Opening New Shopping mall in Mumbai, India

Narendra U
October, 2020

1. Introduction

1.1 Background

In metro cities in India, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, watch movies, dine at restaurants, shop at the various fashion outlets and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. Due to large crowd gathering in shopping malls, it provides great distribution channel for retailers. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. Opening shopping malls allows property developers to earn consistent rental income. Opening a new shopping mall requires serious consideration and is a lot more complicated as we have to select right location of the shopping mall. Location plays major role in determining success of shopping mall.

1.2 Problem Description

Mumbai is the financial, commercial and entertainment capital of India. Mumbai is composed of a number of neighborhoods spread across a total area of 603 sq km and with total population of 20.5million with 73,000 people per sq mile. This project explores, analyzes existing shopping mall information in Mumbai city in order to select best location to open new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer business problem statement of "In the city of Mumbai, India, if a property developer is looking to open a new shopping mall, which is the right place?"

1.3 Interested audience

The target audience for such a project is useful to property developers and investors looking to open or invest in new shopping malls in Mumbai city. The numbers of shopping malls in Mumbai are currently low when compared with the density of population in Mumbai. This provides an opportunity in creating new malls and this project tries to explore right location for creating new Shopping mall in Mumbai city.

2 Data

2.3 Data Sources

Data requirements:
- List of Neighborhoods in Mumbai
- Latitude & Longitude of these neighborhoods
- Venue data to identify number of shopping malls in each neighborhood

Data sources:
- List of Neighborhoods & related latitude, longitude information in Mumbai
  - Wikipedia page
    https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai,
    contains 93 neighborhoods with related Latitude & Longitude information.
    We will use web scraping techniques to extract the data from this
    Wikipedia page with help of Python requests & Beautifulsoup packages.
- Venue data
  - We will use Foursquare API to get the venue data for those
    neighbourhoods. Foursquare has one of the largest database of 105+
    million places and is used by over 150,000 developers. Foursquare API
    will provide many categories of the venue data, we are particularly
    interested in the Shopping Mall category in order to help us to solve the
    business problem put forward.

References:
- About Mumbai:
  - https://en.wikipedia.org/wiki/Mumbai
- Number of shopping malls in Mumbai:
  - https://en.wikipedia.org/wiki/Category:Shopping_malls_in_Mumbai
  - https://www.ixigo.com/shopping-malls-in-at-around-near-mumbai-lp-1140436#:~:text=248%20shopping%20malls%20in%20Mumbai%20%7C%20shopping%20in%20Mumbai
  - https://list.fandom.com/wiki/List_of_shopping_malls_in_Mumbai
- Mumbai population
  - https://populationstat.com/india/mumbai

3 Methodology and Exploratory Data Analysis

As a first step, we have to get list of Neighborhoods in Mumbai City. I extracted Mumbai Neighborhood list from Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai).

Then we will do web scraping using Python requests and Beautifulsoup packages to extract list of Mumbai neighborhood data. This Wikipedia page contains Mumbai neighborhoods and related latitude & longitude values.

After extracting neighborhood data from Wikipedia page, we will populate the data into Pandas dataframe and then visualize the neighborhoods in a map using Folim package.

Next we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. For this we created Foursquare developer account in order to access the data. We then passed each neighborhood latitude & longitude information in order to extract list of Venues, Venue Latitude, Venue Longitude and Venue category.

Using this dataset, we can check how many venues are returned for each neighborhood. Then we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of frequency of occurrence of each venue category. As we are interested in Shopping mall data, we will filter Venue category as 'Shopping mall'.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Shopping Mall".

As a final step, we will analyze these plots to identify and try to draw conclusions on which neighborhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

4 Results:

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for "Shopping Mall":

- Cluster 0: Neighbourhoods with low number to no existence of shopping malls
- Cluster 1: Neighbourhoods with high concentration of shopping malls
- Cluster 2: Neighbourhoods with moderate number of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in blue colour, and cluster 2 in mint green colour.
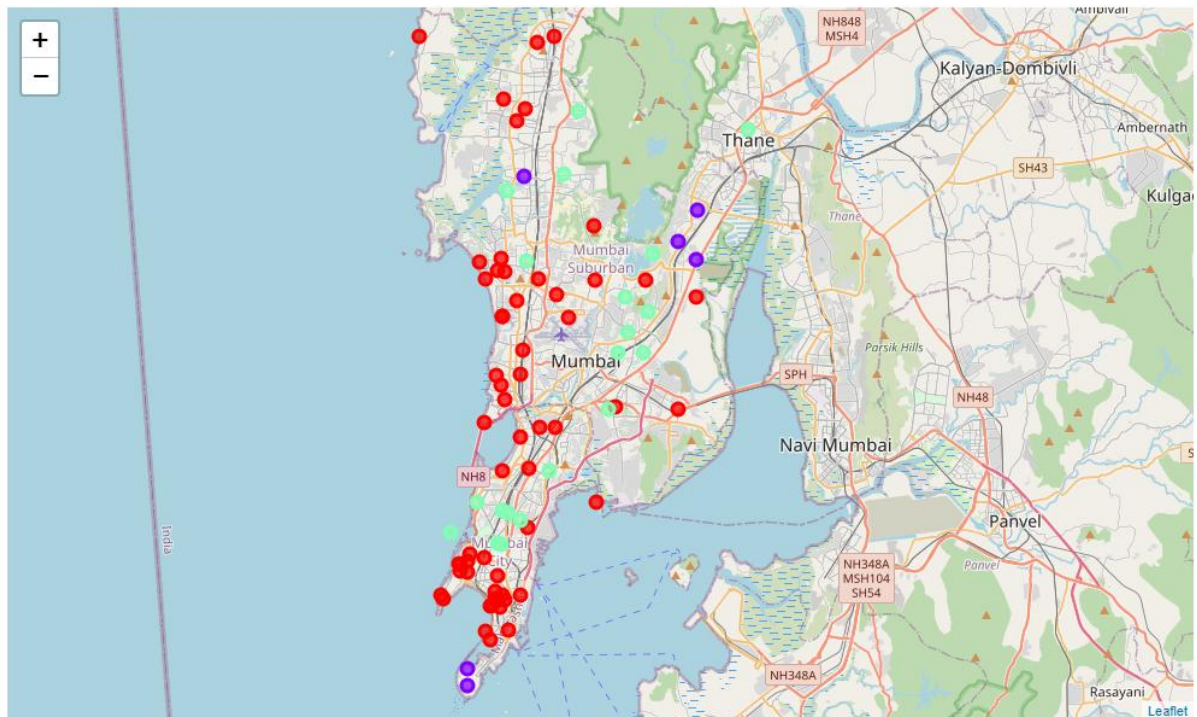
Cluster: 0 (Red)

```
mumbai_merged.loc[mumbai_merged['Cluster Labels'] == 0].count()
```
Neighborhood   65

Cluster: 1 (Blue)

```
mumbai_merged.loc[mumbai_merged['Cluster Labels'] == 1].count()
```
Neighborhood   7

Cluster: 2 (Light green)

```
mumbai_merged.loc[mumbai_merged['Cluster Labels'] == 2].count()
```
Neighborhood   20

5 Discussion:

As per observations noted from the map in the Results section, there are less number of neighbourhoods (7) where we have high concentration of shopping malls as mentioned in Cluster-1. Whereas there are medium number of neighbourhoods (20) having moderate shopping malls in Cluster-2. In majority of the neighbourhoods (65) there are no or minimum shopping malls as mentioned in Cluster-0.

This represents a great opportunity and high potential areas to open new shopping malls in Cluster-0 as there is very little to no competition from existing malls. From another perspective, the results also show there is a scope for creating new shopping mall in Cluster-2 as existing malls may not be able to meet customer demand due to high density of population.

Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 2 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

Limitations:
Please note in this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as income of residents, land cost etc that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project.

6 Conclusion:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is:

"The neighbourhoods in cluster 0 are the most preferred locations to open a new shopping mall".

The 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.