj Lands End | Bandra Lands End | 3919 | 19.043582 | 72.819199 | 4b5dc6e6f964a520166b29e3 | 800.000000 | 4.3 | 1087.0 | |

Merged Dataset

KMeans clustering was used to cluster the restaurants of few neighbourhoods. The restaurants will be grouped into 4 clusters. The result will allow us to identify the differences between the clusters.

Lastly, a dataset of top 5 nearest restaurants within a radius of 3000m of each neighbourhood was created.

| | Area | R1 | R2 | R3 | R4 | R5 |
|---|---|---|---|---|---|---|
| 0 | Amboli | Shawarma Factory | Garden Court | Shetty's Corner | Persia Darbar | Jaffer Bhai's Delhi Darbar |
| 1 | Chakala, Andheri | Faaso's | Hit & Run | Maharaja Restaurant | King chilly | Spirit Kitchen and Bar |
| 2 | D.N. Nagar | Shetty's Corner | Garden Court | Banana Leaf | Shawarma Factory | Tewari Bros Sweets |
| 3 | Four Bungalows | Banana Leaf | Urban Tadka | Gajalee 7 Bungalows | Levo | 015's Café |
| 4 | Lokhandwala | 015's Café | KFC | Shabri Veg Restaurant | Indigo Deli | Indigo Delicatessen |
| 5 | Marol | Benzys | Uttam Da Dhaba | Gurukripa | Eva's Pizza | Burger King |
| 6 | Sahar | Peshawari | Dum Pukth | Pan Asian | Stax | the indian Kebab grill |

Top 5 nearest restaurants

# 5. Results and Discussions:

The restaurants of the selected neighbourhood were grouped into 4 clusters. The description of each cluster is as follows:

a. Cluster 0: High cost, Excellent Score ( Red )
b. Cluster 1: Low cost, Average Score ( Purple )
c. Cluster 2: Moderate cost, Good Score ( Cyan )
d. Cluster 3: Moderate cost, Excellent Score ( Yellow )
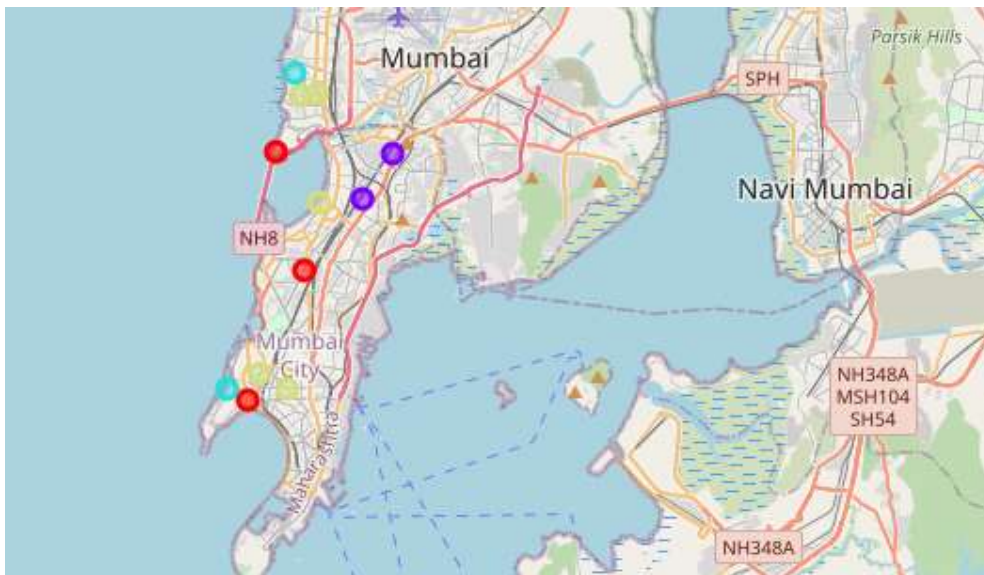


**Example**

1. The restaurants of cluster 0 are recommended if the person wants to have a lavish dining experience in a highly rated restaurant.
2. The restaurants of cluster 1 are recommended if the person wants to have a low cost pocket friendly meal in an average rated restaurant
3. The restaurants of cluster 2 and 3 are recommended if the person wants to have a good dining experience.

## 6. Conclusion:

The concepts of Data science like problem definition, data collection, data cleaning and preparation, exploratory analysis, visualizations, machine learning were used in the project.

In this project, various neighbourhoods of Mumbai were explored. The data obtained from FourSquareApi, Wikipedia and Zomato was used to cluster the restaurants of each neighbourhood. The clustering of restaurants of few neighbourhoods led to the conclusion of grouping the restaurants into 4 different clusters (description mentioned in results and discussion).

The findings of this project will help the user to have a good dining experience in the city of Mumbai.