

Coursera Capstone

IBM Certificate Data Science Course

Opening an Italian Restaurant in Mumbai, India

By Debjyoti Bardhan
June 2020



Introduction

Italian cuisine is prized and enjoyed the world over. It is the perfect food for a nice evening out, with a romantic partner, or with family or with a group of friends. A nice pizza or spaghetti with authentic cheese used in its making can be a part of a grand evening of dining, or a casual hangout session among friends. However, despite the proliferation of eateries - specifically pizzeria - in the city of Mumbai, there is a dearth of restaurants catering to genuine Italian cuisine. Since even eateries serving non-authentic Italian food are garnering good business, it can be surmised that there is genuine demand for such things in the market. The goal of this project is to determine the optimal locations for an Italian restaurant, whether it can be expanded into a chain and whether the possible business fate of the venture can be ascertained.

Business Problem

The primary question is this: In the Indian city of Mumbai, what are the optimal locations for a restaurants dedicated exclusively to Italian cuisine? The secondary question is to gauge whether the outlet will be a success or a failure, and the likelihood of expanding it as a restaurant chain in the city. In order to answer this, we shall use data for existing Italian restaurants and use various data science techniques such as k-means clustering to determine optimal locations.

Data

Data needed

The following data is needed:

1. The different neighbourhoods in Mumbai, along with other details, like population, and their geospatial data.
2. Locations of existing Italian restaurants

Obtaining the Data

Following the numbering used in the previous list:

1. The list of neighbourhoods and the sub-areas in each neighborhood can be obtained from Wikipedia. We can use web scraping to obtain the data from the relevant page, using packages such as BeautifulSoup, in a Python notebook. Wikipedia lists the

-
- latitudes and longitudes of the neighbourhoods and these can be obtained similarly. This will utilise the datascience skills learnt in the course such as web-scraping.
2. One can use data from Zomato or the Foursquare API to explore the different Italian restaurants in each neighbourhood. Clustering of this data can be done using k-means clustering and visualised on a map using Folium.

Methodology

In order to obtain the initial data about the different neighbourhoods in Mumbai, we can simply scrape the information off the Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai#Mumbai_neighbourhood_coordinates). This was done using the **Beautiful Soup** package. The data scraped off includes information about the area name (the same as 'neighbourhood'), the location of the place within the city and the actual latitude and longitude of each area. This information is stored as a **pandas dataframe**. This makes it convenient to use the data later on. The latitude and longitude for the city of Mumbai is just obtained from a simple Google search without the use of the Geoencoder package.

The geospatial information about the different areas in Mumbai can be visualised using a **Folium** map - a snapshot of it is included in the next section.

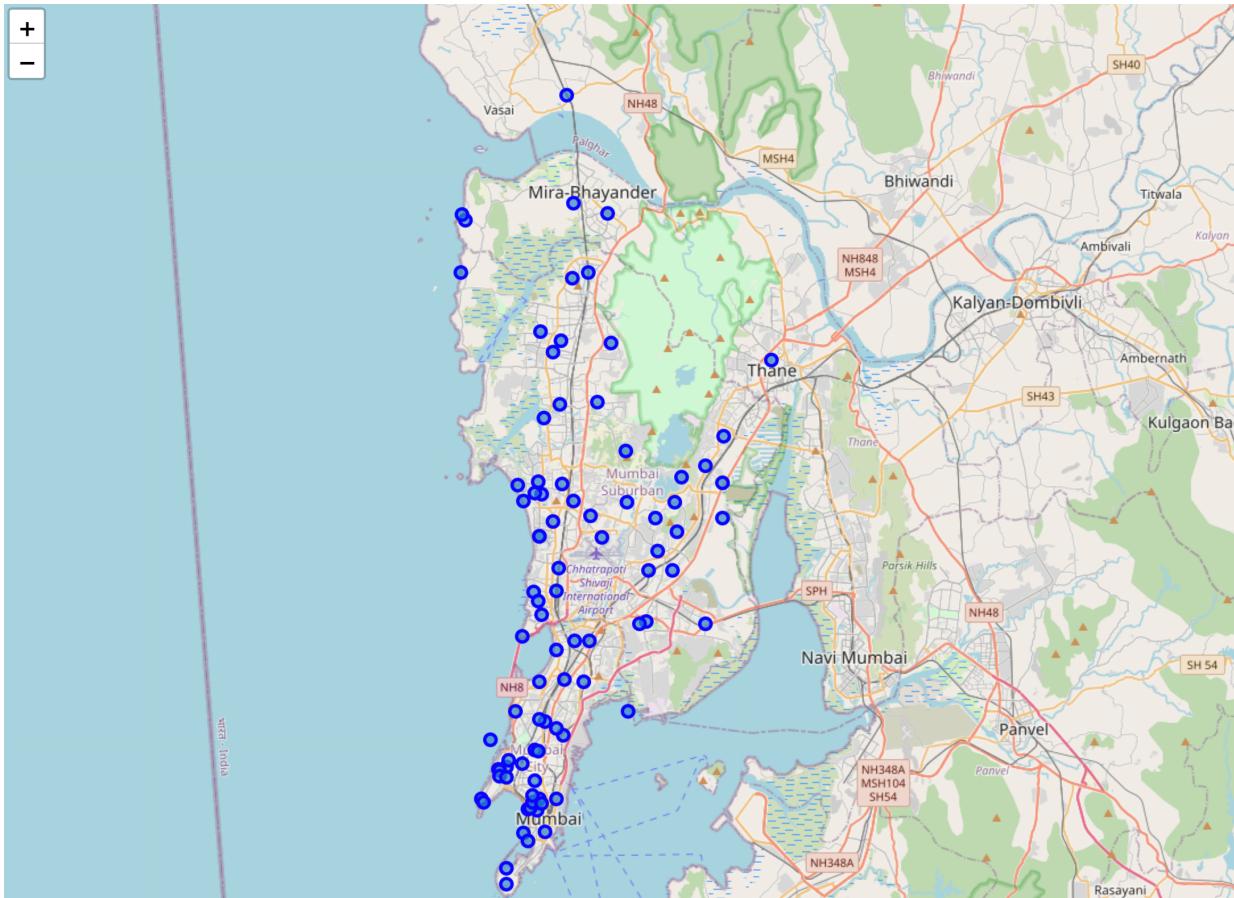
Next, we can use the **Foursquare API** to query the different locations. We want to see the top 100 places within a radius of 2 km of the centre in each area and obtain the total number of different unique categories. This is returned as a JSON file. We extract the information about the name and category of the venue, along with the latitude and longitude of the venue from this file and store it in a dataframe. We can then narrow our investigation to just the category for Italian Restaurants. We can use the geospatial information about the Italian Restaurants and again use Folium to visualise the locations of the restaurants.

Finally, the clustering of these restaurants is done using the *k*-means clustering algorithm. The ***k*-means clustering** groups the different locations into *k* clusters such that each object within the cluster belongs to the closest mean or centroid. This centroid is taken to be the typical point for the cluster. The in-cluster distances between the different objects is minimized. For this problem, we find it easiest to choose **7 clusters** in which to break the data.

All results from the operations described in this section are given in the next section.

Results

The results from the operations described in the previous section are listed here. First we see the locations of the different neighbourhoods in Mumbai.



The city of Mumbai has a predominantly North-South geographical extension. There are more neighbourhoods in the south because there are a larger number of people staying there. There are 93 areas in all.

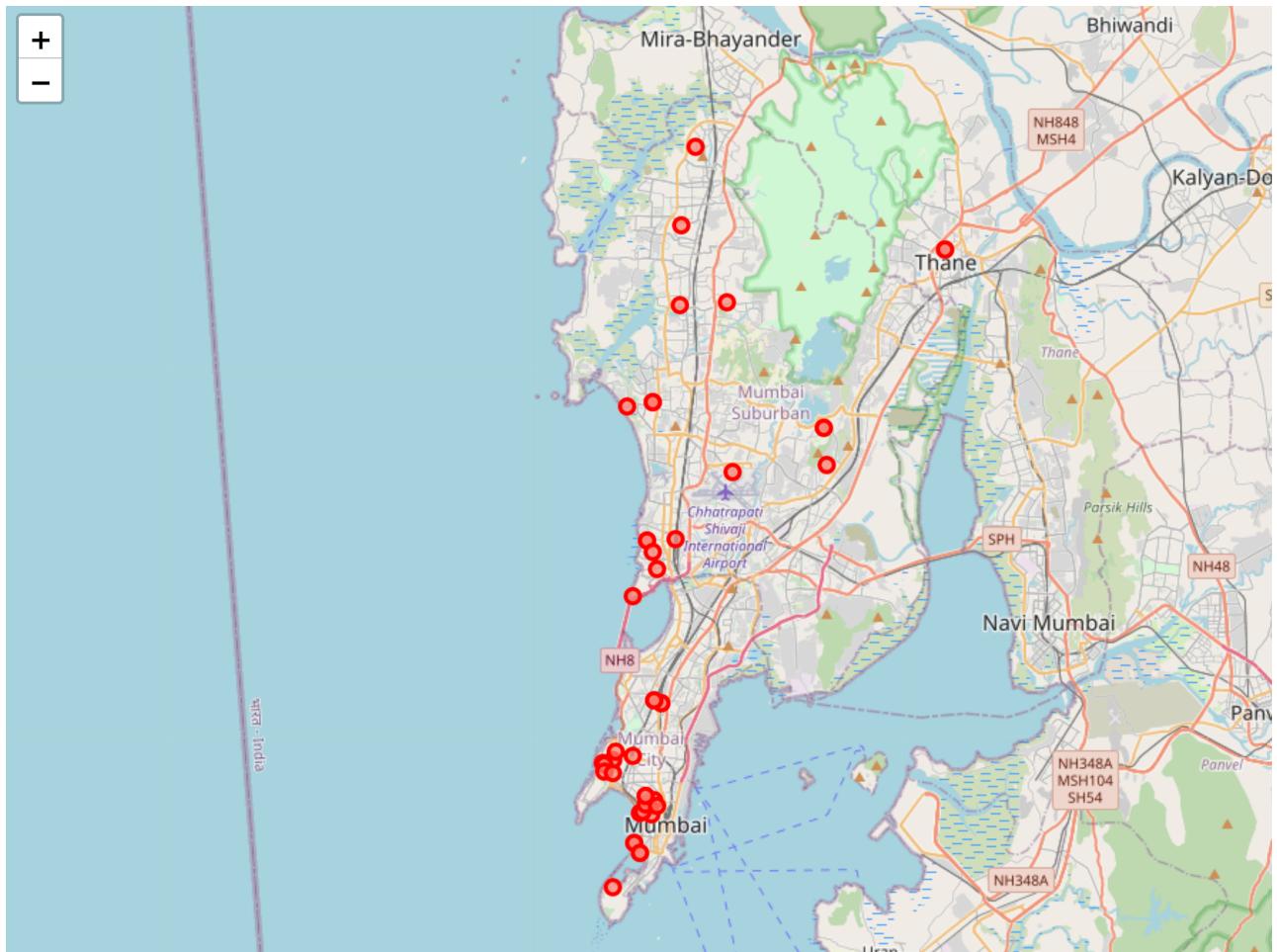
A brief summary of the results of the Foursquare API query is given as:

Area	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory	
0	Amboli	19.1293	72.8434	Shawarma Factory	19.124591	72.840398	Falafel Restaurant
1	Amboli	19.1293	72.8434	Merwans Cake shop	19.119300	72.845418	Bakery
2	Amboli	19.1293	72.8434	Jaffer Bhai's Delhi Darbar	19.137714	72.845909	Mughlai Restaurant
3	Amboli	19.1293	72.8434	Hard Rock Cafe Andheri	19.135995	72.835335	American Restaurant
4	Amboli	19.1293	72.8434	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant

As we can see, the different venues are described by several attributes like VenueName, VenueCategory and the geospatial location. We find that there are 223

different unique categories which exist. We can then choose the category for 'Italian Restaurants' only and we find that there are 141 such places.

It is worthwhile to look at their distribution.



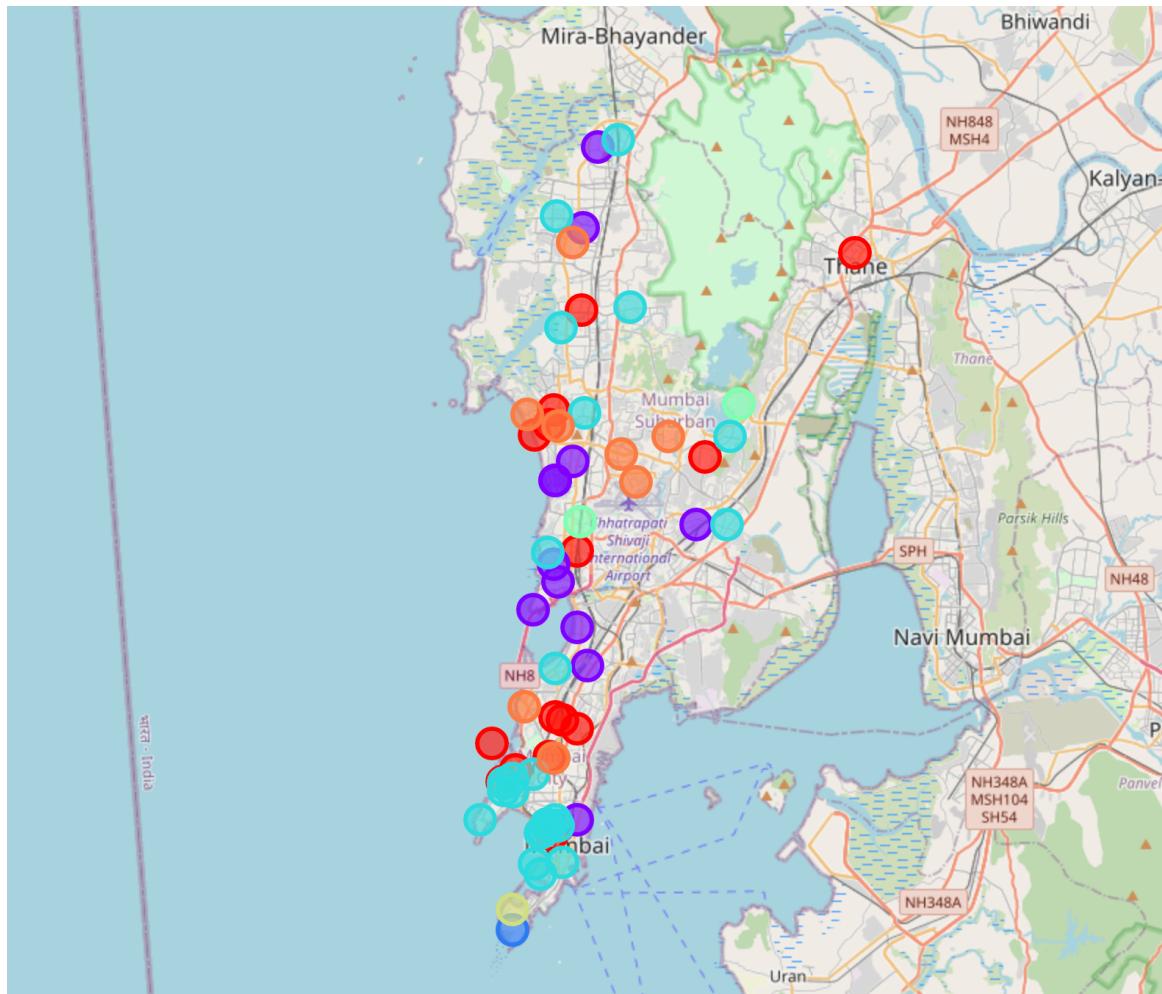
As we can see, the number of Italian restaurants increases as we head to the south of Mumbai. We had seen that the number of neighbourhoods and the population density also increases as we head southwards.

We then use k -means clustering to cluster them into **seven** categories. The relatively high number for the k is because we want a somewhat finer gradation of the density of restaurants in the neighbourhood. The results of the clustering can be seen below. Note that the density of the number of Italian restaurants out of all the available establishments in a certain neighbourhood determines the colour of the cluster.

The typical density of the restaurants in each cluster with respect to the other establishments in the area and the number of such restaurants in the cluster are tabulated below:

Cluster Number	Typical density of Italian restaurants	Colour in the map	Number of areas in the cluster
0	0.03	Red	8
1	0.01	Purple	24
2	0.1	Blue	1
3	0.02	Teal	12
4	0.045	Light Green	15
5	0.076	Yellow	1
6	0.04	Orange	2

Three of these seven categories have either 1 or 2 areas in the cluster.



The details of all the areas in each of the clusters are given in the Appendix.

Discussion

As the map suggests, the majority of the clusters lie in the south of Mumbai. Various clusters even overlap in the general vicinity. Thus there are plenty of Italian restaurants in the area. This is possibly because these areas have a lot of people living there and thus there are notable customer base.

There is intense competition among the various establishments in the south of Mumbai. Rising costs of land and property mean that this might not be the best place to establish a new Italian restaurant.

A notable lack of Italian restaurants can be seen in Thane (to the north east of Mumbai). Thane is growing quite fast in real estate and the area is projected to be an important centre of population growth. The vicinity of the area to the industrial and wealth generating centres in Mumbai and the excellent railway services means that this area will see great growth very soon. This seems like a good place to set up a restaurant and then to grow from this centre.

Conclusion

The main aim of this report is to investigate the possibility of opening an Italian restaurant in the Indian city of Mumbai. The city of Mumbai is a large metropolitan city and a city of great possibilities. In order to ascertain where one can open a new restaurant, we look at the distribution of the areas and then used k -means clustering algorithm to ascertain which areas can be suitable for a new Italian restaurant. After looking at the results, we can say that one good spot for a new restaurant is Thane. This information might be helpful to anyone willing to invest in a new Italian restaurant in the highly competitive market in the city of Mumbai.

Appendix

The areas in the different clusters are listed below:

Cluster 0:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Worli	0.040000	0	South Mumbai	19.000000	72.815000
D.N. Nagar	0.040000	0	Andheri,Western Suburbs	19.124085	72.831373
Dagdi Chawl	0.040000	0	Byculla, South Mumbai	18.977129	72.829131
Marol	0.040000	0	Andheri,Western Suburbs	19.119219	72.882743
Chakala, Andheri	0.040000	0	Western Suburbs	19.111388	72.860833
Sahar	0.040000	0	Andheri,Western Suburbs	19.098889	72.867222
Seven Bungalows	0.040000	0	Andheri,Western Suburbs	19.129052	72.817018
Poisar	0.038961	0	Kandivali West,Western Suburbs	19.204511	72.837639

Cluster 1:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Pant Nagar	0.016667	1	Ghatkopar,Eastern Suburbs	19.080000	72.910000
Pali Hill	0.020000	1	Bandra,Western Suburbs	19.068000	72.826000
Nariman Point	0.020000	1	South Mumbai	18.926000	72.823000
Mumbai Central	0.020000	1	South Mumbai	18.969700	72.819400
Prabhadevi	0.020000	1	South Mumbai	19.016600	72.829500
Marine Lines	0.020000	1	South Mumbai	18.944700	72.824400
Marine Drive	0.020000	1	South Mumbai	18.944000	72.823000
Malabar Hill	0.022727	1	South Mumbai	18.950000	72.795000
Kemps Corner	0.020000	1	South Mumbai	18.962900	72.805400
Hiranandani Gardens	0.024096	1	Powai,Eastern Suburbs	19.118986	72.911767
Gowalia Tank	0.020000	1	Tardeo, South Mumbai	18.962450	72.809703
Dindoshi	0.018519	1	Malad,Western Suburbs	19.176382	72.864891
Dava Bazaar	0.020000	1	South Mumbai	18.946882	72.831362
Kala Ghoda	0.020000	1	South Mumbai	18.930700	72.833100
Charkop	0.018519	1	Kandivali West,Western Suburbs	19.216182	72.830575
Cavel	0.020000	1	South Mumbai	18.947400	72.827200
Bangur Nagar	0.020000	1	Goregaon,Western Suburbs	19.167362	72.832252
Cumbala Hill	0.020000	1	South Mumbai	18.965833	72.805833
Dahisa	0.017241	1	Western Suburbs	19.250069	72.859347
Chira Bazaar	0.020000	1	Kalbadevi, South Mumbai	18.948140	72.825462
Bhuleshwar	0.020000	1	South Mumbai	18.950000	72.830000
Amboli	0.022222	1	Andheri,Western Suburbs	19.129300	72.843400
Churchgate	0.020000	1	South Mumbai	18.930000	72.820000
Altamount Road	0.020000	1	South Mumbai	18.968100	72.809500

Cluster 2:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Navy Nagar	0.1	2	Colaba, South Mumbai	18.9012	72.8101

Cluster 3:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Juhu	0.010000	3	Western Suburbs	19.100000	72.830000
Irla	0.010000	3	Vile Parle, Western Suburbs	19.108056	72.838056
Ballard Estate	0.010753	3	Fort, South Mumbai	18.950000	72.840000
Bandstand Promenade	0.012987	3	Bandra, Western Suburbs	19.042718	72.819132
Mahavir Nagar	0.012821	3	Kandivali West, Western Suburbs	19.211319	72.842737
Mahim	0.010000	3	South Mumbai	19.035000	72.840000
I.C. Colony	0.012500	3	Borivali (West), Western Suburbs	19.247039	72.849830
Vidyavihar	0.010000	3	Eastern Suburbs	19.080000	72.896000
Matunga	0.010000	3	South Mumbai	19.017980	72.844763
Pali Naka	0.010000	3	Khar, Western Suburbs	19.062742	72.829396
Vile Parle	0.010000	3	Western Suburbs	19.100000	72.830000
Kherwadi	0.010000	3	Bandra, Western Suburbs	19.055300	72.831400

Cluster 4:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Sunder Nagar	0.030000	4	Malad, Western Suburbs	19.175000	72.842000
Parel	0.030000	4	South Mumbai	18.990000	72.840000
Thane	0.031250	4	Mumbai	19.200000	72.970000
Versova	0.030000	4	Andheri, Western Suburbs	19.120000	72.820000
Dhobitalao	0.030000	4	South Mumbai	18.943300	72.828600
Carmichael Road	0.030000	4	South Mumbai	18.972200	72.811300
Chandivali	0.030000	4	Powai, Eastern Suburbs	19.110000	72.900000
Mahalaxmi	0.026316	4	South Mumbai	18.983333	72.800000
Lower Parel	0.030000	4	South Mumbai	18.995278	72.830000
Lokhandwala	0.030000	4	Andheri, Western Suburbs	19.130815	72.829270
Khar Danda	0.030000	4	Khar, Western Suburbs	19.068598	72.840042
Four Bungalows	0.030000	4	Andheri, Western Suburbs	19.124714	72.827210
Currey Road	0.030000	4	South Mumbai	18.994000	72.833000
Breach Candy	0.030000	4	South Mumbai	18.967000	72.805000
Agripada	0.030000	4	South Mumbai	18.977700	72.827300

Cluster 5:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Cuffe Parade	0.074074	5	South Mumbai	18.91	72.81

Cluster 6:

Area	Italian Restaurant	Cluster	Location	Latitude	Longitude
Kalina	0.048193	6	Sanctacruz,Western Suburbs	19.081667	72.841389
Indian Institute of Technology Bombay campus	0.046154	6	Powai,Eastern Suburbs	19.133636	72.915358