

Coursera Capstone Project: Applied Data Science

Opening a new Shopping Mall Kolkata, West Bengal, India

BY: SUMAN KUMAR MAL
APRIL 2020

Business Problem

- Since, shopping malls are main interest to various groups therefore they are built and real estate investors invest in these projects. But for shopping malls to attract large crowd there are few major factors, one of those is location. This will specially benefit the real estate builders since Indian retail sector has metamorphosed significantly over last few decades. Rapid urbanization and digitization, rising disposable incomes and lifestyle changes of particularly the middle-class has led to a major revolution in the retail sector, projected to grow from US \$672 billion in 2017 to US \$1.3 trillion in 2020. Evolving rapidly from usual 'kirana shops' to large multi-format stores offering global experience to the e-commerce model that is highly technology-driven, the Indian retail sector has evolved.

Data

Data Required

- List of neighbourhoods in Kolkata. This defines the scope of this project which is confined to Kolkata, the capital city of the state West Bengal, India.
- Latitude and Longitude coordinates of the neighbourhoods. This required to plot the map and get the venue data
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Data Sources

- The data of the neighbourhoods in Kolkata can be extracted from Wikipedia page. (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Kolkata)
- Then the latitude and longitude data can be retrieved from Python geocoder package.
- Then using latitude and longitude data venues can be fetched from Foursquare API.

Methodology

- Web Scraping Wikipedia page for neighbourhood list.
- Get latitude and longitude using Geocoder.
- Use Foursquare API to get venue data.
- Group data based on the frequency of occurrence of each venue.
- Filter the data for Shopping Malls.
- Perform Clustering.
- Visualize the clusters in a map using Folium.

Results

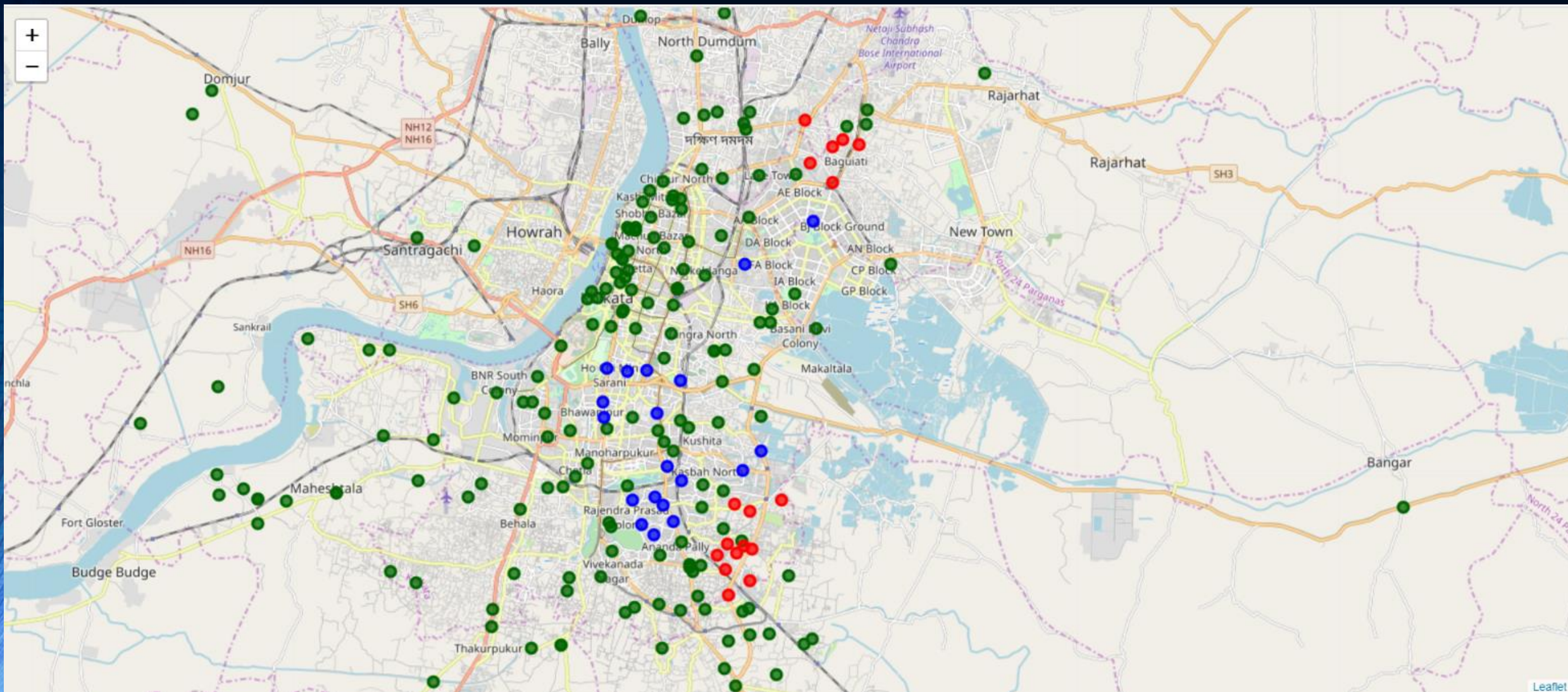
- Results from K-means Clustering show that:

Cluster 0: Neighbourhoods with low to no existence of shopping malls.

Cluster 1: Neighbourhoods with high concentration of shopping malls.

Cluster 2: Neighbourhoods with moderate number of shopping malls.

- Colour Code
- Cluster 0 – Dark Green
- Cluster 1 – Red
- Cluster 2 – Blue



Discussions and Recommendations

- Most shopping malls in the central east, southern and few on the northern part of Kolkata, with the highest number in cluster 1 and moderate number in cluster 2. On the other hand, cluster 0 has very low number to totally no shopping mall in the neighbourhoods
- Property developers are advised to avoid neighbourhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

Conclusion

In this project, we have gone through all the data science methodology. We first identified a business problem, then collected the required data, preparing the data and then performing machine learning by clustering the data into 3 clusters based on their similarities. Finally, we have also provided recommendations to the relevant stakeholders regarding the best locations to open a new shopping mall.

To answer the problem which was raised at the Introduction: The neighbourhoods in cluster 0 are most preferred location to open a new shopping mall.



THANK YOU