Mudita **February 27, 2021**

Mumbai Neighborhood Analysis

Predicting locations for new Shopping Malls & Theaters

1. Introduction

Mumbai is the financial capital and most densely populated city of India. Being a crowded city, people likely to spend most of their leisure time at malls for shopping and food and theaters for watching movies. Based on the current numbers of malls/ theaters, I believe that there is a market for new potential malls/theaters for Mumbai.

Therefore, through this project, I am bringing a solution for the investors while helping them find new potential areas for constructing and opening new shopping malls/theaters in the neighborhood of Mumbai.

This project can help investors to find the new potential areas based on the two main parameters - 1) how many malls/theaters there are in the neighborhood and 2) what the population density is in that neighborhood

2. Data Description

Mumbai Neighborhood Details: To start this project, firstly, I need Mumbai Neighborhood details that include Mumbai area name along with their Latitude and Longitude details. I extracted this information through Wikipedia. [1]

Mumbai Coordinates Details: I extracted Mumbai's coordinates details through 'geolocator' library to plot Mumbai's map

Foursquare API: I used the Foursquare API to extract the venue details around the coordinate of each area. I captured venue details in a radius of 2.5 km from the coordinates of each area and only considered venue details that area related to shopping malls/ theaters. [2]

K-Means Clustering: I analyzed Mumbai's neighborhood using the K-Means clustering machine learning algorithm, computed the best K (number of clusters) for K-Means through Silhouette Score, and used Folium map to visualize clusters into Mumbai's map.

Mumbai GeoJSON details: I extracted Mumbai GeoJson data online and used them in creating a Choropleth map to create boundaries for each area in Mumbai's map [3]

Mumbai Area Wise Population Density: I extracted area wise population density details online and used Choropleth map to populate population density for each area [4]

Python Libraries: I used following python libraries writing my python code.

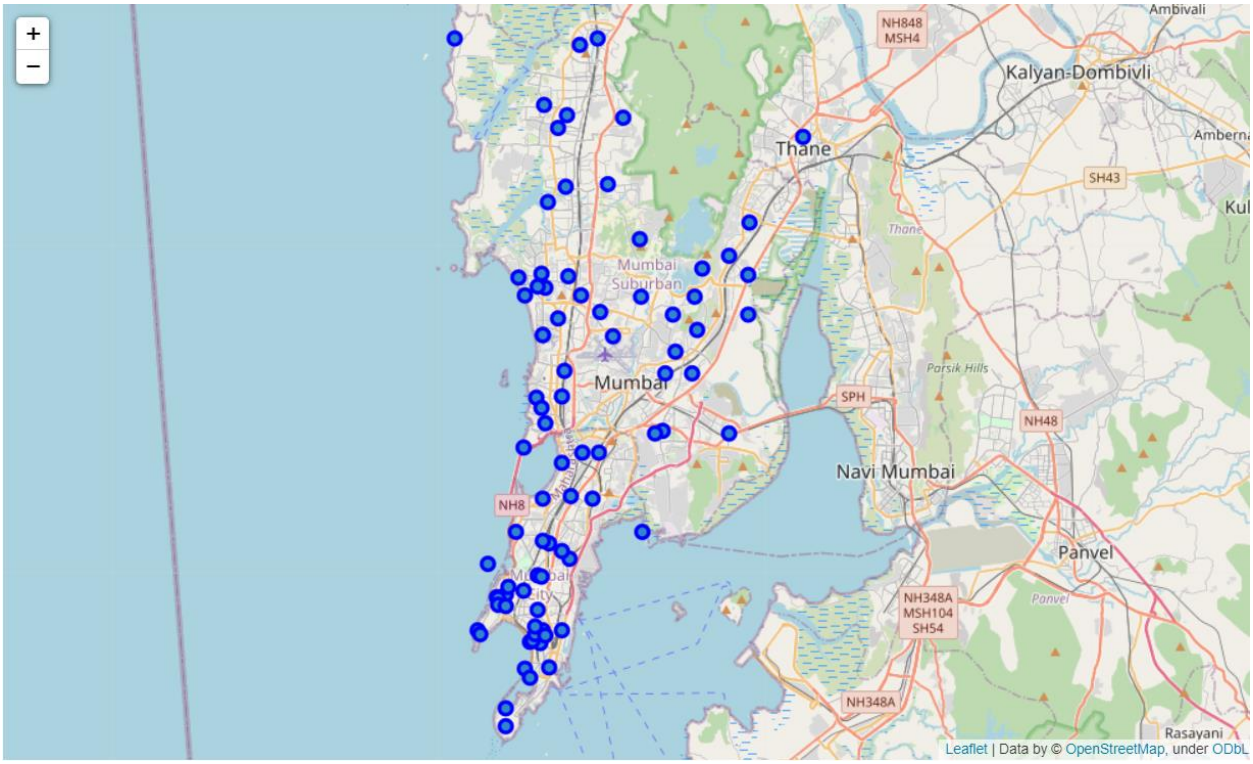
- pandas & numpy : for preparing dataframes
- sklearn.cluster & sklearn.metrics : for K-Mean clustering & Silhouette Score
- matplotlib & Folium: for plotting maps
- geopy: for Mumbai city coordinates
- json: For handling Geojson files
- requests: library to handle requests

3. Methodology

I used Jupyter Notebook to write my python code. Firstly, I extracted Mumbai's neighborhood details online. I prepared master data consisting Area, Location, Latitude and Longitude columns. The master data contains a total of 93 areas within Mumbai.

	Area	Location	Latitude	Longitude
1	Amboli	Andheri,Western Suburbs	19.1293	72.8434
2	Chakala, Andheri	Western Suburbs	19.111388	72.860833
3	D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373
4	Four Bungalows	Andheri,Western Suburbs	19.124714	72.82721
5	Lokhandwala	Andheri,Western Suburbs	19.130815	72.82927
6	Marol	Andheri,Western Suburbs	19.119219	72.882743
7	Sahar	Andheri,Western Suburbs	19.098889	72.867222
8	Seven Bungalows	Andheri,Western Suburbs	19.129052	72.817018
9	Versova	Andheri,Western Suburbs	19.12	72.82
10	Mira Road	Mira-Bhayandar,Western Suburbs	19.284167	72.871111
11	Bhayandar	Mira-Bhayandar,Western Suburbs	19.29	72.85
12	Uttan	Mira-Bhayandar,Western Suburbs	19.28	72.785

Then, I plotted Mumbai neighborhood details through Folium map



In the second step, I extracted Mumbai's area wise venue details using Foursquare API. I passed the Area name, Latitude and Longitude details from the master data and sent the request to Foursquare API to extract the nearby venue details in the range of 2.5 kms from the coordinates. The resulting venue details for each area contains Venue name, Venue Latitude, Venue Longitude, Venue Category. Total 7,371 venues details were extracted for 93 areas.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.1293	72.8434	Shawarma Factory	19.124591	72.840398	Falafel Restaurant
1	Amboli	19.1293	72.8434	Merwans Cake shop	19.119300	72.845418	Bakery
2	Amboli	19.1293	72.8434	Hard Rock Cafe Andheri	19.135995	72.835335	American Restaurant
3	Amboli	19.1293	72.8434	Jaffer Bhai's Delhi Darbar	19.137714	72.845909	Mughlai Restaurant
4	Amboli	19.1293	72.8434	Joey's Pizza	19.126762	72.830001	Pizza Place

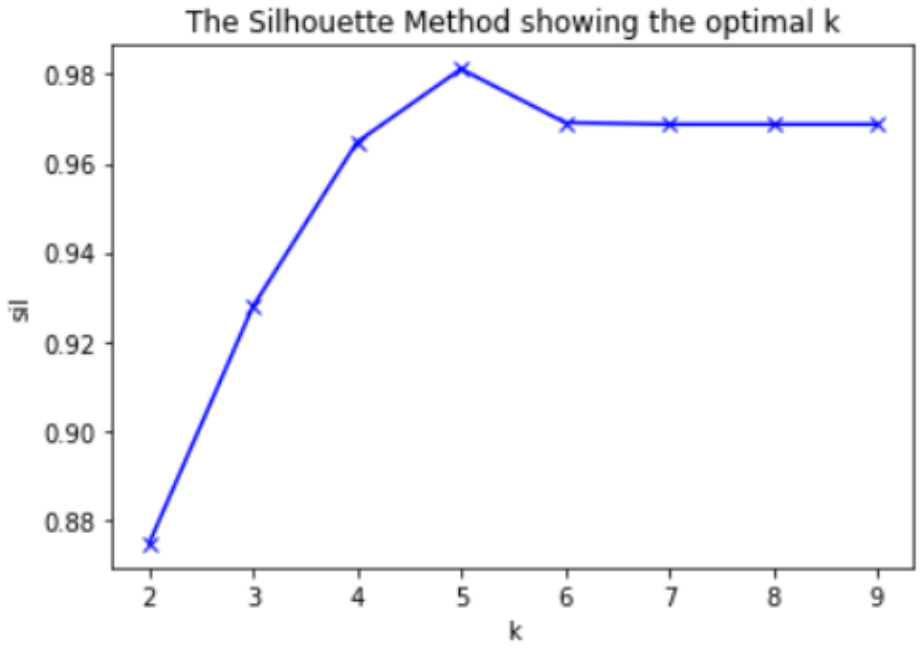
Since our objective is to do analysis around venue related to only shopping malls and theaters. I analyzed data to find out Venue Categories tagged as Shopping mall and Theater and updated consistently as 'Shopping Mall' and 'Theater'.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Amboli	19.1293	72.8434	PVR ICON	19.141531	72.830331	Theater
D.N. Nagar	19.124085	72.831373	Prithvi Theatre	19.106157	72.825810	Theater
D.N. Nagar	19.124085	72.831373	PVR ICON	19.141531	72.830331	Theater
Four Bungalows	19.124714	72.82721	Prithvi Theatre	19.106157	72.825810	Theater
Four Bungalows	19.124714	72.82721	PVR ICON	19.141531	72.830331	Theater
Lokhandwala	19.130815	72.82927	PVR ICON	19.141531	72.830331	Theater
Seven Bungalows	19.129052	72.817018	PVR ICON	19.141531	72.830331	Theater
Versova	19.12	72.82	Prithvi Theatre	19.106157	72.825810	Theater
Mira Road	19.284167	72.871111	Surya Shoping Center	19.272709	72.866619	Shopping Mall
Mira Road	19.284167	72.871111	Inox Cinemas - Thakur Mall - Dahisar	19.263130	72.874483	Theater
Bandstand Promenade	19.042718	72.819132	Le Rêve (Globus Cinema)	19.056014	72.833951	Theater

For analyzing neighborhood wise shopping mall and theater venue details, I used One Hot Encoding. After performing 'group by' by each neighborhood, I found 65 unique neighborhoods in the final result.

	Neighborhood	Shopping Mall_x	Theater_x	Shopping Mall_y	Theater_y
0	Agripada	2	1	0.666667	0.333333
1	Altamount Road	0	1	0.000000	1.000000
2	Amboli	0	1	0.000000	1.000000
3	Amrut Nagar	2	0	1.000000	0.000000
4	Asalfa	2	1	0.666667	0.333333

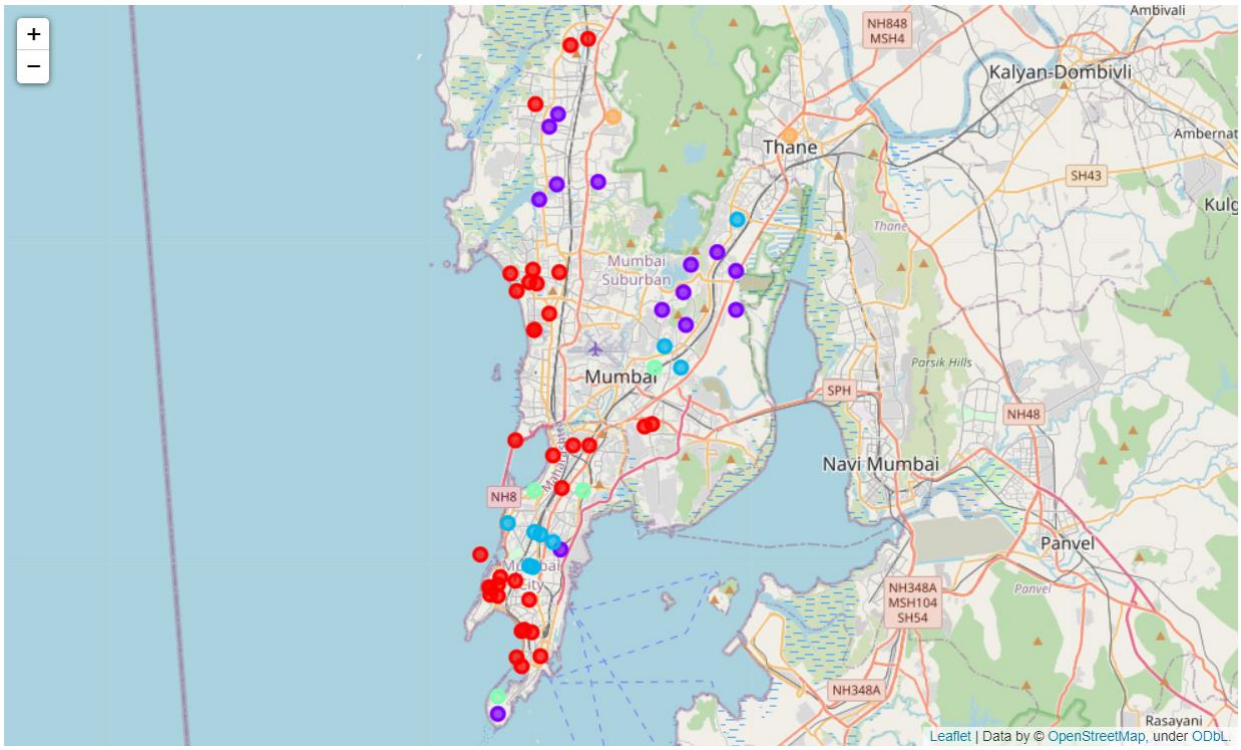
I used K-means machine learning algorithm to cluster Mumbai's neighborhoods, and I used Silhouette matrix to find the optimum value of K (clusters). Form the graph below, it can be seen that optimal value for K is 5 with 98% SIL accuracy.



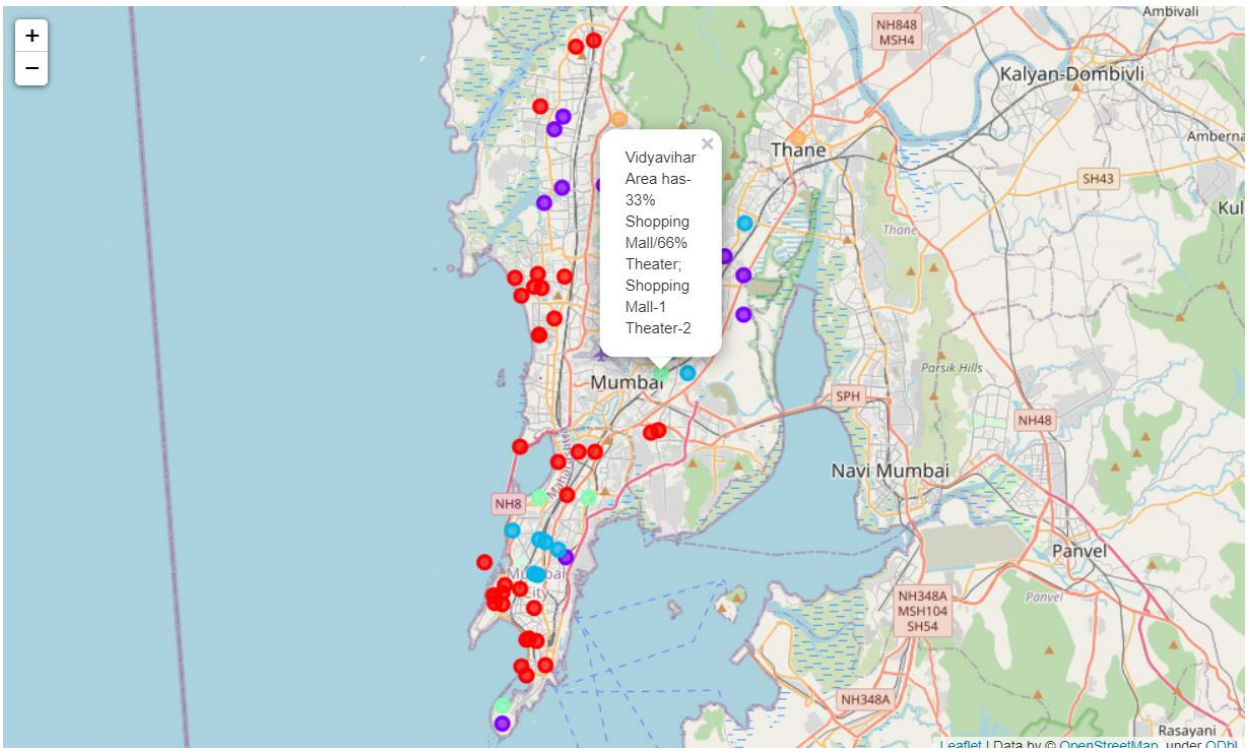
I generated clusters for K=5 through the K-means cluster algorithm, and I labeled each cluster based on the percentage of the shopping malls and theater in each cluster.

Cluster	**Labels**
0	Only Theater
1	Only Shopping Mall
2	66 % Shopping Mall / 33% Theater
3	33 % Shopping Mall / 66% Theater
4	75 % Shopping Mall / 25% Theater

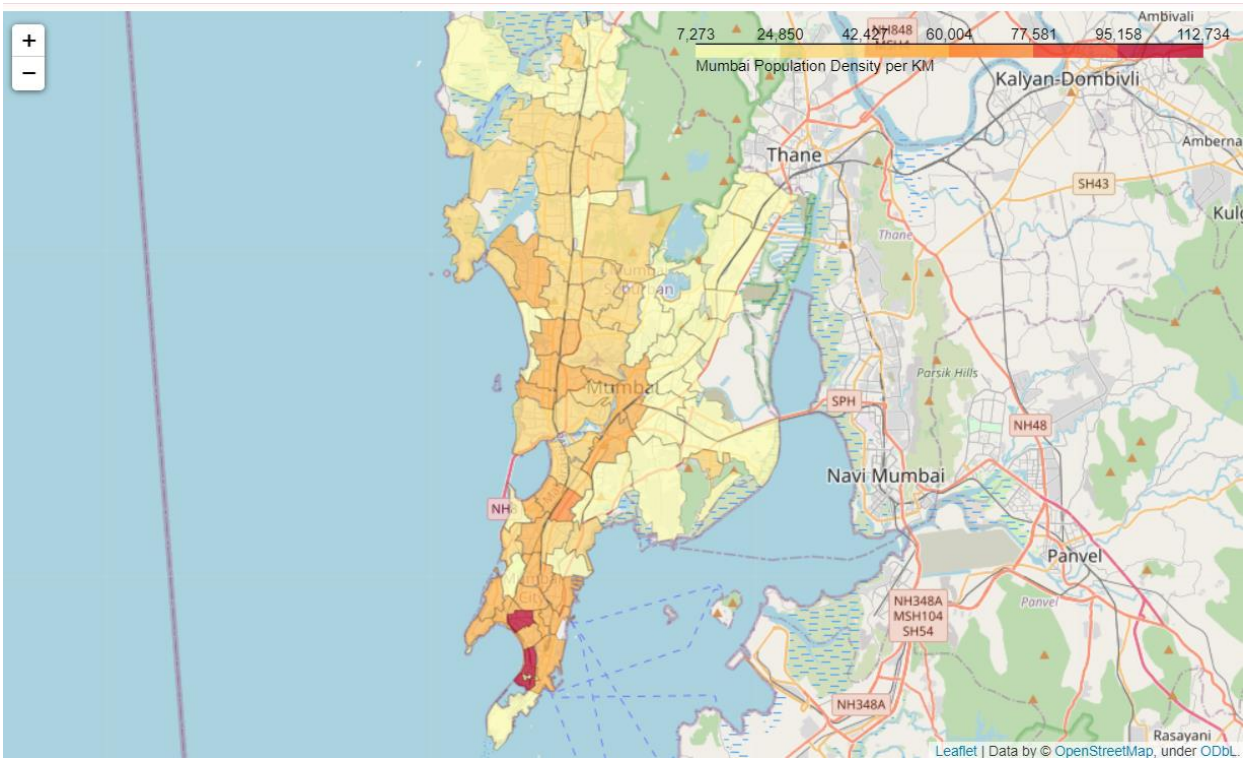
I plotted the clusters on Mumbai's map to analyze these clusters in the Mumbai's neighborhood.



I also added the number of total shopping malls and theaters for each area.



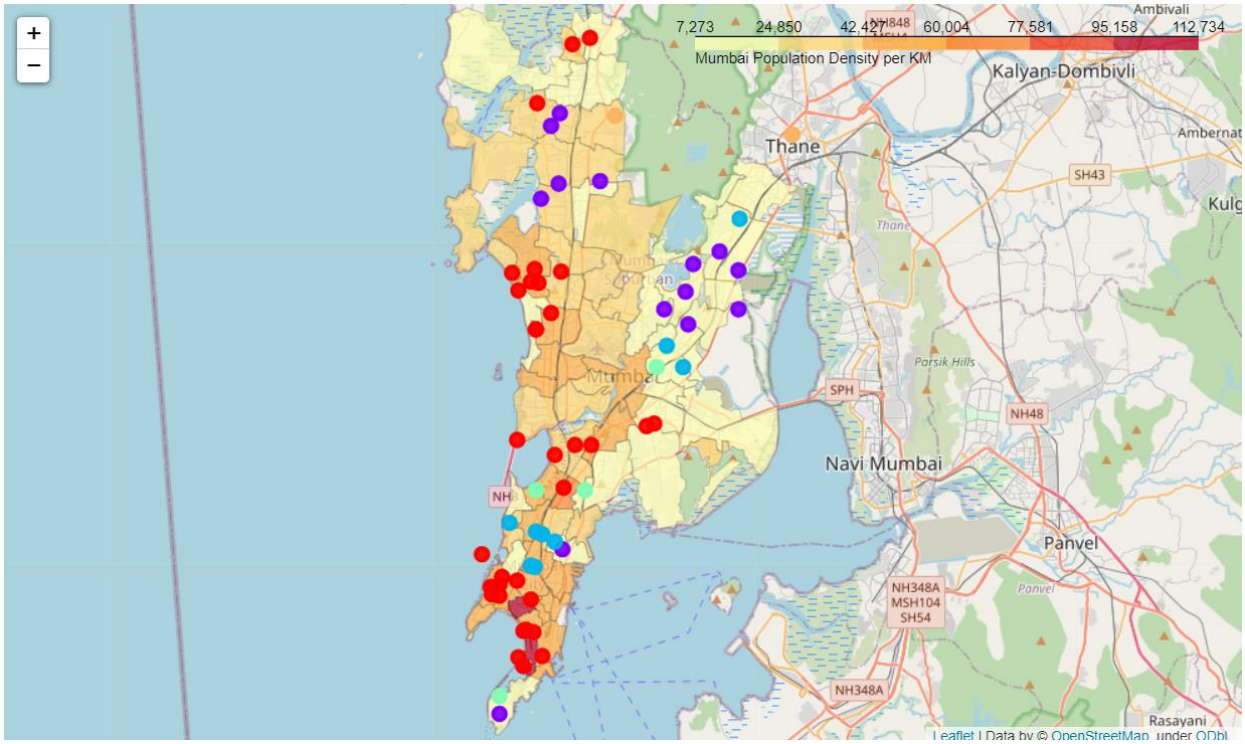
As part of my analysis, I would also like to get information on population density for Mumbai for each neighborhood of Mumbai. For that, firstly, I retrieved Mumbai Geojson data online through a website. Later, I captured density information through Choropleth map. The Mumbai's population density is spread between 7,273/per km and 112,734/ per km.

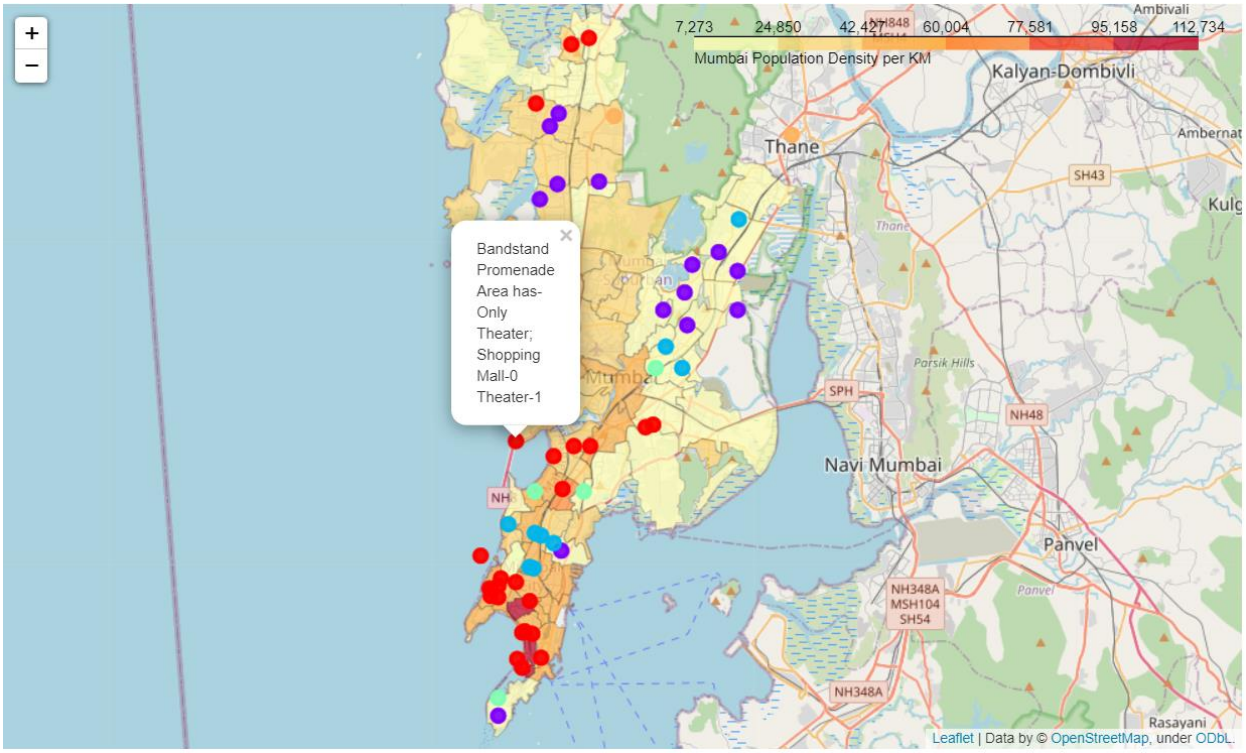


4. Results

Finally, I want to analyze the final result, which has all 5 clusters neighborhood map along with density population map.

I combined above two maps into a single map. Therefore, an investor can look at the information and come up with the best potential location for opening a shopping mall/theater in Mumbai.


7,273
24,850
42,427
60,004
77,581
95,158
112,734
Mumbai Population Density per KM
Kalyan-Dombivli
Thane
Navi Mumbai
Panvel
Parsik Hills
SH43
SPH
NH48
NH348A
MSH104
SH54
Rasayani
Leaflet | Data by © OpenStreetMap, under ODbL


7,273
24,850
42,427
60,004
77,581
95,158
112,734
Mumbai Population Density per KM
Bandstand Promenade Area has- Only Theater; Shopping Mall-0 Theater-1
Kalyan-Dombivli
Thane
Navi Mumbai
Panvel
Parsik Hills
SH43
SPH
NH48
NH348A
MSH104
SH54
Rasayani
Leaflet | Data by © OpenStreetMap, under ODbL.

5. Discussion

After analyzing the final result, we can observe that in the Andheri west areas (Four Bungalows, Seven Bungalows, Versova, Lokhandwala), there are only theaters but not shopping malls. Similarly, we can see that in the Chandivali and Powai areas, there are only shopping malls but not theaters. Also, population density can be considered for these areas to find the optimum locations to build new shopping malls/theaters.

However, this analysis has some limitations. Firstly, I could only extract 100 venue details for each location because Foursquare API only provides 100 venues as free. Secondly, I could not find the complete data for population density for each area. Therefore, I updated density based on the nearby areas' density value.

An optimum result through this analysis can be achieved, if the complete venue details and population density can be retrieved. Also, other parameters can be considered.

6. Conclusion

To conclude, due to people's love for visiting shopping malls/theaters in Mumbai, investors/builders can discover new potentials areas for building shopping malls/theaters within Mumbai through this analysis.

7. References

- **[1]** **Mumbai Neighborhood Details**
- **[2]** **FourSquare API**
- **[3]** **Mumbai GeoJOSN Details**
- **[4]** **Mumbai Area Wise Population Density**