COURSERA CAPSTONE

IBM APPLIED DATA SCIENCE CAPSTONE

OPENING A NEW SHOPPING MALL IN MUMBAI, INDIA

By: Aniruddha Patil

September, 2019

BUSINESS PROBLEM

- The objective of this capstone project is to analyze and select the best locations in the city of Mumbai, India to open a new shopping mall.
- Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question:
- In the city of Mumbai, India, if a property developer or an investor is looking to open a new shopping mall, where would you recommend that they open it?

DATA

To solve the problem, we will need the following data:

- **List of neighborhoods in Mumbai:** This defines the scope of this project which is confined to the city of Mumbai, the financial capital city of the country of India.
- **Latitude and longitude coordinates of those neighborhoods:** This is required in order to plot the map and also to get the venue data.
- **Venue data:** Specifically, the data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

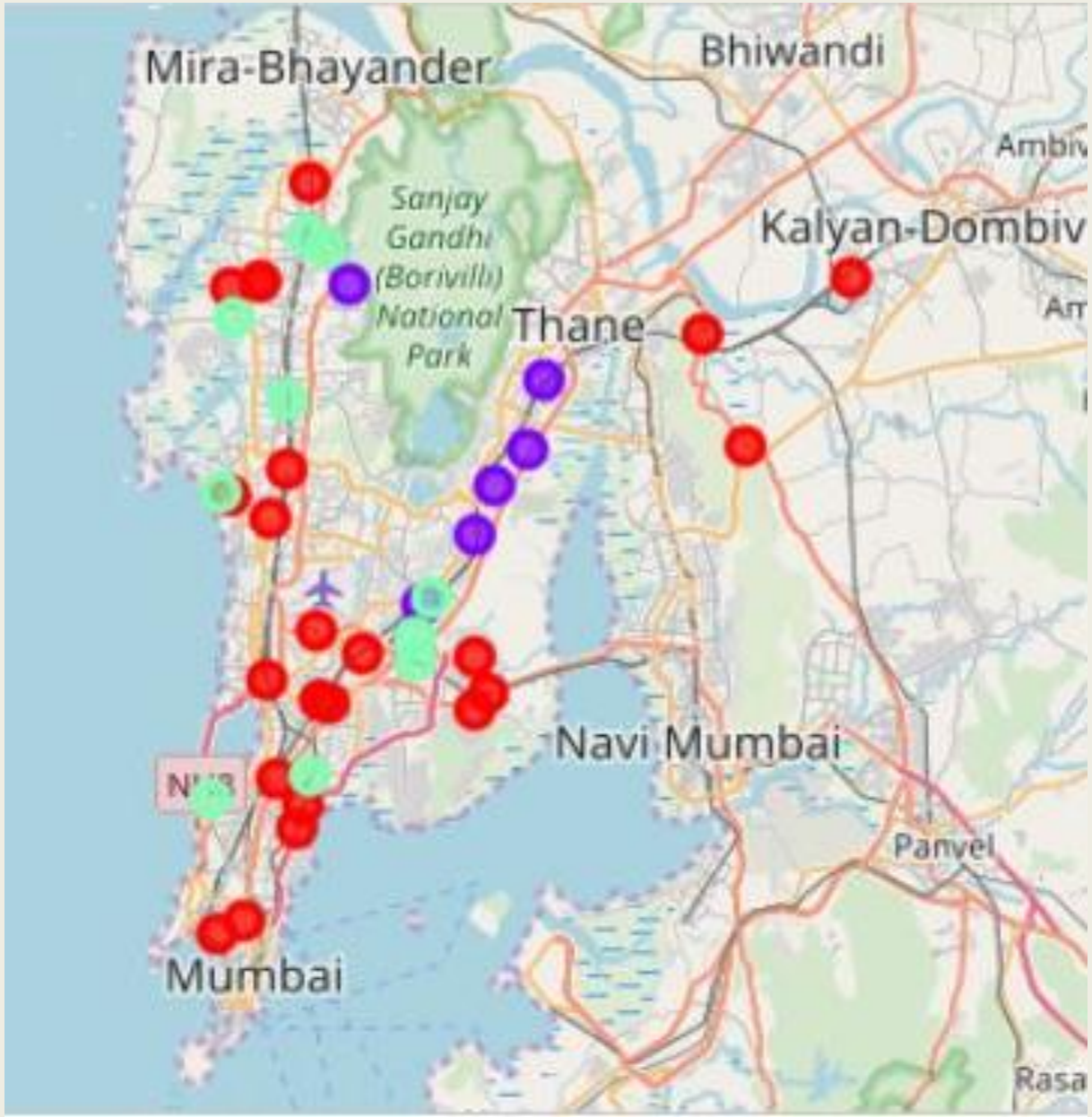
METHODOLOGY

- Web scraping Wikipedia page for neighborhoods list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data
- Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category
- Filter venue category by Shopping Mall
- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

RESULTS

Categorized the neighborhoods into 3 clusters :

- **Cluster 0:** Neighborhood’s with moderate number of shopping malls
- **Cluster 1:** Neighborhood's with low number to no existence of shopping malls
- **Cluster 2:** Neighborhood's with high concentration of shopping malls



DISCUSSION

- Most of the shopping malls are concentrated in the central area of the city.
- Highest number in cluster 2 and moderate number in cluster 0.
- Cluster 1 has very low number to no shopping mall in the neighborhood’s.
- Oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls.

RECOMMENDATIONS

- Open new shopping malls in neighborhoods in cluster 1 with little to no competition
- Can also open in neighborhoods in cluster 0 with moderate competition if have unique selling propositions to stand out from the competition
- Avoid neighborhoods in cluster 2, already high concentration of shopping malls and intense competition

CONCLUSION

- Answer to business question: The neighborhoods in cluster 1 are the most preferred locations to open a new shopping mall
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall