

---

---

# Predicting optimal location for establishing a restaurant in Mumbai city

Applied Data Science Capstone Project by  
Arunit Maity

---

# Problem Statement

To find an optimal location for setting up a restaurant. Specifically, this project is targeted to stakeholders interested in setting up any food-joint/restaurant in Mumbai, Maharashtra, India. We try to detect locations that are not already crowded with restaurants and are as close to the city centre as possible.



# Data Acquisition

**Following data sources will be needed to extract/generate the required information:**

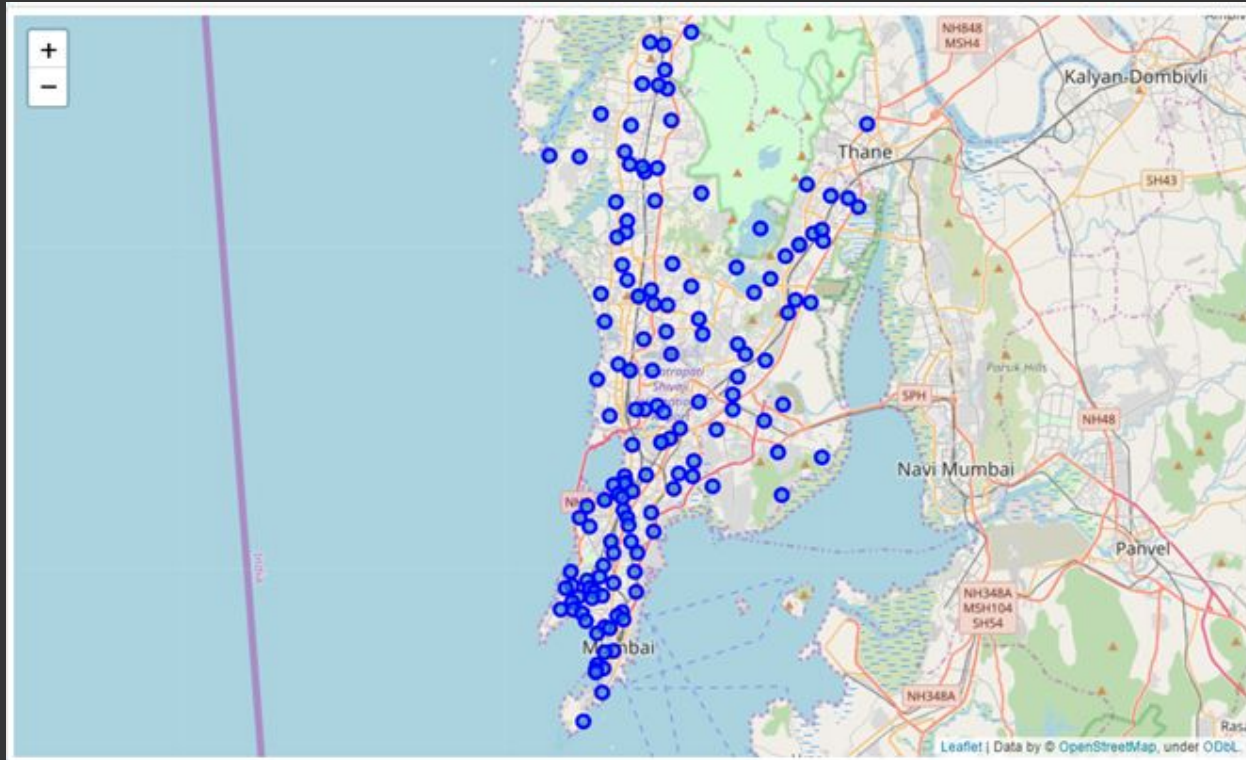
- names of all neighbourhoods will be extracted from the website '<http://pincode.india-server.com/cities/mumbai/>' using Pandas package in python
- approximate addresses of centres of these neighbourhoods will be obtained using GeoPy Geocoder package in python
- number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API

# Data Cleaning

The initial 239 neighbourhoods were filtered down to 103 neighbourhoods with a 3-stage filtering process:

- First, 239 neighbourhoods were filtered down to 172 due to lack of location data.
- Then, they were further filtered down to 145 neighbours with the criterion that only neighbourhoods within 25 Kms from the city centre are to be considered.
- Finally, these neighbourhoods were further filtered down to 103 due to lack of data for proper clustering of neighbourhoods.

# Map of Mumbai with relevant neighbourhoods

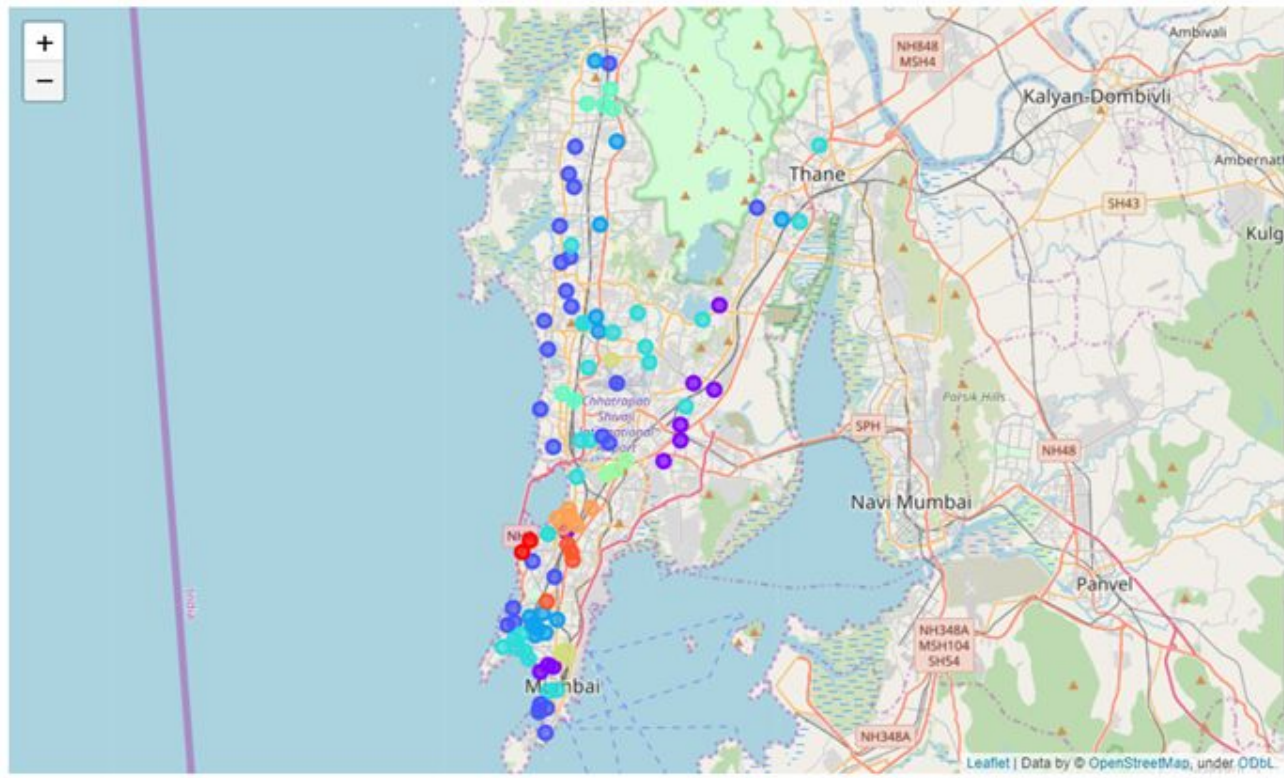


# DataFrame fed to K-Means Clustering Algorithm

	Neighborhood_Name	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	...	Train	Train Station	Travel & Transport	Vegetarian / Vegan Restaurant
0	Agripada	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	...	0.0	0.027778	0.0	0.000000
1	Andheri	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	...	0.0	0.000000	0.0	0.034483
2	Andheri East	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	...	0.0	0.000000	0.0	0.050000
3	Andheri Railway Station	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	...	0.0	0.000000	0.0	0.034483
4	Azad Nagar	0.0	0.0	0.0	0.0	0.029412	0.0	0.0	0.000000	0.000000	...	0.0	0.000000	0.0	0.029412
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
98	Tulsiwadi	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	...	0.0	0.020408	0.0	0.020408
99	V K Bhavan	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.010000	...	0.0	0.000000	0.0	0.010000
100	Worli	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.045455	0.045455	...	0.0	0.000000	0.0	0.000000
101	Worli Colony	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.017241	0.000000	...	0.0	0.000000	0.0	0.000000
102	Worli Sea Face	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.050000	...	0.0	0.000000	0.0	0.000000

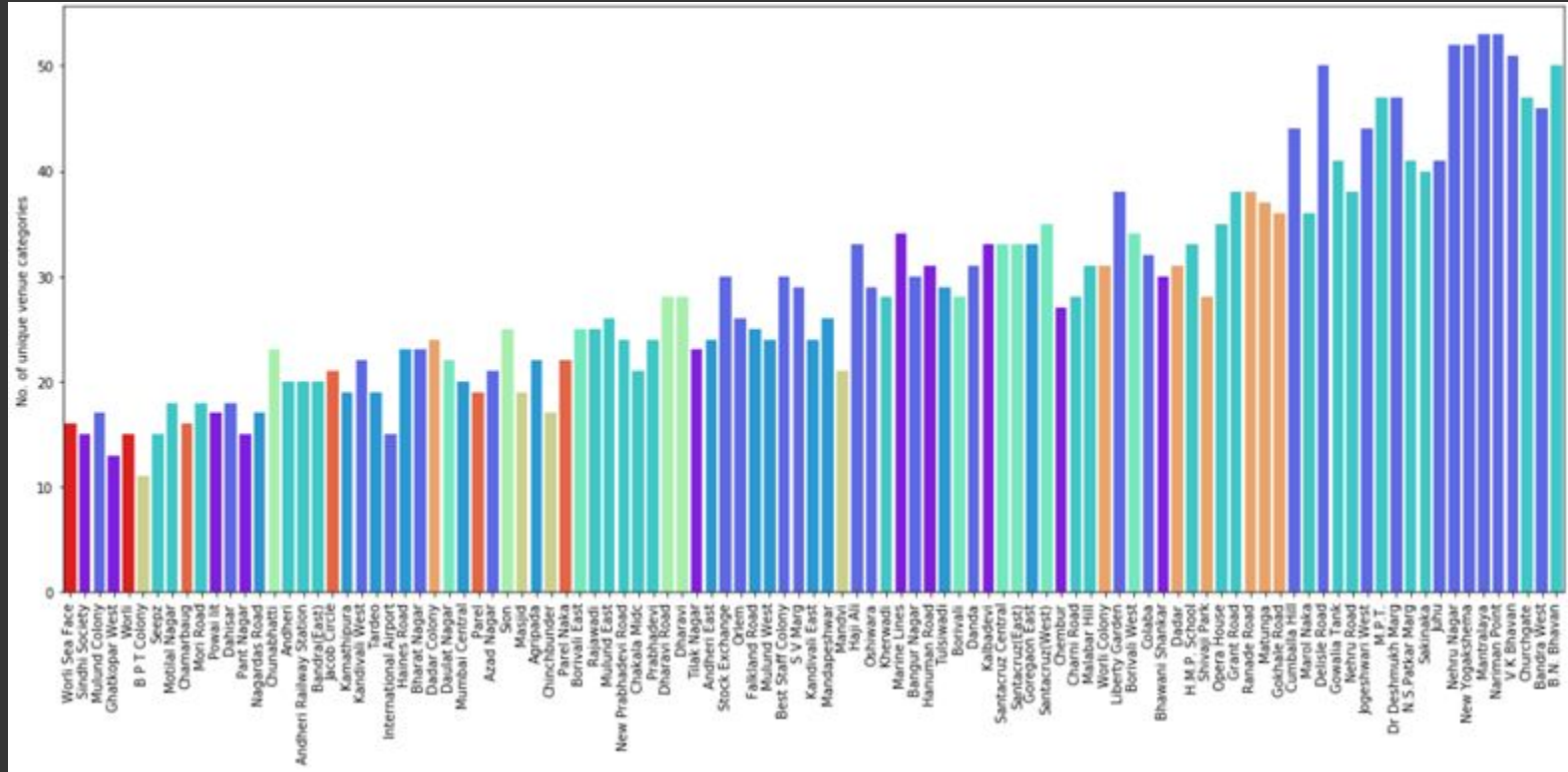
103 rows × 234 columns

# Map of Mumbai with neighbourhoods color-coded according to their clusters.



# Bar graph

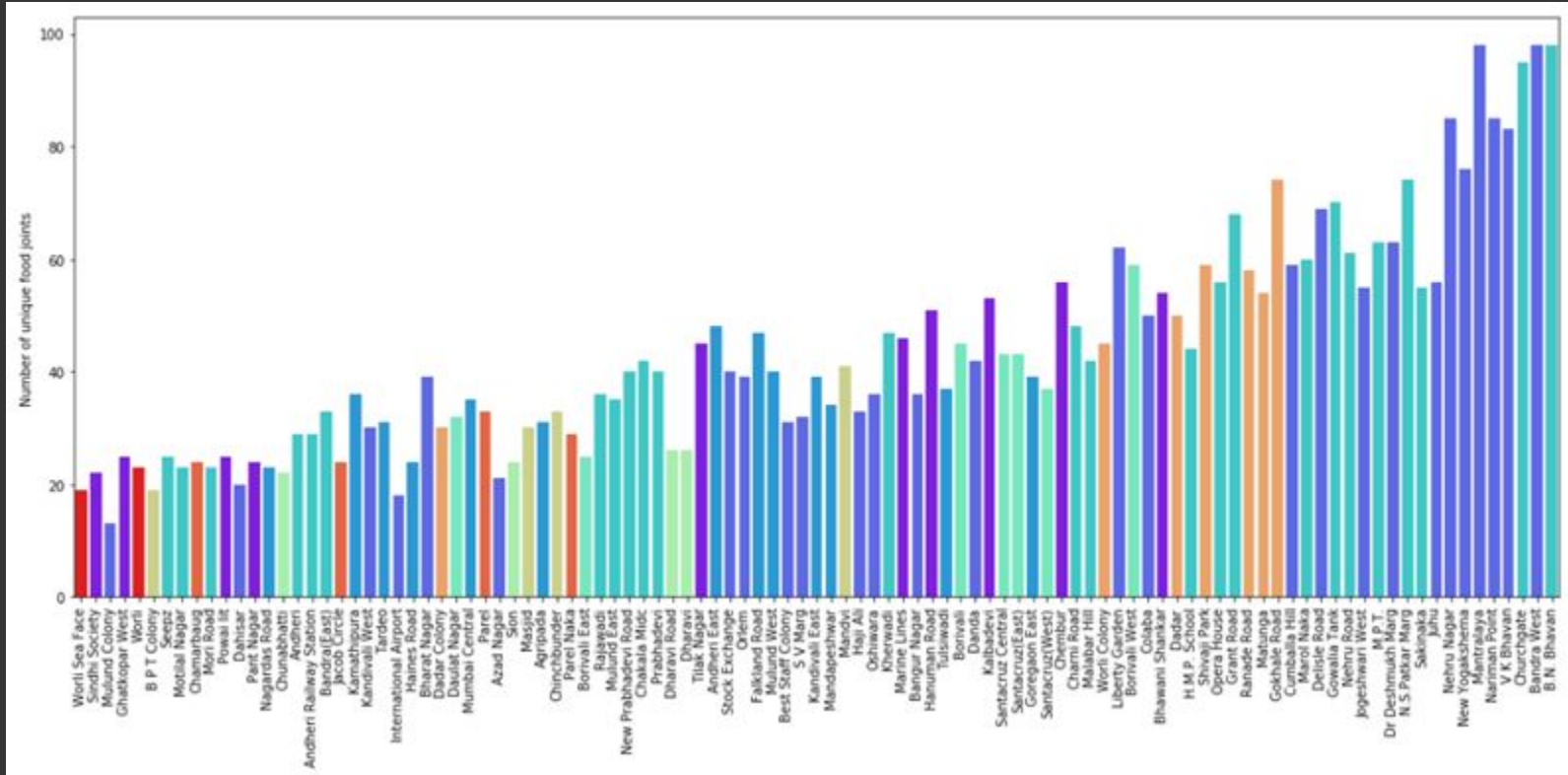
The bar graph below shows the diversity of amenities along Y axis and corresponding neighborhood names along X axis. The bars are color coded according to their cluster number.





# Bar graph

The bar graph below shows the number of food joints along Y axis and corresponding neighborhood names along X axis. The bars are color coded according to their cluster number.



# Conclusion

Optimal locations were identified as those neighbourhoods which are part of the best clusters (having neighbourhoods with most diverse amenities) yet have the least number of food joints. The optimal locations obtained are:

- Mulund Colony
- Dahisar
- Motilal Nagar
- Mori Road
- International Airport

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and their locations, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.