IBM APPLIED DATA SCIENCE CAPSTONE

APPROPRIATE LOCATIONS TO OPEN A NEW SHOPPING MALL IN MUMBAI, INDIA

INTRODUCTION

Shopping malls are a one-stop destination, where shoppers can do various activities, ranging from shopping, eating, gaming and watching movies. It is a great place to visit, and is always buzzing, especially during holidays. For, property developers, it is a great way to make profit out of catering to the demands of the public. However, a lot of thought and consideration goes into opening a shopping mall. One of the most important factors to be considered, is the location.

BUSINESS PROBLEM

The aim of this project is to help property developers in choosing the ideal location for opening a shopping mall, in Mumbai, India, using data science methodology and machine learning techniques like clustering.

TARGET AUDIENCE:

This project aims to help property developers in opening new shopping malls around Mumbai, India. It will help them choose ideal locations minimizing competition and maximizing profit.

DATA AND SOURCES

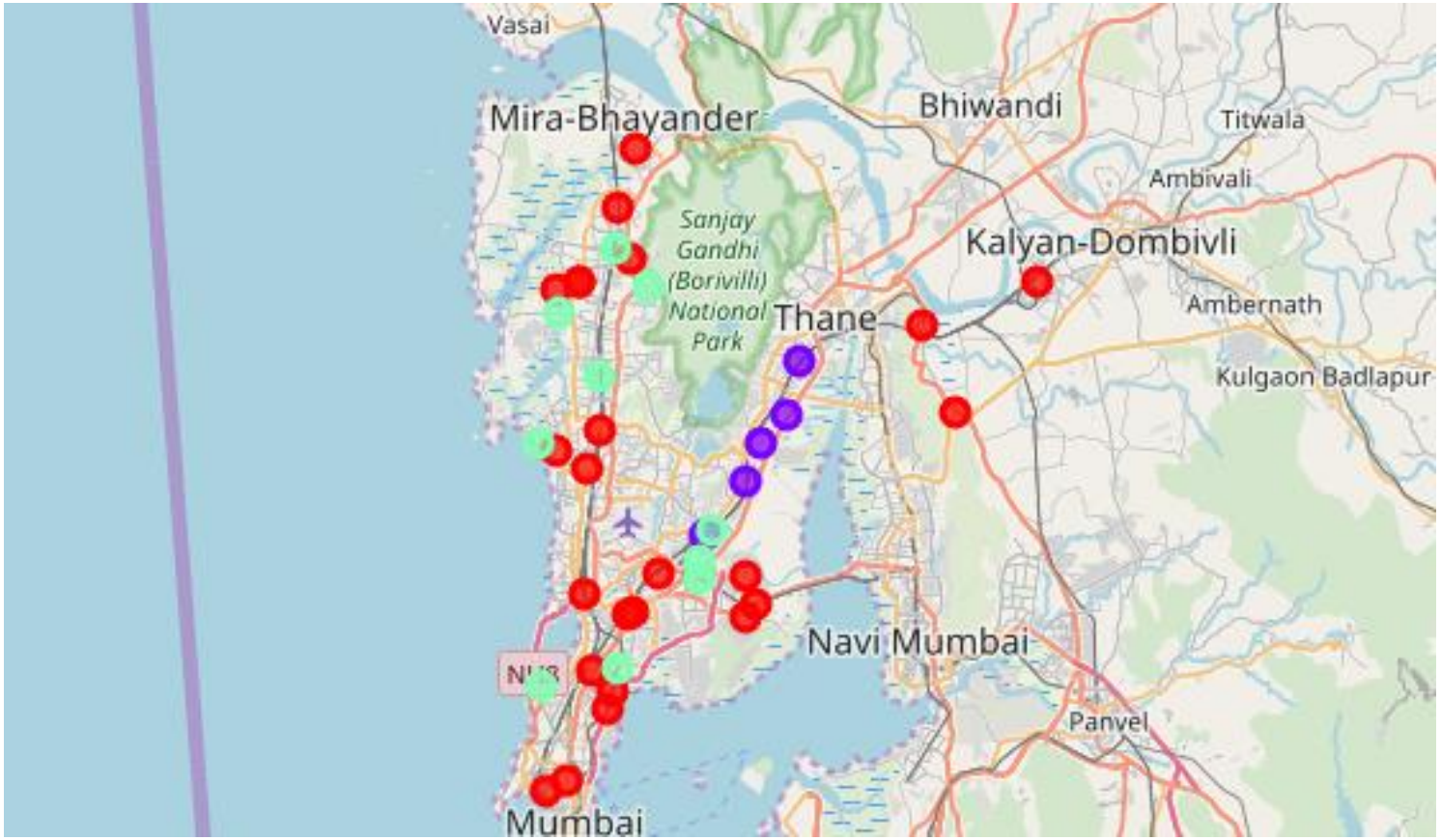
- List of neighborhoods in Mumbai, India
- Latitude and longitude coordinates of those neighborhoods, in order to plot the map
- Data related to shopping malls, in order to perform clustering on the neighborhoods

SOURCES OF DATA:

- Wikipedia page for neighborhoods: https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai
- Geocoder library for coordinates of neighborhoods
- Foursquare API for venue data

METHODOLOGY

- Web scraping Wikipedia page for neighborhoods list
- Using Geocoder to get coordinates of neighborhoods
- Using Foursquare API to get venue data
- Group data by neighbourhood and take the mean frequency of occurrence of each venue category, and filter “Shopping mall”
- Perform k-means clustering on the data
- Visualize the clusters in a map using Folium



RESULTS

Cluster 0 (in red) has no shopping malls.

Cluster 1 (in purple) has the most number of shopping malls.

Cluster 2 (in cyan) has moderate number of shopping malls.

DISCUSSION

Cluster 0 (in red) has no shopping malls. Cluster 1 (in purple) has the most number of shopping malls. Cluster 2 (in cyan) has moderate number of shopping malls.

Therefore, it is advised to open new shopping malls in neighborhoods belonging to cluster 0, where there is no competition. This will also provide the people living in these neighbourhoods with a means of livelihood.

Property developers can consider opening a shopping mall in neighborhoods belonging to cluster 2 if they believe they can stand out from the other malls in the locality and can fight the moderate competition.

Property developers are advised to avoid neighborhoods in cluster 1 which already have high concentration of shopping malls and are suffering from intense competition.

CONCLUSION

We have provided a solution to the business problem by using data science methodology and machine learning techniques. We extracted the data from a webpage and prepared it. We performed machine learning by clustering the data. And we helped relevant stakeholders identify what the best location is to open a shopping mall. From the findings of this project, we advise property developers to open shopping malls in neighborhoods belonging to cluster 0.


THANK YOU