Exploring Hospitals in the neighborhoods of Bangalore City.

1. Introduction

Bangalore, also known as Bengaluru (Kannada) is the capital of the Indian State of Karnataka. Bangalore is nicknamed the Garden City and was once called a Pensioner's Paradise. Located on the Deccan Plateau in the south-eastern part of Karnataka, Bangalore is India's third most populous city.

Today as a large city and growing metropolis, Bangalore is home to many of the most well-recognized colleges and research institutions in India. Numerous public sector heavy industries, software companies, aerospace, telecommunications, and defense organizations are located in the city. Bangalore is known as the Silicon Valley of India because of its position as the nation's leading IT exporter. A demographically diverse city, Bangalore is a major economic and cultural hub and the fastest growing major metropolis in India.

2. Problem Description

Bangalore has many hospitals. The main objective of the problem is to explore the neighborhoods of Bangalore and find the number of hospitals in each neighborhood need to make a data analysis of the number of hospitals in each neighborhood using suitable clustering algorithm.

The idea here is to recognize those areas in Bangalore having minimum number of hospitals and guide the stake holder in constructing the hospitals in these areas.

2.1. Target Audience

This analysis is very much helpful to the business people and government who can use this analysis as the basis for identifying the places having minimum number of hospitals and plan for construction.

2.2. Dataset:

I am using Bangalore neighborhoods dataset downloaded from Kaggle 'https://www.kaggle.com/rmenon1998/bangalore-neighborhoods' which has the location co-ordinates of each region of Bangalore.

3. Methodology:

- We need to first collect the data corresponding to all areas in Bangalore which will have location co-ordinates.
- Next using this data ,we have to use the Foursquare API to explore these neighborhoods and visualize them on a map.
- Further hospital statistics need to be computed in each neighborhood.

- We have to cluster the neighborhoods based on the number of hospitals.

- The information in these clusters will guide the stake holder to detect the optimal areas for construction of hospitals.

4.1 Data Preprocessing:

In this step I first remove the unwanted columns from the dataframe.
Secondly I also remove duplicate redundant data if present.
The dataset which I have selected does not have missing values.
Hence this step is not carried.

Next in order to carry Location based activities such as checking the neighbourhood information and filtering the information based on the category 'Hospitals' in Bangalore and plotting maps, I have used the very popular FourSquare API.

4.2 Overview of FourSquare API:

Foursquare is a technology company that built a massive dataset of location data. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others, and is currently being used by over 100,000 developers, and this number is only growing.

Communicating with the Foursquare database is really very easy, all thanks to their RESTful API. You simply create a uniform resource identifier, or URI, and you append it with extra parameters depending on the data that you are seeking from the database. Any call request you make is composed of, we can call this base URI, which is api.foursquare.com/v2, and you can request data about venues, users, or tips. But, every time you make a call request, you have to pass your developer account credentials, which are your Client ID and Client Secret as well as what is called the version of the API, which is simply a date. It is designed to give developers the freedom to adapt to Foursquare API changes on their own schedule.

I have used the Folium Library to plot maps.

4.3 Exploratory Data Analysis:

Here we try to group our data based on the neighbourhood column and get the count of hospitals in that column

```
count_hosp=pd.DataFrame(hospital_df.groupby(['Neighborhood'],as_index=False).count())

count_hosp.head()

#count_hosp.columns
```

	Neighborhood	ID	Name	Latitude	Longitude
0	Adugodi	4	4	4	4
1	Agram	24	24	24	24
2	Amruthahalli	3	3	3	3
3	Anekal	1	1	1	1
4	Banaswadi	13	13	13	13

Further I have also used one hot encoding to transform categorical values which helps in clustering and plotting.

```
hosp_onehot['Neighborhood'] = hospital_df['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [hosp_onehot.columns[-1]] + list(hosp_onehot.columns[:-1])
hosp_onehot = hosp_onehot[fixed_columns]

hosp_onehot.head()
```

	Neighborhood	4th floor BGS hospital	A. V. Multispeciality Hospital	ACTS College	AVD 53 NÄL	AVD 64 NÄL	Aarthi Scans	Acura Speciality Hospital	Aditya nethralaya	Akshaya Nethralaya	...	pobbati maternity home	pristine consultation and diagnostics	ra matern hospi
0	Agram	0	0	0	0	0	0	0	0	0	...	0	0	
1	Agram	0	0	0	0	0	0	0	0	0	...	0	0	

4.4 K -means Clustering Based on groups

Here I am doing 2 types of analysis.

1. First type of clustering is creating cluster groups containing nearby neighborhoods(**Places closer to each other)** with their hospital information. This is mainly to see the area statistics with hospital information
2. Second type of clustering is grouping the neighborhoods based on the count of hospitals.
   a) Areas with very less number of Hospitals are put in1 cluster.
   b) Areas with maximum number of hospitals are put in another cluster and so on

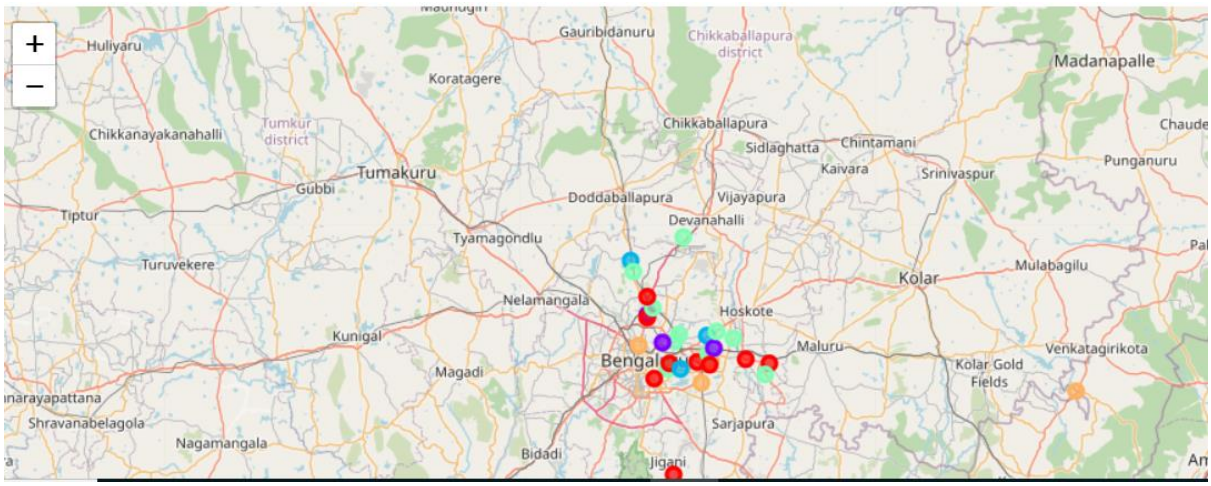
4.5 Overview of K-means Clustering Algorithm

K-means clustering algorithm computes the centroids and iterates until we it finds optimal centroid. ... In this algorithm, the data points are assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroid would be minimum.

5. Results:

I have tried to do the analysis by defining K=5 (5 clusters) and group the neighbourhoods based on the count of hospitals.

Map showing Clustering using K-Means algorithm.



Cluster 1 Results

	Neighborhood	Longitude
0	Adugodi	4
7	Byatarayanapura	6
8	Chickpet	5
9	Chikkalasandra	6
11	Doddakallasandra	5
19	Hosur	6
27	Konanakunte	4
29	Kundalahalli	4
32	Mallathahalli	4
33	Mathikere	5
37	Nagarbhavi	4
44	Vimanapura	5
45	Yelachenahalli	6
46	Yelahanka	4

Cluster 2 Results

	Neighborhood	Longitude
1	Agram	24
22	Indiranagar S.O (Bangalore)	24
24	Jayanagar H.O	25
28	Koramangala	19

Cluster 3 Results

	Neighborhood	Longitude
5	Basaveshwaranagar	9
15	Girinagar S.O (Bangalore)	10
35	Msrit	9
41	Sadashivanagar	8
43	Vijayanagar S.O (Bangalore)	8

	Neighborhood	Longitude
2	Amruthahalli	3
3	Anekal	1
6	Bhattarahalli	1
10	Deepanjalinagar	1
12	Doddanekkundi	1
14	EPIP	1
16	Gottigere	1
17	Hessarghatta	2
18	Horamavu	2
20	Hunasamaranahalli	1
21	Huskur	1
23	Jalahalli H.O	2
25	Kamakshipalya	1
30	Laggere	1
31	Lingarajapuram	2
38	Nayandahalli	1
39	Peenva Dasarahalli	1

Cluster 4 Results

	Neighborhood	Longitude
4	Banaswadi	13
13	Domlur	13
26	Kathriguppe	12
34	Mavalli	11
36	NAL	12

Cluster 5 Results

6. Discussions:

It is very much clear from the observations that cluster4 has the least number of hospitals and these are the areas for which hospital construction recommendations can be given.
Further for better analysis we can consider the population in each area and accordingly do the analysis based on population and hospitals.
Further hospital size, number of beds,ICU count and other features can be explored for better analytical results and providing more valuable insights.

7. Conclusion:

This is just a pilot study focusing on the need for hospital construction in areas of Bangalore using the very popular FourSquare API.