CLUSTERING LOACTIONS FOR RESTAURANTS

IN MUMBAI

BY - CHIRAG SHETTY

CONTENTS

1

INTRODUCTION

INTRODUCTION

BACKGROUND

Mumbai is the most populated city in the world and fifth most densely populated city. Its is the financial capital of India and its 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

PROBLEM

- 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.

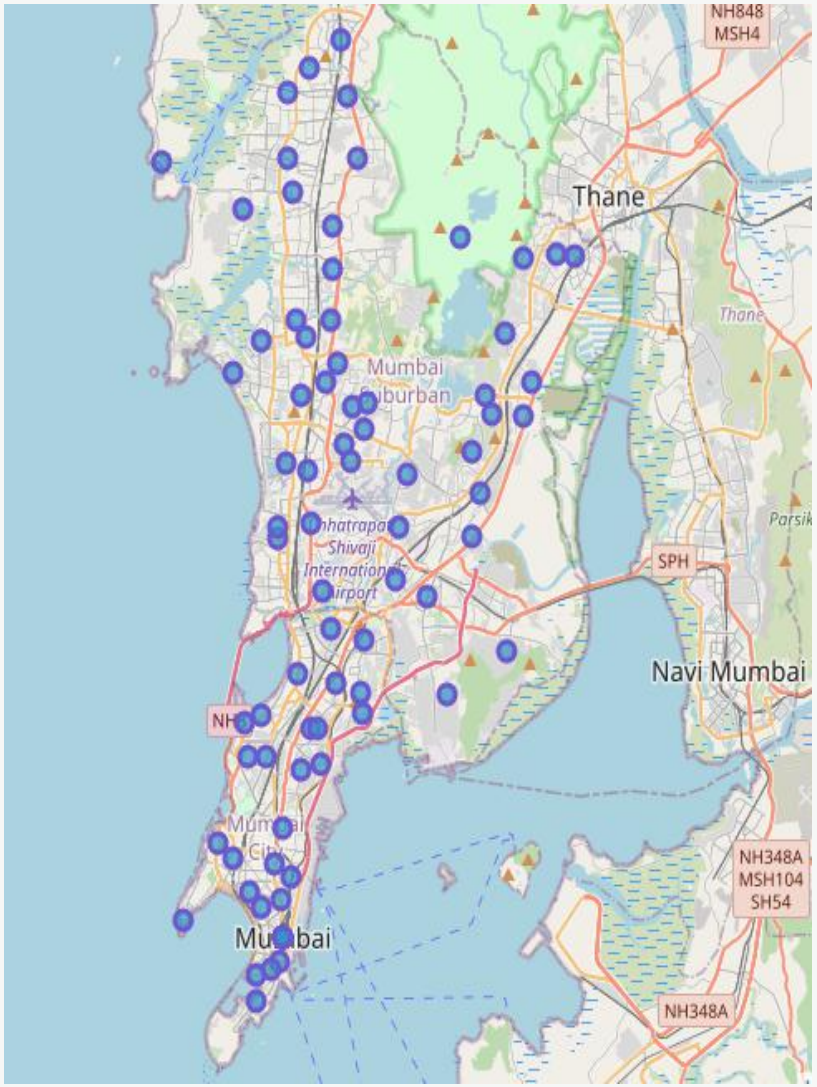
- In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Cafe in Mumbai, India.

- 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.

2

DATA

DATA



FOURSQUARE API

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.

The data-set has locations of 8500 restaurants in Mumbai along with their name and the category of restaurant.

DATA

	Borough	Pincode	Latitude	Longitude
0	Bazargate	400001	18.938771	72.835335
1	Kalbadevi	400002	18.949439	72.826324
2	B.P.Lane	400003	18.951606	72.834797
3	Ambewadi	400004	18.954329	72.821730
4	Asvini	400005	18.916863	72.824315
5	Malabar Hill	400006	18.944790	72.793776
6	Bharat Nagar	400007	18.965888	72.814797
7	Falkland Road	400008	18.963839	72.831471
8	Chinchbunder	400009	18.959645	72.838526
9	Dockyard Road	400010	19.053122	73.099739
10	Agripada	400011	19.016016	72.867872

	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

GIVEN THIS DATA-FRAME

FINAL DATA-FRAME

3

METHODOLOGY

METHODOLOGY

Finding a cluster of similar Neighborhoods.

Finding number of existing Restaurants & Cafe's in neighborhood

Assesing the accessibility and benefits of a location over others

Neighborhoods will be clustered based on the similarity in their most common restaurant types

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

METHODOLOGY

1

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.

2

EXPLORATORY 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.

3

MODELLING

Clustering algorithms are applied to the data. Hierarchical clustering algorithm like Agglomerative clustering is applied along with K-means for better results

EXPLORATORY ANALYSIS

----Bangur Nagar----

	venue	freq
0	Indian Restaurant	9.0
1	Café	7.0
2	Hotel	6.0
3	Bar	5.0
4	Coffee Shop	5.0
5	Seafood Restaurant	4.0
6	Ice Cream Shop	4.0
7	Bakery	3.0
8	Donut Shop	3.0
9	Brewery	3.0

----Barve Nagar----

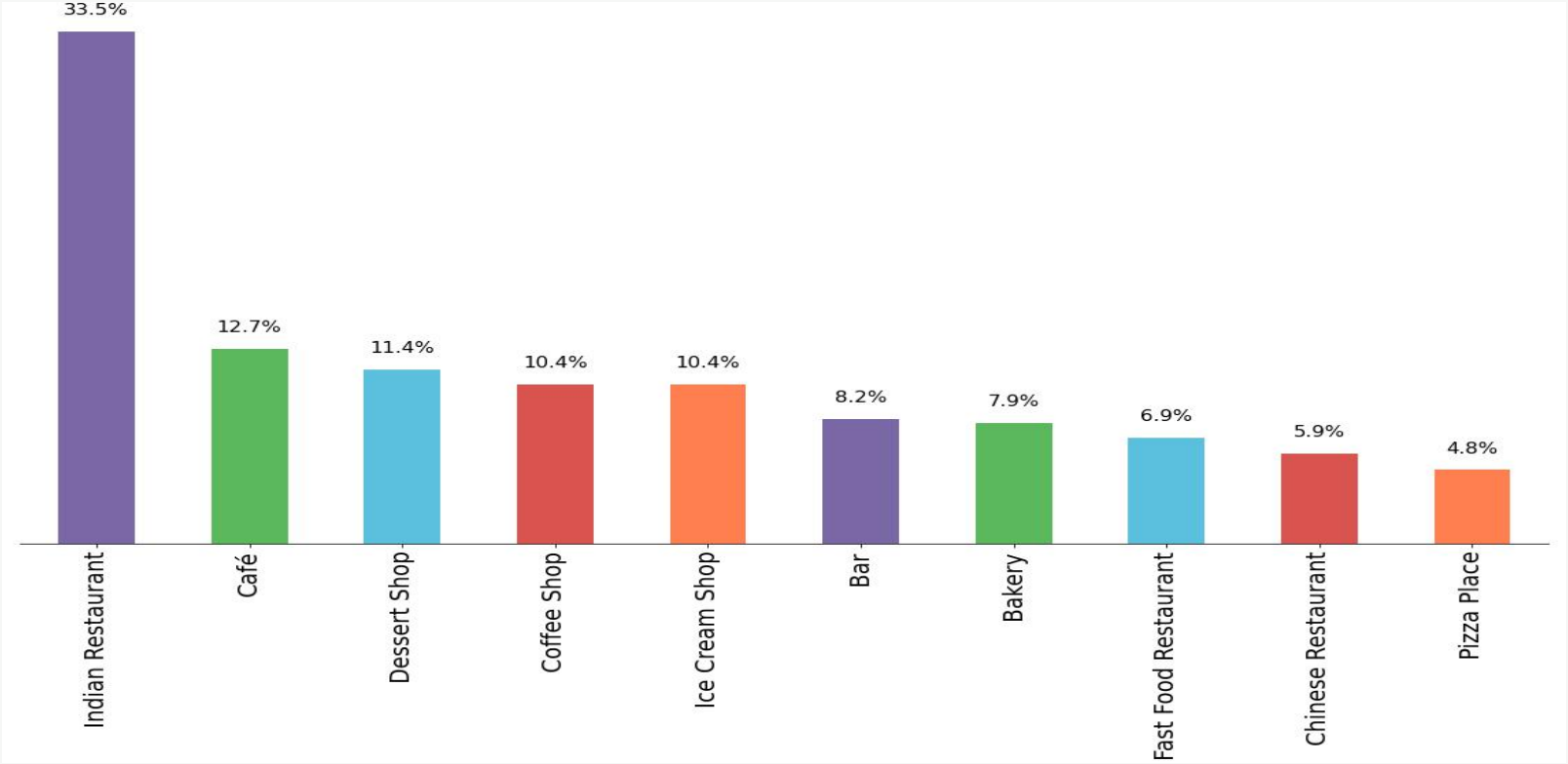
	venue	freq
0	Indian Restaurant	21.0
1	Hotel	7.0
2	Coffee Shop	6.0
3	Italian Restaurant	6.0
4	Restaurant	6.0
5	Lounge	5.0
6	Café	3.0
7	Ice Cream Shop	3.0
8	Dessert Shop	3.0
9	Asian Restaurant	3.0

----Ambewadi----

	venue	freq
0	Indian Restaurant	7.0
1	Hotel	6.0
2	Dessert Shop	5.0
3	Café	3.0
4	Ice Cream Shop	3.0
5	Fast Food Restaurant	3.0
6	Pizza Place	3.0
7	Diner	2.0
8	Pub	2.0
9	Lounge	2.0

----Andheri----

	venue	freq
0	Indian Restaurant	9.0
1	Café	7.0
2	Bar	5.0
3	Hotel	5.0
4	Seafood Restaurant	4.0
5	Ice Cream Shop	4.0
6	Coffee Shop	4.0
7	Bakery	3.0
8	Donut Shop	3.0
9	Brewery	3.0



Number of restaurants in Mumbai in Percentage%

MODELLING

K-MEANS CLUSTERING

The basic algorithm of K-means clustering.

- starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids

- It halts creating and optimizing clusters when either:

- The centroids have stabilized — there is no change in their values because the clustering has been successful.The defined number of iterations has been achieved.

AGGLOMERATIVE CLUSTERING

The basic algorithm of Agglomerative clustering.

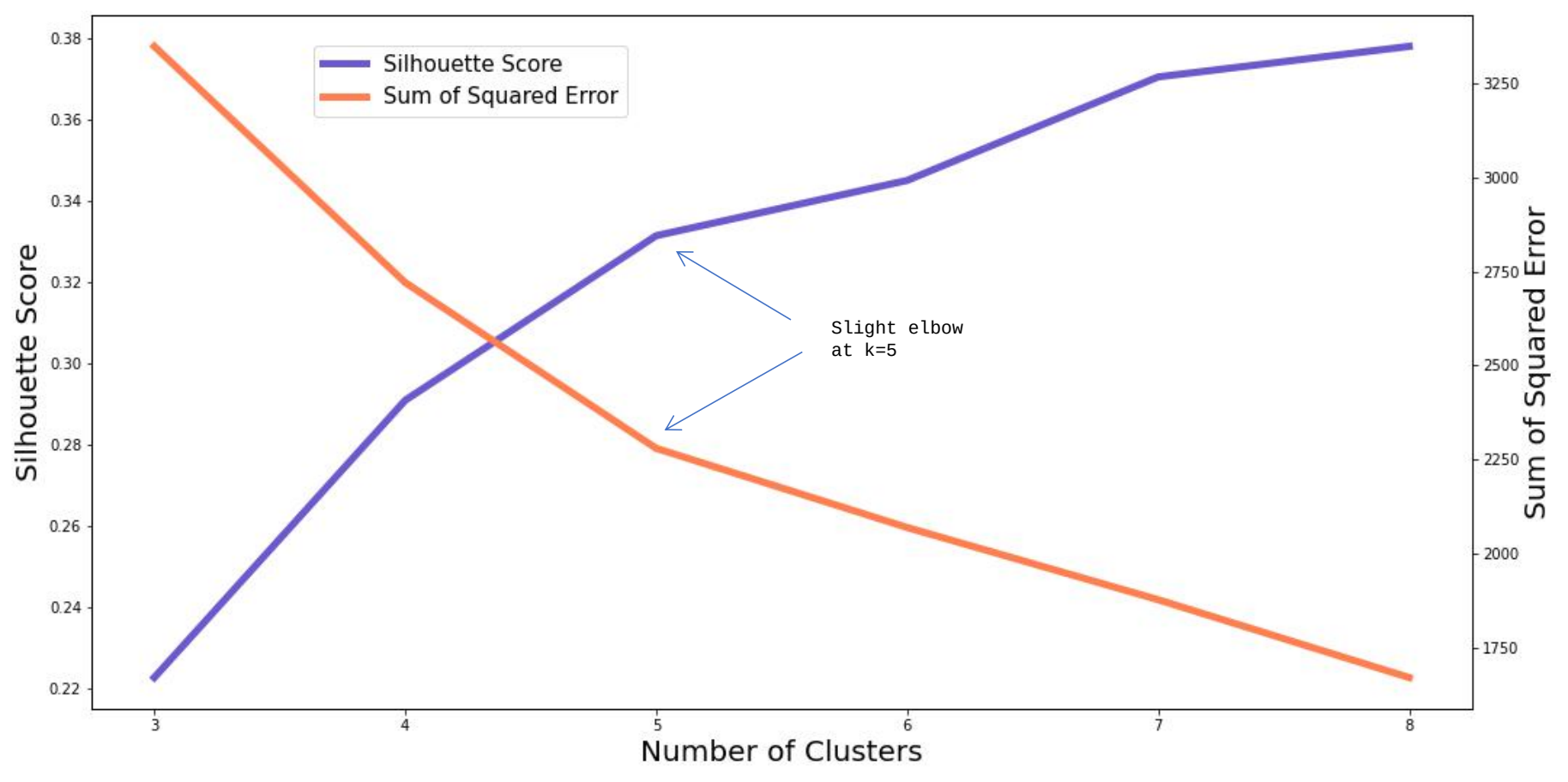
- Compute the proximity matrix

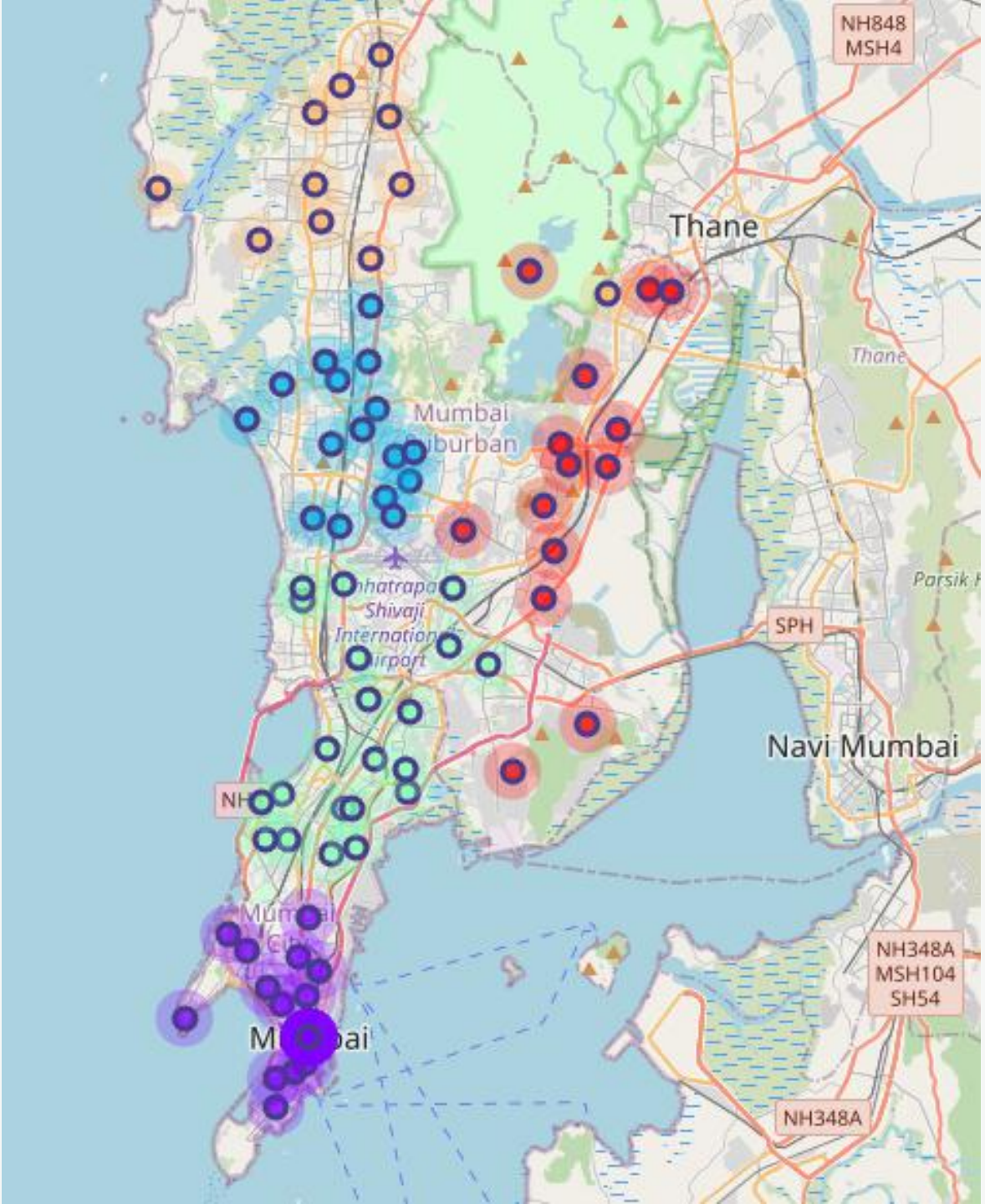
- Let each data point be a cluster

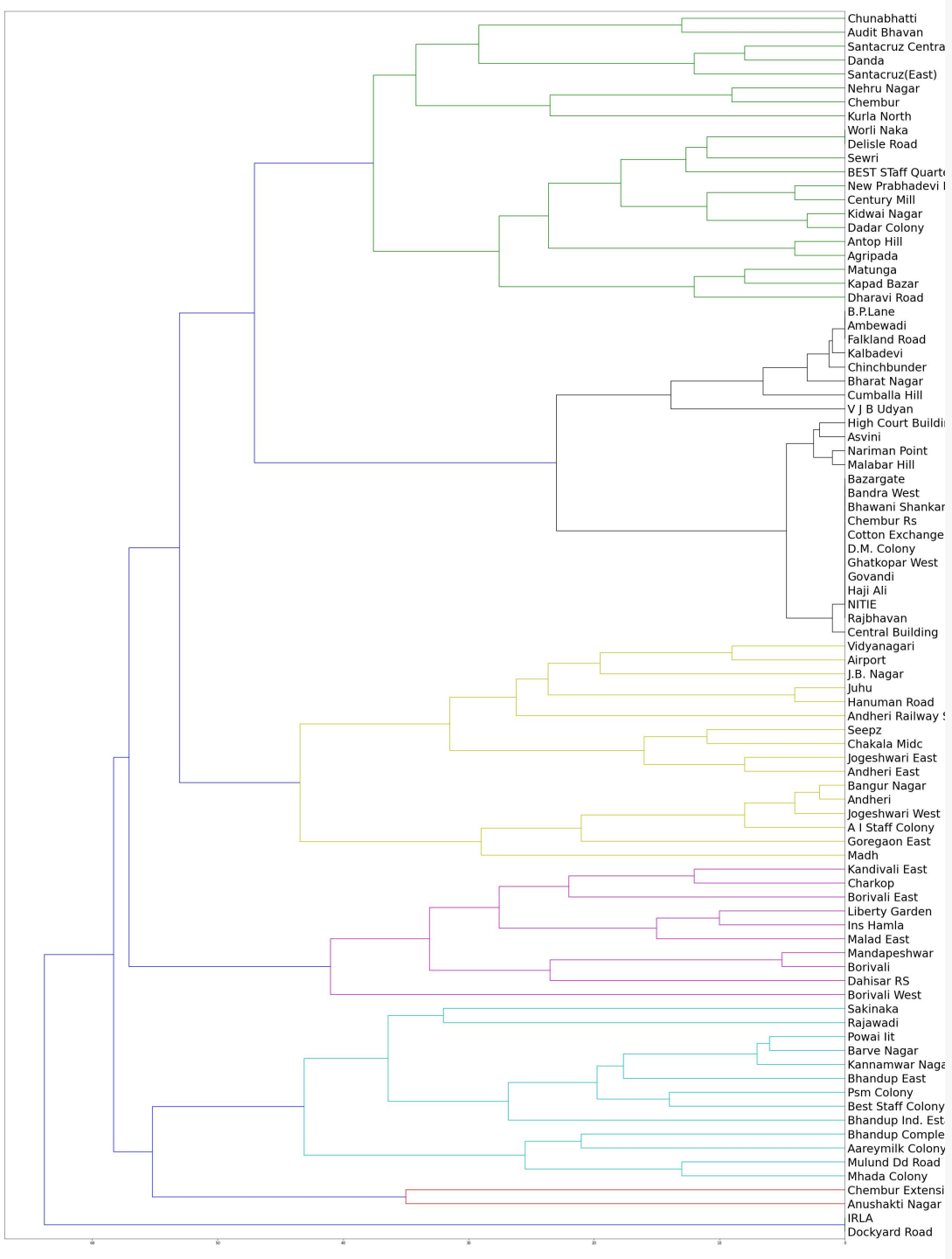
- Repeat: Merge the two closest clusters and update the proximity matrix

- Until only a single cluster remains

Finding the number of cluster for K-means using the elbow method




NH848
MSH4
Thane
Mumbai
Suburban
Navi Mumbai
SPH
NH348A
MSH104
SH54
NH348A


Chunabhatti
Audit Bhavan
Santacruz Centra
Danda
Santacruz(East)
Nehru Nagar
Chembur
Kurla North
Worli Naka
Delisle Road
Sewri
BEST STaff Quarte
New Prabhadevi
Century Mill
Kidwai Nagar
Dadar Colony
Antop Hill
Agripada
Matunga
Kapad Bazar
Dharavi Road
B.P.Lane
Ambewadi
Falkland Road
Kalbadevi
Chinchbunder
Bharat Nagar
Cumballa Hill
V J B Udyan
High Court Buildi
Asvini
Nariman Point
Malabar Hill
Bazargate
Bandra West
Bhawani Shankar
Chembur Rs
Cotton Exchange
D.M. Colony
Ghatkopar West
Govandi
Haji Ali
NITIE
Rajbhavan
Central Building
Vidyanagari
Airport
J.B. Nagar
Juhu
Hanuman Road
Andheri Railway
Seepz
Chakala Midc
Jogeshwari East
Andheri East
Bangur Nagar
Andheri
Jogeshwari West
A I Staff Colony
Goregaon East
Madh
Kandivali East
Charkop
Borivali East
Liberty Garden
Ins Hamla
Malad East
Mandapeshwar
Borivali
Dahisar RS
Borivali West
Sakinaka
Rajawadi
Powai Iit
Barve Nagar
Kannamwar Naga
Bhandup East
Psm Colony
Best Staff Colony
Bhandup Ind. Est
Bhandup Comple
Aareymilk Colony
Mulund Dd Road
Mhada Colony
Chembur Extensi
Anushakti Nagar
IRLA
Dockyard Road

These Locations seem to be highly co-related since they have zero distances in the Distance Matrix

Bazargate
Bandra West
Bhawani Shankar
Chembur Rs
Cotton Exchange
D.M. Colony
Ghatkopar West
Govandi
Haji Ali
NITIE
Rajbhavan

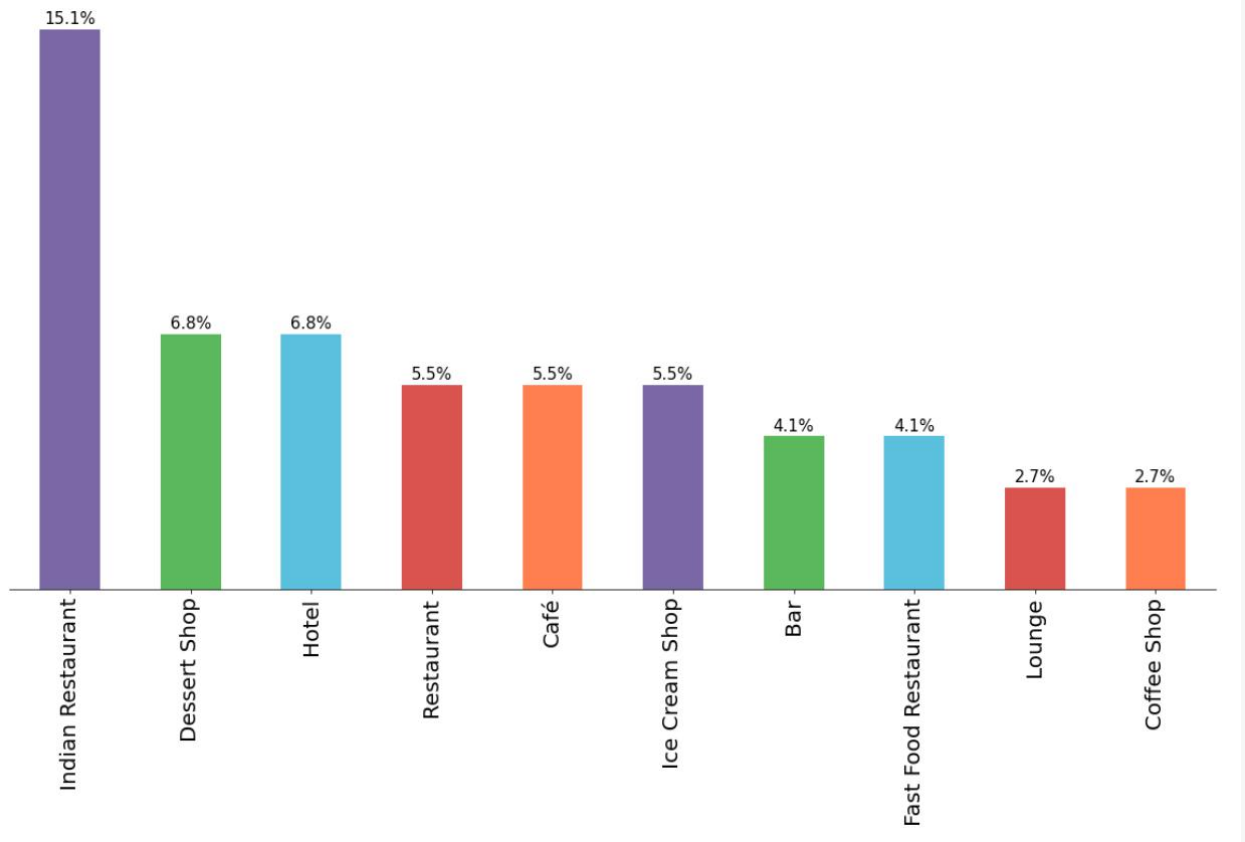
We will be narrowing our search to this loaction, due to its proximity to the centre of the city, and it being well connected to the rest of the city

We can expand our business in any of the other loactions in future by giving it a closer look

Why Bandra Kurla Complex?

- *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.*
- *There are approximately 400,000 people working in various offices throughout the BKC, which makes it ideal to set up a restaurant.*

EXPLORATORY ANALYSIS OF BANDRA



Why Cafe?

- *Cafe, perfectly complements our chosen location.*

- *The main customers of a cafe in ascending order are :*

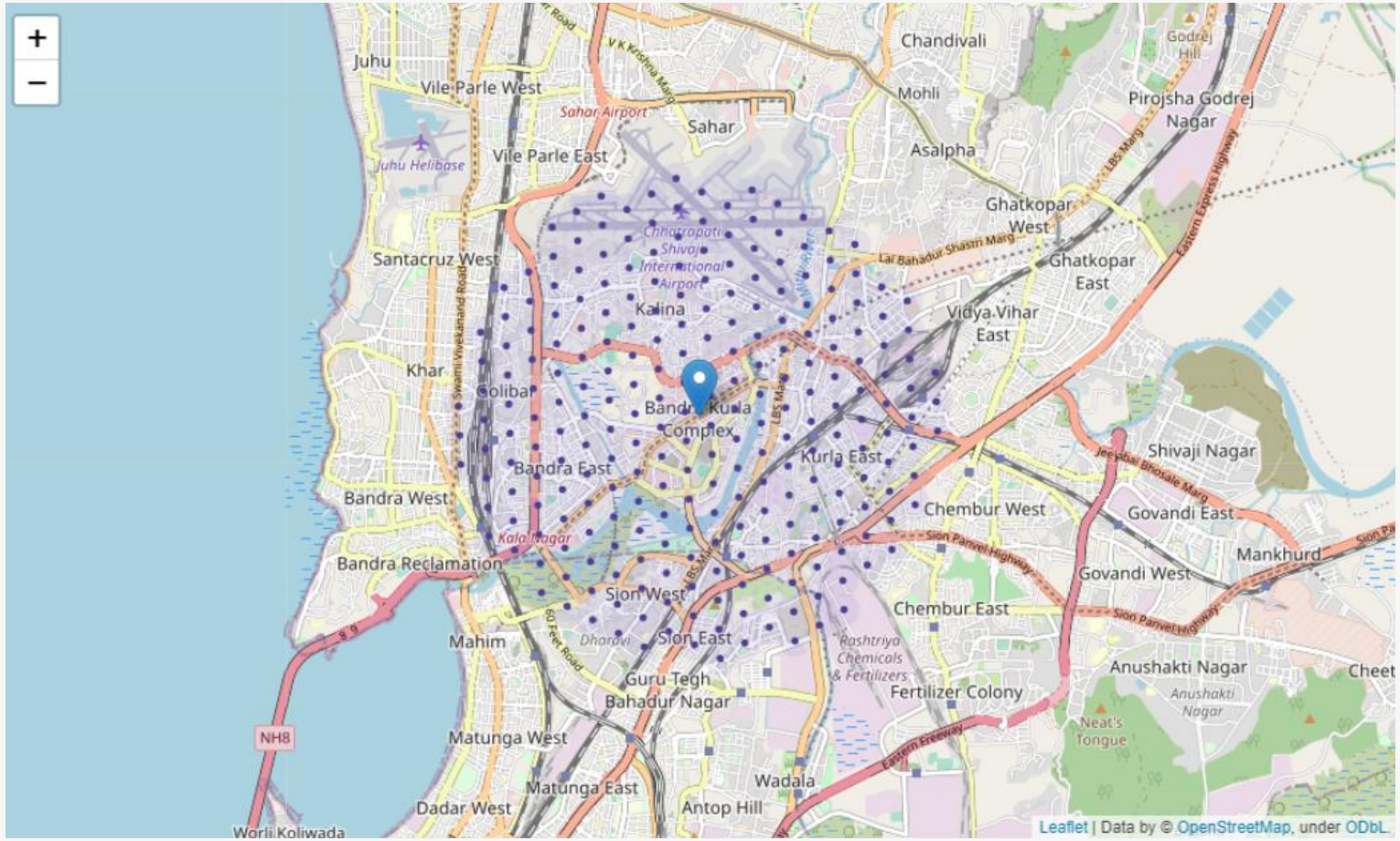
    1. *Employees or Working class(Proximity to BKC)*

    2. *Couples (married & unmarried both)*

    3. *youngsters (17–30)(Proximity to Mumbai University)*

- *Cafe's aren't as common as Indian restaurants, hence less risk of over saturation and less competition.*

Narrowing down to Bandra Kurla Complex from previous clustering



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 due various railway lines and highways near it.

Applying the K-means Algorithm to get centres of cluster of good locations from the grid points.

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 400098, India	1.2km
Green field apartments, Parshiwadi Road, Mathuradas Colony, H/E Ward, Zone 3, Mumbai 400098, India	2.4km

4

RESULTS & CONCLUSIONS

RESULTS

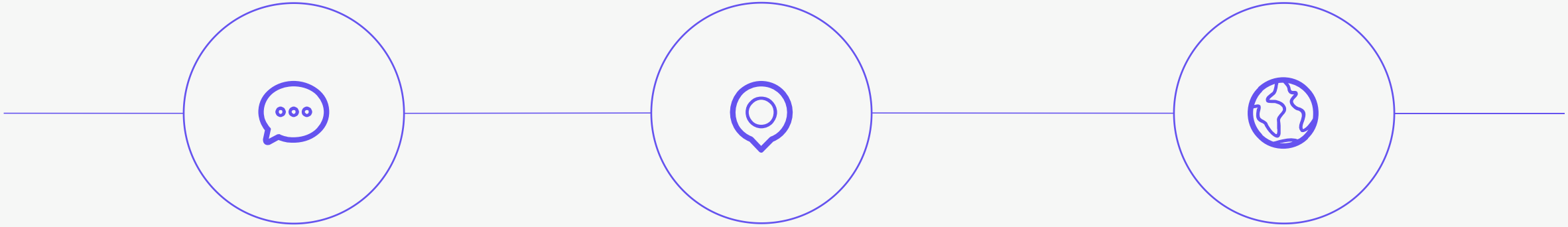
.The analysis shows that although there is a great number of restaurants in Mumbai (>8500 in our initial area of 7.5x7.5km around each pin-code), there are pockets of low restaurant density fairly close to Bandra Kurla Complex

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

We were able to cluster neighborhoods based on how common a type of restaurant is in that locality. The result of which we chose Bandra as our area odf interest.

We 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.

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

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.

THANK YOU