Final Capstone Project Report:

An Analysis of Mumbai

Introduction about Mumbai:

Mumbai is busy, Mumbai is chaotic, Mumbai is loud. Also called the ‘Maximum City’, Mumbai throbs with a harsh, edgy energy. There is absolutely no sense of personal space, yet it gives you the space to be yourself.

21.5 million people call the greater Mumbai metropolitan area, their home. With over 25,000 people per square kilometer it is one of the most densely populated spaces in the world. Asia’s largest slum, Dharavi, packs over 300,000 people per square kilometer. Mumbai is also one of the wealthiest cities in India, accounting for 25% of industrial output and over 70% of maritime trade. The city is a financial, money making, and showbiz capital of India. ‘Bollywood’ the heart of the Indian film industry is firmly anchored in Mumbai. Most of India’s major television networks and publishing houses are located here as well.

This concentration of cultural and financial institutions attracts migrants from all over the country and creates a high level of diversity. It is a city of dreams for some, but a tough life on unforgiving streets for many others. No wonder BBC calls Mumbai the most cosmopolitan city on the subcontinent. In the cosmopolitan chaos of this modern city is a community that thrives. Call them Mumbaikars or Bombayites – they live in a unique ecosystem not found. All these facts make Mumbai a very interesting place for this project.

Business problem:

In this study, we'll aim to answer some questions through data analysis and clustering:

- Visualize all important places in the city with a top existing infrastructure.
- Revealing the best locations in the city according to the infrastructure.
- Revealing the areas of the city that need or have the potential for infrastructure improvements.
- Detecting areas that lack infrastructure facilities.
- Concluding the overall best place to stay within the city.

Target Audience:

This project aims to help 2 types of people:

- Normal individuals who are looking for places with specific facilities in mind (for example, someone who likes to eat outside would search for places with many restaurants in a tight radius).
- Investors who would like to build new facilities in the city and looking for good potential locations to do so (for example, someone who would like to build a hotel would most likely avoid building it near other hotels to avoid competition).

Data description:

This concentration of cultural and financial institutions attracts migrants from all over the country and creates a high level of diversity. It is a city of dreams for some, but a tough life on unforgiving streets for many others. In the cosmopolitan chaos of this modern city is a community that thrives. Call

them Mumbaikars or Bombayites – they live in a unique ecosystem not found. For this project, we will use the following:

- Mumbai postal codes from: https://mumbai7.com/postal-codes-in-mumbai/
- Mumbai city data.
- Different kinds of infrastructures (hotels, restaurants, coffee shops,... etc) in each neighborhood in the city. We will use the Foursquare API for that.
- Geospace date. By using it we will be able to get the latitude and longitude of the different facilities in the city.

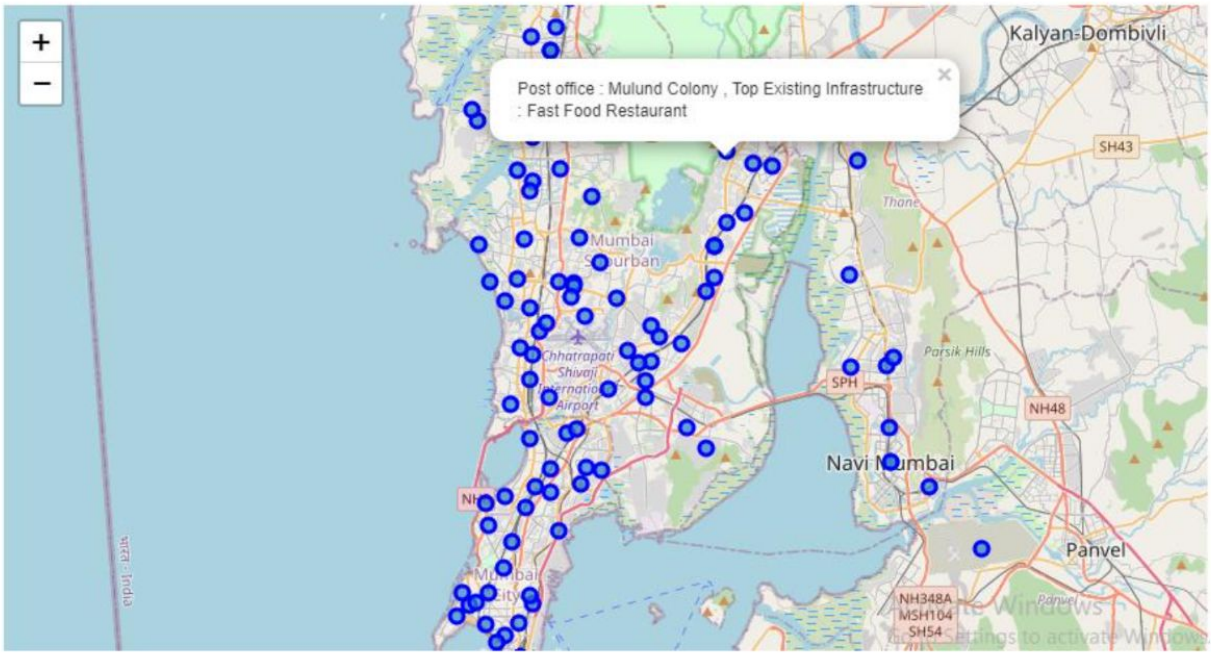
Using these different sets of data and API will serve to answer the questions discussed in the part 1 document in this project.

Methodology:

Visualizing the neighbourhood in the map using the Folium package. Finding a set number of infrastructures within a set radius. Lastly, performing clustering on the data using the k-means algorithm.

Results:

display the top existing infrastructure for each postal office:



Best locations as per infrastructure:

Post Office	Bandra (West)
Pin Code	400050
City	Mumbai
Airport Terminal	0
Bank	0
Bus Station	0
Business Service	0
Café	10
College Auditorium	1
Electronics Store	1
Farmers Market	1
Garden	0
Government Building	0
Gym / Fitness Center	3
Hotel	1
Indie Movie Theater	1

Indie Movie Theater	1
Light Rail Station	0
Market	0
Monument / Landmark	0
Park	1
Pharmacy	0
Playground	0
Resort	0
Restaurant	1
Shopping Mall	1
Theater	0
Train Station	0
Total infrastructure	21

All areas that lack infrastructure facilities:

Post Office	Pin Code	City
Agashi	401301	Thane
Anu Shakti Nagar	400094	Mumbai
Bassien	401201	Thane
Bhandup (East)	400042	Mumbai
Bhayander (East)	401105	Thane
Boisar	401501	Thane
Ghansoli	400701	Navi Mumbai
Jacob Circle	400011	Mumbai
Jakegram	400606	Thane
Jawhar	401603	Thane
Jawhar	401603	Thane
Kopri Colony	400603	Thane
Krishi Utpanna Bazar	400705	Navi Mumbai
Mahim	400016	Mumbai
Nerul Mode	400706	Navi Mumbai
Santacruz P&T Colony	400029	Mumbai
Sopara	401203	Thane
Tagore Nagar	400083	Mumbai
Talasari	401606	Thane
Umbarpada	401102	Thane
Uran	400702	Navi Mumbai
Vasai East I/E	401208	Thane
Wadala	400031	Mumbai

Areas with potential for infrastructure development:

```
These are infrastructures with highest potential in Mantralaya area :
Airport Terminal
Bank
Bus Station
Business Service
College Auditorium
Farmers Market
Garden
Government Building
Indie Movie Theater
Light Rail Station
Market
Monument / Landmark
Park
Pharmacy
Playground
Resort
Train Station
```

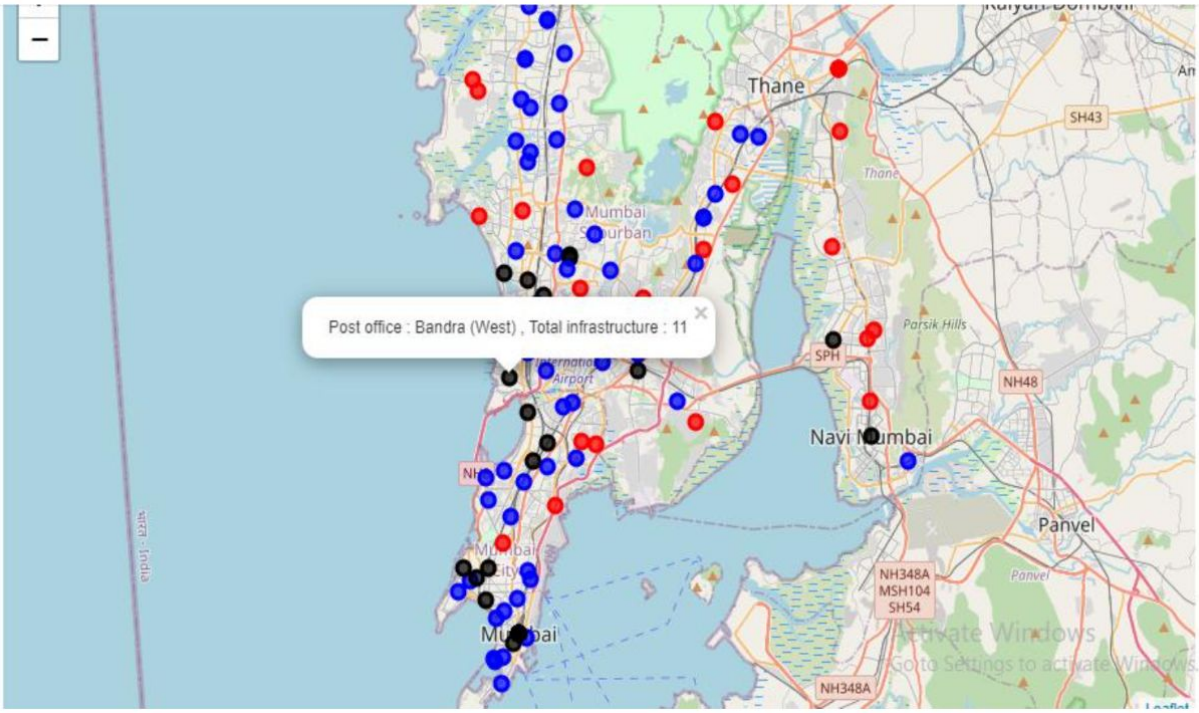
Best places to stay within the city:

	Post Office	Total infrastructure
18	Bhavani Shankar Road	13
28	Council Hall	13
29	Cumballa Hill	13
30	Dadar	13
34	F C I Mumbai	13
35	Ganeshpuri	13
38	Girgaon	13
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:

Most of the infrastructures are concentrated in the Southern areas of Mumbai city, with the highest number in cluster 0 and moderate number in cluster 2. On the other hand, cluster 1 has a very low number of infrastructures in the neighborhoods. This represents a great opportunity and high potential areas to open new infrastructures as it is very little to no competition from existing varied infrastructures. Meanwhile, one can specifically check the infrastructure of choice against the postal office choice area.

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 neighborhoods in cluster 2 with moderate competition and supporting adequate no. of infrastructures. Lastly, people planning to settle in the city are advised to start in cluster 0 which already has a high concentration of infrastructures.



Observations:

Most of the infrastructures are concentrated in the Southern areas of Mumbai city, with the highest number in cluster 0 and moderate number in cluster 2. On the other hand, cluster 1 has a very low number of infrastructures in the neighborhoods. This represents a great opportunity and high potential areas to open new infrastructures as it is very little to no competition from existing varied infrastructures. Meanwhile, one can specifically check the infrastructure of choice against the postal office choice area.

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 neighborhoods in cluster 2 with moderate competition and supporting adequate no. of infrastructures. Lastly, people planning to settle in the city are advised to start in cluster 0 which already has a high concentration of infrastructures.

Acknowledgement:

In this project, we 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.

Special thanks:

I want to thank IBM, Coursera and anyone who made it this far.

Thanks and have a nice day!