Capstone Project - The Battle of Neighbourhoods (Week 2)

**Tanvik V P**

**Data Analyst and Database Developer - Accenture**

Jan – 16th 2021

A large building with many people in front of it

Description automatically generated with low confidence

# **1.Introduction**

**1.1 Background**

Mumbai (previously known as Bombay) is the biggest metropolis of India. A city that is full of life and is also known for its well-known tourists places, commercial hubs restaurants and government bodies. It is also known as the financial capital of India. The city is located on the western part of the India and is the capital of Maharashtra. Restaurants across the city are in cahoots to bring out the food connoisseurs with a range of food showcasing the art of fine dining to mouth-watering street food. The dining experience at an upscale restaurant in Mumbai is more or less the same as elsewhere in the world.

Whenever a person searches for a venue in a new city, they’re highly interested in the best places that the city has to offer. The person might want to know how good a given restaurant is or the price range it falls under. This extra information would help decide which venue to choose amongst the many venues in the city. Combining the location of the venues in the city with their price and rating information would surely help visitors in a city make better informed decisions about the places they should visit.

**1.2 Target Audience**

The concept of theme based restaurant is growing at a very fast pace. Every entrepreneur wants to be different from the rest so that a new element can be presented to the guests. Distinction is seen in form of cuisine offered, décor, menu designing, overall concept etc. Customers have very high demands, fulfilling of which can make a restaurant successful in the business. Since the majority of guests are from upper-middle class to rich category, they are likely to be the ones to try new cuisines. It would be good for the restaurant owners/managers to know what people want and what they are searching for.

# **2.DATA**

**2.1 Data Sources**

To get location and other information about various venues in South Mumbai, I used two APIs and decided to combine the data from both of them together. As Zomato has more influence and Restaurant reviewers than FourSquare in India.

Using the Foursquare’s explore API (which gives venues recommendations), I fetched venues up to a range of 6 kilometres from the centre of South Mumbai and collected their names, categories and locations (latitude and longitude).

I used the Zomato search API to fetch venues from its database using the Latitude and longitude values of South Mumbai. This API helps to find venues based on search criteria (usually the name), latitude and longitude values and more. As the data from FourSquare and Zomato API’s did not align properly, Data cleansing was required.

From Foursquare API (<https://foursquare.com/developers>):

I retrieved the following for each venue:

* **Name**: The name of the venue.
* **Category**: The category type as defined by the API.
* **Latitude**: The latitude value of the venue.
* **Longitude**: The longitude value of the venue

From Zomato API (<https://developers.zomato.com/api>) :

* **Name**: The name of the venue.
* **Address**: The complete address of the venue.
* **Rating**: The ratings as provided by many users.
* **Price** **range**: The price range the venue belongs to as defined by Zomato.
* **Price** **for** **two**: The average cost for two people dining at the place. Conversion of PF2 is by getting average price per person by multiplying by 2.
* **Latitude**: The latitude value of the venue.
* **Longitude**: The longitude value of the venue.

# **3.Methodology**

This project aims at identifying the venues in South Mumbai based on their rating and average costs. This would enable any visitor to identify the venues customer wants to visit based on their rating and cost preference.

As a first step, we retrieved the data from two APIs (Foursquare and Zomato). We extract venue information from the center of mumbai, upto a distance of 6 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

We then explored the data retrieved from the two APIs on the map and identified the top category types. The data from the two sources is carefully combined based on the name, latitude and longitude values from the two sources. The final dataset would include the rating and price values for each venue.

Next, we'll analyse the data that we created based on the ratings and price of each venue. We'll identify places where many venues are located so that any visitor can go to one place and enjoy the option to choose amongst many venue options. We'll also explore areas that are high rated and those that are low rated while also plotting the map of high and low priced venues. Lastly, we'll cluster the venues based on the available information of each venue. This will allow us to clearly identify which venues can be recommended and with what characteristics.

Finally, we'll discuss and conclude which venues to be explored based on visitor requirement of rating and cost.

**3.1 Data Ingestion and API call’s**

1. Latitude and longitude is given to a parameter which gets midpoint location of South Mumbai.
2. The Latitude and longitude is passed to folium to get the map.

|  |  |
| --- | --- |
| |  | | --- | | Map  Description automatically generated | |

1. These Parameters are passed to Zomato API and Foursquare API to get the venues in the 6 kms distance from the central point shown above.
   1. **Data Cleansing and Merging Foursquare and Zomato API’s**
2. The data from multiple resources might not always align. Thus, it is important to combine the data retrieved from multiple resources properly.
3. Plotting the venues/data from foursquare API into the map(green markers)

|  |
| --- |
|  |

1. Plotting the venues/data from Zomato API into the map(red markers)

|  |
| --- |
|  |

1. Combining the data set of FourSquare and Zomato API using latitude and longitude values of each corresponding venue
2. calculating the difference between the corresponding latitude and longitude values and see if the difference is less than 0.0004 which should ideally mean that the two locations are same.
3. The final cleansed data set of 76 venues when combined is as shown below.

|  |
| --- |
|  |

**4.Analysis:**

Data Set is cleansed and is in the final form We will inspect these venues based on their rating. The rating of a venue are based on user reviews and belongs to a range from 1 to 5. We'll also analyse the venues based on their price per person as well as the price range.

# **4.1** **Categories**

We have various types of venues in the final dataset. We will take a look at the venues and check which are the majority venue categories in the list.

The below chart shows us the count of venues grouping by the type of restaurants.

Chart, histogram

Description automatically generated

In this above chart we can see the majority venues are Indian restaurant, and at second comes Bar, Cafe, Coffee shop and Ice cream shops. So, if as a tourist, you're looking for these places, you're in luck.

# **4.2 Rating**

Rating of a venue is an important factor on which a visitor decides whether it is worth it to visit the place. To get to this, we will first see what is the average rating for all the venues in the city. Next, we will plot the venues on the map and colour code them.

We'll first identify the various rating values and plot them as a bar plot with their counts to see the most common rating

**Average Ratings Bar chart:**

Chart, histogram

Description automatically generated

From the above chart we can see that the majority of places has rating from 3.5 to 4.3. For this creating bins for the rating would be a good choice and would be easier to cluster.

**Map based on Rating/Reviews:**

Map

Description automatically generated

* **Red** : 1 to 2 (Very low Rating)
* **Orange** : 2 to 3 (Average Rating)
* **Green**: 3 to 4 (High Rating)
* **Dark Green**: 4 to 5 (Very High Rating)

As you can see from the map that there is no places with **Red**(very low rating), there are 4 places with **Orange** (Average rating) and many places with **Green**(High) and **Dark Green** (Very high )rating.

**Map based on Price Range:**

**Map

Description automatically generated**

* **Red** : 4.0 (Very High Price)
* **Orange** : 3.0 (High Price)
* **Green**: 2.0 (Average Price)
* **Dark Green**: 1.0 (Very low Price)

Venues with **Red** has high price range, followed by **orange** with high price and **Green** and **Dark Green** with low price ranges.

# **4.3 Clustering**

We will now cluster all these venues based on their price range, location and more to identify similar venues and the relationship amongst them. We'll cluster the venues into two separate groups.

Map

Description automatically generated

Here we can see the clusters near Mumbai Naval base has relatively more price than the western side of the naval base.

# **Result and Discussion**

Based on our analysis above, we can draw a number of conclusions that will be useful to aid any visitor visiting the city of Mumbai, India.

After collecting data from the Foursquare and Zomato APIs, we got a list of 168 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as names to combine them and remove all the outliers. This resulted in a total venue count of 76.

We identified that from the total set of venues, majority of them were Hotel and Pizza Places.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 4.01. On the other hand, there are few venues which are high priced and have average rating of 4.28.

A consumer and food explorer can use this information to explore or research and a company can use this information to build up an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price).

# **Conclusion**

The purpose of this project was to explore the places that a person visiting to South Mumbai . The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are so many places where one can visit during his stay in South Mumbai. Based on the visitor's venue rating and price requirements, they can choose the places.

Note: As GitHub doesn’t allow python based maps to be rendered when pushing to the main branch to view it later, Providing you the link to my IBM note book for detailed code, explanation and maps for better understanding.

Link to the NoteBook:

<https://jp-tok.dataplatform.cloud.ibm.com/analytics/notebooks/v2/d5c73a8d-0e8c-4d59-b665-7a9d5cf31ce8/view?access_token=9c0694477e71f7b9941abbe5c90eada9e3291272606ccd1e1a3d64a356ba8d89>