A Tourist's Guide to Restaurants in Mumbai

This project has been submitted in partial fulfilment for the IBM Data Science specialisation offered on Coursera

PROJECT BY: SAURABH SUMAN

Table of Contents

1. INTRODUCTION

The City of Dreams – Mumbai, attracts a wide range of aspirants. Some come to the city with the hope of gaining some meaningful employment, some pass by the city, some come as tourists and some come as students. The diversity of aspirations of the footsteps setting in Mumbai is perhaps as wide as their demography.

However, when one is far away from home, the person often tries to explore corners in the new city they are in, where they can find that corner, which can give them solace and exchange the warmth of their home city.

As a student having lived in Mumbai for a period of two years, food exploration became one of my key experiences and led the foundation to several friendships, and widened my ability to embrace greater cultural differences. This remains one prime reason as to why I have chosen Mumbai, and have chosen to classify neighbourhoods on the basis of restaurants in Mumbai.

The importance of this project could be its contribution to the wonderful experiences one can have in Mumbai, especially when cuisine is concerned. The project can especially be useful to travel bloggers, students, as well as tourists, and the common category across all – the foodies of course. This will also help in enhancing the lived experiences, and might contribute in adding to the list of familiar places one has in a particular city, guided by the zeal of exploration.

For the purpose of identification of these neighbourhoods with the greatest number of restaurants, Foursquare location data, along with making use of geocodes, finally

topped with creation of a map will be carried out. The same have been explained in detail in the next section. Stay tuned.

2. DATA REQUIREMENTS

In order to address the goals of the project, the data required can be divided into three parts:

1. Data for the list of neighbourhoods in Mumbai, Maharashtra - to be retrieved from Wikipedia.

2. The coordinates for the city and the neighbourhoods (latitudes and longitudes) – to be extracted with the help of the 'geocode' function in python.

3. The data regarding the various 'Restaurants' in the city – **Foursquare** location data will be leveraged to extract this information.

These information points combined will help in making the necessary classifications for the neighbourhoods. K-Clustering method will be utilized to form cluster, followed by their mapping.

3. METHODOLOGY

First and foremost, it was ensured that all the required packages are installed and imported. Thereafter, the data extraction process began with scrapping of data from Wikipedia. The data was extracted using python inbuilt libraries and beautiful soup packages.

Following the data extraction and cleaning, a data frame with the data for neighbourhoods in Mumbai was created. Following this, the coordinates for the respective neighbourhoods were imported using the 'geocoder' package. This helped in aligning the neighborhoods with their latitudes and longitudes. After this, a map was created using 'Folium' to cross check the neighbourhoods and their location in Mumbai city.

Hereafter, the **Foursquare** location data played a crucial role in fulfilling the project goals. API calls were made in a loop to gather the venue detail such as name and category in each neighbourhood. Later, these categories were filtered for only 'Restaurants'. The 'Restaurants' category is exhaustive and hosts 36 categories of restaurants.

Namely - 'Afghan Restaurant', 'American Restaurant', 'Asian Restaurant', 'Brazilian Restaurant', 'Chinese Restaurant', 'Comfort Food Restaurant', 'Dim Sum Restaurant', 'Falafel Restaurant', 'Fast Food Restaurant', 'French Restaurant', 'German Restaurant', 'Gluten-free Restaurant', 'Goan Restaurant', 'Halal Restaurant', 'Hawaiian Restaurant', 'Indian Restaurant', 'Italian Restaurant', 'Japanese Restaurant',

'Korean Restaurant', 'Maharashtrian Restaurant', 'Mediterranean Restaurant', 'Mexican Restaurant', 'Middle Eastern Restaurant', 'Modern European Restaurant', 'Molecular Gastronomy Restaurant', 'Mughlai Restaurant', 'Multicuisine Indian Restaurant', 'New American Restaurant', 'Punjabi Restaurant', 'Restaurant', 'Seafood Restaurant', 'South American Restaurant', 'South Indian Restaurant', 'Sushi Restaurant', 'Thai Restaurant', 'Vegetarian / Vegan Restaurant'.

This helps in identifying the cultural plurality of the neighbourhoods, and wide range of choices available for the city dwellers as well as tourists.

The K-means clustering technique was used for the identification of the different clusters based on the types of restaurants. These were superimposed with the data frame for the neighbourhoods to understand the neighborhoods based on the number of different restaurants.

The results have been further elaborated upon in the next section.

4. RESULTS

A final map was created using the 'Folium' package to visualise the clusters. The inferences on the basis of these clusters have been made following the illustration:

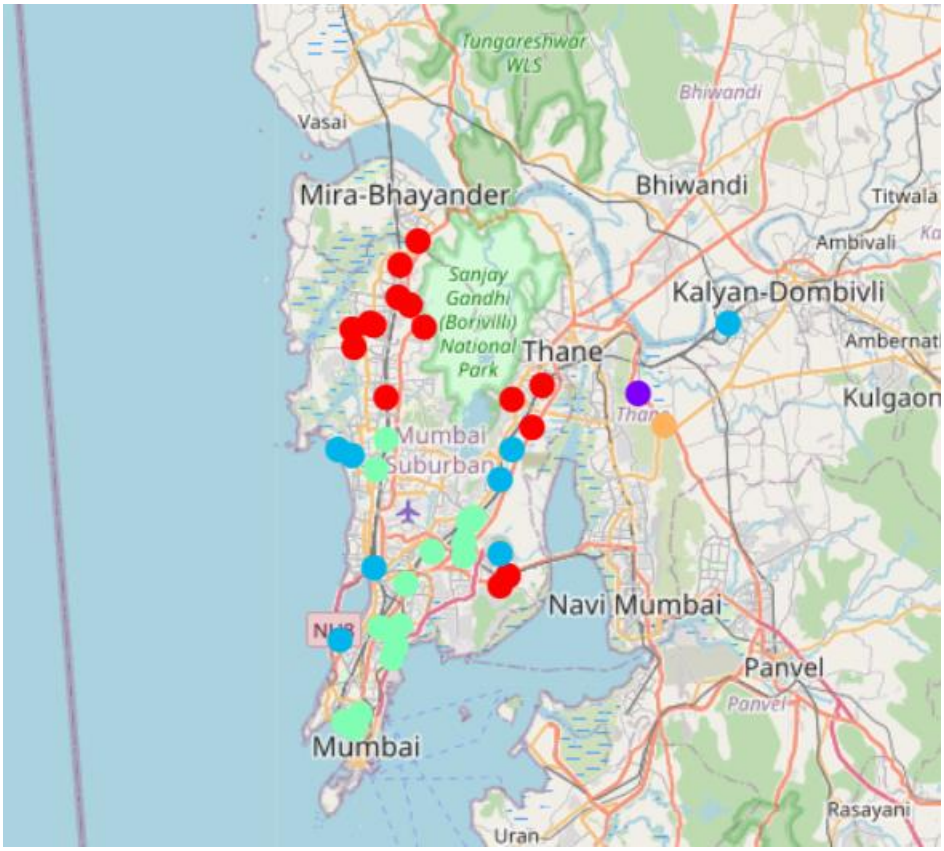

Figure 1: Neighborhood Clusters with Restaurants

In the above figure, the clusters can be visualised based on the density of different types of restaurants. The red colour marks clusters with '0' occurrence, purple with '1', blue with '2', green with '3' and orange with '4'.

The clusters red and purple, which represent an occurrence of 0 and 1 may indicate a lower level of cuisine diversity in the neighbourhood.

Whereas, the clusters with the blue, green and orange colors, which indicate the occurrence as 2, 3 and 4 respectively might be representative of a higher cuisine diversity in the neighbourhoods.

It can also be seen from the map that the greatest number of restaurants offering different cuisine are located towards the southern part of Bombay, and those with lower number of cuisines offered are located in the northern part of Bombay.

The probable utility of the results is further discussed in the next section.

5. DISCUSSION

The map with the cluster of neighbourhood and restaurants can be very revealing. It clearly depicts the neighbourhoods with a higher concentration of restaurants with a wider range of cuisine choices. At the same time, it neatly indicates the localities that the stakeholder population can choose from to build up on their food experiences. The southern part of Bombay can be seen clearly offering more varieties of food, as compared to the other parts of Mumbai.

The map is somewhere a good depiction of the ground, as having lived two years in Mumbai, and having explored the city for food, I can affirm from personal experiences that South Bombay does offer cultural and cuisine diversity. This also is an evidence of the authentic nature of the **'Foursquare'** location data.

In addition to these, the neighbourhood clusters are also indicative of the high concentration of restaurants in a particular neighbourhood. This can also help the

potential investors identify spots which are not so crowded with restaurants and can help them narrow down an investment area.

Another potential of the project and the exercise is that it can help entrepreneurs further filter the data of the types of restaurants and the cuisines they offer, and help them decide which restaurant and what cuisine would they like to offer in case they want to venture in a neighbourhood with already high number of restaurants.

6. CONCLUSION

The project uses the Foursquare data on an as it is basis. Therefore, the categories which had Restaurant included in them have been included for the assessment. The study is good as a pilot, but more rigour needs to be added to further refine the results for more robust conclusions.

The neighbourhoods do indicate the presence of several restaurant categories (in context of cuisine), however, some eateries or cafes, or even restaurants could have remained excluded due to a different description of venue. Such categories need to be refined and added for a better representation of the ground reality.

The project model can lay a basis for further analysis with addition of more data points, and with integration of a richer location data. The scope of the study is limited for a project, and has also been subject to factors such as a restricted number of calls due to the personal and free developer profile.

Nonetheless, the project has provided for a great learning exercise, and gives the foundation and basis for bigger project capabilities.

7. REFERENCES

1. Wikipedia - https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai
2. Foursquare API