

**IBM Applied Data Science  
Capstone Project**

**Battle of Neighborhoods**

**Final Project Report**

**By**

**Shreyas Dronamraju**

# **Introduction**

A shopping mall is a modern, chiefly North American, 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. Shopping malls these days are on stop destinations due to their versatility of covering eating, shopping, entertainment and other leisure activity.

Due to this popularity of shopping malls, constructors are building many more such malls in Mumbai to cater to the demands. Shopping malls act as a consistent rental medium to the owners.

Various parameters are taken into concern when building a shopping mall. One of the most important of these parameters is the location of the mall. Hence this report will be dealing with analyzing and selecting the best location to open a mall in Mumbai. The project will be useful for builders to select a location to build a mall. A secondary benefit is that it can also help people look at the areas with the most malls in the city.

## **Business Problem**

Which is the best location to open a mall in Mumbai?

# **Data**

## **Data Needed**

1. List of neighborhoods in Mumbai
2. Latitude and Longitude coordinates
3. Data of venues

## **Data Collection**

1. List of neighborhoods in Mumbai data via Wikipedia defines the scope of this project which is confined to the city of Mumbai.
2. Python Geocoder package for latitude and longitude coordinates of the neighborhoods in order to plot the map and also to get the venue data.
3. Foursquare API to get the venue data for the neighborhoods.
4. Python Folium Package for Map Visualization.

# Methodology

**Step 1:** Import the libraries going to be used

**Step 2:** Data collection

- We obtain the list of neighborhoods of Mumbai using web scrapping from Wikipedia page: [https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Mumbai](https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai)
- We also obtain the latitudes and longitudes of the neighborhoods from Foursquare API.
- Plot these neighborhoods onto a map of Mumbai using folium.
- Next, we use Foursquare to get the top 100 venues in a radius of 2000 meters. This is done by using our Foursquare dev account to make calls to Foursquare API
- Once we receive the venue data, we find the number of unique categories of venues are present in the data set.

**Step 3:** Analysis

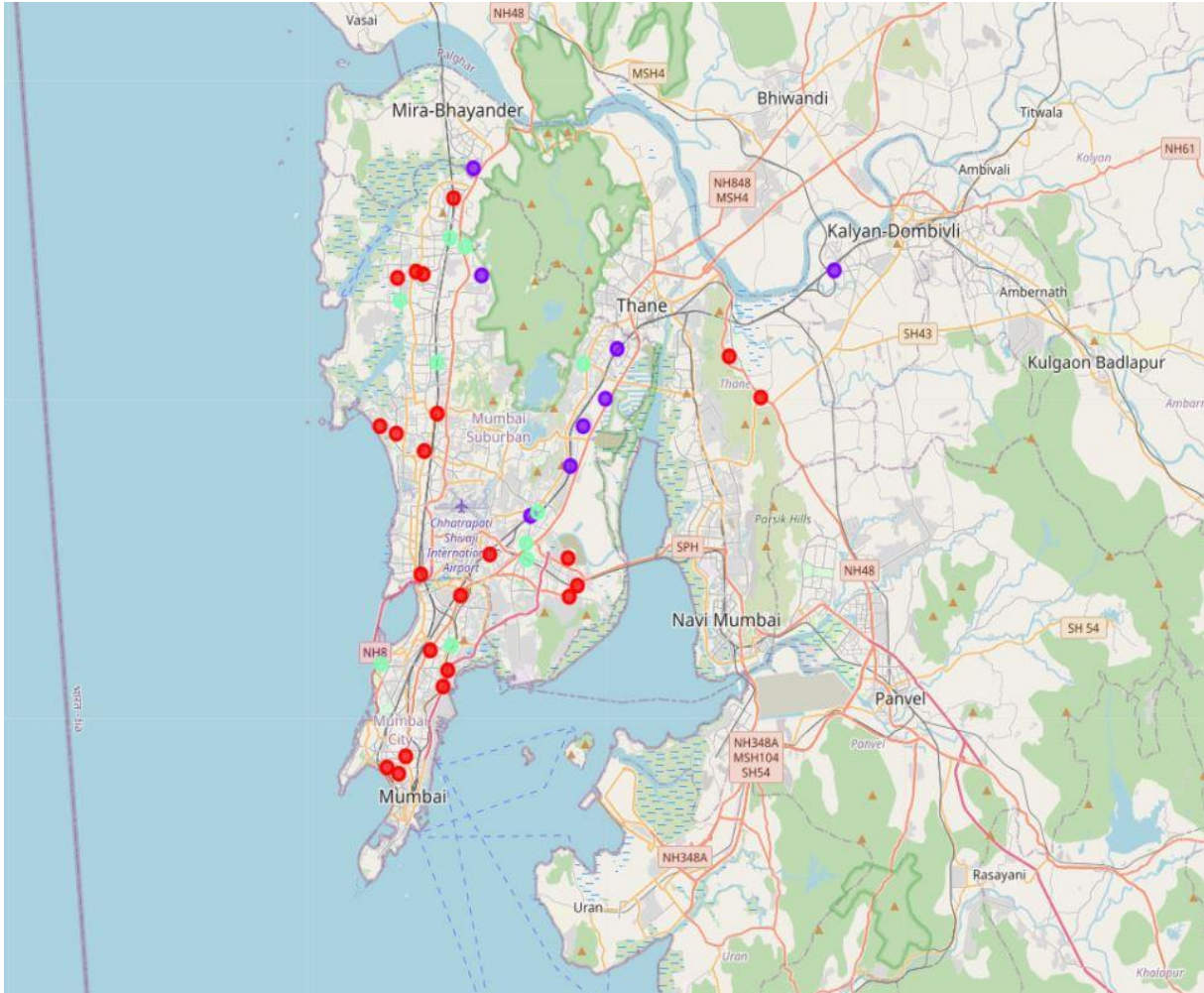
- We analyze each neighborhood and group the neighborhoods based on th frequency of occurrence of each venue category.
- After grouping we filter the venue category “shopping malls” in each neighborhood to study shopping malls data.

**Step 4:** Implementation

Finally, we apply clustering on the data using k-means clustering method. We divide it into 3 clusters based on the frequency of occurrence for shopping malls.

## Results

1. Red cluster (0): Neighborhoods with negligible to no number of malls.
2. Purple cluster (1): Neighborhoods with highest number of malls.
3. Green cluster (2): Neighborhoods with a moderate number of shopping malls.



Map of Mumbai with the clusters marked on it.

Most of the malls are concentrated in the Western and Southern areas of Mumbai while the Eastern and Central part of Mumbai consist of lesser number of malls.

A. Hence cluster 1 will is overpopulated in terms of malls which causes intense competition. Developers are recommended to avoid these areas. These areas include mainly the western and a part of southern parts of Mumbai.

B. Cluster 2 is not over populated but the presence of competition will be noticed. Developers with new and innovative projects can be recommended this cluster to deal with the moderate competition.

C. Cluster 0 has negligible competition due to lack of malls. hence developers are strongly recommended to take advantages of these zones. These are the eastern and centrals parts of Mumbai.

## **Conclusion**

In this project we use data science to answer the business problem stated in the start. We use create clusters based on the similarities in the property of frequency of occurrence to provide viable recommendations using the data.

To answer the question stated in the start of the project, the red cluster or cluster one is the most viable option to open a mall due to the lower concentration of shopping malls in the areas.