

Capstone Project – The Battle of Neighborhoods

TITLE: ANALYSIS OF SCHOOLS IN BANGALORE

INTRODUCTION

- In this project I have tried to analyze the distribution of schools in Bangalore. It will help us to identify the areas where the infrastructure, in terms of school, needs to be developed.
- It is essential for every child to have quality education. Bangalore is one the IT-hub of India, which means more and more people are migrating to Bangalore to have a better job opportunity. An influx in migration will in turn lead people to search for a neighborhood which have good schooling access.

BACKGROUND

- We would like to look at how schools are clustered in the neighborhood of Bangalore .
- Long term aspect of the project demonstrates that of identifying those regions which lacks infrastructure in terms of schools .
- While, in the short run we would like to know if a family reallocates to Bangalore then what are the likely neighborhoods with most number of schools.

DATA

- Using Bangalore's Neighborhood Wikipedia page I have scraped data for different localities of Bangalore via beautifulsoup4 package.
- Next, by using geocoder I have obtained the Latitudes and Longitudes of these neighborhoods.
- Lastly, FourSquare API has been used to obtain the location of various schools in each neighborhood setting the limit to 100 for each neighborhood and a radius of 1km.
- *Data Link:*
"https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore"

DATA PREPROCESSING

- Unwanted entries such as 'driving schools', 'schools of music', 'schools of engineering' etc. were removed .
- Duplicate data because was dropped out because a school was within 1000 meters distance from each neighborhood may appear twice with the same latitude and longitude.
- Several schools whose category were mismatched ,are grouped into four broad spectrums as -high schools, primary or play schools, student centers or not mentioned category.

First 10 elements of the processed DataFrame

	Neighbourhood	Latitude	Longitude	School	School_Latitude	School_Longitude	School_category
0	Adugodi	12.94402	77.60800	Mary Immaculate School	12.945718	77.599380	High School
1	Adugodi	12.94402	77.60800	Indian Retail School	12.942211	77.608254	Not Mentioned
2	Adugodi	12.94402	77.60800	Dairy Colony Govt. School	12.938816	77.606491	Not Mentioned
3	Arekere	12.88568	77.59668	AECS Magnolia Maaruti Public School	12.885073	77.596503	Not Mentioned
4	Arekere	12.88568	77.59668	Aradhana School	12.884844	77.592087	Primary/Play
5	Arekere	12.88568	77.59668	Heritage Academy School	12.879765	77.594310	Primary/Play
6	Arekere	12.88568	77.59668	BGS National Public School	12.877473	77.599268	Not Mentioned
7	Arekere	12.88568	77.59668	nakshatra pre school	12.879890	77.603584	Primary/Play
8	Arekere	12.88568	77.59668	Cambridge School	12.886161	77.585502	Not Mentioned
9	Arekere	12.88568	77.59668	Brigade School @ JP Nagar	12.891879	77.585289	High School

Each row corresponds to a particular school and its information, i.e. neighborhood it belongs to, its category, name of the school and its location. There are a total of 389 entries in the DataFrame.

LIBRARIES

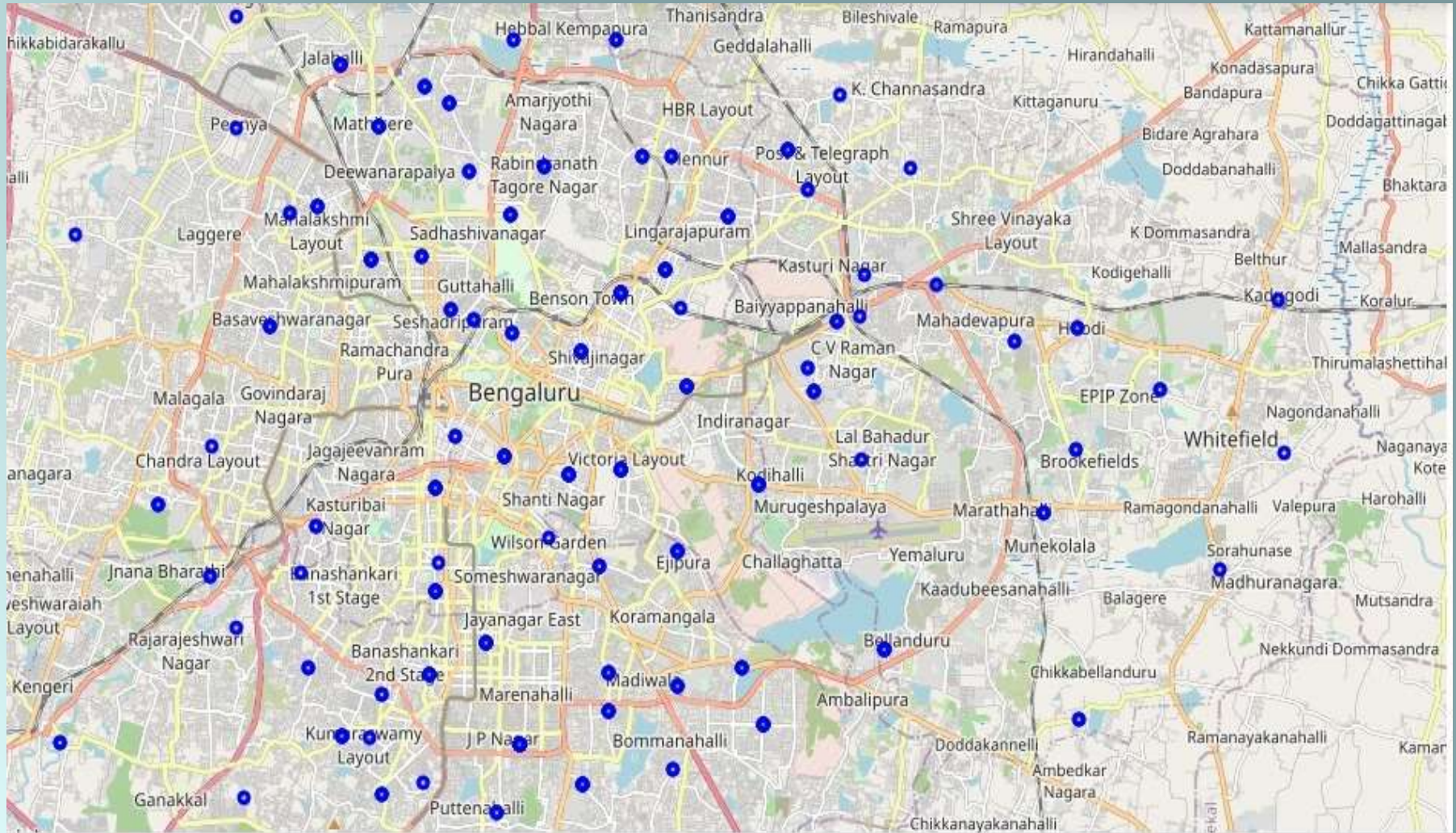
- ***Pandas***: For creating and manipulating dataframes.
- ***Folium***: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- ***Scikit-Learn***: For importing k-means clustering.
- ***JSON***: Library to handle JSON files.
- ***XML***: To separate data from presentation and XML stores data in plain text format.
- ***Geocoder***: To retrieve Location Data.
- ***Beautiful-Soup and Requests***: To scrap and library to handle http requests.
- ***Matplotlib***: Python Plotting Module.

METHODOLOGY

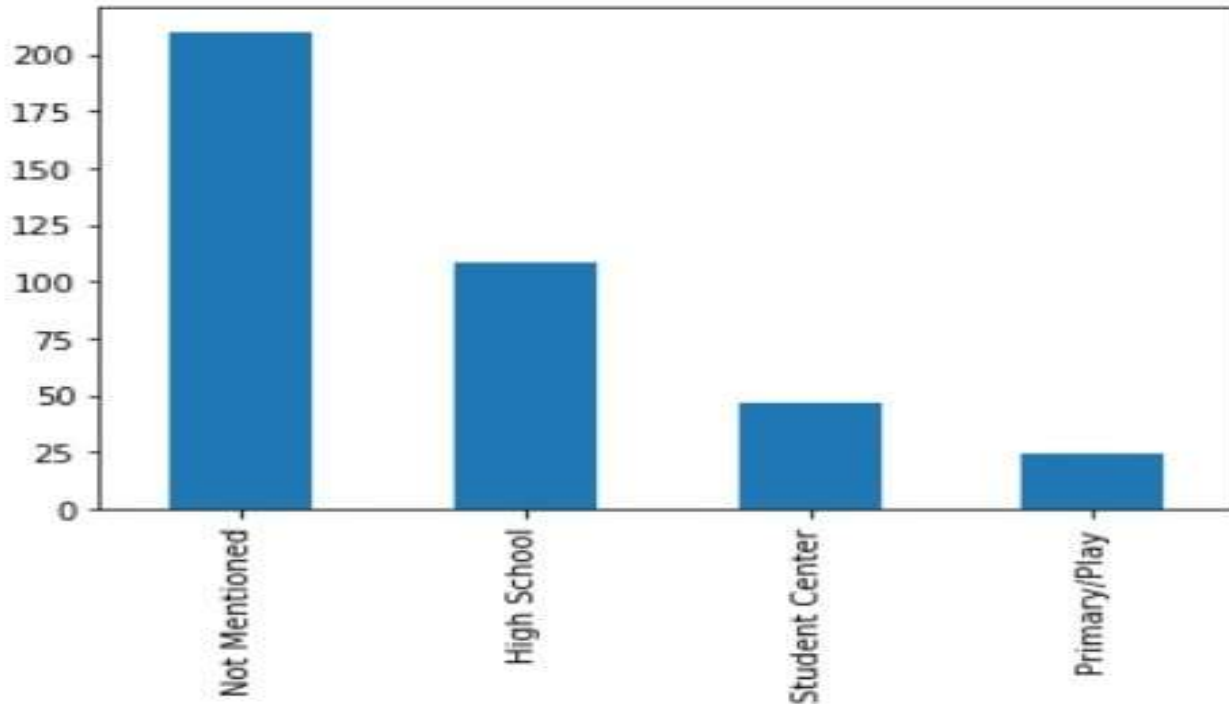
- Followed by data pre-processing , schools are explored within a radius of 1km for each neighborhood with a limit set to 100.
- Next, we analyze the data by exploring the various categories to which the schools belong.
- For those neighborhoods where there are no schools we find the distance to the closest school by incrementing the radius by 100 meters in each iteration until a school is found using Folium.

- On obtaining this refined data, consisting of the distance to nearest school for each neighborhood and the number of schools in each neighborhood we cluster this data using K-Means.
- This helps us to identify those regions where there is a need for improvement and those neighborhoods where the influx of migrant can be reallocated.

EXPLORATORY DATA ANALYSIS



- The figure, in the previous slide shows distribution of schools in Bangalore which have been obtained using Folium.
- It can be observed that the schools are clustered more in the centre of the city especially near areas such as Kormangala, Wilson Garden .
- We can also see that Whitefield which is a prominent destination for IT-hubs lacks infrastructure in terms of schools.
- Towards the outskirts of the city, the density of schools decreases.



- For the ease of analysis I have broadly categorized schools into four groups namely High School, Primary, Student Centre and Not-Mentioned.
- Most of the schools have not mentioned their category because a lot of these schools provide education from kindergarten level to high school.
- The number of High Schools are quite more in number than in case of Primary Schools.

CLUSTER ANALYSIS

- I have clustered the neighborhoods using K-Means clustering algorithm into 3 clusters based on distance to nearest school and number of schools in each neighborhood.
- Regions having no schools within the radius of 1km, are identified the distance to nearest schools using Folium by incrementing the radius in steps by 100 meters until 2km

First five elements showing the nearest distance

	Neighbourhood	Latitude	Longitude	Near_school
0	Agara, Bangalore	12.84283	77.48759	2000
1	Ananthnagar	12.95408	77.54135	1100
2	Anjanapura	12.85811	77.55909	1300
3	Bommasandra	12.81753	77.67879	1100
4	Devara Jeevanahalli	13.01444	77.59951	1100

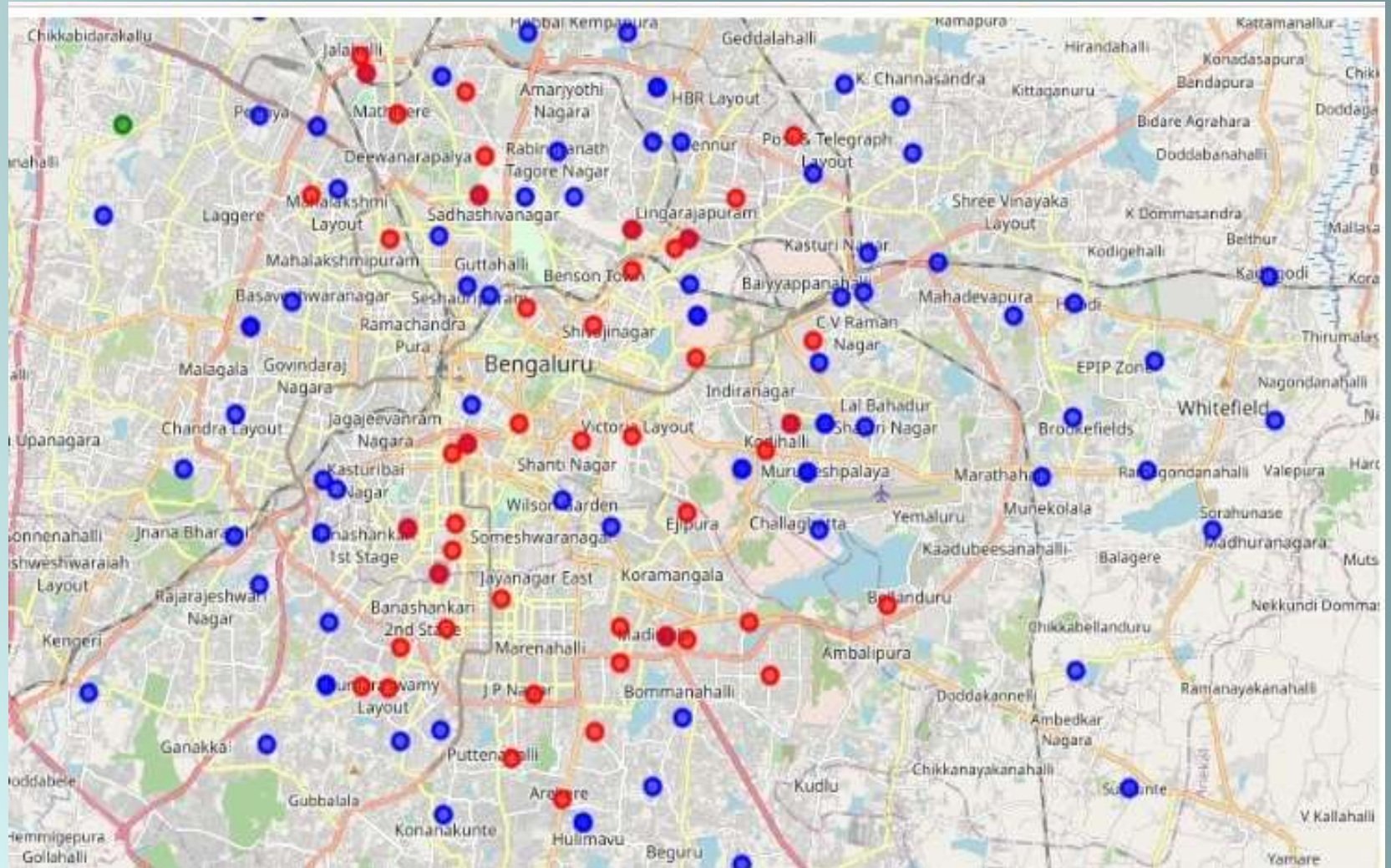
- The number of schools are counted in each neighborhood and a data frame is created based on this information.
- In this method I have taken nearest distance to school to vary from 1km to 2km.

Number of schools within 1km radius under each cluster

	Latitude	Longitude	Near_school	num_schools_in_1km
label				
0	12.975056	77.610372	1034.862385	1.871560
1	12.936445	77.491425	1750.000000	0.000000
2	12.964732	77.597203	1000.000000	8.978723

The average of the distance to nearest school and the number of schools in each neighborhood has been obtained when grouped by the label assigned in the clustering algorithm.

Distribution of schools in different clusters



Here, the blue circles correspond to label 0, the green ones to 1 and the red ones to 2.

RESULT

- Among the schools that have been observed in Bangalore, most of them are standard schools which provide education from kindergarten level itself till 12th grade.
- Observing the distribution of schools ,we find that most of them are concentrated towards the center of the city and are less dense in the outskirts.
- Number of High schools exceeds the Primary schools.

- From cluster analysis, we find that the that the neighborhoods labeled 2 perform best with about an average of 9 schools in 1km radius.
- In comparison the neighborhood 1 has about an average of 2 schools within 1km radius.
- The neighborhoods labeled 1 are the ones that require improvements as the average distance to nearest schools is close to 2km with no schools within 1km of the neighborhood.

CONCLUSION

- The study concludes, Bangalore lacks Primary schools which is the first building block of education. Considering the mass of young work-force government along with private organization must build schools near to the neighborhood of major IT-hubs.
- Secondly, people migrating to Bangalore should try to reallocate primarily in the label 2 cluster as shown in the analysis which provides good schooling facilities. The neighborhoods near the vicinity of label 1 clusters require massive help from government to lay the educational infrastructure.