

Location Clustering for Restaurants in Mumbai City.



By - Chirag Shetty
(VIT University)

April, 2020

contact - chirag4798@gmail.com

Table of Contents

Title	Page Number
Abstract	3
Introduction	3
Data	4
Methodology	5
Results and Discussion	15
Conclusion	16

Abstract: Mumbai is the restaurant capital of India, but choosing a location for a restaurant and the type of restaurant is no easy task. A new approach to solve this problem has been adopted. Using various clustering algorithms can help identifying similarities in neighborhoods that can be exploited to address the target demographic. Clustered neighborhoods can be further used for expansion of the business, helpful for multinational restaurant franchises. Locations can be optimized for their distance from other competitors and also from the potential customer.

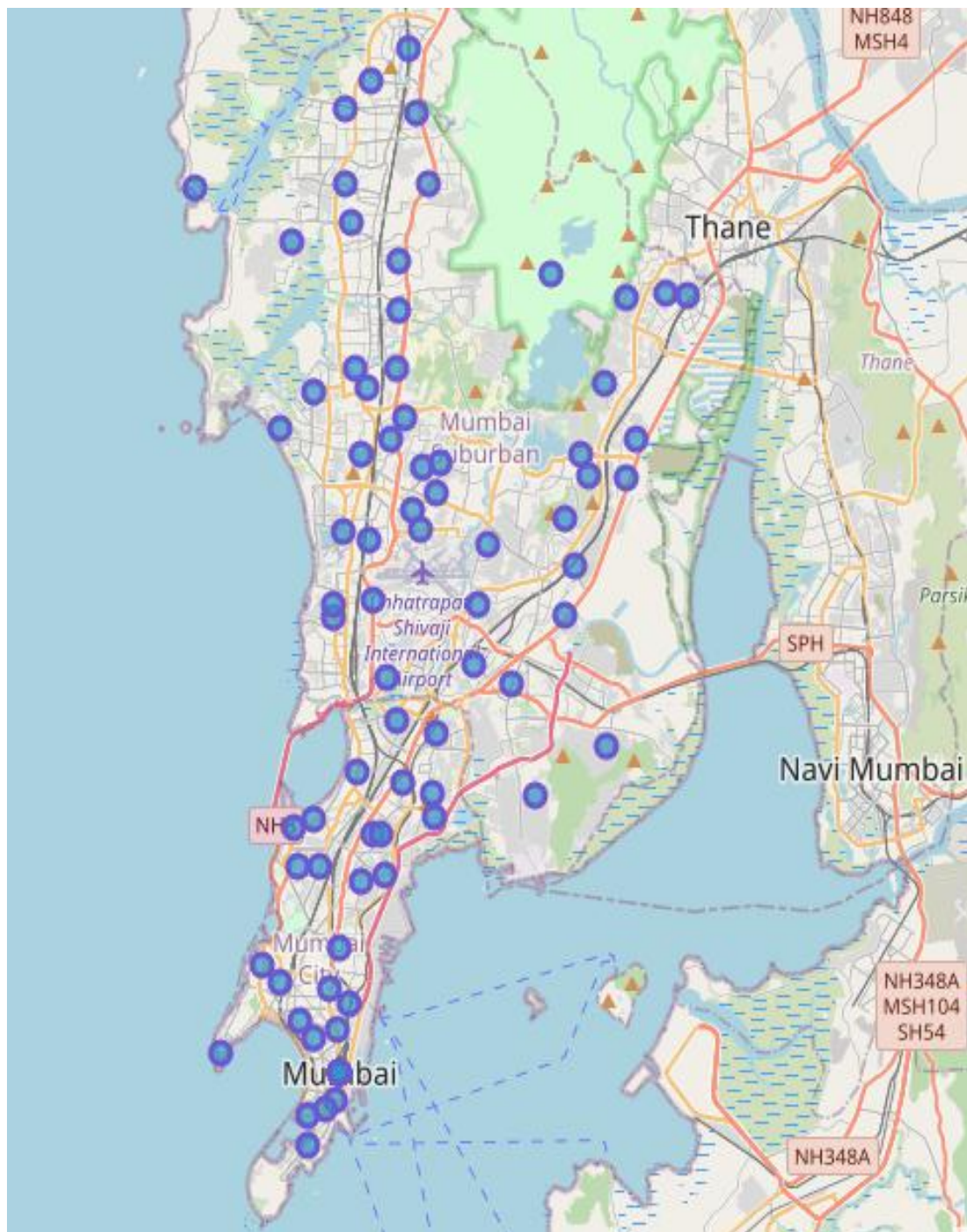
Introduction:

Mumbai is the most populated city in the world and fifth most densely populated city. It is the financial capital of India and it is no surprise that Mumbai is also the restaurant capital of India. The restaurant business has flourished with the local food and cuisine as well as multinational franchises. With such over-crowdedness comes cut throat competition. One of the main aspects of starting or expanding a restaurant business is choosing its location. Location of a restaurant decides the amount of traffic in a restaurant which makes it necessary to optimize.

A restaurant franchise owner should understand that expanding their business might be limited due to the parent company's territorial restrictions. Parent companies do not want multiple franchises competing with each other, so spatial and geographic growth might be limited for an owner. Not only is there competition from same franchises but also from already established restaurants. Such restrictions should be considered while setting up a location. Businesses thriving in a certain neighborhood can be expanded easily in similar neighborhoods in a city. Restaurants have a very high risk associated with them while starting out in a new location. If clustering of neighborhood is possible, the restaurant would have lower risk and better chances of survival. Stepping into the highly competitive restaurant industry can be both thrilling and intimidating to new franchise owners. Being able to see the benefits, rewards, and potential failures of a neighborhood, will help prospective owners decide whether or not opening a restaurant franchise is the right decision for them. In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a Cafe's in Mumbai, India. Since there are lots of restaurants in Mumbai we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Cafe's in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

Data:

Data associated with restaurants is collected using the Foursquare API. It is an excellent tool to collect venues at a given co-ordinate. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server. To use the Foursquare API, we first need a list of boroughs in Mumbai along with their pin-codes.



After which the latitude and longitude value for the given pin-codes are required. This can be achieved by using a geocoder. We can begin our analysis by getting the details of restaurant venues in 7.5 km radius of the given location, this will ensure major parts of the city are covered without worrying about overlapping locations. Data cleaning is required to get rid of the venues which are not restaurants or venues that are not related to food. Once the resulting data frame is obtained the process of data collection can be considered complete and we can proceed with exploratory data analysis. The data-set has locations of 8500 restaurants in Mumbai along with their name and the category of restaurant.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bazargate	18.938771	72.835335	Food for Thought	18.932031	72.831667	Café
1	Bazargate	18.938771	72.835335	Wankhede Stadium	18.938792	72.825944	Cricket Ground
2	Bazargate	18.938771	72.835335	Royal China	18.938715	72.832933	Chinese Restaurant
3	Bazargate	18.938771	72.835335	Starbucks	18.932190	72.833959	Coffee Shop
4	Bazargate	18.938771	72.835335	Britannia & Co.	18.934683	72.840183	Parsi Restaurant
5	Bazargate	18.938771	72.835335	Marine Drive	18.941221	72.823261	Scenic Lookout
6	Bazargate	18.938771	72.835335	Trishna	18.928619	72.832356	Seafood Restaurant
7	Bazargate	18.938771	72.835335	Natural's Ice Cream Parlour	18.934892	72.824222	Ice Cream Shop
8	Bazargate	18.938771	72.835335	Jehangir Art Gallery	18.927606	72.831464	Art Gallery
9	Bazargate	18.938771	72.835335	Taj Mahal Palace & Tower	18.922306	72.833578	Hotel
10	Bazargate	18.938771	72.835335	Nariman Point	18.929183	72.822232	Scenic Lookout

Methodology:

Based on definition of our problem, factors that will influence our decisions are:

- Finding a cluster of similar Neighborhoods.
- Number of existing restaurants in the neighborhood (any type of restaurant).
- Number of and distance to Cafe's in the neighborhood, if any.
- Accessibility of neighborhood.

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using reverse geocoding

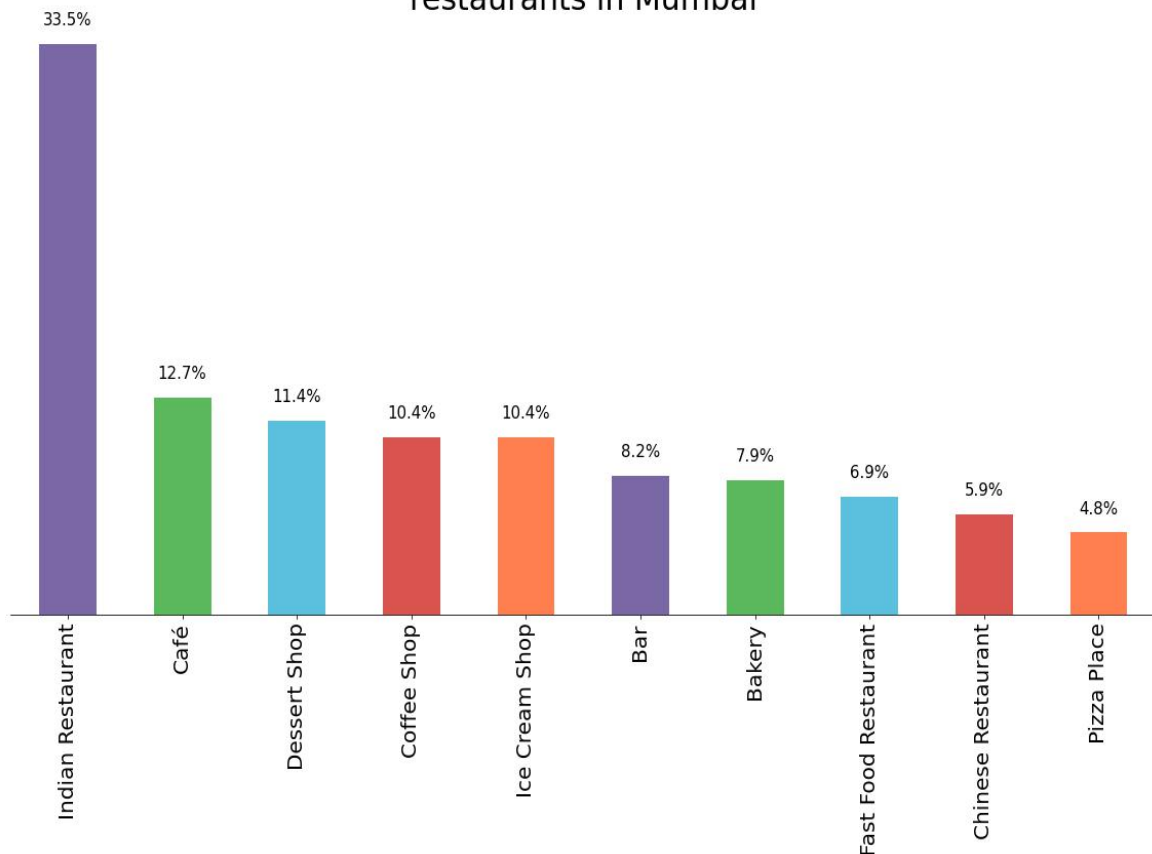
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- Coordinates will be obtained using geocoding of well known Borough pin-codes.

Feature Extraction: Relevant features have to be extracted from the data, the data-set contains noisy features like sports stadiums, scenic lookouts and monuments. Features with key words restaurant in their category have to be extracted. Extracted variables have to be converted to categorical variables, where each category will form an attribute with binary values.

	Neighborhood	Indian Restaurant	Chinese Restaurant	Bar	Parsi Restaurant	Coffee Shop	Café	Bakery	Seafood Restaurant	Ice Cream Shop	Fast Food Restaurant
0	Bazargate	0	0	0	0	0	1	0	0	0	0
1	Bazargate	0	0	0	0	0	0	0	0	0	0
2	Bazargate	0	1	0	0	0	0	0	0	0	0
3	Bazargate	0	0	0	0	1	0	0	0	0	0
4	Bazargate	0	0	0	1	0	0	0	0	0	0

Exploratory Data Analysis: Exploring the data gives us valuable insights before modelling the data for machine learning. Examining the market share of various restaurants shows that Indian restaurants are the most common restaurant type in Mumbai with about 33% of the market share. People in India are very culturally rooted and prefer having food with family in familiar setting which is the reason for the popularity of Indian restaurants. But the younger generation have more inclination towards Cafe's and Coffee places. These places have trendy ambience and younger people like to hangout there after work and college with their friends. The popularity is in an upward trend with Cafe's and Coffee houses, already having a hefty market share. The market may have been saturated with Indian restaurants but there is still room left for Cafe's and Coffee places for expansion.

Market share of
restaurants in Mumbai



Looking at common venues in each neighborhood also concludes that Indian restaurants are the most common restaurants in Mumbai, but other restaurants are gaining popularity steadily. The closest to Indian restaurants in popularity are the Café's and Coffee places. When combined they cover nearly 23% of the market share. Dessert shops in India are generally sweet/confectionery shops and not restaurant per-say, but are quite popular too.

----Ambewadi----			----Bangur Nagar----		
	venue	freq		venue	freq
0	Indian Restaurant	7.0	0	Indian Restaurant	9.0
1	Hotel	6.0	1	Café	7.0
2	Dessert Shop	5.0	2	Hotel	6.0
3	Café	3.0	3	Bar	5.0
4	Ice Cream Shop	3.0	4	Coffee Shop	5.0
5	Fast Food Restaurant	3.0	5	Seafood Restaurant	4.0
6	Pizza Place	3.0	6	Ice Cream Shop	4.0
7	Diner	2.0	7	Bakery	3.0
8	Pub	2.0	8	Donut Shop	3.0
9	Lounge	2.0	9	Brewery	3.0

----Andheri----			----Barve Nagar----		
	venue	freq		venue	freq
0	Indian Restaurant	9.0	0	Indian Restaurant	21.0
1	Café	7.0	1	Hotel	7.0
2	Bar	5.0	2	Coffee Shop	6.0
3	Hotel	5.0	3	Italian Restaurant	6.0
4	Seafood Restaurant	4.0	4	Restaurant	6.0
5	Ice Cream Shop	4.0	5	Lounge	5.0
6	Coffee Shop	4.0	6	Café	3.0
7	Bakery	3.0	7	Ice Cream Shop	3.0
8	Donut Shop	3.0	8	Dessert Shop	3.0
9	Brewery	3.0	9	Asian Restaurant	3.0

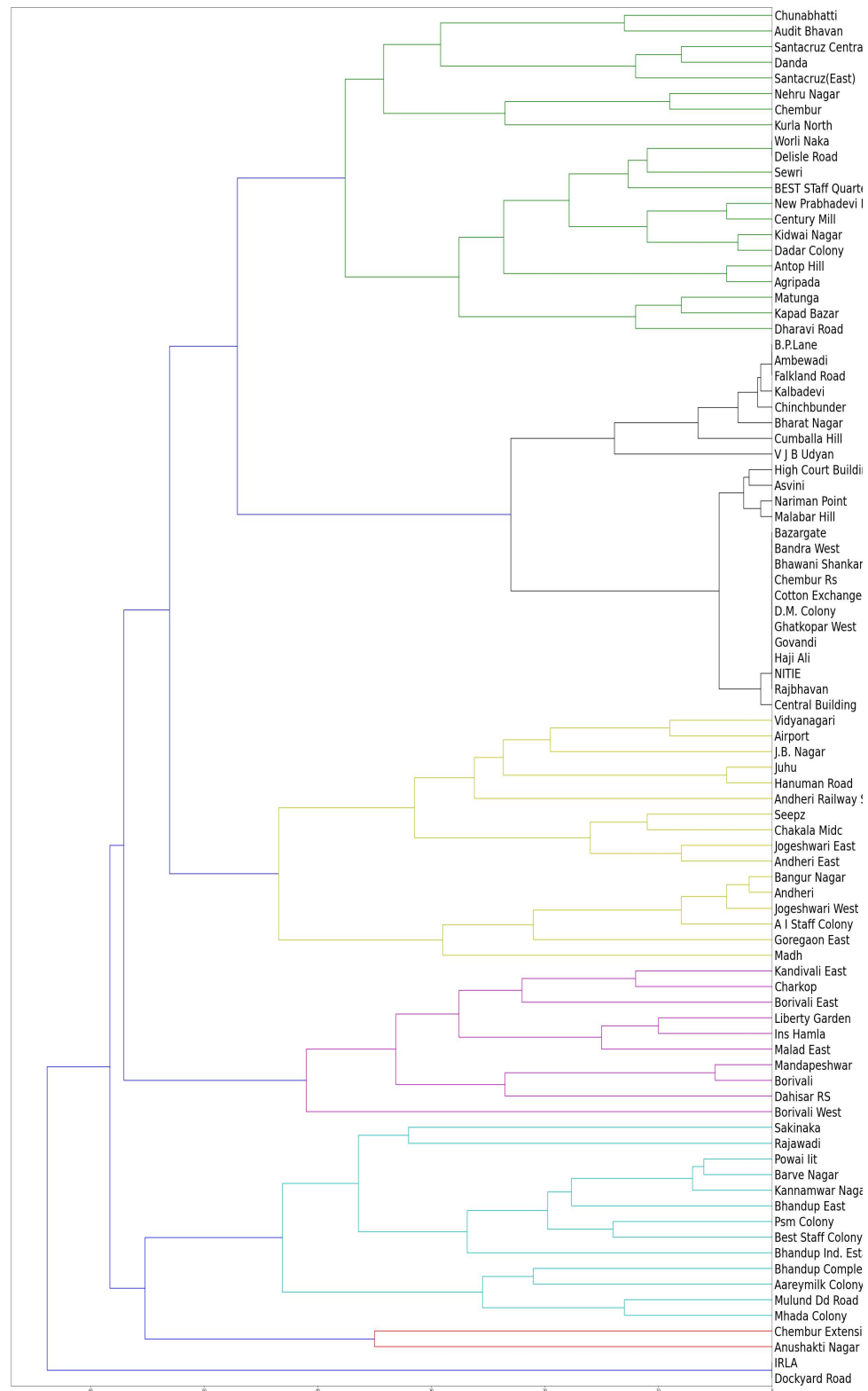
Modelling: For a clustering problem the attributes with binary data are prepared. The attributes contain information about the type of restaurant in each neighborhood and how common are those restaurant. Agglomerative clustering algorithm is used to get a rough idea about the number of clusters that can be extracted from the data without losing useful information. A distance matrix is created for each point with a randomly assigned cluster centre. An agglomerative cluster can be visualized using a dendrogram. The x-axis on a dendrogram represents the distance between individual clusters. The y-axis is a list of boroughs with coloured lines indicating the cluster which it belongs. The names on the y-axis are arranged in such a manner that similar neighbors are arranged closer to each other. The affinity parameter is set to '11' and the linkage is set to 'average'. Average uses the average of the distances of each observation of the two sets. Once we have a rough idea about the number of clusters we can tune our Kmeans model.

	Indian Restaurant	Chinese Restaurant	Bar	Parsi Restaurant	Coffee Shop	Café	Bakery	Seafood Restaurant	Ice Cream Shop	Fast Food Restaurant	Dessert Shop	Asian Restaurant	Breakfast Spot
Neighborhood													
A I Staff Colony	11	2	5	0	5	6	3	4	4	0	2	0	2
Aareymilk Colony	12	1	2	0	6	4	3	1	3	2	3	1	0
Agripada	12	2	4	0	2	4	4	4	3	1	4	1	0
Airport	9	1	3	0	5	5	6	4	3	1	5	1	0
Ambewadi	7	2	2	1	2	3	1	2	3	3	5	1	0

Agglomerative Hierarchical clustering Technique: In this technique, initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed.

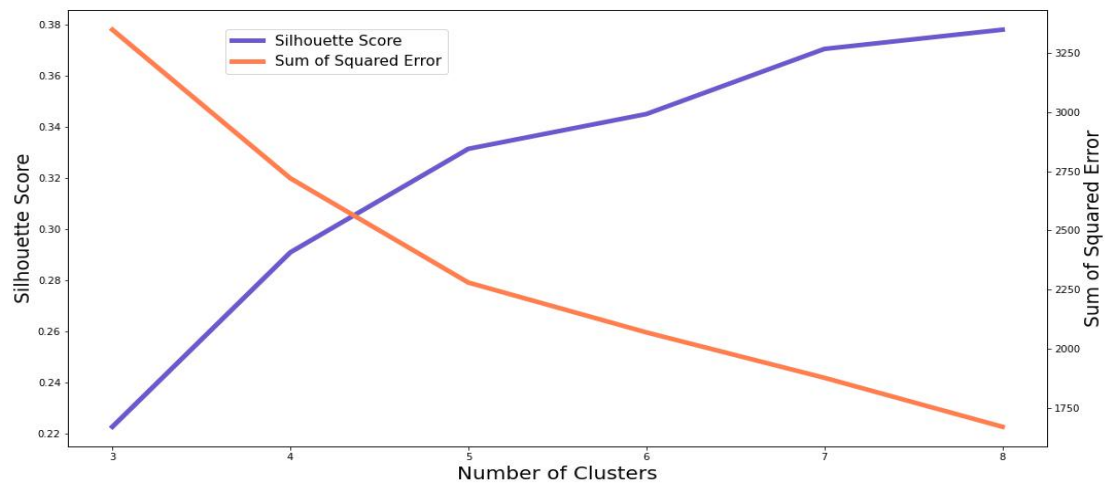
The basic algorithm of Agglomerative is straight forward.

- Compute the proximity matrix
- Let each data point be a cluster
- Repeat: Merge the two closest clusters and update the proximity matrix
- Until only a single cluster remains

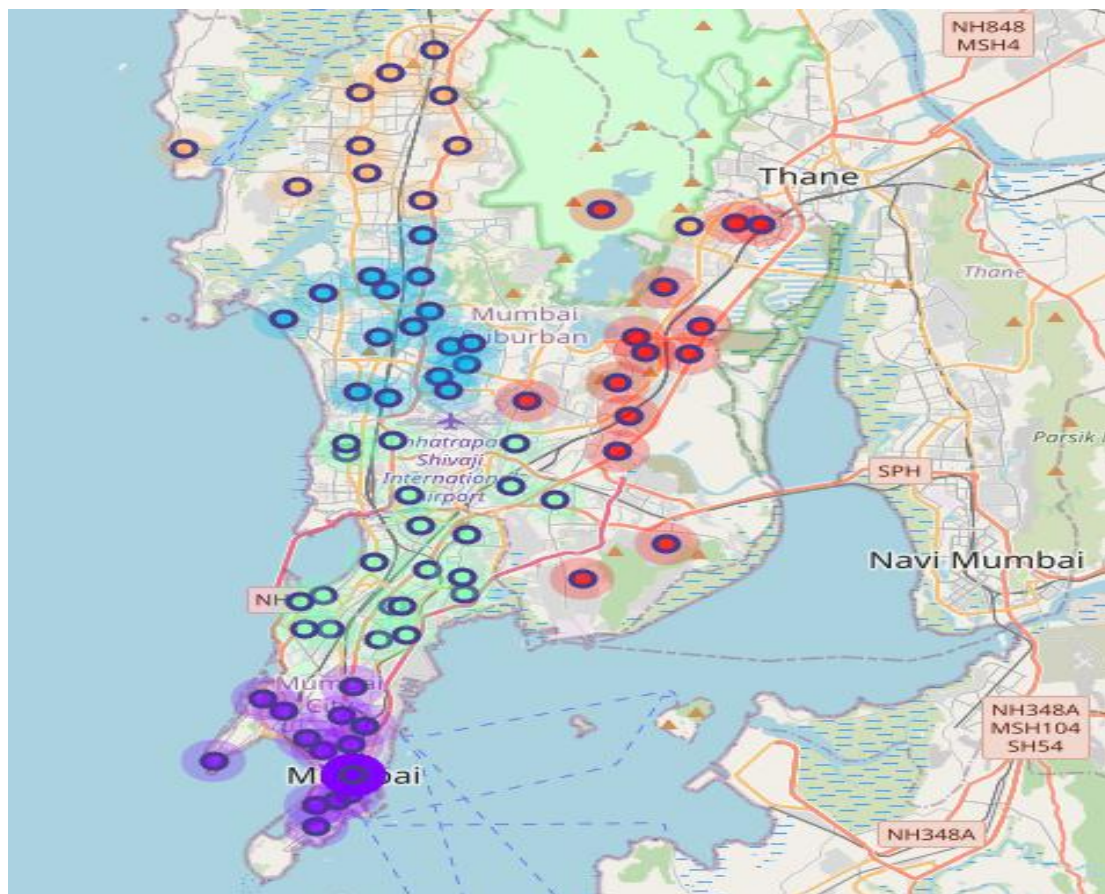


A highly co-related neighborhood with virtually zero distance between them in the distance matrix can be seen, the locations being Bazargate, Bandra, Bhawani Shankar Road, Chembur, Cotton Exchange, DM Colony, Ghatkopar, Govandi, Haji Ali, NITIE and Raj Bhavan areas. The number of clusters for Kmeans algorithm can also be optimized using the elbow method and calculating the sum of squared error among clusters

and the silhouette scores of the model for different values of k . A minor elbow can be observed at $k=5$ in an otherwise smooth curve.



For $k=5$ the neighborhoods are clustered in this manner. Even though the model wasn't trained on any location based data, it's not a surprise that closer neighborhoods are more similar to each other, this indicates a good model.



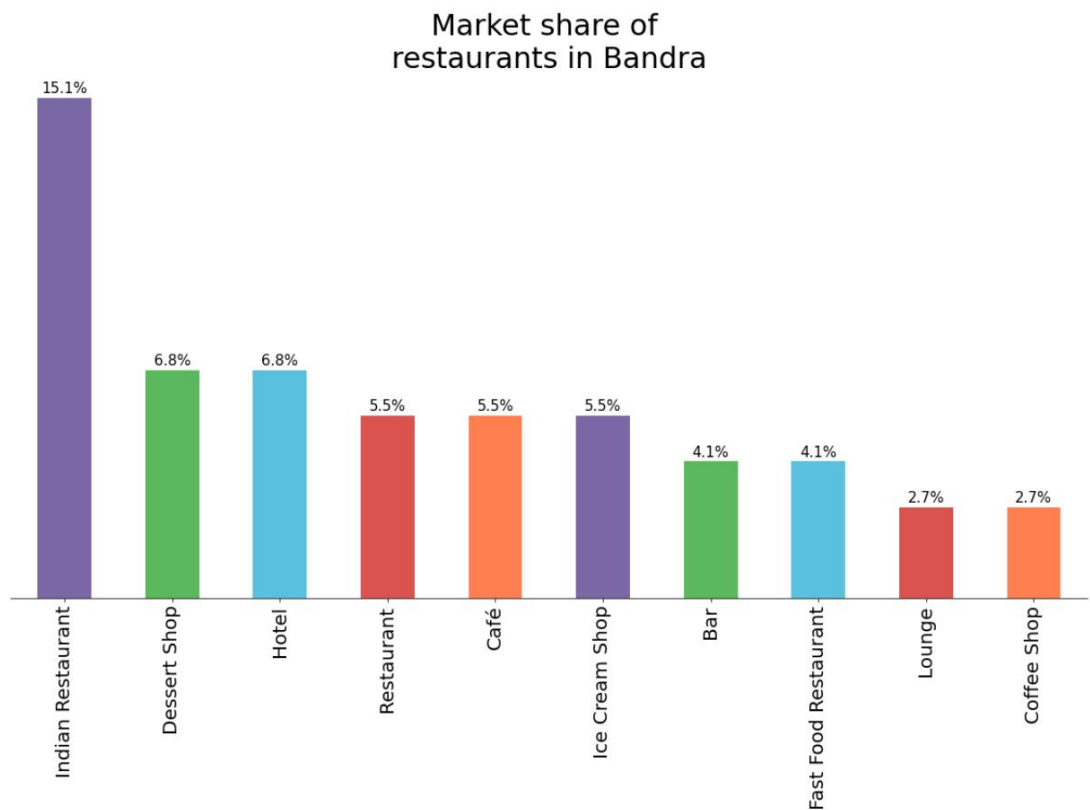
This group of neighborhoods have a lot in common as both Agglomerative and Kmeans clustering clustered them into the same group. Once we start out at any one of these locations it would be easier

to expand in other locations. Here the overall trend is followed, that Indian restaurants are popular, other than that Dessert shops are quite popular too. Dessert shops in India tend to be Sweet shops and confectioneries, they are not a restaurant per-say but are quite popular. Lets focus on the Coffee shops and Cafe's here. The area is not saturated with it but can be seen as quite popular. Lets narrow our search on the Bandra region of the cluster. Bandra is located at the heart of the city with lot of foot traffic around it. The Bandra Kurla Complex in Bandra is considered the financial hub of the city.

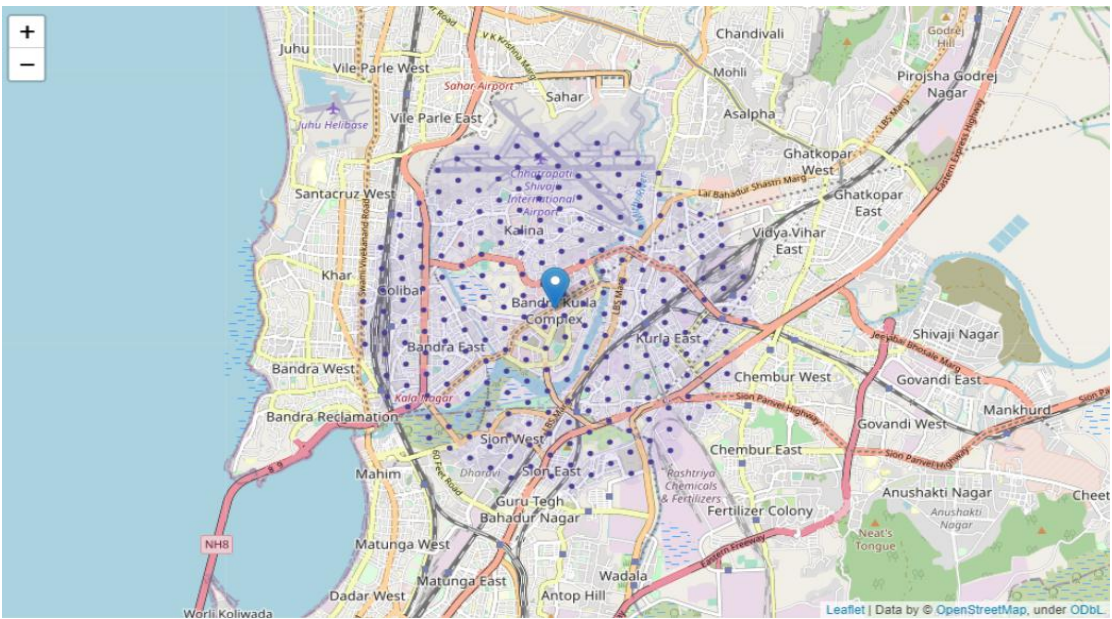
	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bazargate	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
1	Kalbadevi	0	Indian Restaurant	Hotel	Dessert Shop	Café	Pizza Place
2	B.P.Lane	0	Indian Restaurant	Hotel	Dessert Shop	Café	Pizza Place
3	Ambewadi	0	Indian Restaurant	Hotel	Dessert Shop	Café	Pizza Place
4	Asvini	0	Indian Restaurant	Dessert Shop	Café	Hotel	Ice Cream Shop
5	Malabar Hill	0	Indian Restaurant	Ice Cream Shop	Dessert Shop	Café	Hotel
6	Bharat Nagar	0	Indian Restaurant	Hotel	Dessert Shop	Café	Fast Food Restaurant
7	Falkland Road	0	Indian Restaurant	Hotel	Dessert Shop	Café	Pizza Place
8	Chinchbunder	0	Indian Restaurant	Hotel	Dessert Shop	Café	Pizza Place
9	Dockyard Road	0	Indian Restaurant	Pizza Place	Café	Fast Food Restaurant	Ice Cream Shop
19	Central Building	0	Indian Restaurant	Hotel	Dessert Shop	Café	Ice Cream Shop
20	Nariman Point	0	Indian Restaurant	Dessert Shop	Café	Hotel	Ice Cream Shop
24	Cumballa Hill	0	Indian Restaurant	Hotel	Dessert Shop	Café	Fast Food Restaurant
25	V J B Udyan	0	Indian Restaurant	Hotel	Dessert Shop	Restaurant	Café
26	Bhawani Shankar Rd	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
30	High Court Building	0	Indian Restaurant	Hotel	Dessert Shop	Café	Ice Cream Shop
31	Cotton Exchange	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
32	Haji Ali	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
33	Rajbhavan	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
36	D.M. Colony	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
38	Bandra West	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
44	IRLA	0	Indian Restaurant	Pizza Place	Café	Fast Food Restaurant	Ice Cream Shop
71	Ghatkopar West	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
72	NITIE	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
73	Govandi	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant
74	Chembur Rs	0	Indian Restaurant	Hotel	Dessert Shop	Café	Restaurant

Bandra Kurla Complex is a business and residential district in Bandra, Mumbai. It is a prominent commercial hub in India. According to MMRDA, the complex is the first of a series of "growth centers" created to "arrest further concentration" of offices and commercial activities in South Mumbai. It has aided to decongest the CBD in South Mumbai while seeding new areas of planned commercial real estate in the metropolitan region. There are approximately 400,000 people working in various offices throughout the BKC, which makes it ideal to set up a cafe. Considering a radius of around 5kms around the BKC, equidistant grid points are created about 100m apart from each other. These grid points act as potential candidates for setting up the cafe. The market share graph of Bandra indicates that Cafe's are not as common in this region as they

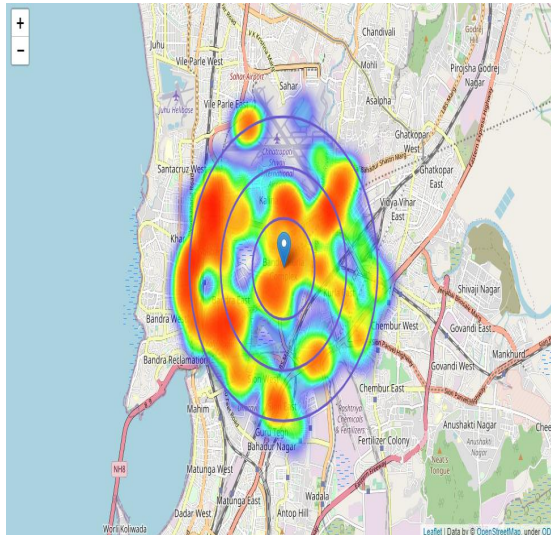
are in the rest of the city. This is a good indication that there is opportunity to expand in this region.



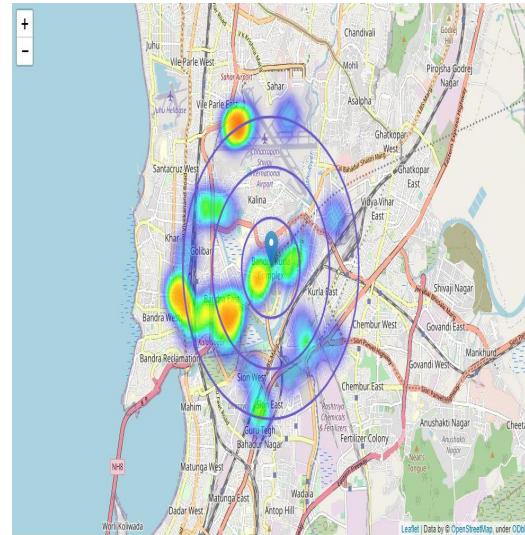
The equidistant grid-points can be visualized using a map. The points stretch from Dharavi in the south to the Airport in the north. The locations abundantly cover the centre of the city. This part of the city is one of the most accessible.



The density of restaurants in a region can be visualized using a heat-map. It can be seen that very less area is left untouched by a restaurant, but the same map for cafe tells a completely different story. It shows that only about 15% of the restaurants are cafes. Further analysis revealed that a Cafe is at a distance of approximately 1km from each grid-point whereas a restaurant is at an average distance of 400m.

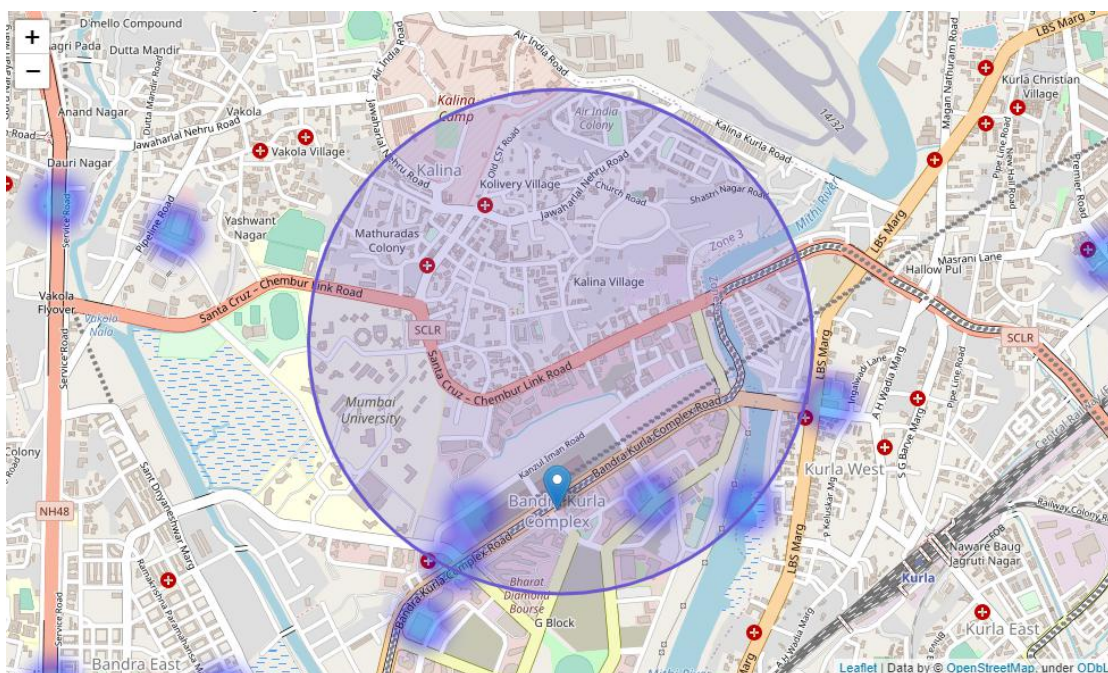


Density of Restaurants

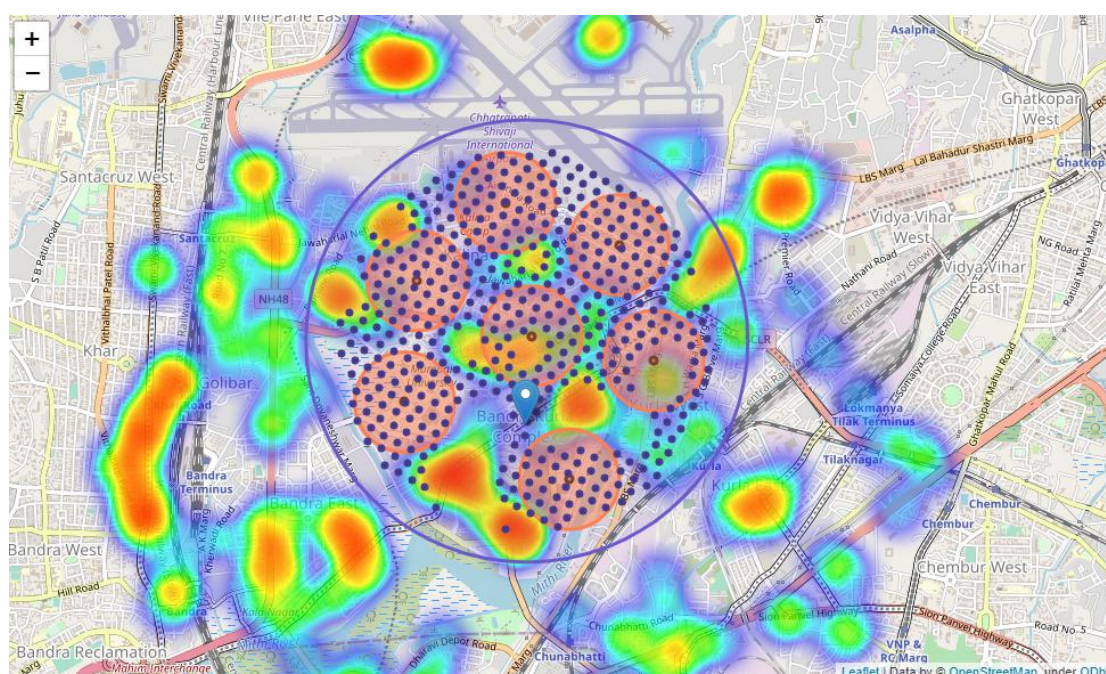


Density of Cafes

Lets further narrow our region to the place just north of BKC, where the region looks fairly empty of Cafes. This region is in the vicinity of Mumbai University. It gives an added advantage to the Cafe business due to the presence of large population of the target demographic.



The region looks devoid of competitors. The procedure of making grid-points is repeated with slight variation. The grid-points now only cover regions where there are no Cafes in a 400m radius and no more than 2 restaurants in its vicinity. This step will ensure that the business will have little to no competition from its surroundings. The points of interest can be further clustered into regions using the k-means algorithm. The centers of the clustered locations can be considered the final suggestions for setting up a Cafe. We have created 6 addresses representing centers of zones containing locations with low number of restaurants and no Cafe's nearby, all zones being fairly close to Bandra Kurla Complex. Although zones are shown on map with a circle, their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.



Address	Distance from BKC
Mumbai University, Sant Dnyaneshwar Marg, RPF Colony, H/E Ward, Zone 3, Mumbai 40051, India	1.6km
Kolivery Village, H/E Ward, Zone 3, Mumbai 400098, India	2.7km
H/E Ward, Zone 3, Mumbai 400 051, India	0.9km
LBS Marg, Hallow Pul, L Ward, Zone 5, Mumbai 400070, India	1.9km
Kalina Village, H/E Ward, Zone 3, Mumbai 4000	1.2km

98, India	
Green field apartments, Parshiwadi Road, Mathura das Colony, H/E Ward, Zone 3, Mumbai 400098, India	2.4km

Results and Discussion:

The analysis shows that although there is a great number of restaurants in Mumbai (>8500 in our initial area of interest which was 7.5x7.5km around each pin-code), there are pockets of low restaurant density fairly close to Bandra Kurla Complex. Highest concentration of restaurants was detected west, south from Bandra Kurla Complex, so we focused our attention to areas north, near the Mumbai University, where a number of pockets of low restaurant density was found. After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Bandra Kurla Complex) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Cafe's closer than 400m were removed. Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 6 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Cafe's particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Bandra Kurla Complex but not crowded with existing restaurants (particularly Cafe) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion:

Purpose of this project was to identify clusters similar neighborhoods in Mumbai with low number of restaurants (particularly Cafe's) in order to aid stakeholders in narrowing down the search for optimal location for a new Cafe. Clustering Neighborhoods in Mumbai and selecting highly co-related neighborhoods which can be used in future for expansion, by calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Bandra), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, 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 neighborhood etc.