Coursera

Starting up a Restaurant - Mumbai

Applied Data Science Capstone Project

Varun Mahesh
7-8-2020

Index

1. Introduction

The objective of this project is to find a suitable region to set up a restaurant in Mumbai, India.

By acquiring the geographical data online, and using Foursquare API, we can find information regarding existing venues around each neighborhood in Mumbai. These venues shall later be filtered categorically for restaurants, and clustered using K-Means Clustering. Ultimately the aim is to obtain a visual representation of the restaurant distribution in Mumbai. This would help us identify areas with relatively sparse presence of restaurants, which would indicate potential spaces to set up a restaurant of our own.

The audience for this project is any prospective entrepreneur exploring the restaurant space in Mumbai.

2. Data Acquisition and Wrangling

The preliminary data regarding the neighborhoods in Mumbai have been taken from the following Wikipedia Page:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

After converting this into a dataframe and cleaning up the data, we get a dataframe like the one attached below:

```
df.head()
```

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

Fig 1. Primary Dataframe containing Geographical Data on Mumbai Neighborhoods

In addition to this, we shall be using Foursquare API to extract information regarding the venues around each neighborhood. Each Foursquare API call shall be Limited to 100 venues, and to a radius of ~1000m around each neighborhood.

The following URL shall be used for making calls to the Foursquare API:

```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
```

VERSION,

area_latitude,

area_longitude,

radius,

LIMIT)

3. Methodology

The methodology applied for this project has been briefly summarized below:

- ❑ Create primary dataframe from Wikipedia Data
- ❑ Call Foursquare API for venues around each neighborhood
- ❑ Filter Venues by Restaurants
- ❑ Cluster Restaurants using k-Means Clustering (10 Clusters)
- ❑ Visualize distribution through Folium
- ❑ Identify potential spaces for setting up restaurant

After filtering the Foursquare API Calls for restaurants, we receive a dataframe like the one attached below:

	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.129300	72.843400	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
1	Amboli	19.129300	72.843400	Cafe Arfa	19.128930	72.847140	Indian Restaurant
2	Amboli	19.129300	72.843400	Nukkad Food Bistro	19.126058	72.846618	Fast Food Restaurant
3	Chakala, Andheri	19.111388	72.860833	Faaso's	19.113938	72.862330	Fast Food Restaurant
4	Chakala, Andheri	19.111388	72.860833	Hit & Run	19.107787	72.863333	Falafel Restaurant

Fig 2. Dataframe obtained after filtering data from Foursquare API Calls

4. Assumptions and Limitations

A few preliminary assumptions have been made while creating this project. These have been clarified below:

- It has been assumed that property rents are uniform throughout the city of Mumbai. Although this parameter would vary from region to region, due to lack of accessible online data this has been excluded from our analysis
- The purchasing power of the population of Mumbai has been assumed to be evenly distributed throughout all neighborhoods. In reality, the southern part of Mumbai would possibly have higher purchasing power than the northern part, on an average. Similarly, many such inconsistencies would be observed. However, due to lack of authentic data, this parameter has been excluded from the analysis
- Procurement channels of the restaurant owner have not been considered. Ideally, the restaurant would be set up in close proximity to the procurement center, if possible, to cut distribution costs. However the entrepreneur could also set up a convenient supply chain after finalizing the location of his restaurant. While this could have been a useful parameter to consider, ignoring its effect on this analysis would be a fair assumption.

While it has been attempted to make this analysis exhaustive, there remains scope for delving deeper into this problem with additional data. A few possible extensions to this analysis have been mentioned below:

- Restaurants could be filtered and sorted as per the restaurant type, to give the entrepreneur a better idea about his direct competition.
- Accessibility to public and private transport could be factored into this analysis. If our restaurant is set up close to a metro station or a popular

shopping complex on a highway, it would be visited more frequently as opposed a setup on the outskirts of the city.

- ❑ Proximity to residential buildings could also be considered to improve this model.

5. Results and Discussions

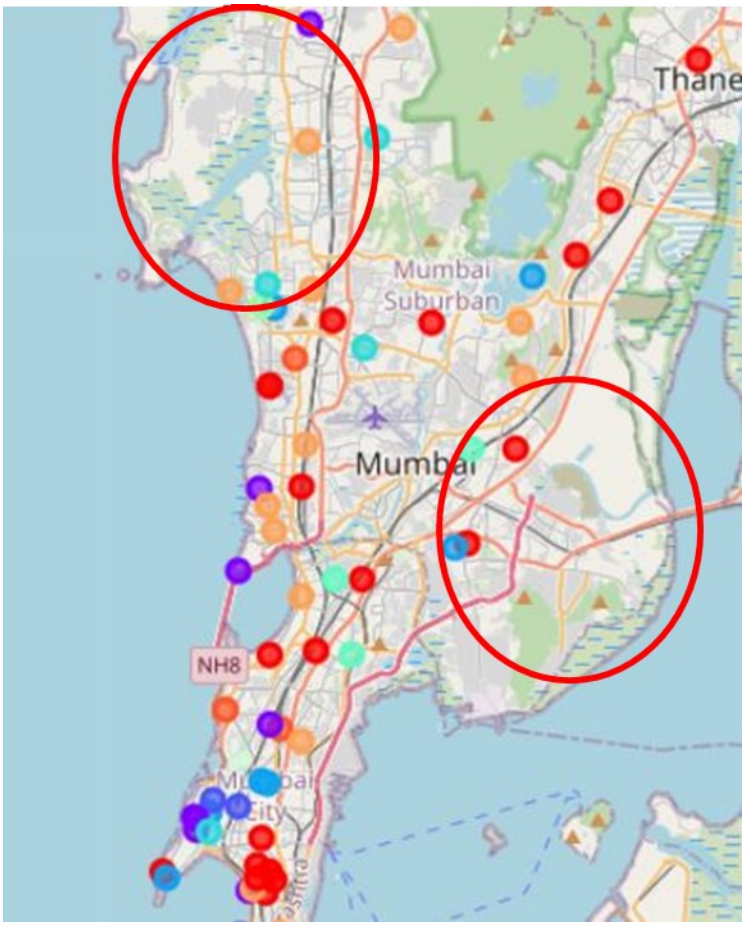

Fig 3. Map obtained after performing K-Means Clustering

The results obtained indicate the following:

- ❑ High Density of restaurants have been observed in South and West Mumbai.

- Regions highlighted in the figure have sparsely distributed restaurants.
- This agrees with the conventional notion that South and West Mumbai are the developed areas of Mumbai, whereas North and East Mumbai are the developing regions.
- Indication that there is high potential to set up a restaurants in Northern and Eastern Mumbai, specifically around the areas highlighted in the figure above.

6. Conclusion

There seems to be a high potential to set up restaurants in the Northern and Eastern Parts of Mumbai. While these regions contain a significant distribution of the residential areas, they are still part of developing Mumbai, and there is good scope to set up a restaurant business over there. The map obtained in this analysis highlights the regions which have a sparse distribution of restaurants. The data analysis and visualization presented agrees with the conventional wisdom about the various neighborhoods in Mumbai. Further scope for refining this project has also been detailed in this report, for the benefit of any prospective entrepreneurs exploring the restaurant business in Mumbai.