Capstone Project: The battle of neighbourhoods (week 1)

TABLE OF CONTENTS

INTRODUCTION: BUSINESS PROBLEM

Problem description: Mumbai city situated in Maharashtra, India is a populous city comprising of nearly 21 million people closely huddled within a meagre 603.4 km² area. The aim of this project is to find a reasonably densely populated area to open up a new supermarket. This would mean that Mumbai City as a whole will be segmented and clustered to meet the desired goal. By leveraging the foursquare location data, each area will be analysed in this process.

Background discussion: Mumbai, also called Bombay, is the capital city of the state of Maharashtra in India, and it's the most populous city in India. As the 4th most populous city in the world and one of the populous urban regions in the world, Mumbai has a metro population of about 20,185,064 in 2019. Mumbai's demographics relate to us that the city is considered a melting pot due to all of the migrants that relocate to the city for employment opportunities. With workforce readily available the major issue to opening up a supermarket is the tiresome process of acquiring large lands; which is a tedious job given the population explosion. Assuming that stakeholders involved have influence, power and funds to readily acquire a piece of land of their choice, this project will help facilitate them to find an optimum location for setting up the supermarket. The key influencing factors would be-

1. How populated the area is? A reasonably populated area would mean more client base.
2. Is there another supermarket in the location of interest? One would definitely want less competition or none at all.
3. Is the area of interest in plush locality? This will target customers with higher income for a greater profit margin and revenue generation.

Keeping the above key points in mind, the exploratory analysis of Mumbai City will be undertaken to affix a suitable location for the said supermarket.

DATA

Data description and how it will be used to solve the problem:

1. **Statistical information on Mumbai's population and its density.**
   http://worldpopulationreview.com/world-cities/mumbai-population/
2. **A map rendering detailed classification of population density in Mumbai with key highlights:**
   Census data reveals that population density varies noticeably from area to area. Small area census

-Tahoora Qureshi

data do a better job depicting where the crowded neighborhoods are. In this map, areas of highest density exceed 30,000 persons per square kilometer. Very high density areas exceed 7,000 persons per square kilometer.   High density areas exceed 5,200 persons per square kilometer. The last categories break at 3,330 persons per square kilometer, and 1,500 persons per square kilometer.
https://www.arcgis.com/home/webmap/viewer.html?webmap=d39c316b61364e86918fe566aaccf54e

3. **A complete list of boroughs and neighbourhoods in Mumbai city:** This data will be used to construct dataset comprising of boroughs, neighbourhoods, latitude and longitude co-ordinates of corresponding neighbourhoods.
https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

4. **Latitude and longitude co-ordinates of various neighbourhoods in Mumbai city:** This data will be used to construct dataset comprising of boroughs, neighbourhoods, latitude and longitude co-ordinates of corresponding neighbourhoods.
https://www.latlong.net/place/mumbai-maharashtra-india-27236.html

5. **Data pertaining to venues of specific location:** The foursquare location data will be leveraged for all our location data requirements.
https://foursquare.com/

6. **Data on real estate pricing for each neighbourhood of Mumbai City:** This data will be used to identify and categorize customers on the basis of their income. The real estate pricing of neighbourhoods will be a key indicator of that.
https://www.makaan.com/price-trends/property-rates-for-buy-in-mumbai

METHODOLOGY

1. **Compiling dataset:** The process of shortlisting the most suitable areas where a supermarket can be located starts off by formulating a dataset which comprises of basic information like the name of borough, neighbourhood, its associated latitude and longitude coordinates, population density and real estate pricing. It's interesting to note that the data pertaining to population density and real estate pricing was available on the internet in the form of categories namely Highest, Very High, High, Low and Very Low. However each of the categories had well defined ranges as seen in the coming section. This dataset will be labelled "mumbai".

The rubrics defined for parameters "Population Density" and "Average real estate pricing" along with the main dataset for Mumbai City are as follows:

	Parameter	Highest	Very High	High	Low	Very Low
0	Population Density/sq.km	> 30000	>7000	>5200	3300	1500
1	Average real estate pricing/sq.km	>80000	>60000	>40000	>20000	<20000

Fig 1: Rubrics for "Population Density" and "Average real estate pricing" for Mumbai City.

-Tahoora Qureshi

"Predicting ideal location for setting up a new supermarket in the city of Mumbai"

	Borough	Neighbourhood	Latitude	Longitude	Population Density	Average real estate pricing
0	Western Suburb	Andheri	19.120371	72.848043	Highest	High
1	Western Suburb	Bandra	19.054979	72.840220	Very High	Very High
2	Western Suburb	Borivali	19.228739	72.856877	Very High	Low
3	Western Suburb	Dahisar	19.257178	72.857536	Very High	Low
4	Western Suburb	Goregaon	18.153715	73.295064	Very High	Low
5	Western Suburb	Jogeshwari	19.134899	72.848820	Very High	Low
6	Western Suburb	Juhu	19.107021	72.827528	Very High	High
7	Western Suburb	Kandivali west	19.208380	72.842227	Highest	Low
8	Western Suburb	Khar	19.072458	72.833707	Very High	Very High
9	Western Suburb	Malad	19.184677	72.835807	Very High	Highest
10	Western Suburb	Santacruz	19.081522	72.841756	Highest	Very High
11	Western Suburb	Vasai	19.440284	72.878656	Highest	Very Low
12	Western Suburb	Virar	19.440284	72.878656	Highest	Low

Fig 2: The main dataset "Mumbai".

2. **Explore 'mumbai' dataset:** Next step will be to explore the dataset for mumbai city first by visualizing its entire neighbourhoods on a map. We use the folium library to render the map.
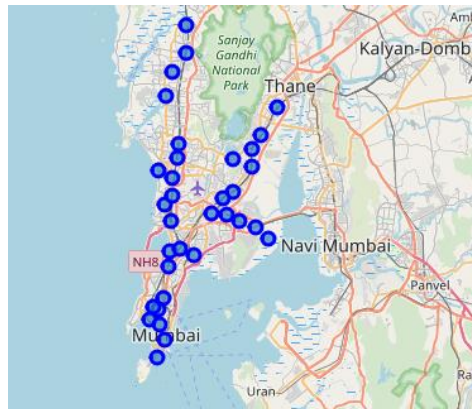

Fig 3: Mumbai city with all its neighbourhoods.

3. **Find existing supermarkets:** All of mumbai city will be scanned for supermarkets with their location. We take the help of foursquare for this step. Observe the borough which has least supermarkets.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	
0	Suryodaya	Convenience Store	NaN	IN	Mumbai	India	Opp. Churchgate Station	1026	[Opp. Churchgate Station, Mumbai, Mahārāshtra,...	[{'label': 'display', 'lat': 18.93356949853369...	18.933569	72.827
1	Dhanraj Supermarket	Department Store	Kohinoor City	IN	Mumbai	India	Off LBS Marg, Kurla West	16808	[Kohinoor City (Off LBS Marg, Kurla West), Mum...	[{'label': 'display', 'lat': 19.08212271265058...	19.082123	72.885
2	Ratna Supermarket	Supermarket	Haware Parekh Chamber, Sion Trombay Road, Chembur	IN	Mumbai	India	Opposite Union Park	14418	[Haware Parekh Chamber, Sion Trombay Road, Che...	[{'label': 'display', 'lat': 19.049532, 'lng':...	19.049532	72.906
3	Society Supermarket	Miscellaneous Shop	Shop 6+7 Beach View Appts.	IN	Mumbai	India	77 Chimbai Rd, Bandra West	13324	[Shop 6+7 Beach View Appts. (77 Chimbai Rd, Ba...	[{'label': 'display', 'lat': 19.05796130691874...	19.057961	72.823
4	Sarvodaya Supermarket	Grocery Store	Dadar	IN	Mumbai	India	Ranade Road	9164	[Dadar (Ranade Road), Mumbai 4000028, Mahārāsh...	[{'label': 'display', 'lat': 19.02096631437195...	19.020966	72.840
5	Dhanraj's Supermarket	Shopping Mall	Kohinoor City	IN	Mumbai	India	Kurla West	8905	[Kohinoor City (Kurla West), Mumbai 400070, Ma...	[{'label': 'display', 'lat': 19.01847117947931...	19.018471	72.842

Fig 4: Existing supermarkets in Mumbai city.

-Tahoora Qureshi

"Predicting ideal location for setting up a new supermarket in the city of Mumbai"

There are a total of 50 supermarkets in Mumbai City. On closely observing the above data frame one would realize that majority of supermarkets exist in 'Western Suburb' borough. This is conducive to our predicament that **'South Mumbai'** could be the ideal place to set up the new supermarket.

4. **Filter a borough:** After closely observing a borough, we filter out a suitable borough based on population density. We want it to be in the higher range. Also the borough should have least number of supermarkets to avoid unwanted competition.

|	Borough	Neighbourhood	Latitude	Longitude	Population Density	Average real estate pricing
0	South Mumbai	Antop Hill	19.020761	72.865256	Highest	Low
1	South Mumbai	Byculla	18.976622	72.832794	Highest	High
2	South Mumbai	Colaba	18.915091	72.825969	Highest	Very High
3	South Mumbai	Dadar	19.023823	72.839427	Highest	High
4	South Mumbai	Fort	18.933267	72.834515	Very High	Low

Fig 5: Dataset of neighbourhoods in South Mumbai.
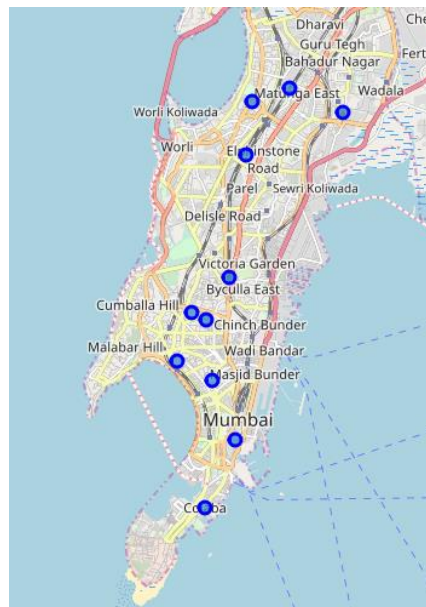
Fig 6: South Mumbai with all its neighbourhoods.

5. **Explore the borough:** Explore and visualize the borough, check out each neighbourhood and find out popular venues. The foursquare location data will be extracted in this step.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Antop Hill	Indian Restaurant	Pizza Place	Smoke Shop	Hostel	History Museum	Diner	Donut Shop	Electronics Store	Hotel	Fast Food Restaurant
1	Byculla	Indian Restaurant	Restaurant	History Museum	Hotel	Park	Pharmacy	Zoo	Bar	Bakery	Beach
2	Colaba	Chinese Restaurant	Restaurant	Coffee Shop	Brewery	Indian Restaurant	German Restaurant	Mediterranean Restaurant	Men's Store	Middle Eastern Restaurant	Pizza Place
3	Dadar	Indian Restaurant	Ice Cream Shop	Chinese Restaurant	Bar	Movie Theater	Electronics Store	Arcade	Breakfast Spot	Burger Joint	Café
4	Fort	Indian Restaurant	Café	Fast Food Restaurant	Seafood Restaurant	Bookstore	Irani Cafe	Lounge	Bakery	Park	Vegetarian / Vegan Restaurant

Fig 7: Exploratory dataset of South Mumbai

-Tahoora Qureshi

6. **Cluster neighbourhoods:** The next step would be to cluster the neighbourhoods by running k means clustering algorithm. This will bring together similar neighbourhoods and bring apart dissimilar neighbourhoods.

	Neighbourhood	Longitude	Population Density	Average real estate pricing	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Byculla	72.832794	Highest	High	0	Indian Restaurant	Restaurant	History Museum	Hotel	Park	Pharmacy	Zoo	Bar	Bakery	Beach
2	Colaba	72.825969	Highest	Very High	0	Chinese Restaurant	Restaurant	Coffee Shop	Brewery	Indian Restaurant	German Restaurant	Mediterranean Restaurant	Men's Store	Middle Eastern Restaurant	Pizza Place
3	Dadar	72.839427	Highest	High	0	Indian Restaurant	Ice Cream Shop	Chinese Restaurant	Bar	Movie Theater	Electronics Store	Arcade	Breakfast Spot	Burger Joint	Café
4	Fort	72.834515	Very High	Low	0	Indian Restaurant	Café	Fast Food Restaurant	Seafood Restaurant	Bookstore	Irani Cafe	Lounge	Bakery	Park	Vegetarian / Vegan Restaurant
5	Girgaon	72.817908	Highest	Low	0	Indian Restaurant	Ice Cream Shop	Juice Bar	Pizza Place	Italian Restaurant	Fast Food Restaurant	Indie Movie Theater	Grocery Store	Coffee Shop	Multicuisine Indian Restaurant
6	Kalbadevi	72.827938	Highest	High	0	Indian Restaurant	Fast Food Restaurant	Jewelry Store	Food	Cheese Shop	Café	Snack Place	Food Court	Electronics Store	Flower Shop
8	Matunga	72.850147	Highest	High	0	Indian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Train Station	Snack Place	Bar	Café	Gym	Jewelry Store	Department Store
9	Parel	72.837661	Highest	Very High	0	Food Court	Coffee Shop	Bakery	Jewelry Store	Food & Drink Shop	Restaurant	Sandwich Place	Dessert Shop	Flower Shop	Pizza Place

Fig 8: Cluster 1

	Neighbourhood	Longitude	Population Density	Average real estate pricing	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Antop Hill	72.865256	Highest	Low	1	Indian Restaurant	Pizza Place	Smoke Shop	Hostel	History Museum	Diner	Donut Shop	Electronics Store	Hotel	Fast Food Restaurant

Fig 9: Cluster 2

|	Neighbourhood	Longitude	Population Density	Average real estate pricing	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Kamathipura	72.826435	Highest	Low	2	Department Store	Fast Food Restaurant	Hotel	Smoke Shop	Donut Shop	Chinese Restaurant	Middle Eastern Restaurant	Dessert Shop	Arts & Crafts Store	German Restaurant
10	Tardeo	72.822151	Highest	Highest	2	Department Store	Fast Food Restaurant	Train Station	Ice Cream Shop	Movie Theater	Restaurant	Food	Diner	Donut Shop	Electronics Store

Fig 10: Cluster 3
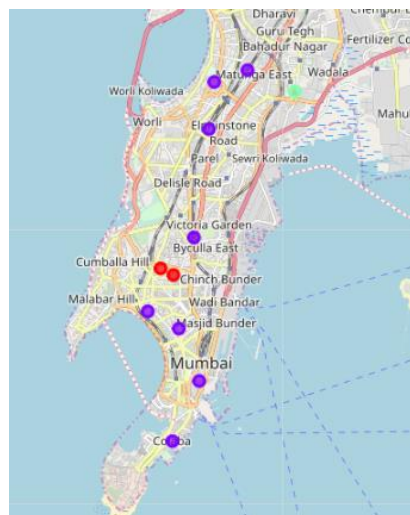
Fig 11: Clustered neighbourhoods.

7. **Observe and predict:** The clusters will then be observed to identify most suitable traits of neighbourhoods which will be conducive to the prospect of starting a new supermarket.

-Tahoora Qureshi

ANALYSIS

Cluster 1:

For our analysis we might as well ignore Fort and Girgoan as their real estate pricing is "Low" and we are targeting high end plush localities. Following table summarizes our observations for cluster 1.

Neighbourhood	Key venues	Population Density	Average real estate pricing	Key observations
Byculla	History Museum, Hotel, Park, Zoo, Beach	Highest	High	Consists of inter-neighbourhood popular venues. Place attracts huge chunks of people from all over Mumbai/India.
Colaba	Indian, german, chinese, mediterrenean, middle eastern restaurants	Highest	Very High	Consists of inter-neighbourhood popular venues. Place attracts reasonable number of people.
Dadar	Movie theatre, arcade	Highest	High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.
Kalbadevi	Indian Restaurant, Fast Food Restaurant	Highest	High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.
Matunga	Departmental Store, Train station	Highest	High	An existing popular departmental store.
Parel	Food court, Pizza place	Highest	Very High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.

Cluster 2

Since the average real estate pricing for Antop Hill is low, we might as well ignore it too.

Cluster 3

Both of the above mentioned neighbourhoods have Departmental Store as their popular venue. This means more competition for the supermarket. Hence, we ignore these neighbourhoods too.

-Tahoora Qureshi

"Predicting ideal location for setting up a new supermarket in the city of Mumbai"

RESULTS AND DISCUSSION

The aim of this project was to determine an ideal location to set up a new supermarket given a set of pre-requisites. They are summarised as follows:

1. The location should be **densely populated** to attract more crowds in order to have a good client base.
2. The location should not have **competitors** in the form of existing departmental stores/ malls/supermarkets etc.
3. Location should be plush and high end to target customers with high income. This information is sort from the **real estate pricing** which is a key indicator of income groups of people.

Keeping the above set of information in mind, the overall count of large scale supermarkets in Mumbai City was derived using foursquare location data. It was observed majority of supermarkets were in the 'Western borough' of Mumbai city. That being said, it was also observed that the population density was highest in 'South Mumbai' borough. Both the observations supported the idea that the new supermarket should be located in 'South Mumbai' but the question was where? This necessitated a detailed analysis of all neighbourhoods in 'South Mumbai'. Top 10 venues for each neighbourhood were extracted and the neighbourhoods were clustered as follows:

Cluster no.	Details	Cluster Name
Cluster 1	Neighbourhoods with restaurants as their most popular venue.	Food Centric Venues
Cluster 2	Neighbourhood with smoke joint and hostel.	Students centric venue
Cluster 3	Neighbourhoods with Departmental stores as their most popular venue.	Consumer goods centric venues

The obvious focus was on cluster 1 namely "Food centric Venues". A detailed analysis of neighbourhoods in cluster 1 is as follows along with its conclusion:

Neighbourhood	Key venues	Population Density	Average real estate pricing	Key observations
Byculla	History Museum, Hotel, Park, Zoo, Beach	Highest	High	Consists of inter-neighbourhood popular venues. Place attracts huge chunks of people from all over Mumbai/India.
Colaba	Indian, german, Chinese, mediterrenean, middle eastern restaurants	Highest	Very High	Consists of inter-neighbourhood popular venues. Place attracts reasonable number of people.
Dadar	Movie theatre, arcade	Highest	High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.

-Tahoora Qureshi

Kalbadevi	Indian Restaurant, Fast Food Restaurant	Highest	High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.
Matunga	Departmental Store, Train station	Highest	High	An existing popular departmental store.
Parel	Food court, Pizza place	Highest	Very High	Consists of intra-neighbourhood popular venues and hence attracts the nearby crowd.
*From the above discussion, it is clear that "**Byculla**" and "**Colaba**" seems to be the ideal places to set up a new super market.*				

The choice among the two will be left onto the stakeholders involved. The key differences between the two most suitable neighbourhoods are:

Byculla	• Real estate pricing is 'High' (40,000 to 60,000 per sq. ft.)
Colaba	• Real estate pricing is 'Very High' (60,000 to 80,000 per sq. ft.)

Depending on amount of funds the stakeholders are ready to invest; either of the neighbourhoods can be selected as an ideal location for setting up the new supermarket.

CONCLUSION

This project demonstrated that an ideal location for supermarket can be arrived at given certain requirements. First the city of interest was decided, which in our case was Mumbai. Then a data frame was formulated which consisted of boroughs, neighbourhoods, latitude, longitude, real estate pricing and population density of each neighbourhood in Mumbai. An exploratory analysis was conducted to filter out the most suitable borough and then the most suitable neighbourhood given the fact that our aim was to target areas of high population density and high average real estate pricing. It was thus concluded that the neighbourhoods "Byculla" and "Colaba" were most suitable as our specific requirements were successfully met.

-Tahoora Qureshi