IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

APPLIED DATA SCIENCE CAPSTONE PROJECT

Opening a New Shopping Mall in Bangalore, India

By: -

J PRINCE KELVIN

12-Jun-2021

Introduction

Shopping malls are the place where people tend to go very frequently in a city like Bangalore. Mostly Working individuals will go to malls during weekends and holidays; So, visiting shopping malls has become a great way to get relaxed and enjoy themselves.

Shopping malls are like a one-stop destination for all types of shoppers, for example a customer can do Grocery shopping, watch movies and shows, can dine at a restaurant, shop at various fashion outlets and can perform many more activities.

When it comes to retailers, the Central Location which attracts lots of customers at malls provides a great distribution channel to market their product and services. Property Developers and Real Estate sectors are also taking advantage of this trend to build more shopping malls.

As a result, there are many shopping malls in the city of Bangalore any many more are being built.

Opening a new shopping mall in a city like Bangalore is really a Challenging task, it requires serious considerations and its lot more complicated than it seems. Particularly, finding a good location for the shopping mall is the one of the most important decisions that will determine whether the mall will be a success or failure.

This project will be particularly useful to property developers and investors who are looking to open or invest in new shopping mall in the Silicon Valley of India i.e., Bangalore.

Business Problem

The main objective of this capstone project is to analyze and select the best locations in the city of Bangalore, India to open a new shopping mall. This project aims to provide solutions to answer the Business Question – In Bangalore, if a property developer is looking to open a new shopping mall, where would you recommend them to open?

Description of the Data

To solve our business problem, we need the following data:

- Bangalore Pincode with coordinates.
  - Data source:
    https://www.kaggle.com/hegdetapan/bangaloreareaspincodewithlatitudelongitude
  - Description: Contains a list of pin codes, area names, and latitude and longitude coordinates.
- Shopping Malls in Neighborhoods of Bangalore.
  - Data source: Foursquare API
  - Description: By using this API we will get all the venues in each postal code. We can get filter these venues to get data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Using this data will allow exploration and examination to answer the questions. This is a project that will make use of many data science skills, from Data Collection, working with Foursquare (API), Data cleaning, Data wrangling and map visualization using Folium and to machine learning (K-means clustering).

Methodology

Data Exploration –

First, we need to get the list of neighborhoods in Bangalore, India. Fortunately, we have a dataset containing all the data needed on Kaggle Website (Source Link given above). We need to download that Dataset. However, this dataset contains area name(neighborhood), postal code, latitude & longitude coordinates.

Data Geocoding-

Even though our Dataset contains the geographical coordinate values, we will use geocoder to generate and get accurate Coordinate values of our Neighborhoods (Area). We need to get the geographical coordinates in the form of latitude and longitude in order to use Foursquare API.

Data Visualization-

Visualize the neighborhoods in a map using Folium package. This allows us to perform a check to make sure that the geographical coordinate's data returned by Geocoder are correctly plotted in the city of Bangalore.

Using Foursquare API-

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 3 Kilometers. We need to create a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category.

Data Wrangling-

By doing so, we are also preparing the data for use in clustering. Since we are analyzing the "Shopping Mall" data, we will filter the "Shopping Mall" as venue category for the neighborhoods.

Data Clustering –

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 neighborhoods into **5 clusters** based on their frequency of occurrence for '**Shopping Mall**'.

The results will allow us to identify which neighborhoods have higher concentration of shopping malls while which neighborhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new shopping malls.

Results

The results from the k-means clustering shows that we can categorize the neighborhoods into 5 clusters based on the frequency of occurrence for "Shopping Mall":

- Cluster 1 & 3: - Neighborhoods has high concentration of shopping malls.
- Cluster 2 & 4: - Neighborhoods has low number to no existence of shopping malls.
- Cluster 5: - Has Moderate number of shopping malls.

The results of the clustering are visualized in the map below
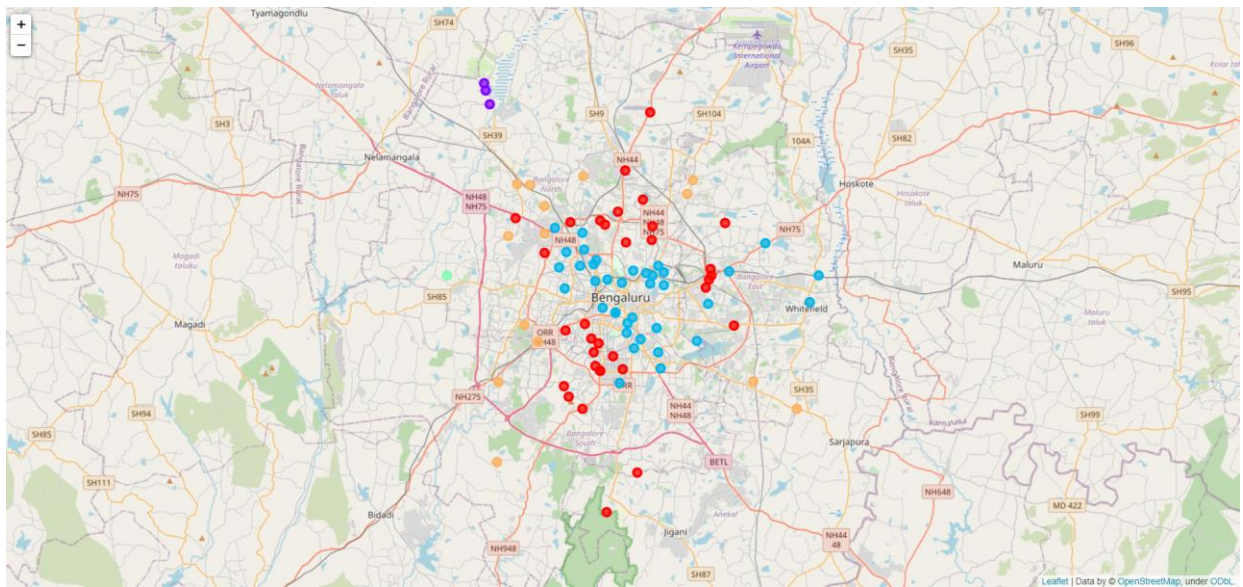
Clusters	Color
Cluster 1	Red
Cluster 2	Purple
Cluster 3	Sky Blue
Cluster 4	Green
Cluster 5	Orange

Discussion

By observing the map in the Results section, we can notice that, most of the shopping malls are concentrated in the central area of Bangalore city, with the highest number of malls in Cluster 1 and 3 (more than 30 malls) and moderate number of malls in Cluster 5(which is around 14 malls). On the other hand, Cluster 2 and 4 has very low number of malls. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in Cluster 1 and 3 are suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, this also shows that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls.

Therefore, this project recommends Property Developers & Investors to capitalize on these findings to open new shopping malls in neighborhoods in Cluster 2 and 4 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in Cluster 5 with moderate competition. Lastly, Property Developers & Investors are advised to avoid neighborhoods in Cluster 1 and 3 which already have high concentration of shopping malls and suffering from intense competition.

Limitations and Suggestions for future Research

In this project, we only consider one factor; that is, frequency of occurrence of shopping malls, there are many other factors such as population, income of residents, and road connectivity are all those that could influence the decision on location of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighborhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall.

In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned.

Future research could make use of paid account to bypass these limitations and obtain more results.

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 5 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 neighborhoods in cluster 2 & 4 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.

References

Foursquare Developers Documentation. Foursquare. Retrieved from https://developer.foursquare.com/docs

Appendix

Original Data Set

	Area	Pincode	Latitude	Longitude
0	BangaloreG.P.O.	560001	12.887035	77.74105
1	BangaloreCityH.O.	560002	12.966180	77.58690
2	Malleswaram	560003	13.006320	77.56840
3	Vyalikaval	560003	13.037920	77.61863
4	Basavangudi	560004	12.941120	77.57236

Bangalore Map
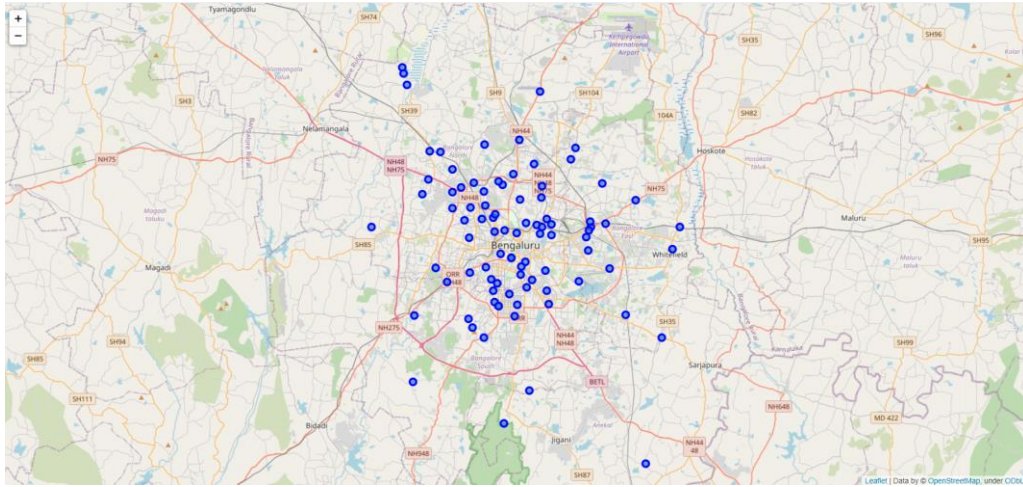
Clusters

Cluster 1

Total number of Shopping malls is = 33

	Neighborhood	Shopping Mall	Cluster Labels	Pincode	Latitude	Longitude
44	JayanagarSouth	0.010000	0	560041	12.930140	77.584960
25	Dooravaninagar	0.022222	0	560016	12.997445	77.668300
28	Gandhinagar	0.000000	0	560009	13.026430	77.617510
29	GavipuramExt.	0.000000	0	560019	12.945280	77.566340
30	GokulaExtension	0.019608	0	560054	13.041120	77.548290
33	HebbalAgri.Farm	0.000000	0	560024	13.049810	77.589030
37	HospitalTownWest	0.023810	0	560015	12.905700	77.542720
43	JayanagarEast	0.010000	0	560069	12.919330	77.592690
87	Yelhanka	0.000000	0	560064	13.084356	77.594929
23	Doddakallasandra	0.000000	0	560062	12.886610	77.558720
53	Mahadevapura	0.000000	0	560048	12.994090	77.666330
61	Nagasandra	0.100000	0	560073	13.044870	77.501510
62	Nagashettyhalli	0.000000	0	560094	13.042500	77.573960
63	NandhiniLayout	0.025641	0	560096	13.015749	77.526730
67	RTNagar	0.000000	0	560032	13.024450	77.595830

Cluster 2 - (Lesser Concentration of shopping malls)

Total number of Shopping malls is = 3

	Neighborhood	Shopping Mall	Cluster Labels	Pincode	Latitude	Longitude
34	Hesaraghatta	0.0	1	560088	13.13928	77.47976
35	HesaraghattaLake	0.0	1	560089	13.15029	77.47639
36	HospitalTownEast	0.0	1	560014	13.15665	77.47514

Cluster 3

Total number of Shopping malls is = 38

	Neighborhood	Shopping Mall	Cluster Labels	Pincode	Latitude	Longitude
69	RichmondTown	0.010000	2	560025	12.962330	77.601230
68	Rajajinagar	0.020000	2	560010	13.005440	77.556930
22	DharmaramCollege	0.010000	2	560029	12.936750	77.602640
8	BangaloreUniversity	0.030000	2	560056	12.996940	77.618430
59	Mathikere	0.010000	2	560054	13.032340	77.558640
58	Maruthisevanagar	0.034091	2	560033	13.000130	77.627790
6	BangaloreCityH.O.	0.020000	2	560002	12.966180	77.586900
65	NewTippasandra	0.010000	2	560075	12.973600	77.665650
74	St.ThomasTown	0.030000	2	560084	13.005020	77.623250
71	Shanthinagar	0.010000	2	560027	12.957560	77.596930
72	Srirampuram	0.020000	2	560021	12.992590	77.569470
73	St.Johns	0.030000	2	560034	12.990285	77.616107

Cluster 4- (Lesser Concentration of shopping malls)

Total number of Shopping malls is = 1

	Neighborhood	Shopping Mall	Cluster Labels	Pincode	Latitude	Longitude
52	MagadiRoad	0.0	3	560023	12.996875	77.443065

Cluster 5

Total number of Shopping malls is = 14

	Neighborhood	Shopping Mall	Cluster Labels	Pincode	Latitude	Longitude
38	Indiranagar	0.125000	4	560038	13.030060	77.495260
20	Chikkabanavara	0.000000	4	560090	13.072135	77.513750
64	Nayandahalli	0.000000	4	560039	12.942050	77.521000
49	Kothanur	0.000000	4	560077	13.064560	77.647820
7	BangaloreG.P.O.	0.000000	4	560001	12.887035	77.741050
46	Kengeri	0.000000	4	560060	12.908680	77.487180
80	Vijayanagar	0.000000	4	560040	13.076000	77.652400
17	Carmelram	0.000000	4	560035	12.909638	77.704379
41	Jalahalli	0.166667	4	560013	13.054500	77.526580
60	Nagarbhavi	0.000000	4	560072	12.956240	77.509360
1	Agaram	0.000000	4	560007	12.842920	77.485820
51	Kumbalgodu	0.000000	4	560074	13.072610	77.503210
66	PeenyaS.I.	0.111111	4	560058	13.031880	77.526540
79	Vidyaranyapura	0.000000	4	560097	13.079290	77.559120