

Coursera Capstone
IBM Applied Data science

Title: Battle of Neighborhoods

“Opening A New Restaurant in Mumbai, India”

By: Vipin Uniyal

May 2020

INTRODUCTION

According to a report, the restaurant industry in India is growing at the rate of 7% of which the organized sector is growing more rapidly at 16% as compared to the unorganized sector. Out of the 1.5 million eating outlets, only a very small fraction of outlets is part of the organized sector. The organized segment is slated to reach Rs.22, 000 crore by 2017. Quick service restaurants will show maximum growth followed by casual dining, cafes and fine dining. Most Indians eat out only twice a month and if they eat out just a little more often, the market opportunity will increase tremendously. Mumbai has alone 1500+ licensed restaurants. The restaurant business is all set to reach new heights with the rise in disposal incomes, nuclear families and increase in working population.

The location of your food service business will impact its success nearly as much as the menu. If your restaurant is in the wrong place, you won't attract the number of customers you will need in order to stay in business. The same is true if your location is inundated with competitors, or has poor visibility, or is hard to find. There are many things that must be considered as you look for a location in which to open your business.

Problem Statement

The location is one of the most important parameters to consider while discussing the idea as it can surely determine the success of the business. The Aim of this capstone project is to analyse and select the best locations in the Mumbai city, India to open a new restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the Mumbai city of Maharashtra, India, if any individual person or group are looking to start their restaurant business and want to open a new restaurant, where would you recommend them to open it?

Target Audience of this project

Any individual or group of people who want to open a new restaurant in Mumbai city and are finding a suitable location for their business.

DATA DESCRIPTION

Mumbai City District is a district of Maharashtra in Konkan Division. As a city district, it has no headquarters or subdivisions. It, along with the Mumbai Suburban District, makes up the metropolis of Mumbai. The city area is called the "island city" or South Mumbai or Old Mumbai. It extends from Colaba in the south to Mahim and Sion in the north. The city has an area of 157 km² and a population of 3,085,411.

Data Sources

- Data of postal code and areas of Mumbai districts is taken from:
<https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/>
- Latitude and longitude data of all the areas is fetched using google's api and python geocoder library.
- The Restaurant data of each neighbourhood is taken using foursquare api.
- To generate map python's geopy library is being used.

METHODOLOGY

Firstly, import all the important libraries.

Importing datasets

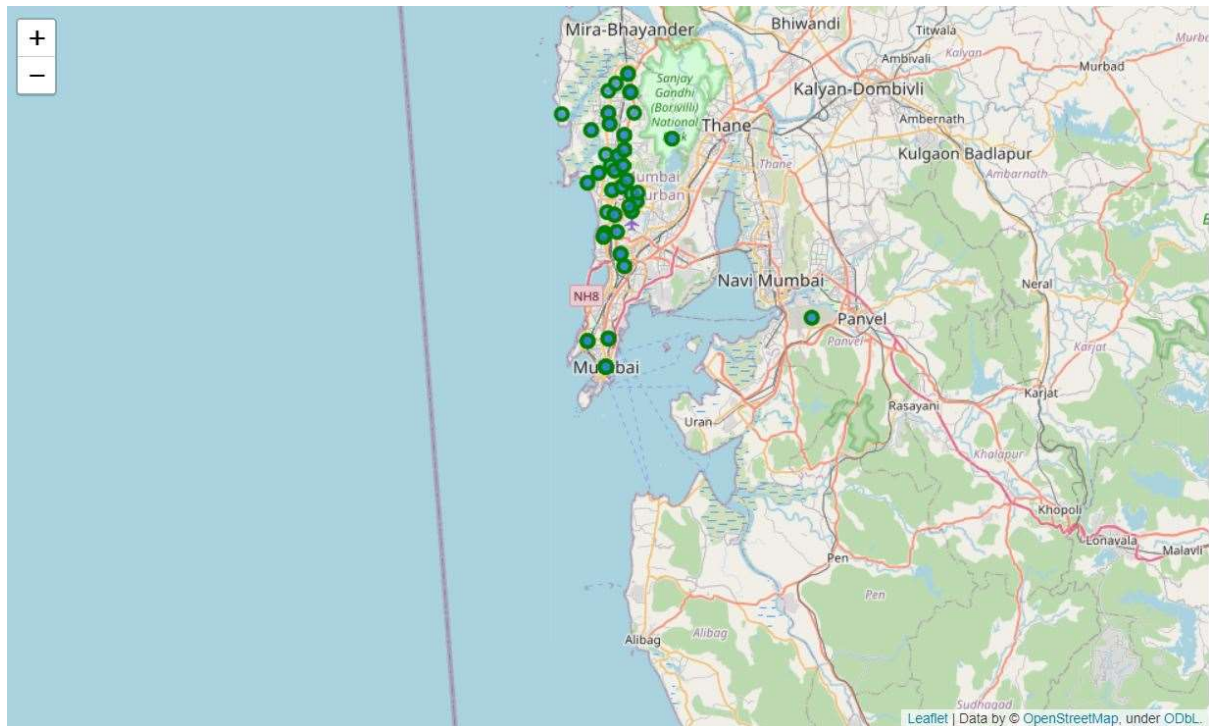
I used python beautiful soup4 to crawl the neighbourhood in Mumbai data with postal codes from the website, and converted it into pandas dataframe. It looks like this:

Out[10]:	PostalCode	Neighborhood	District
1	400029	A I staff colony	Mumbai
2	400065	Aareymilk Colony	Mumbai
3	400011	Agripada	Mumbai
4	400099	Airport	Mumbai
5	400004	Ambewadi	Mumbai
6	400053	Andheri	Mumbai
7	400069	Andheri East	Mumbai
8	400058	Andheri Railway station	Mumbai
9	400037	Antop Hill	Mumbai
10	400005	Asvini	Mumbai

Then I cleaned it and used python library of google geocoder to fetch the location coordinates latitude and longitude. The new dataframe looks like this:

Out[5]:	PostalCode	Neighborhood	District	Latitude	Longitude
0	400029	A I staff colony	Mumbai	19.151622	72.854981
1	400065	Aareymilk Colony	Mumbai	19.180136	72.908812
2	400011	Agripada	Mumbai	18.938771	72.835335
3	400099	Airport	Mumbai	19.103066	72.863560
4	400004	Ambewadi	Mumbai	18.938771	72.835335
5	400053	Andheri	Mumbai	19.144420	72.826612
6	400069	Andheri East	Mumbai	19.130182	72.853080
7	400058	Andheri Railway station	Mumbai	19.125732	72.842675
8	400037	Antop Hill	Mumbai	18.938771	72.835335
9	400005	Asvini	Mumbai	18.938771	72.835335
10	400053	Azad Nagar	Mumbai	19.144420	72.826612

After getting the data I created Mumbai city map using folium library



Then I used foursquare API to fetch the nearby venues in each neighbourhood. After that I created a new data set which looks like below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A I staff colony	19.151622	72.854981	China Ming	19.154614	72.856497	Chinese Restaurant
1	A I staff colony	19.151622	72.854981	Pizza Hut	19.149885	72.856264	Pizza Place
2	A I staff colony	19.151622	72.854981	Peshawari Grill	19.154210	72.856668	Indian Restaurant
3	A I staff colony	19.151622	72.854981	McDonald's	19.154511	72.856462	Fast Food Restaurant
4	A I staff colony	19.151622	72.854981	Golden Chariot	19.154864	72.856495	Indian Restaurant

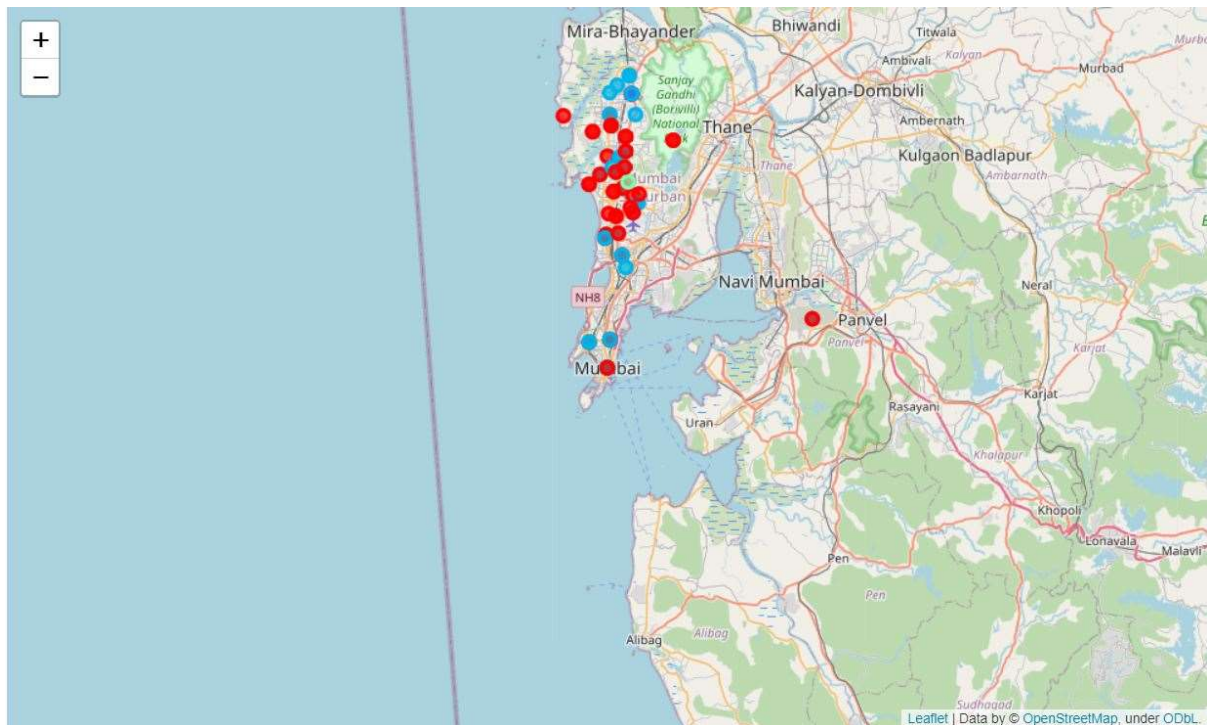
After that I created a data set for top 10 venues in the neighbourhoods and the table looks like below:

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A I staff colony	Indian Restaurant	Chinese Restaurant	Food Truck	Café	Pizza Place	Sporting Goods Shop	Fast Food Restaurant	Plaza	Lounge	Multiplex
1	Agripada	Indian Restaurant	Bar	Irani Cafe	Sandwich Place	Hotel	Seafood Restaurant	Café	Multiplex	Lounge	Train Station
2	Airport	Hotel	Gym	Café	Vegetarian / Vegan Restaurant	Falafel Restaurant	Cupcake Shop	Dance Studio	Department Store	Design Studio	Dessert Shop
3	Ambewadi	Indian Restaurant	Bar	Irani Cafe	Sandwich Place	Hotel	Seafood Restaurant	Café	Multiplex	Lounge	Train Station
4	Andheri	Café	Indian Restaurant	Bakery	Chinese Restaurant	Coffee Shop	Park	Cupcake Shop	Dessert Shop	Gym / Fitness Center	Fast Food Restaurant

After creating venues dataframe I analysed the data and clustered it using kmeans clustering algorithm and assigned label to each cluster.

	PostalCode	Neighborhood	District	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	400029	AI staff colony	Mumbai	19.151622	72.854981	2.0	Indian Restaurant	Chinese Restaurant	Food Truck	Café	Pizza Place
1	400065	Aareymilk Colony	Mumbai	19.180136	72.908812	NaN	NaN	NaN	NaN	NaN	NaN
2	400011	Agripada	Mumbai	18.938771	72.835335	0.0	Indian Restaurant	Bar	Irani Cafe	Sandwich Place	Hotel
3	400099	Airport	Mumbai	19.103066	72.863560	3.0	Hotel	Gym	Café	Vegetarian / Vegan Restaurant	Falafel Restaurant
4	400004	Ambewadi	Mumbai	18.938771	72.835335	0.0	Indian Restaurant	Bar	Irani Cafe	Sandwich Place	Hotel

Finally, I visualized the clusters in Mumbai city map.



RESULT

The cluster with most neighbourhood is cluster 0. It consists following categories in decreasing order of popularity:

- Indian Restaurant
- Bar
- Irani Café
- Sandwich Place
- Hotel
- Seafood Restaurant
- Café
- Multiplex

The clusters having following details:

```
print("The biggest cluster is cluster {} and its size is {}".format(compare.index(max(compare)),max(compare)))
print("cluster 0 has {} neighbourhoods".format(compare[0]))
print("cluster 1 has {} neighbourhoods".format(compare[1]))
print("cluster 2 has {} neighbourhoods".format(compare[2]))
print("cluster 3 has {} neighbourhoods".format(compare[3]))
print("cluster 4 has {} neighbourhoods".format(compare[4]))
```

```
The biggest cluster is cluster 0 and its size is 98
cluster 0 has 98 neighbourhoods
cluster 1 has 2 neighbourhoods
cluster 2 has 67 neighbourhoods
cluster 3 has 6 neighbourhoods
cluster 4 has 2 neighbourhoods
```

So, the appropriate location to start the restaurant business should be all the neighbourhoods falling in cluster 0.

DISSCUSSION

The location is one of the most critical factors which determine the success or failure of any restaurant. The takeaway restaurant should ideally be in a densely populated area, keeping the target audience in mind. Accessibility and visibility of the area are also essential to attract customers. Ground floor shops, located at the front are preferred for QSRs and Takeaways.

CONCLUSION

The end of this project brings forth the following conclusions:

- The best locations for opening restaurants in Mumbai city are falling under cluster 1 which can be seen in jupyter notebook.
- Indian restaurants are most popular in Mumbai city followed by bar and then sea food.
- The clusters in map are visualized.
- Foursquare API is used and it made easy to fetch the venues in all neighbourhoods.