CUISINES OPPROTUNITIES TO INVEST IN MUMBAI

PARTH BELADIYA

(20 Mar 2020)

INTRODUCTION

Background

Mumbai being a financial capital of India has a large population and lot of businesses. Mumbai also being a tourist place welcomes lot of visitors both foreign and from the country. People have shown interest in experiencing the world cuisines since past years and business in international cuisines have increased apart from local cuisines. The competition is high, and a restaurant needs to unique in its area to succeed.

Problem

Although Mumbai has lot of restaurants already, the idea of new restaurant to invest upon is lucrative. This analysis focuses on analyzing the cuisine thriving in Mumbai and exploring new opportunities to invest in a different cuisine which is more promising compared to ones nearby.

Interest

This report interests the investors looking for a new restaurant to invest upon. The pre-existing owners of restaurants have an extra source of information to further expand their businesses. New entrepreneurs looking at restaurants as their target business can obtain a perspective and identify opportunities from the report.

DATA COLLECTION & PREPARATION

Data Sources

1. **Mumbai Geo JSON**: We would be using the Geo-JSON data for Mumbai to analyze the areas and map the restaurants with the neighborhoods.
   https://github.com/datameet/Municipal_Spatial_Data/blob/master/Mumbai/BMC_Wards.geojson
   Data which will be useful from the source : Area Names, Geo Coordinates, Area Boundaries
2. **FourSquare API**: The Foursquare API would be used to obtain restaurant's details especially the cuisine data in a nearby area.
   Data which will be useful from the source: Restaurant ID, Cuisine ID, Cuisine Name, Location Coordinates
3. **FourSquare Categories** : https://developer.foursquare.com/docs/build-with-foursquare/categories
   We will obtain the list of Categories under "Food", filter the required ones and create a CSV file which will be used for API calls.
   Data which will be useful from the source : Category IDs

Feature Selection

Now that I was ready with the required data, I then proceeded to gather data as follows.

1. Obtain Names of Area Locations

I started my data collection with obtaining the coordinates of Mumbai location. Next, I obtained the geoJson file for Mumbai using the URL mentioned above and converted to required format for further process. The contents of the file were not accurate or streamlined. Some of the location names contained white spaces and the casing was not aligned. Moreover, the coordinates Latitude and Longitude were swapped. The problems were considered and to be corrected post data gathering. A list of area was extracted from the geoJson to process further.

2. Obtain All Categories Data from FourSquare

The FourSquare API was used to query all the geo location coordinates with the above collected category filters and a range of 2.5 km. All the content obtained was stored in a data frame. The contents extracted from JSON responses included the Restaurant Id, Restaurant Name, Location, Coordinates, Category, Category ID.

3. Obtain All Restaurant Data from FourSquare

Once the main location and area content was obtained, I created as set of random points on the boundaries of a location along with projected center of the area. This data would be used to query FourSquare API to get restaurants nearby. The random points helped in covering most of the area in the neighborhood and thereby obtaining maximum possible cuisine restaurants.

I used the FourSquare category details and drafted a csv file. A set of cuisine categories & subcategories up to level 3 down were finalized and uploaded in the notebook. A set of unwanted categories were marked as ignored in the dataset csv. The content considered was Category Name, Category ID, Primary Category (if the category is a subcategory).

Data Cleaning

Although the data was available, there was lot of redundancy since we collected data from multiple points and the same place can be a part of the 2.5km range. I performed the below steps in order to make data formatted and cleaned.

- The names of Area were trimmed and made to upper case for consistency.
- Removed Duplicate Content from Four Square responses
- Removed Spaces in all Content
- Removed Nan values if present
- Join Category to Primary Category Data
- Remove Indian and Chinese and Asian restaurants
- Converted the categories which were not mapped to a cuisine but belonged to a cuisine.
    1. Dim Sum was assigned with Chinese
    2. New American was changed to American
    3. Tapas was changed to Spanish
    4. Modern and Easter European was changed to European
    5. Burritos was changed to Mexican

6. Falafel was changed to Middle Eastern
7. Tex-Mex was changed to Mexican
8. South American was changed to American
9. Japanese Curry was changed to Japanese

- Removed restaurants not belonging to required cuisine and generic categories. The list was as follows :'Bar','Hotels','Pizza','Lounge,'Diner','Seafood','Vegetarian/Vegan','Bistro', 'Deli','Arcade','Buffet','Coffee Shop', 'Wine Bar',' Pub', 'Café','Halal'.
- Renamed Subcategory to Primary Category
- Filled missing Category data with Primary Category data were not available

Data Preparation

I created a data set using the cuisine data to generate list with the counts of all cuisine types.

Another data set comprised of a list of area and total restaurant in the area. This was prepared by using the geo location of a restaurant and mapping the location to each polygon area form geo json and finding the are which lies inside the polygon coordinates.

The list of cuisine was identified and attached to the above generated data set. This data comprised of total counts of restaurants in Each Area, And Each Cuisine Type. The areas with no restaurants were removed from the data set.

DATA ANALYSIS

The aim of the analysis is to study the distribution of cuisines across Mumbai and identify the counts and richness of an area. A projection of what cuisine is present in what area and how many, would be our target. I started with plotting of all restaurants and created a heat map to view how the restaurants are more in count.

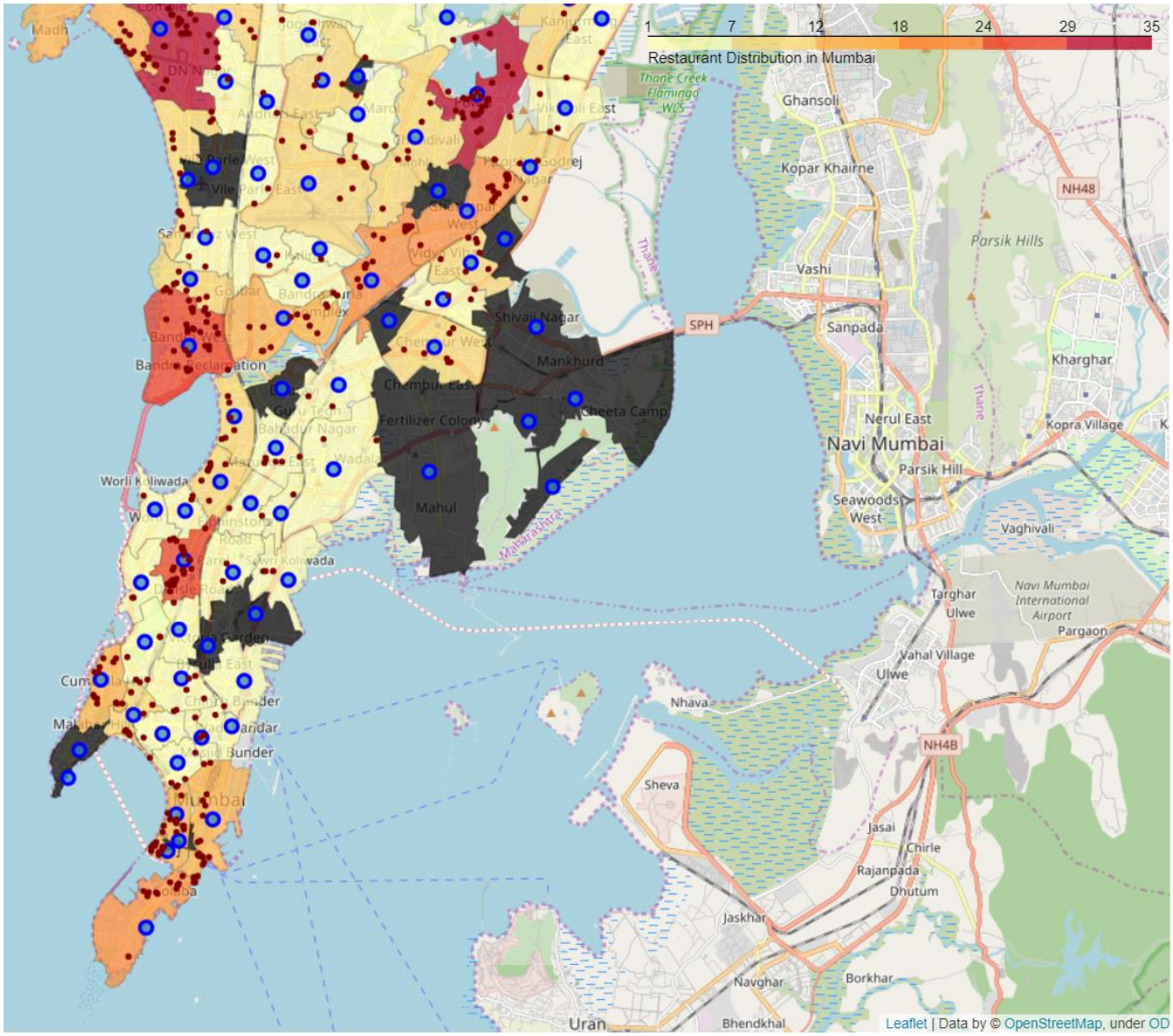
Next I created a chart representing all the areas and the restaurant counts in the area.

Then I created a chart to identify what are the cuisines most available in Mumbai and represent in a bar chart showing the ones with presence of more than 5 across Mumbai.

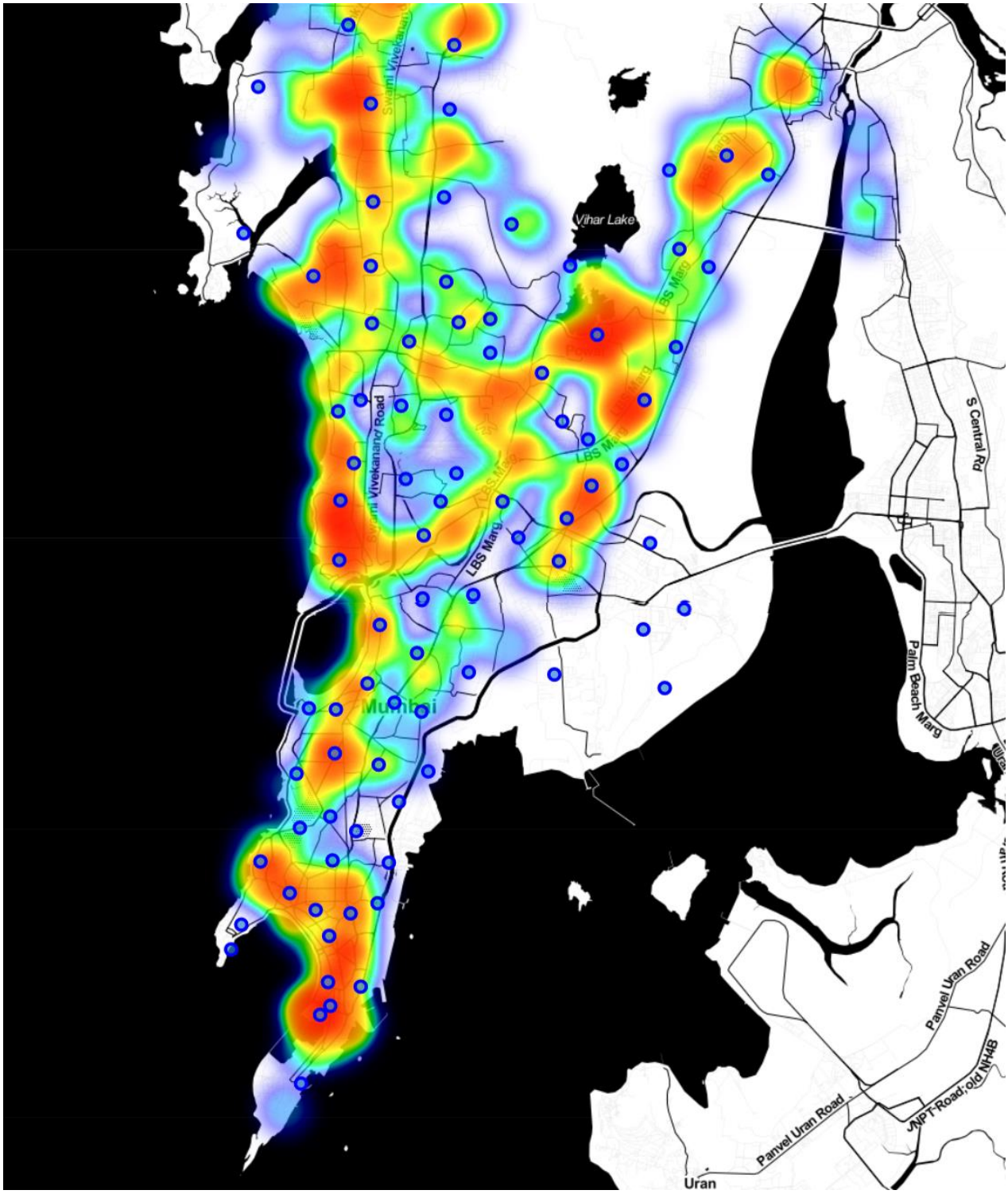
Also, I created a data to find what are the highly renowned places for restaurants and chart the top 15 places amongst them.

Now we know that there are restaurants and cuisines in many areas, but the richness of cuisines is a good way to analyze the areas. I created an index and calculation considering the variety of cuisine in an area and the count of available cuisine in the area. This index was then plotted for all the areas across Mumbai featuring the vividness of restaurants.

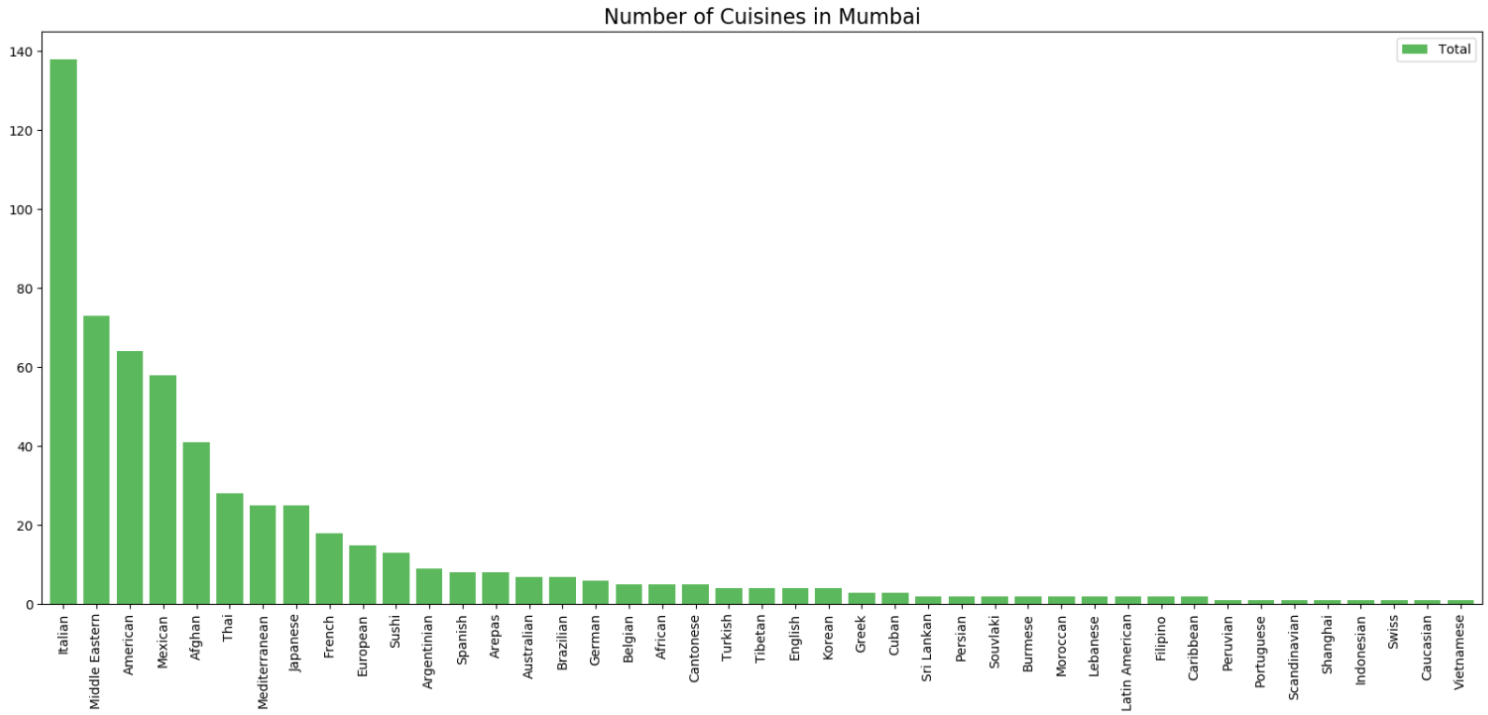
1. Distribution of Restaurants in Mumbai



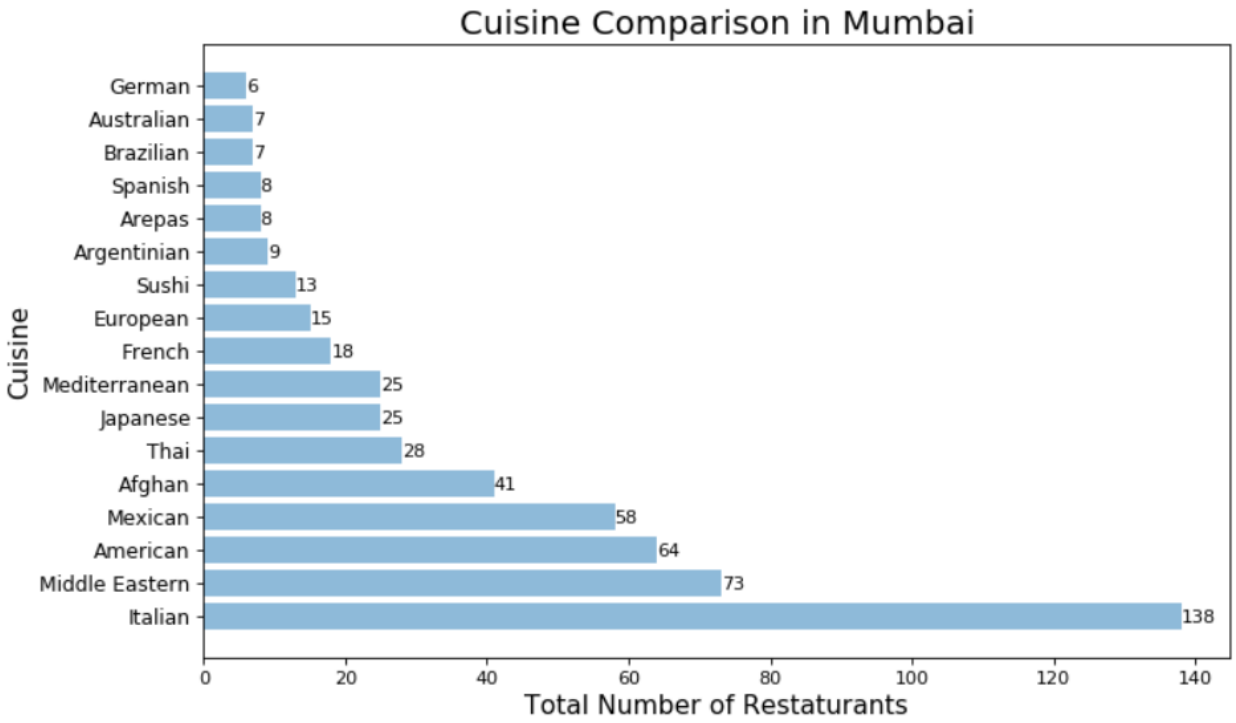
1.2 More Insight on distribution using a different Visualization of Heat Map



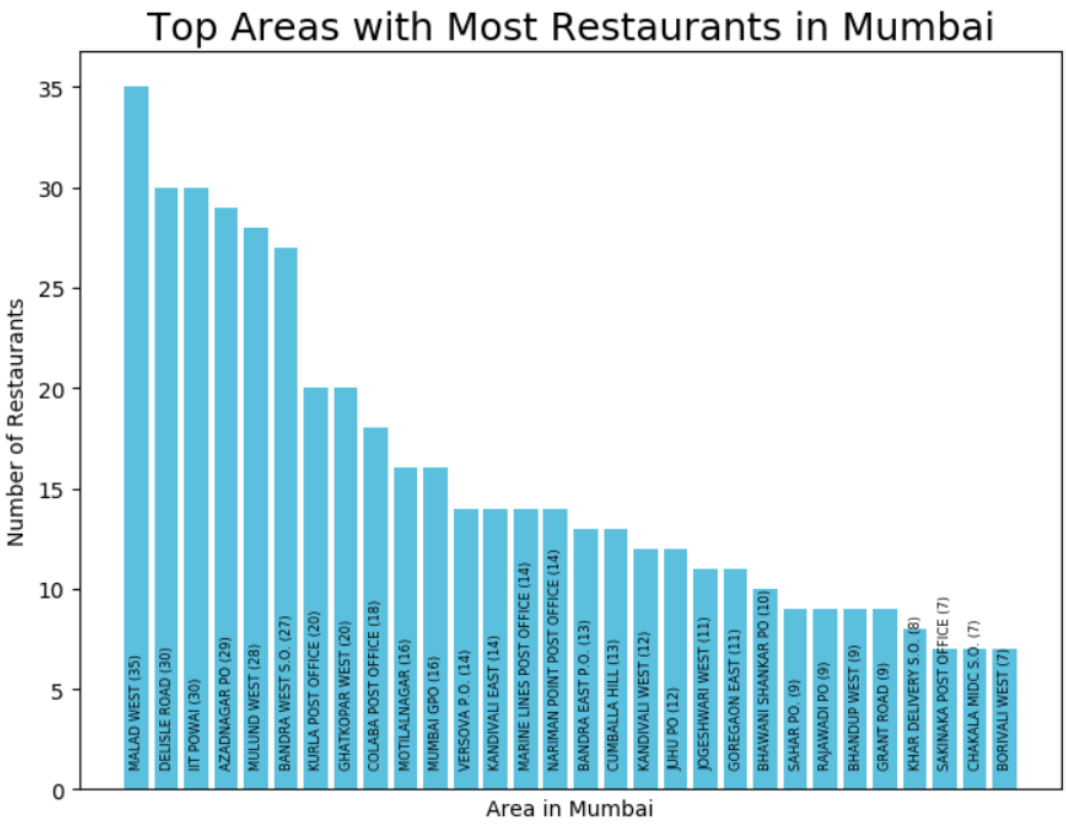
2. Number of Cuisine Restaurants in Mumbai



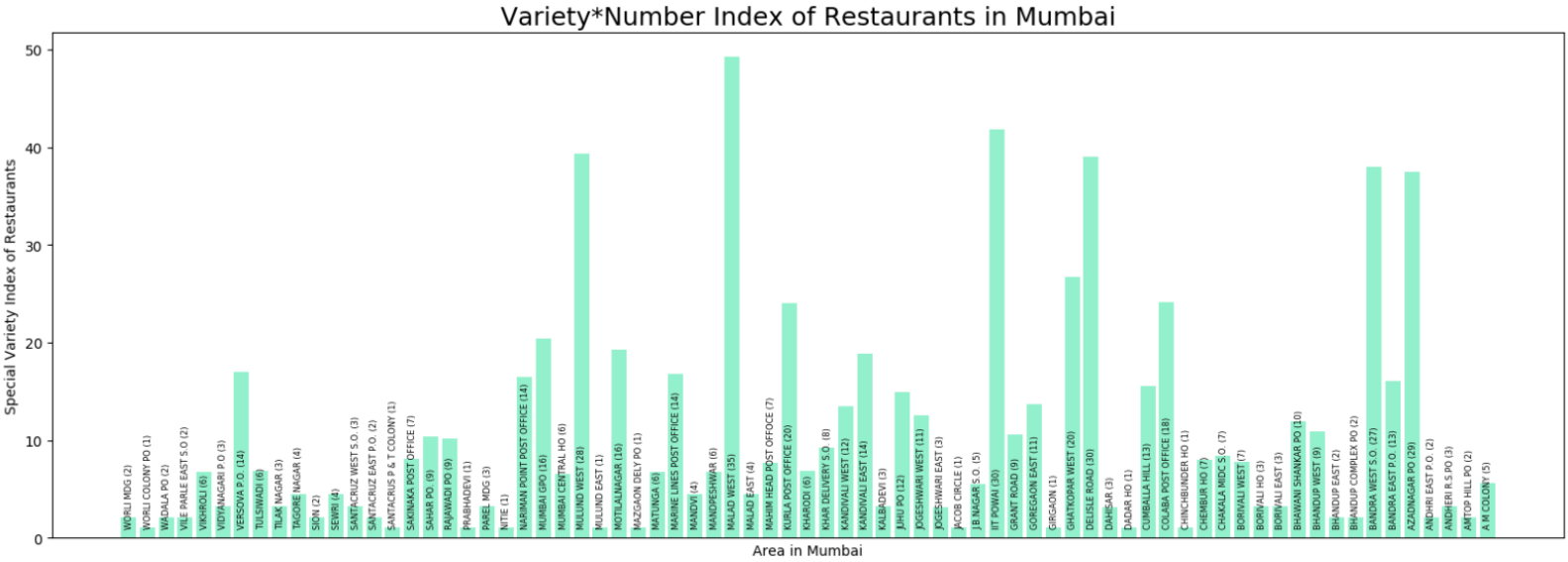
3. Top Cuisine Restaurants with presence of more than 5 across Mumbai compared with other cuisines



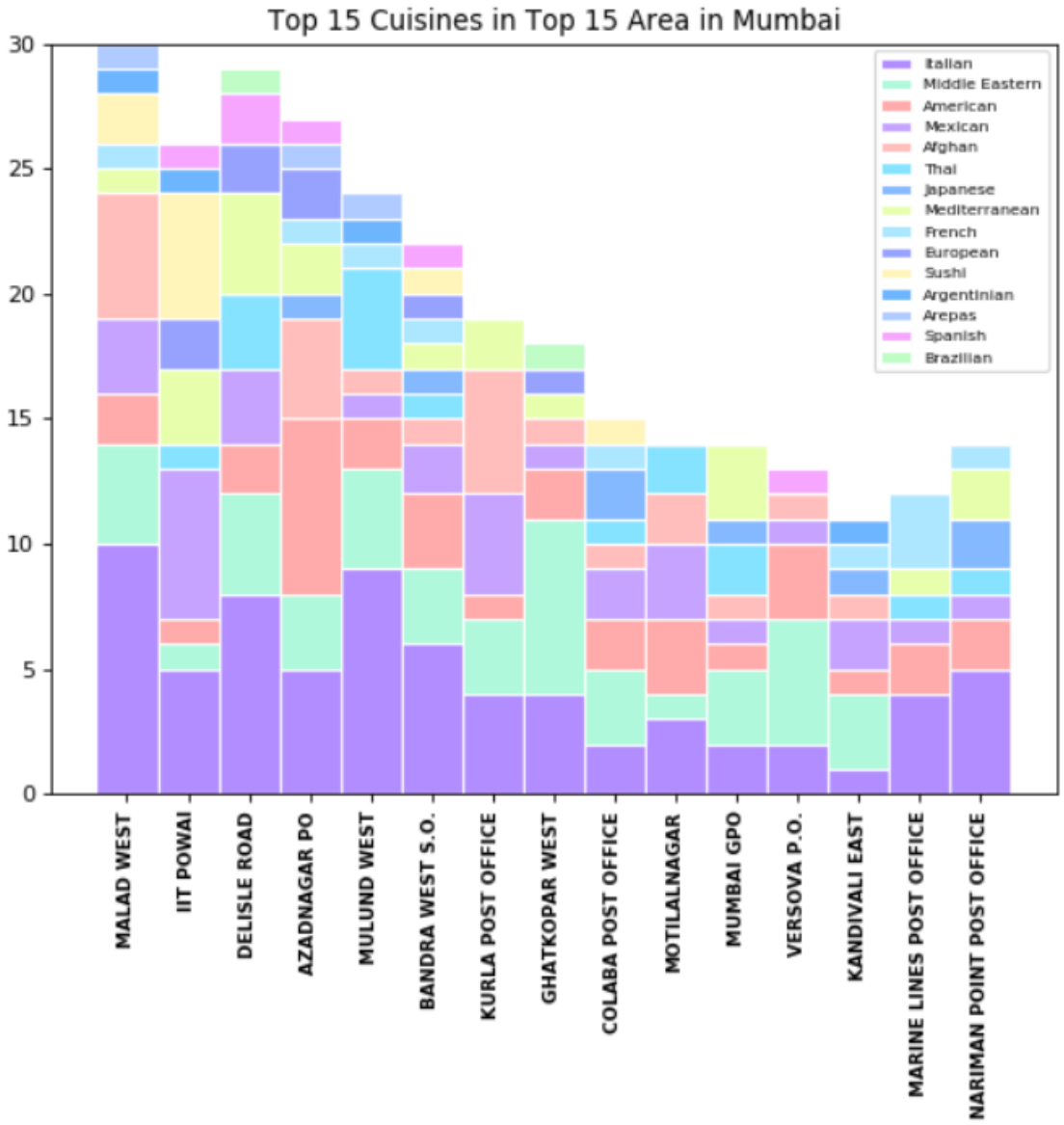
4. Areas with most restaurants



5. A Variety Index representation of Restaurant count against Number of Types of Cuisines



6. Top 15 Cuisines against Top 15 Areas



ANALYSIS

As per our analysis, we have explored the neighborhoods of Mumbai for multiple cuisine types. We considered only the foreign cuisines with exclusion of Chinese, Pizza and Indian. We have focused on the content which are restaurants serving a cuisine and not generic eatery places. We have considered almost all restaurants whose data was available from FourSquare. Most of the areas are on the West coast where there are a greater number of restaurants. The variety index which was calculated based on the number of the total number of restaurants and the variety in cuisines in an area to show how flourishing the area is. Mumbai being a financial capital and having a population of more than 1.5 shows interest of people in cuisines of all types. However, Italian, Middle Eastern, American and Mexican lead in terms of interests.

RESULTS AND DISCUSSION

Opening a new restaurant is challenging as the competition is high and the interest of the people also matters. We filter the categories for international cuisines and generated a dataset for all cuisine restaurants. We specifically removed Indian and Chinese restaurants and we want to target cuisine which are not predominant. We further removed generic restaurant data which may be mapped under categories we are not targeting. Once the data is ready, we generate counts for the cuisine restaurants. This gives us information on what are the top categories in which the cuisines are more developed in Mumbai.

We then map the cuisines with the areas in Mumbai. On representing the data on a heat map, we can see what areas are concentrated with cuisine and with what intensity. Also, a bar chart representation shows what are the top 15 areas where we can find cuisine culture.

Our analysis shows that there are many restaurants for Italian cuisine. Although we have excluded the category for Pizza which very much targets a specific kind of food, the restaurant is the most flourishing. The areas Azadnagar, Malad, Delislie seem to have a greater number of restaurants. Bandra seems to be most exquisite place having most varieties of cuisines.

Italian, Middle Eastern, American are more popular among the prominent areas. But apart from these, the cuisine spreading is very unequal in many areas which leaves a scope of business opportunities for investors.

Areas like Malad, Delislie, Powai and Bandra have a rich culture of restaurant cuisines with more variety and numbers.

CONCLUSION

This project is meant to analyze and obtain details on the cuisines present in Mumbai and how they are distributed among various areas. It gives an insight to the investors looking to invest in restaurants, an idea of the various neighborhoods and the concentration of food junctions. The analysis data graphs represent the total distribution of restaurants and cuisines obtained from FourSquare, including the counts, the category of cuisine and the neighborhood where they are present. The data obtained is mapped with the geographical coordinates to represent them on a map to have a better idea of their distribution. Further, we represent the data in bar charts to visualize the numbers pertaining to the neighborhoods. We finally cross reference the counts of neighborhoods and individual cuisine types to create a whole visualization for most flourishing areas.

Considering the above data, one can decide upon the probable cuisine to invest upon and what area to target which poses less competition and more business. It is ultimately the stake holders who needs to decide upon where to invest considering the richness, flourishing area, interest of the people in cuisines and competition.

FUTURE SCOPE

A further analysis can be made using the selected data on factors affecting the flourishing of restaurants. However, the required data needs to be gathered for the analysis like real estate, population density, cost of living in areas, work culture and workplaces nearby.