COURSERA – IBM

APPLIED DATA SCIENCE CAPSTONE

PREDICTING A SUITABLE LOCATION TO OPEN A SHOPPING MALL

IN BAGALORE, KARNATAKA

By: ANANDAKRISHNAN
Date: 18th APRIL 2020

INTRODUCTION

Shopping malls are very common and needy thing now days like a city were IT persons are living.

If they are coming to a Shopping Mall they get everything from Groceries, Household items, they can watch movies, Restaurants and many more.
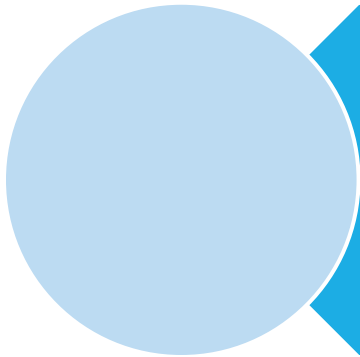
Here we are going to predict where to open a new shopping mall in Bangalore where over sampling has happened in many places

You will come to now in detail as you go forward

BUSINESS PROBLEM
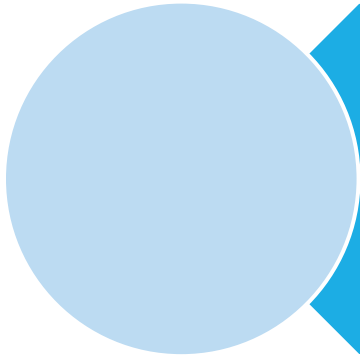
The objective of this capstone project is to analyse and select the best locations in the city of BANGALORE, KARANATAKA to open a new shopping mall.

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question.

BUSINESS QUESTION

In the city of BANGALORE, KARANATAKA if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

TARGET AUDIENCE OF THIS PROJECT

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the IT HUB of INDIA, i.e, BANGALORE KARNATAKA. This project is timely as the city is currently suffering from oversupply of shopping malls.

DATA

To solve the problem, we will need the following data:

- List of neighborhoods in Bangalore. This defines the scope of this project which is confined to the city of Bangalore.

- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

METHODOLOGY

▶ Web Scraping Wikipedia page for neighborhood list

▶ Get latitude and longitude coordinates using Geocoder

▶ Using Foursquare API to get venue data

▶ Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category

▶ Filter venue category by shopping malls

▶ Perform clustering on the data frame using k-means clustering

▶ Visualize the clusters in a map using Folium

DATA FRAME CREATED FROM DATA IN WIKIPEDIA

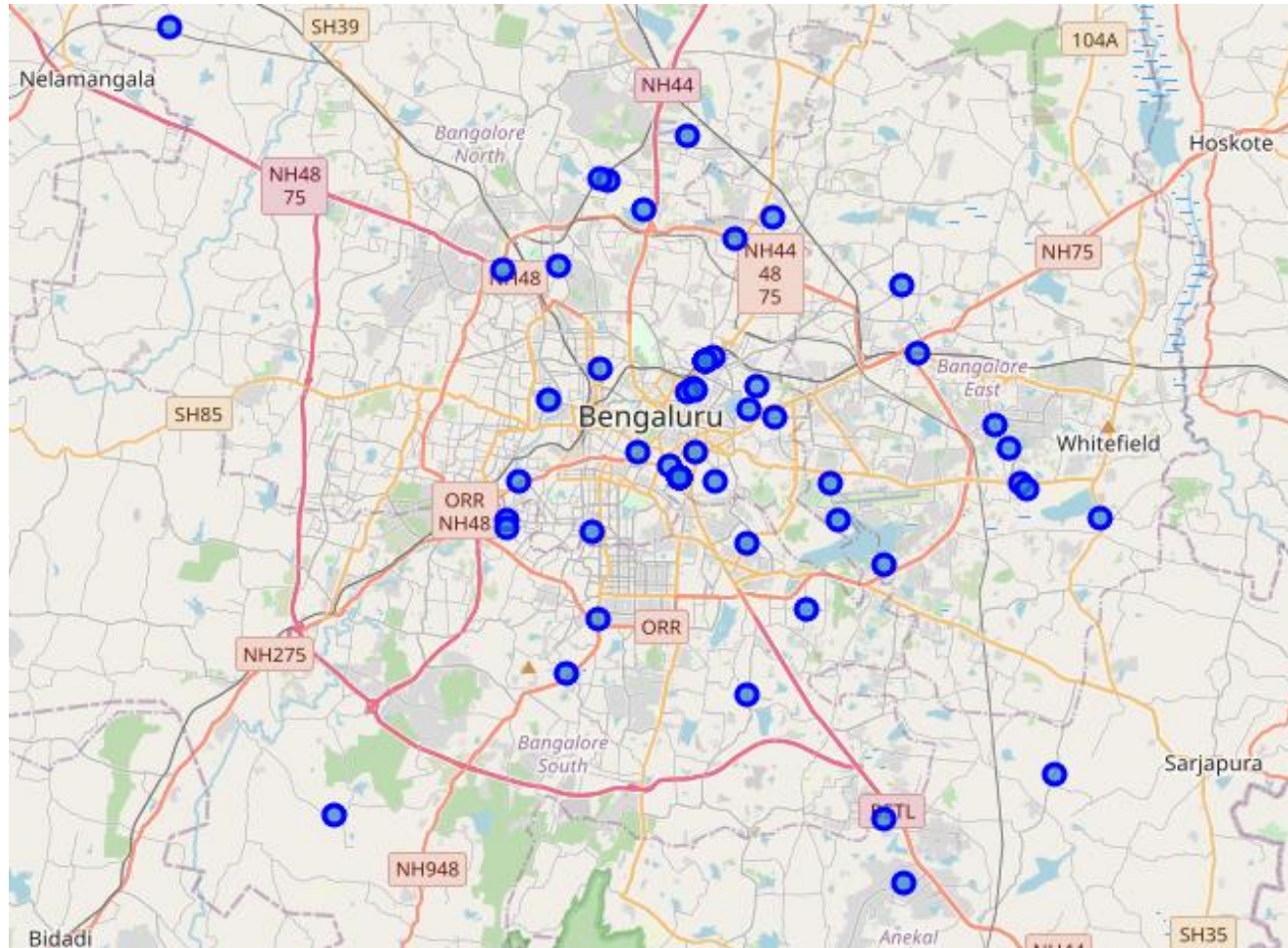|	Neighborhood
0	← Agara, Bangalore (2 C, 6 F)
1	← Arekere (5 F)
2	← Banashankari (1 C, 4 F)
3	← Banaswadi (2 F)
4	← Basavanagudi (5 C, 11 F)

FINDING THE GEOGRAPHICAL COORDINATES

```
]:  [[12.841269941511115, 77.4815905253185],
     [12.997973409660101, 77.61047545115127],
     [12.90876015015293, 77.57288203719202],
     [12.997820409660104, 77.61038642561601],
     [12.938980000000072, 77.57137000000006],
     [12.882450000000063, 77.62475000000006],
     [12.927350000000047, 77.67185000000006],
     [12.981176497085812, 77.62506386955391],
     [12.966180000000065, 77.58690000000007],
     [12.958034857060447, 77.6009343090484],
     [12.817530000000033, 77.67879000000005],
     [12.966235877087087, 77.6067910877087],
     [12.855496201771626, 77.73111547336825],
     [13.250110000000063, 77.70788000000005],
     [12.997856409660098, 77.61040795115127],
     [12.943290000000047, 77.65602000000007],
     [12.839884329698537, 77.67221065989045],
     [12.998940000000061, 77.61276000000004],
     [12.942790000000059, 77.54122000000007],
     [13.047455598389988, 77.63327823980035],
     [13.049810000000036, 77.58903000000004],
     [12.957454844018699, 77.60090848251485],
     [12.912160000000029, 77.64490000000006],
     [12.978220000000022, 77.63397000000003]
```

ADDING LATITUDE & LONGITUDE TO DATAFRAME

	Neighborhood	Latitude	Longitude
0	► Agara, Bangalore (2 C, 6 F)	12.841270	77.481591
1	► Arekere (5 F)	12.997973	77.610475
2	► Banashankari (1 C, 4 F)	12.908760	77.572882
3	► Banaswadi (2 F)	12.997820	77.610386
4	► Basavanagudi (5 C, 11 F)	12.938980	77.571370

CREATED A MAP OF BANGALORE WITH NEIGHBORHOODS SUPERIMPOSED ON TOP

AFTER USING FOURSQUARE API WE HAVE MODIFIED THE DATAFRAME

:	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	➤ Agara, Bangalore (2 C, 6 F)	12.841270	77.481591	Art of Living International Center	12.844607	77.507343	Spiritual Center
1	➤ Agara, Bangalore (2 C, 6 F)	12.841270	77.481591	rachenamadu	12.850793	77.505317	Nature Preserve
2	➤ Agara, Bangalore (2 C, 6 F)	12.841270	77.481591	sнιvα sαι ιη∂υsтяy αη∂ тяα∂εяs	12.816927	77.491872	Outdoor Supply Store
3	➤ Arekere (5 F)	12.997973	77.610475	Mangalore Pearl - Seafood Restaurant	12.994472	77.615551	Seafood Restaurant
4	➤ Arekere (5 F)	12.997973	77.610475	Naturals Icecream	12.996912	77.611268	Ice Cream Shop

EXAMINING CLUSTER 0

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
29	► Kettohalli (1 C)	0.010000	0	12.957455	77.600908
27	► Jayanagar, Bangalore (1 C, 7 F)	0.010000	0	12.961606	77.597723
57	► Yeshwantpur (1 C, 6 F)	0.018519	0	13.029540	77.540220
31	► Konanakunte (1 F)	0.018868	0	12.890437	77.561763
35	► Madiwala (1 C, 6 F)	0.020000	0	12.956603	77.613550
36	► Magadi (2 C, 10 F)	0.020000	0	12.988227	77.605822
37	► Mahadevapura (2 C)	0.010000	0	12.958035	77.600934
38	► Majestic (Bangalore) (1 C)	0.010000	0	12.957455	77.600908
39	► Malleswaram (4 C, 2 F)	0.020000	0	12.995000	77.573460
40	► Marathahalli (8 C, 1 P, 30 F)	0.020619	0	12.955740	77.719419
25	► J. P. Nagar (2 C)	0.010000	0	12.958035	77.600934
41	► Mathikere (1 C)	0.010000	0	13.030327	77.559672

EXAMINING CLUSTER 1

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
48	► Seetharampalya (1 C, 14 F)	0.0	1	13.113200	77.424630
46	► Ramamurthy Nagar (1 C, 20 F)	0.0	1	13.023820	77.677850
45	► Rajarajeshwari Nagar, Bangalore (1 C, 4 F)	0.0	1	12.940380	77.541539
10	► Bommasandra (33 F)	0.0	1	12.817530	77.678790
12	► Chandapura (4 F)	0.0	1	12.855496	77.731115
13	► Devanahalli (5 C, 2 F)	0.0	1	13.250110	77.707880
3	► Banashankari (1 C, 4 F)	0.0	1	12.908760	77.572882
0	► Agara, Bangalore (2 C, 6 F)	0.0	1	12.841270	77.481591
16	► Electronics City (2 C, 34 F)	0.0	1	12.839884	77.672211
32	► Koramangala (1 C, 12 F)	0.0	1	12.935130	77.624450
18	► Girinagar (1 C, 11 F)	0.0	1	12.942790	77.541220
30	► Kodihalli, Bangalore (1 C, 4 F)	0.0	1	13.059765	77.576729

EXAMINING CLUSTER 2

|	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
1	► Arekere (5 F)	0.030000	2	12.997973	77.610475
55	► Whitefield, Bangalore (4 C, 1 P, 22 F)	0.050000	2	12.975970	77.710317
24	► Ittamadu (3 F)	0.030000	2	12.997856	77.610408
14	► Dhobi Ghat (Bangalore) (3 F)	0.030000	2	12.997856	77.610408
17	► Fraser Town, Bangalore (1 C, 10 F)	0.030000	2	12.998940	77.612760
34	► Kundalahalli (96 F)	0.030000	2	12.967520	77.715000
33	► Krishnarajapura (3 C, 3 F)	0.030769	2	13.000390	77.683680
28	► Jeevanbheemanagar (3 F)	0.030000	2	12.997856	77.610408
4	► Banaswadi (2 F)	0.030000	2	12.997820	77.610386
11	► Brigade Road, Bangalore (3 C, 8 F)	0.030000	2	12.966236	77.606791

18th APRIL 2020

RESULT

Categorize the neighborhoods into 3 clusters:

Cluster 0: Neighborhoods with moderate number of shopping malls

Cluster 1: Neighborhoods with low number to no existence of malls

Cluster 2: Neighborhoods with high concentration of shopping malls
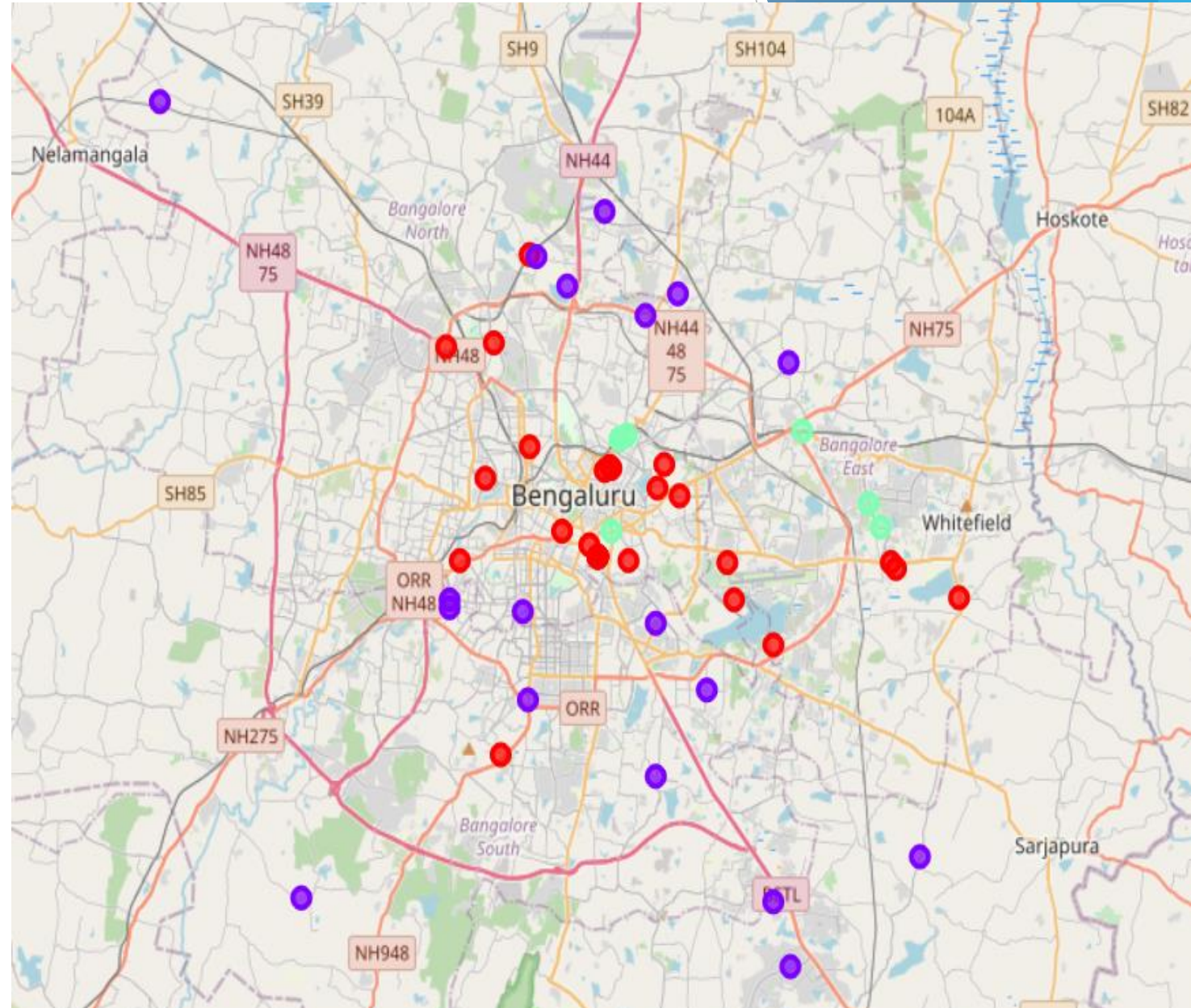
LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

- We only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall.

- 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.

- 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

▶ 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

▶ To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 1 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.