IBM DATA SCIENCE

Applied Data Science Capstone

The Battle of Neighbourhoods (Part-1)

Finding Best Venues, Finding Best Neighbourhoods, Segmenting and Clustering Neighbourhoods in Mumbai City

FINAL REPORT

Introduction

Mumbai is the capital city of the Indian state of Maharashtra. According to UN, as of 2018, Mumbai was the second most populous city in India after Delhi and the seventh most populous city in the world with a population of almost 20 million. In 2008, Mumbai was named an alpha world city. It has the highest number of millionaires and billionaires among all cities in India. Mumbai is home to three UNESCO World Heritage Sites: the Elephanta Caves, Chhatrapati Shivaji Maharaj Terminus, and the city's distinctive ensemble of Victorian and Art Deco buildings.

Mumbai is the financial, commercial and entertainment capital of India. It is also one of the world's top ten centres of commerce in terms of global financial flow, generating 6.16% of India's GDP and accounting for 25% of industrial output, 70% of maritime trade in India (Mumbai Port Trust and JNPT), and 70% of capital transactions to India's economy. Mumbai's billionaires had the highest average wealth of any city in the world in 2008. The city houses important financial institutions and the corporate headquarters of numerous Indian companies and multinational corporations. It is also home to some of India's premier scientific and nuclear institutes. The city also houses India's Hindi (Bollywood) and Marathi cinema industries. Mumbai's business opportunities attract migrants from all over India. Tourists from various countries around the world visit Mumbai.

With such a huge population and many venues for various purposes its often hard to choose the best one. Therefore, this project will help user decide best neighborhoods, best venues, etc.

Problem:

Note: Best is selected for the category of venue given by user as input.

To find the answers to the following questions:

* What are the places of Mumbai City that have top rated venues of the given category?
* What are the top-rated venues of Mumbai City of the given category?
* Which are the best places to stay if you prefer particular type of venues such as Gyms, Pubs, Italian Restaurant, Chinese Restaurant, etc. ?
* How are the neighborhoods around the Mumbai clustered?

Data Section

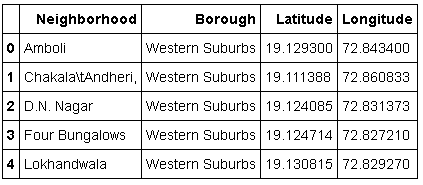
Mumbai is the financial, commercial and entertainment capital of India. It is also one of the world's top ten centres of commerce in terms of global financial flow, generating 6.16% of India's GDP. Therefore, there are venues of various categories and given below are the sources which will be used to collect data.

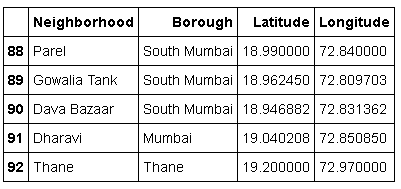
Data needed for this project is as follows:

* Mumbai City data that contains list Boroughs, Neighbourhoods along with their Latitude and Longitude.
  + Data source : <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai>
  + Description : This website contains the required information. After cleaning this data, it will be used to explore various neighbourhoods of Mumbai City.
* Venues of certain category given by user, details of venues of certain category and top venues in each neighbourhood of Mumbai City.
  + Data source : Foursquare API
  + Description : By using this API required information can be fetched and after converting it to data frames it can be cleaned and used to get required data.
* Locating the coordinates of addresses and cities.
  + Data source : geopy-geocoders (Third party geocoders and other data sources)
  + Description : This will be used to fetch co-ordinates of locations.

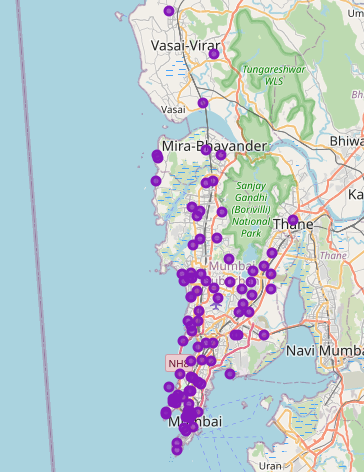
Methodology:

1. Importing necessary libraries and defining functions to get venues of certain type, get details of venue and get venues around neighborhood.
2. Download dataset containing of Postal Code, Boroughs and Neighborhoods from link.
3. Cleaning of dataset obtained above.

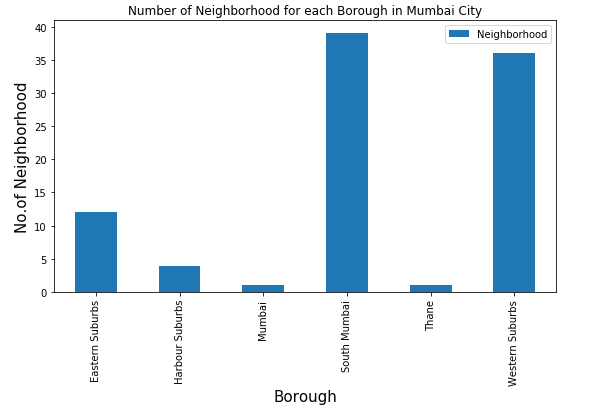




1. Fetching co-ordinates of Mumbai using geopy.
2. Visualizing Map of Mumbai City with Neighborhoods superimposed on top.



1. Bar plot of Number of Neighborhoods in each borough of Mumbai.



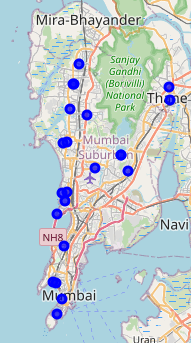
1. Get input to find venues of certain category.



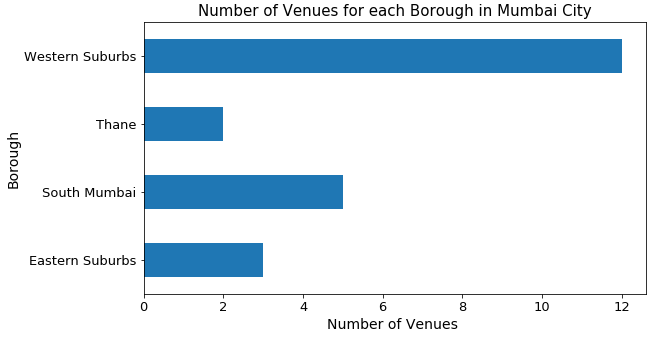
1. Using Foursquare API to download data of venues containing of category given as input.
2. Converting data of venues to data frame and cleaning it.



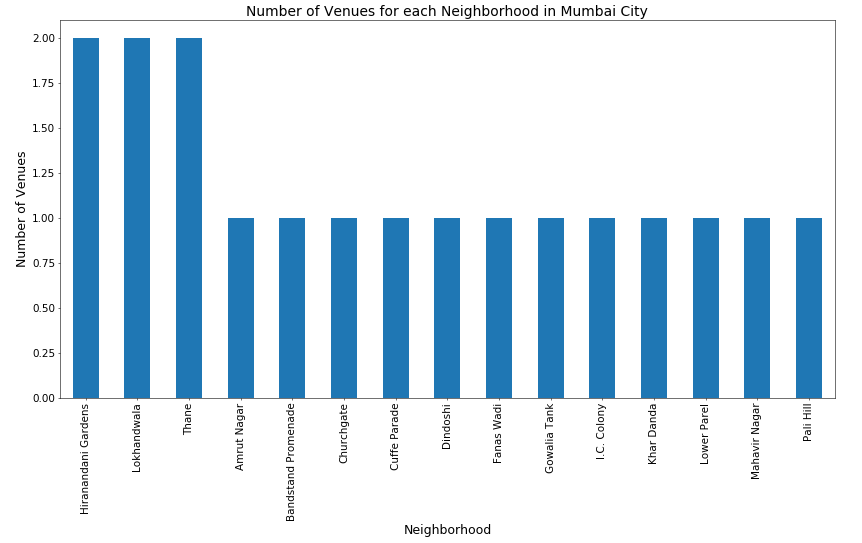
1. Visualizing Map of Mumbai City with Venues superimposed on top.



1. Bar plot of number of venues in each Borough of Mumbai



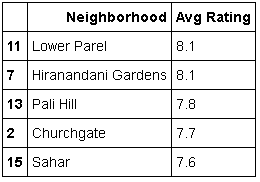
1. Bar plot of number of venues in top 15 neighborhoods of Mumbai



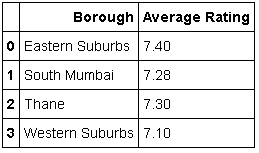
1. Using Foursquare API to download data of details of venues containing of category given as input



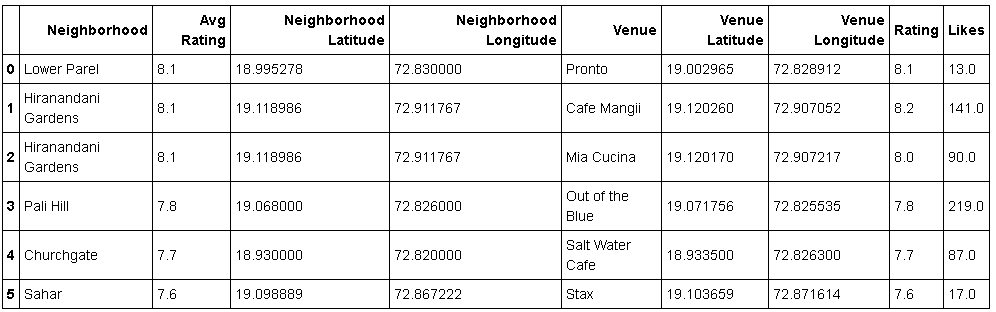
1. Creating dataframe containing neighborhood and average rating



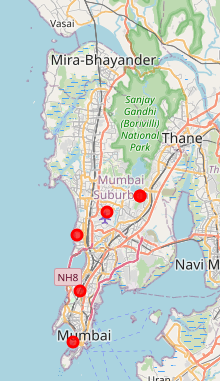
1. Creating dataframe containing borough and average rating



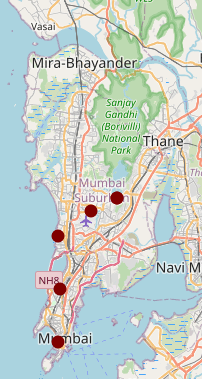
1. Creating dataframe best neighborhood to stay



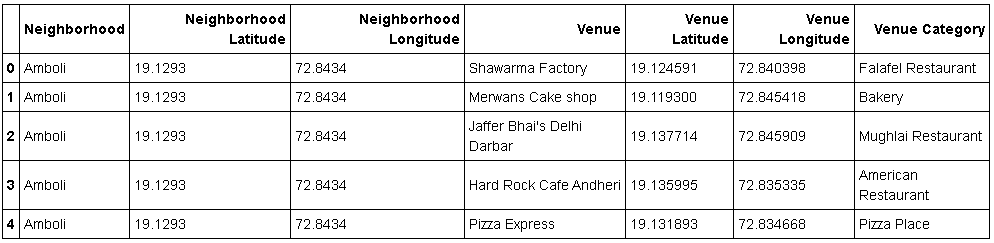
1. Visualizing Map of Mumbai City with Best Neighborhood to stay superimposed on top.



1. Visualizing Map of Mumbai City with Top Rated Venues superimposed on top.



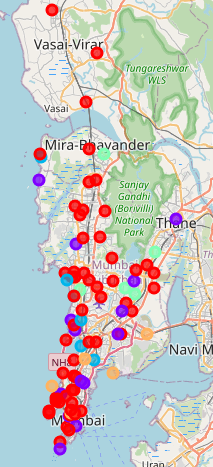
1. Using Foursquare API to download data of top venues containing of in neighborhoods and converting it to dataframe.



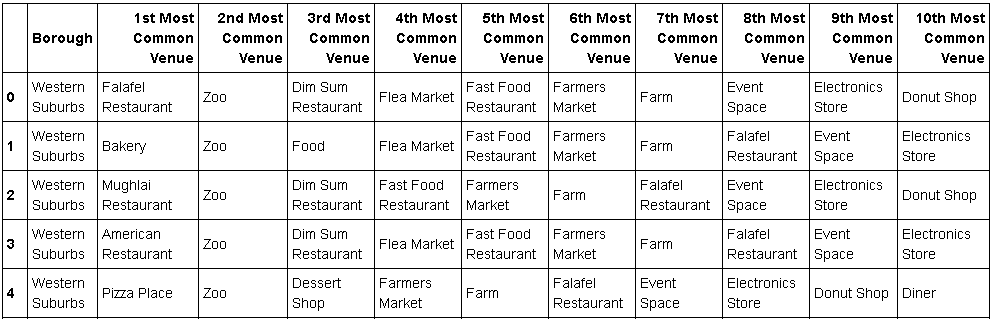
1. Finding number of unique categories and one-hot encoding according to those categories
2. Creating Dataframe consisting of Top 10 venue categories in each Neighborhood



1. Applying KMeans on top 10 venues and adding cluster labels to top 10 venues dataframe.
2. Visualizing Map of Mumbai City with Clusters superimposed on top.



1. Displaying Neighborhoods in each cluster.
2. Cluster 1:



1. Cluster 2:



1. Cluster 3:



1. Cluster 4:



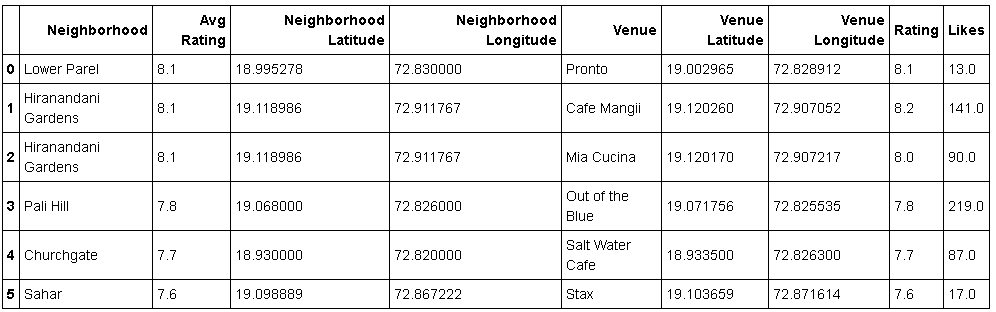
1. Cluster 5:



Result

Solutions to the problems given above in the Problems section:

1. The neighborhoods of Mumbai City that have top rated venues of the given category.
   1. Lower Parel
   2. Hiranandani Gardens
   3. Pali Hill
   4. Churchgate
   5. Sahar



1. The top-rated venues of Mumbai City of the given category are:
   1. Pronto
   2. Café Mangli
   3. Mia Cucina
   4. Out of the Blue
   5. Salt Water Café
   6. Stax
2. Best places to stay if you prefer particular type of venues:
   1. Lower Parel
   2. Hiranandani Gardens
   3. Pali Hill
   4. Churchgate
   5. Sahar
3. Neighborhoods around Mumbai can be clustered into 5 types with top visited venues:

Cluster 1: Mixed top visited venues

Cluster 2: Indian Restaurant, Zoo, Dhaba, Fast Food Restaurant, etc.

Cluster 3: Chinese Restaurant, Dhaba, Fast Food Restaurant, Farmers Market, etc.

Cluster 4: Pub, Zoo, Dessert Shop, Farmers Market, etc.

Cluster 5: Hotel, Zoo, Dhaba, Fast Food Restaurant, etc.

Conclusion

Best venues and neighborhoods for the category of venue given as Input were found and successfully visualized on maps and also using plots. KMeans Clustering model was built and used to successfully cluster neighborhoods around the Mumbai City.

There is always room for improvement and hence the solution provided above can also be improved to get best results depending upon the data.