Determining The Best Neighborhoods To Open A New Shopping Mall

COURSERA CAPSTONE PROJECT FOR IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Niladri Deb

March 2019

Business Problem

- Determining the location of the Shopping Malls and the frequency in each neighborhood is very important to know before selecting a neighborhood to build a Shopping Mall.
- The Objective of this project is to determine the best neighborhoods in Mumbai, India, where property developers could build a new Shopping Mall.
- I would say that this is an appropriate project in context of the time as in a recent article in the Economic Times says that 100 new malls are going to be constructed by the end of 2020 in India, and 36 will be built in the western region, which includes the financial capital Mumbai.
- Business Problem: If a property developer is looking to open a mall in Mumbai, India, where would you recommend them to open one?

Data

- The following data was needed for this project:

1. List of Neighborhoods in Mumbai
2. The co-ordinates of the Neighborhoods of Mumbai
3. The data on venues, especially Shopping Malls

- Data Sources:

1. Wikipedia page on the List of Neighborhoods in Mumbai (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai)
2. Geocoder library to retrieve the co-ordinates
3. Foursquare API to extract the data on venues

Methodology

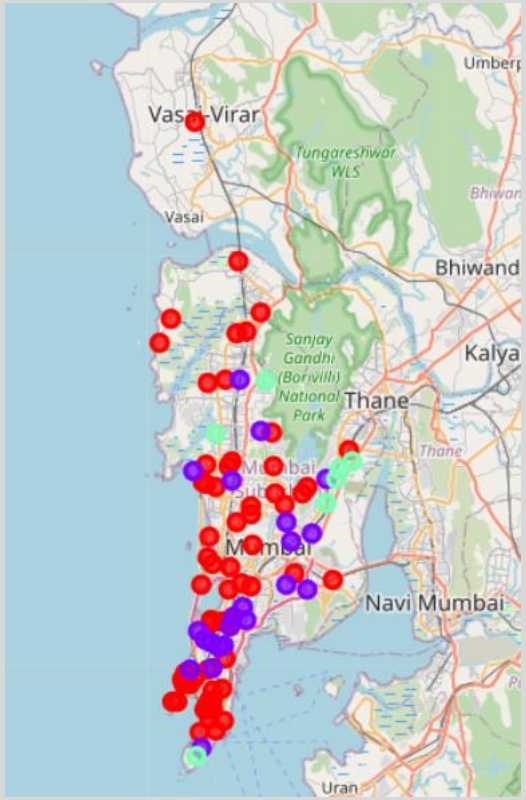
- Extract the data in the form of a table at the end of the Wikipedia page for the List of Neighborhoods in Mumbai and store them in a Data Frame.
- Get the co-ordinates using Geocoder and drop the co-ordinates taken from the Wikipedia page.
- Use the Foursquare API to get the data for the venues in each Neighborhood.
- Group the data according to Neighborhood with the ‘one-hot encoding’ values in a Data Frame and then take the mean of the frequencies of each of the venues’ occurrence.
- Create a Data Frame with the data of Neighborhoods and the mean of the frequencies’ value for Shopping Malls.
- Cluster the data values using ‘k-means clustering’ and add the co-ordinates
- Sort the Data Frame in the order of the Cluster Labels and finally plot them on the Map to analyze the data.

Results

- Cluster 0, which is in red, shows the Neighborhoods with no Shopping Malls.
- Cluster 1, which is in purple, shows the Neighborhoods with a moderate number of Shopping Malls
- Cluster 2, which is in green, shows the Neighborhoods with a large number of Shopping Malls.

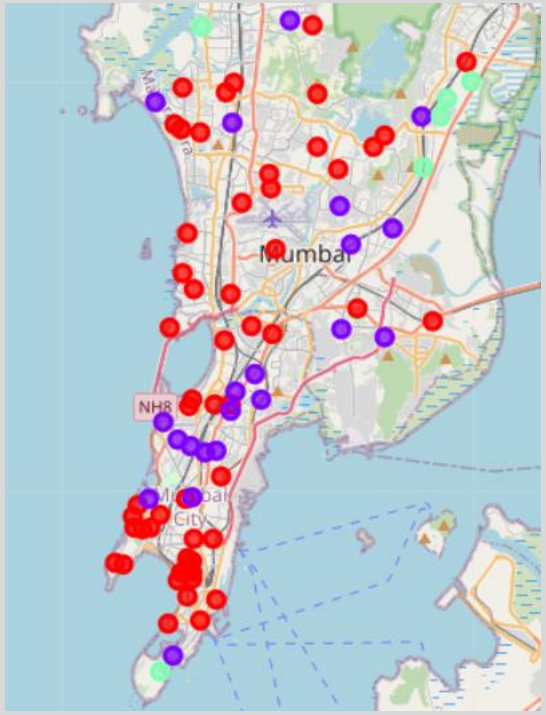
- The next slides contain the pictures of the map. It has been divided into three parts for the convenience of reading it.

The Complete Map



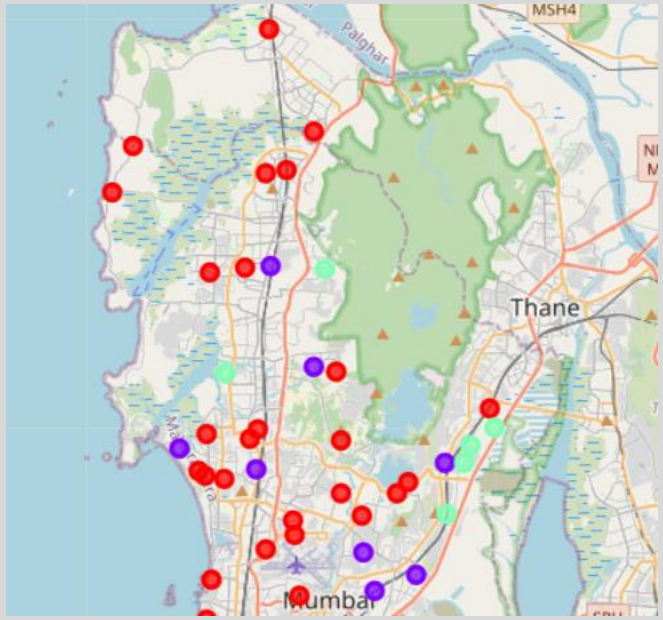
The full Map of Mumbai with the Clusters

First Part of the Map



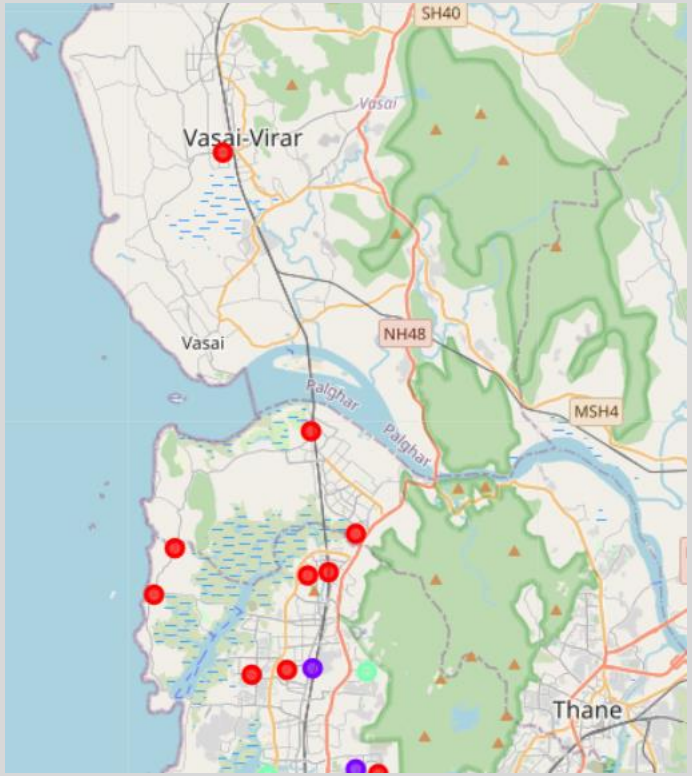
Southern Part of Mumbai with all the clusters

Second Part of Map



A little to the North of Southern Part of Mumbai with all the clusters

Third Part of Map



Northern Part of Mumbai with all the clusters

Discussion

- Neighborhoods with either high number of malls, moderate number of malls or even no malls are scattered across the city and are not specifically clustered in a certain area of the city.
- Highest number of Shopping Malls in Neighborhoods of Cluster 2.
- Moderate number of Shopping Malls in Neighborhoods of Cluster 1.
- No Shopping Malls in Neighborhoods of Cluster 0.
- Neighborhoods in Cluster 0 contain data from a lot of the Southern Parts of Mumbai and a lot from the Northern Parts of Mumbai, and Cluster 1 also shows a similar too. However, there are very few Neighborhoods in Cluster 2 but they are also scattered across the city.

Recommendations

- To open a new Shopping Mall, Neighborhoods in Cluster 0 is highly recommended because it shows the presence of no other Shopping Malls, and hence, no competition.
- Cluster 1 could be recommended as there will be some amount of competition, but some other features and factors will affect the success or failure of building a Shopping Mall in these Neighborhoods.
- It would be recommended to avoid Cluster 2 because of the presence of higher number of Shopping Malls in these Neighborhoods and the competition will be very high in order to succeed or to gain profits.

Conclusion

- To answer the Business Question: Neighborhoods in Cluster 0 will be highly recommended to build a new Shopping Mall
- The relevant stakeholders, in this case property developers, will find the results of this project useful for determining Neighborhoods for building new Shopping Malls in order to avoid (Tough) Competition and higher chances of succeeding depending on the location and frequency of the already built Shopping Malls in the Neighborhoods.

THANK YOU!