

IBM Capstone Project

Opening a shopping mall in Mumbai (India)

Rohit Garg
March 2020



Introduction

Mumbai (also known as Bombay the official name until 1995) is the capital city of the Indian state of Maharashtra. According to United Nations, as of 2018, Mumbai was the second most populous city in India after Delhi and the seventh most populous city in the world with a population of 19.98 million. Mumbai is the financial, commercial and entertainment capital of India. It is also one of the world's top ten centres of commerce in terms of global financial flow,[36] generating 6.16% of India's GDP and accounting for 25% of industrial output, 70% of maritime trade in India (Mumbai Port Trust and JNPT), and 70% of capital transactions to India's economy. Mumbai's billionaires had the highest average wealth of any city in the world in 2008.

A shopping mall is a modern term for a form of shopping precinct or shopping center in which one or more buildings form a complex of shops with interconnecting walkways, usually indoors. From the late 20th century, entertainment venues such as movie theaters and restaurants began to be added. As a single built structure, early shopping centers were often architecturally significant constructions, enabling wealthier patrons to buy goods in spaces protected from the weather.

Business problem

Location plays an important role in the success of a shopping mall. In this exercise we will analyse the best location/s to open a shopping mall in Mumbai (India).

Target audience

This analysis will be very helpful for property developers and investors looking to open a shopping mall in Mumbai (India).

Data

- List of neighbourhoods in Mumbai
https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai
- Co-ordinates of these neighbourhoods
[Python – geocoder package](#)
- Venue data
[Foursquare API](#)

Methodology



1. Get data from Wiki

- Get neighbourhood information from Wikipedia page - https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai
- We will do web scraping using beautiful package

2. Get coordinates

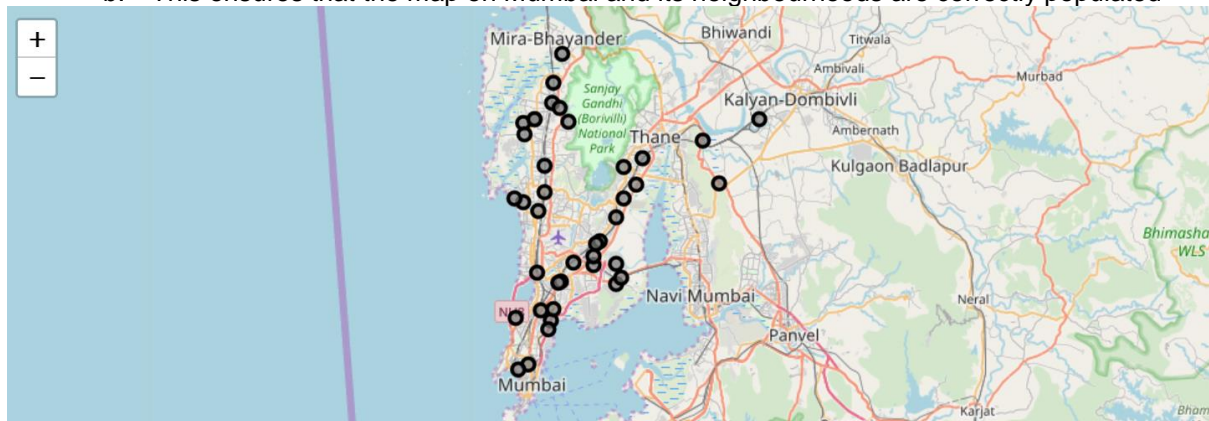
- We will use the geocoder package
- Create dataframe containing the neighbourhood and the corresponding longitude and latitude

3. Get coordinates for Mumbai

- We will use the geocoder package
- The geographical coordinate of Mumbai, India: 18.9387, 72.8353

4. Map of Mumbai

- We will use the folium package
- This ensures that the map on Mumbai and its neighbourhoods are correctly populated



5. Foursquare - top 100 venues in a radius of 1000 meters

- We will use the Foursquare API to get the top 100 venues in a radius of 1000 meters (1 km).
- We will make API calls to Foursquare by passing the geographical co-ordinates of the neighbourhood.
- We will then check how many venues are returned for each neighbourhood.

6. Analyse each area

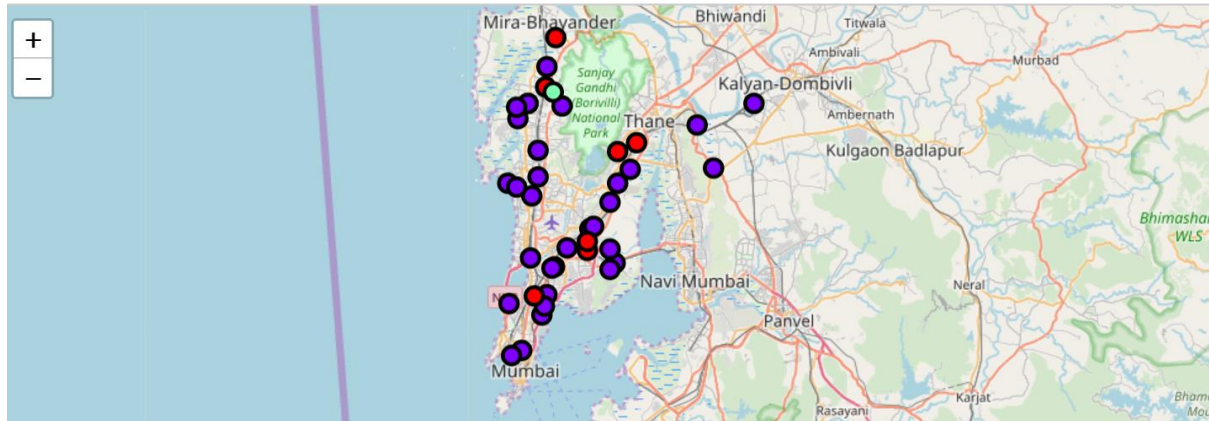
- We will do one-hot encoding for each venue category
- Then we will analyse each area by taking the frequency of occurrence of each venue category
- We will filter "Shopping Mall" as venue

7. Cluster Areas - kmeans

- Next we will perform k-means clustering
- We will cluster neighbourhoods into 3 clusters based on the frequency of occurrence for "Shopping Mall"
- The result will help us to identify the areas with lower concentration of shopping malls and the areas with higher concentration of shopping malls

8. Cluster map

- We will use the folium package
- The map will help us to answer which neighbourhood is best for opening the shopping mall



Results

The k-means clustering algorithm creates 3 clusters

- Cluster 0 (in blue – ref above map): No shopping mall
- Cluster 1 (in red – ref above map): Moderate concentration of shopping mall
- Cluster 2 (in green – ref above map): High concentration of shopping mall

Cluster 0	Andheri, Western Suburbs (Mumbai), Kurla, Mahavir Nagar (Kandivali), Mankhurd, Matharpacady, Mumbai, Mumbra, Seven Bungalows, Shil Phata, Sion, Mumbai, Thakur village, Tilak Nagar (Mumbai), Uttan, Vashi, Vikhroli, Wadala, Kanjurmarg, Kandivali, Kausa, Worli, Jogeshwari, Grant Road, Goregaon, Ghatkopar, Eastern Suburbs (Mumbai), Anushakti Nagar, Baiganwadi, Dombivli, Bandra, Kalyan, Dahisar, Bhandup, Charkop	Min (average shopping mall in 1 km) = 0 Max (average shopping mall in 1 km) = 0
Cluster 1	Borivali, Juhu, Chembur, Pestom sagar, Mulund, Mira Road, Sonapur, Bhandup	Min = 0.02 Max = 0.03
Cluster 2	Devipada	0.08

Discussion

- Most of the shopping malls are located in the dense residential areas with the highest concentration in cluster 2 and lowest in cluster 1. There is good opportunity to open shopping mall in cluster 1:
- cluster 2 - likely tough competition
- cluster 0 - likely low demand

Therefore we recommend property dealers and investors to look for opportunities for the neighbourhoods that are in cluster 1.

Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
Borivali	0.021277	1	19.22936	72.85751
Juhu	0.02439	1	19.01492	72.84522
Chembur	0.02439	1	19.06218	72.90241
Pestom sagar	0.025641	1	19.07064	72.90217
Mulund	0.030303	1	19.17183	72.95565
Mira Road	0.041667	1	19.280032	72.867932
Sonapur, Bhandup	0.034483	1	19.16394	72.93544

Limitations

We considered only the density of shopping malls in a particular neighbourhood. The following limitations were identified:

- Other factors such as ease of transportation, average income of the neighbourhood, average spending of the neighbourhood, etc were studied.
- We could have also converted this study into a supervised learning algorithm (propensity of mall is determined by various factors) rather than limiting ourselves to unsupervised learning algorithm (that is k-means).

Conclusion

In this project we have addressed the question – analyse the best location/s to open a shopping mall in Mumbai (India). This analysis will be very helpful for property developers and investors looking to open a shopping mall in Mumbai (India).

The neighbourhoods in cluster 1 are the most preferred locations to open a new shopping mall.