MUMBAI -The Commercial City to Excel the Opportunity

Introduction:

Mumbai is the commercial capital of India. It is also known as the city that never sleeps. Mumbai is the perfect blend of culture, customs, and lifestyles. Mumbai is India's most cosmopolitan city, its financial powerhouse, and the nerve centre of India's fashion industry. Mumbai is also dotted with plenty of architectural landmarks from the Victorian era and the days of Raj. Mumbai is also the birthplace of Indian Cinema. Located on Maharashtra's coast, Mumbai is India's most-populous city, and it is one of the largest and most densely populated urban areas in the world. Mumbai developed a highly diversified infrastructure. It suffers, however, from some of the perennial problems of many large expanding industrial cities: air and water pollution, widespread areas of substandard housing, and overcrowding. With its diverse society, comes diverse infrastructure which decides the quality of living. There are many infrastructures in Mumbai, each belonging to different categories like Drinking Water? Plant, Waste Water/ Sewage, Hospitals, Schools, Colleges, Railway Network, Electricity Power Plants, Telecommunication Support, Bank, Shopping malls, Supermarket, Gas Station, Hotels, Police Station, Café, medical shops, grocery shops, theatre, etc. One of the main problems, when one moves to a new city, is where to find a good area to build and grow prosperously.

Business Problem:

The questions I aim to answer in this project are the following:

1. List and visualize all major parts of Mumbai City with top existing infrastructure.
2. What are the best locations in Mumbai as per infrastructure?
3. Which areas have the potential for the development of infrastructure of different kinds?
4. Which all areas lack the infrastructure facilities?
5. What is the best place to stay within a city for all vital infrastructure facilities?

Data Description:

Mumbai City's demographics show that it is a large and ethnically diverse metropolis. With its diverse society, comes diverse infrastructure. There are many different kinds of infrastructure in Mumbai City, each belonging to different categories like Hospitals, Schools, Colleges, Hotels, etc.

For this project we need the following data:
Mumbai Pin code (Scraped from web source)

- Data source: https://mumbai7.com/postal-codes-in-mumbai/

• Description: Contain a list of pin codes, postal office names, city which can be used to discover all postal office of Mumbai.
Mumbai City data contain list pin codes, postal office names, city along with them latitude and longitude.
• Data Source: https://mumbai7.com/postal-codes-in-mumbai/
• Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Mumbai City.
Different kinds of infrastructures in each neighbourhood of Mumbai City.
• Data source: Foursquare API
• Description: By using this API we will get all the venues in each postal office.
We can filter these venues to get different infrastructures and venues.
GeoSpace data • Data source: https://github.com/geospace-code/pymap3d
• Description: By using this geospace data, we will get the latitude and longitude coordinates of the postal office of Mumbai.

Using this data will allow exploration and examination to answer the questions. This is a project that will make use of many data science skills, from web scraping working with API (Foursquare), data cleaning, data wrangling and map visualization and to machine learning (K-means clustering)

FINDING TOP INFRASTRUCTURE EXPLORATION:

Next, we make use of Foursquare API to get the top venues that are within a radius of 600 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data inJSON format and we will extract the venue name, venue category, venue latitude, and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category.

DATA WRANGLING:

We are also preparing the data for use in selection. Based on the occurrence of infrastructures in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new infrastructures and which neighbourhoods are most suitable to visitors to stay.

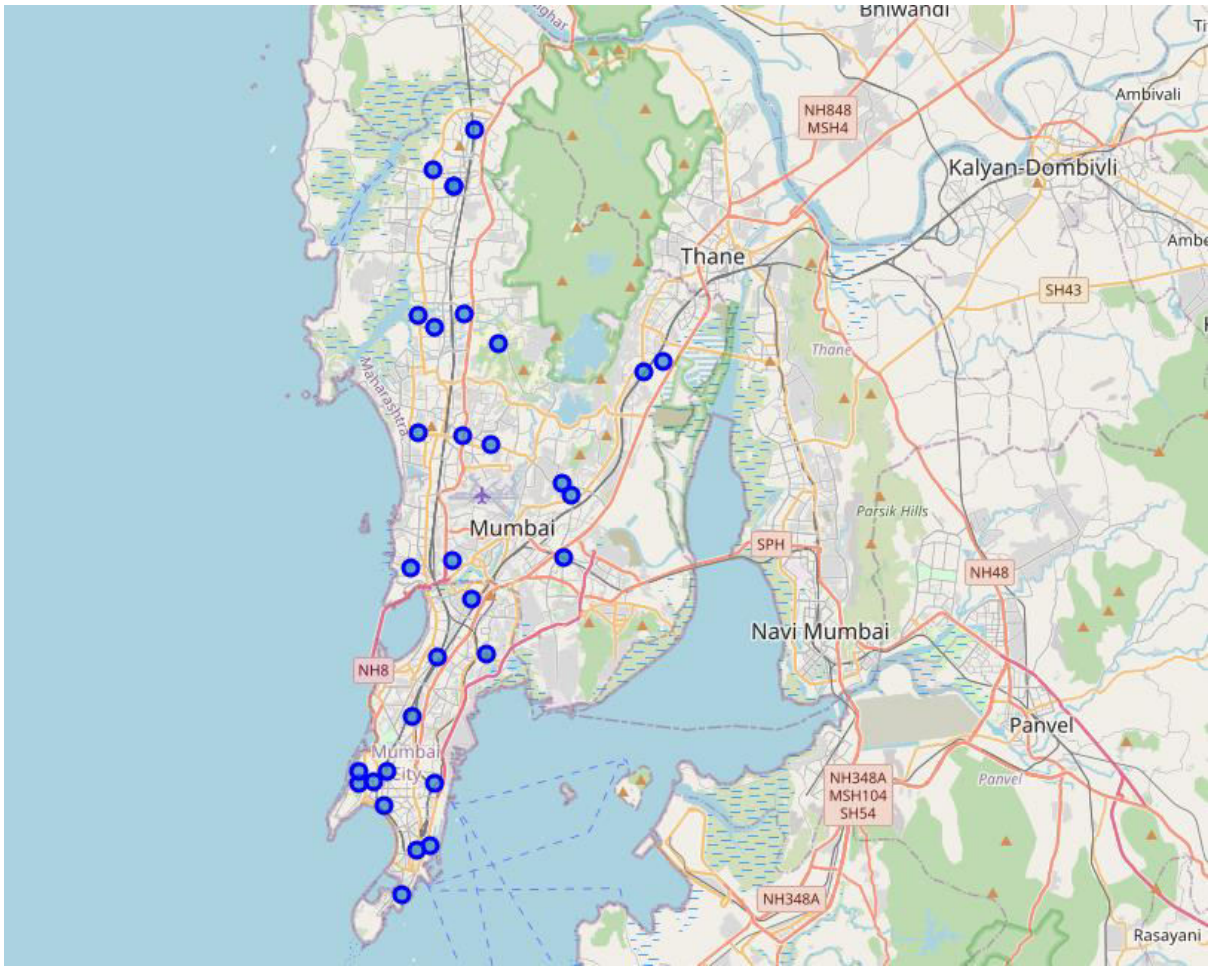
DATA CLUSTERING:

Finally, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to

the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “no. of existing infrastructures”. The results will allow us to identify which neighbourhoods have higher, medium and lower concentration of infrastructures. Based on the occurrence of infrastructures in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new infrastructures.

Results:

1.Top Existing Infrastructure



2. lack the infrastructure facilities:

```
badquality = qualitymumbai_grouped[qualitymumbai_grouped['Total infrastructure'] == qualitymumbai_grouped['Total infrastructure'].min()]
badquality
```

53]:

	Post Office	Pin Code	City	Bank	Bus Station	Business Service	Café	College Auditorium	Electronics Store	Farmers Market	...	Indie Movie Theater	Light Rail Station	Monument / Landmark	Park	Pharmacy	Restaurant	Shopping Mall	Theater	Train Station	Total infrastructure
2	Antop Hill	400037	Mumbai	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
10	Bhandup (East)	400042	Mumbai	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	1
18	Council Hall	400039	Mumbai	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
23	Ghatkopar (West)	400086	Mumbai	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0	1
28	I I T Mumbai	400076	Mumbai	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	1

5 rows x 23 columns

2. Best Location as per infra:

Post Office	Bandra (West)
Pin Code	400050
City	Mumbai
Bank	0
Bus Station	0
Business Service	0
Café	9
College Auditorium	1
Electronics Store	0
Farmers Market	1
Garden	0
Gym / Fitness Center	2
Hotel	1
Indie Movie Theater	1
Light Rail Station	0
Monument / Landmark	0
Park	1
Pharmacy	0
Restaurant	2
Shopping Mall	1
Theater	0
Train Station	0
Total infrastructure	19

4. Highest Potential Antop Hill:

```
# making data frame from csv file
print("These are infrastructures with highest potential in" , yourchoicearea, "area : " )
for idx in range(len(infraqualitychoice)) :
    if (infraqualitychoice.iloc[idx,1] == 0):
        print (infraqualitychoice.iloc[idx,0])
```

```
These are infrastructures with highest potential in Antop Hill area :
Bank
Bus Station
Business Service
Café
College Auditorium
Electronics Store
Farmers Market
Garden
Hotel
Indie Movie Theater
Light Rail Station
Monument / Landmark
Park
Pharmacy
Restaurant
Shopping Mall
Theater
Train Station
```

5.Best Place to stay within a city:

43	I I T Mumbai	13
44	J B Nagar	13
45	JNPT Town Ship	13
67	Manor	13
79	Mumbai G P O	13
80	N I T I E	13
88	Papdi	13
92	Rajbhavan	13
97	Santacruz (East)	13

Clustering Based on Total Infrastructure:

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “no. of existing infrastructures”:

- Cluster 1: Neighbourhoods with a low number to no existence of infrastructures.
- Cluster 0: Neighbourhoods with a high number of infrastructures
- Cluster 2: Neighbourhoods with a moderate number of infrastructures

Discussion:

This project recommends a person who is planning to build infrastructure to capitalize on these findings to open new Infrastructures in neighbourhoods in cluster 1 with little to no competition. A person who is planning to build infrastructure with unique selling propositions and lives prosperously to stand out from the competition can also open new infrastructures in neighbourhoods in cluster 2 with moderate competition and supporting adequate no. of infrastructures. Lastly, people with planning to settle in the city are advised to start in cluster 0 which already has a high concentration of infrastructures

Limitations:

This project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of a paid account to bypass these limitations and obtain more results.

Conclusion:

- In this project, I have gone through the process of identifying the business problems, specifying the data required, extracting and preparing the data, visualizing the results, performing machine learning by clustering the data into 3 clusters based on their frequency similarities, tackling and reaching to a definitive solution to business problems (mentioned in results).
- Lastly, the project is providing recommendations to the relevant stakeholders i.e., business developers regarding the best locations to open a new infrastructure. The project also provides visitors and immigrants to the city regarding postal office areas for growth and living prosperously