

Segmentation and Clustering of Medical Facilities in the Greater Accra Region of Ghana

Edward Deladem Kobla Lampoh

February 10, 2020

1. INTRODUCTION

1.1 Background:

The Greater Accra region of Ghana in spite of having the smallest land mass among the other 16 regions in this West African country, is the 2nd most densely populated region in the country[1]. The most populated region being the Ashanti Region. It is no surprise that the statistics are this way since this rather small region is also the largest commercial region of the country and home to the capital city, Accra. It is worth noting that in the wake of recent developments in the continent of Africa, Ghana has been elected the commercial capital at the center of the Africa Continental Free Trade Agreement.

1.2 Problem:

More commercial activity would mean more offices and more people engaging this commercial terrain. This presents the challenge of increasing the already high traffic congestion across the region. High traffic congestion in cities stresses the populace as they spend longer time on the road commuting from one location to the other. Proximity to vital locations such as medical facilities need to be looked at, so that traffic situations do not delay access to such facilities in case of emergencies. It's also worth noting that the people's choice of medical facility to visit when the need arises is not only influenced by proximity. This project seeks to explore and provide insight about the distribution of medical facilities across the region showing the areas that need medical facilities and the kind that the people will find most useful.

1.3 Interest:

The Ministry of Health, the Metropolitan Assemblies as well as Real Estate developers will find this insight useful and inform their decision in addressing the medical needs of the people in the Greater Accra. Entrepreneurs who want to set private hospitals in the region would be better informed as the best locations to set up their facilities which would meet the needs of the people.

2. DATA ACQUISITION AND CLEANING

2.1 Data sources

The list of the various districts and their capitals within the Greater Accra Region were obtained from [here](#). The list of hospitals and medical facilities that can be found within a 500 meter radius along with their location data and user rating stats were obtained using Foursquare API.

2.2 Data Cleaning

The data on the medical facilities were obtained in json format using the API. There was very little challenge with the data that was obtained as desired features were readily available except for the opening hours of these facilities. The other issue that had to be dealt with was the fact that some of the data obtained from the api were not medical facilities, that is outliers which do not belong to the target of the call I made to the api, however these were very few and were dropped from the data. I defined a function that made a count of the total number of hospitals found within 500 meters radius of each district capital and assigned the total number to each respective district. It became apparent that the distribution of medical facilities was uneven which is exploratory data analysis session. It is worth noting that though the data source for the list of districts and their capitals also included the coordinates of each district capital which could be scrapped, it was easier and more direct to pull that information using the API.

A total number of 29 district capitals exist across the greater Accra region with a total number of 517 medical facilities.

Th tables below are samples of the data collected:

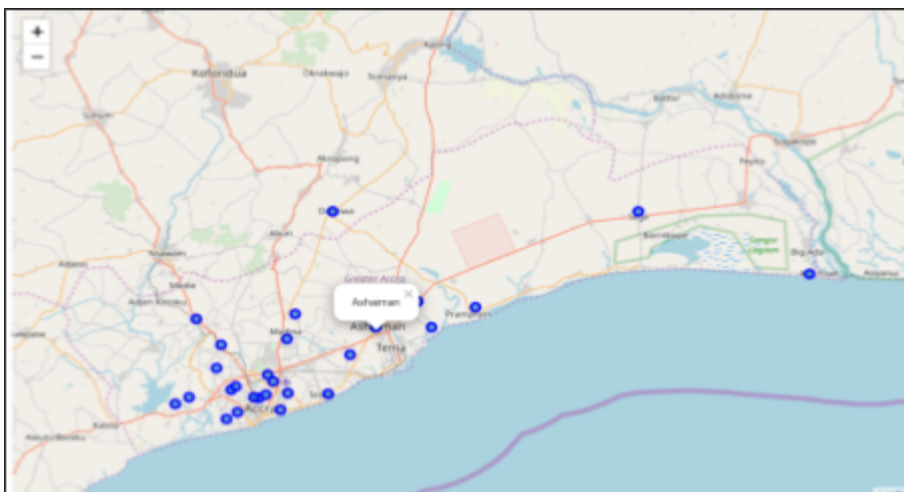
District	Capital	Number of Medical Facilities
Ablekuma Central Municipal	Lartebiokorshie	20
Ablekuma North Municipal	Darkuman	20
Ablekuma West Municipal	Dansoman	20
Accra Metropolitan	ACCRA	20
Ada East	Ada-Foah	20
Ada West	Sege	3
Adentan Municipal	Adenta	20
Ashaiman Municipal	Ashaiman	20

NAME OF HOSPITAL	RATING	TOTAL RATINGS	LATITUDE	LONGITUDE
Barnor Hospital	3.8	21	5.549374	-0.240991
Holy Care Specialist Hospital	4.4	9	5.549925	-0.244662
Loving Life Chinese Clinic	4.2	6	5.548236	-0.232870
Beth's Clinic	4	1	5.550399	-0.233826
Robert Medical Centre	4.7	6	5.547638	-0.233553
D-Health Enterprise	3.5	2	5.541845	-0.239605

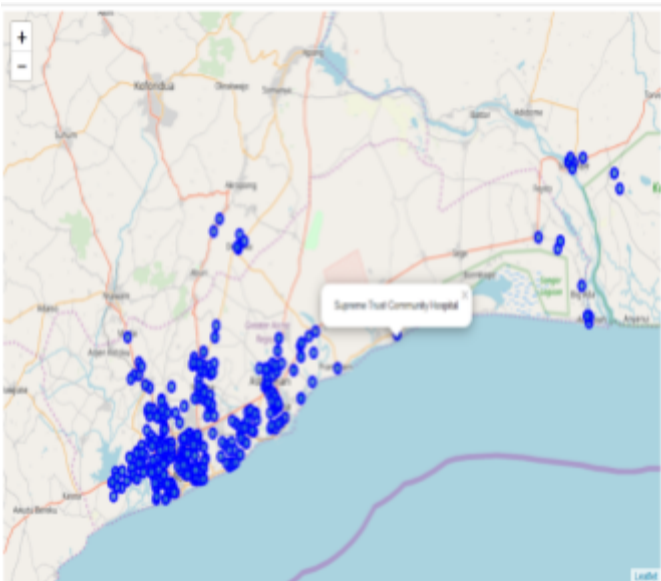
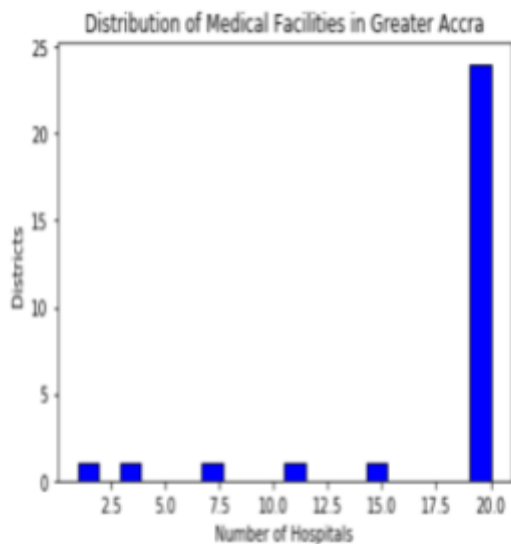
3. EXPLORATORY DATA ANALYSIS

3.1 Distribution of district capitals and hospitals:

Using the matplotlib library to plot histograms and folium to plot maps, a better appreciation of the distribution of the various features. In total 29 district capitals are present in the greater Accra region. The map below shows the location and distribution of the districts across the region.

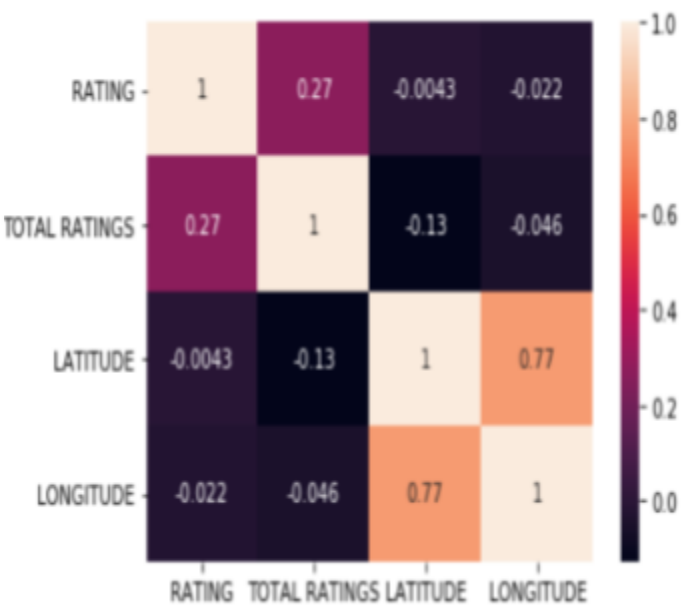
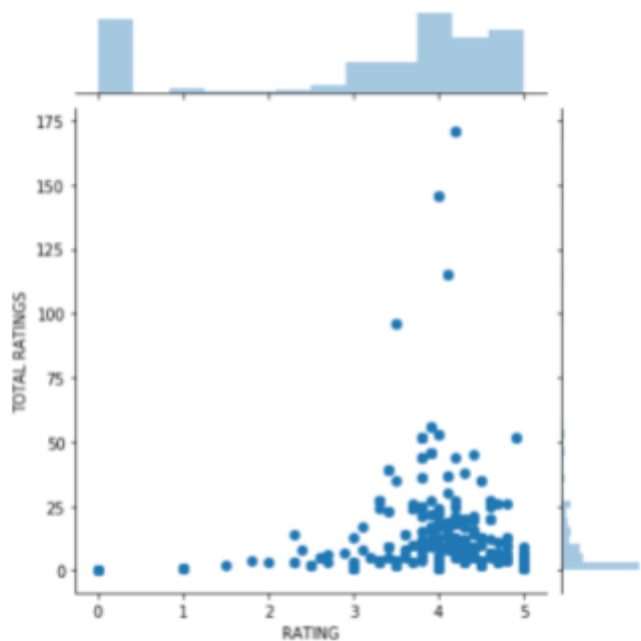


Districts are demarcated based on population density among other factors. A total of 419 medical facilities can be found within the 500 m radius of these district capitals, however not evenly distributed. Some locations were found to have at least 20 medical facilities within the defined proximity, whereas others had as low as one and even none. The histogram below shows the distribution of the number of hospitals across the region. Followed by a map showing the locations of the various medical facilities.



3.2 Correlation between total ratings and average rating:

A scatter plot was made with the ratings and the total number of people who rated the medical facilities to get an understanding of any relationship that exists between the two. The correlation between these features were further emphasized using a heat map annotated with the correlation between the features, using the seaborn library. Below are the graphical representations showing the correlations:

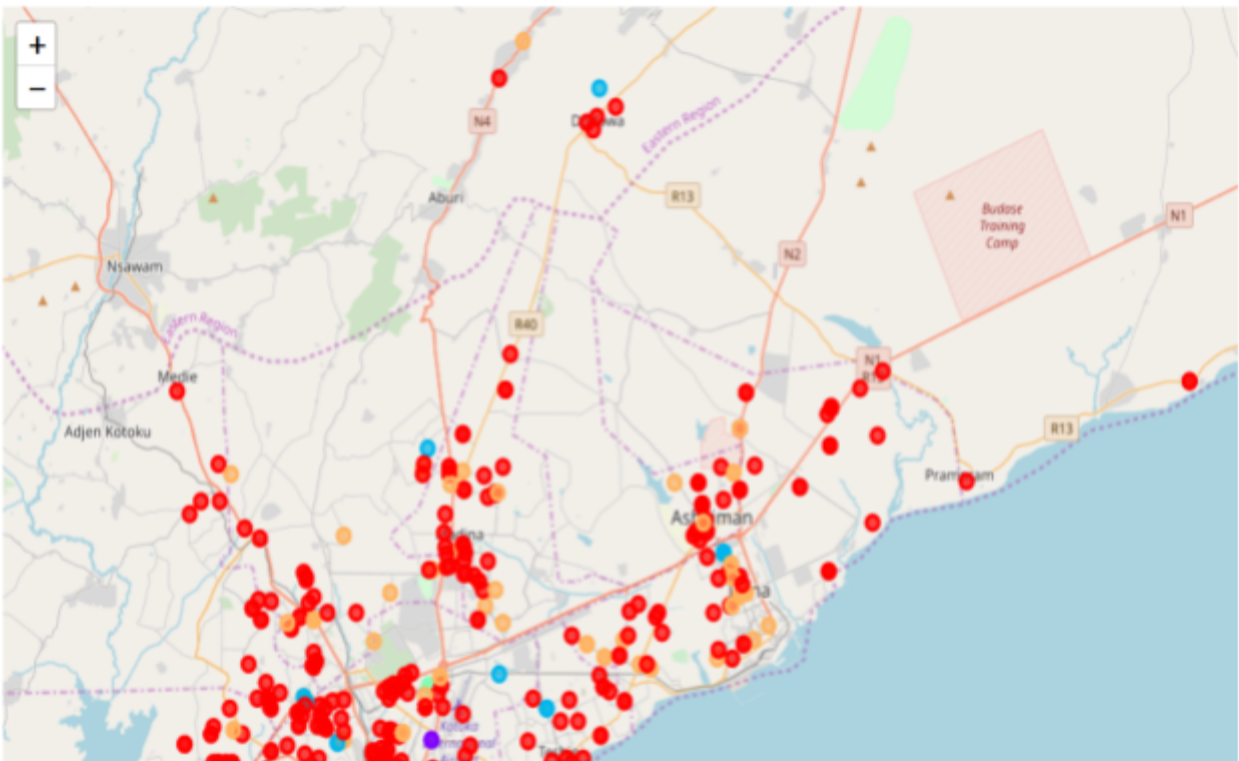


3.3 Modelling: Cluster Analysis:

Finally a cluster analysis was conducted using the K-means Clustering algorithm. A total of 5 clusters were defined for the algorithm to map out of the data. All features including location coordinates were selected. However the names of the hospitals were dropped. Not only because they were categorical features but because the names of the medical facilities do not necessarily have a bearing on significant insight needed to segment or cluster the various medical facilities. After the clustering was complete, the cluster labels were assigned to the various hospitals. Below is a sample:

	NAME OF HOSPITAL	RATING	TOTAL RATINGS	LATITUDE	LONGITUDE	Cluster Labels
0	Barnor Hospital	3.8	21	5.549374	-0.240991	4
1	Holy Care Specialist Hospital	4.4	9	5.549925	-0.244662	0
2	Loving Life Chinese Clinic	4.2	6	5.548236	-0.232870	0
3	Beth's Clinic	4	1	5.550399	-0.233826	0
4	Robert Medical Centre	4.7	6	5.547638	-0.233553	0

A folium map was generated with the locations of the various medical facilities plotted with color markers based on the class generated by the cluster analysis. This is shown below:



4. DISCUSSION AND RESULTS

It was seen the district capitals themselves were not evenly distributed as certain areas were more densely populated than others, among other factors. It was observed that most of the capitals were as a matter of fact clustered around the regional capital Accra. With respect to the distribution of medical facilities, it was observed that some locations were found to have at least 20 medical facilities within the defined proximity, whereas others had as low as one and even none. This implies that whereas the densely populated are well resourced with medical facilities, it is not case in other suburbs. As a result, residents in these areas will be victims of the heavy traffic conditions and suffer delays during life threatening situations requiring immediate medical attention. The cluster analysis further revealed significant insights into the preferences of people when it comes the choice of medical facilities they visited. It was seen first of all from the scatter plot a curved linear relationship between the ratings and the number of ratings. It was clear that more people rated the facilities when they preferred those facilities. Facilities that were not preferred were largely not rated altogether.

5. RECOMMENDATIONS

The cluster analysis did a good job uncovering relationships and classifying the medical facilities based on those relationships. However, it is worth that there was not sufficient amount of data on the various facilities to uncover more underlying relationships, to better explain the reasons why people preferred their choices. For instance, factors such as resources available at medical centers, cost of medication, whether it was a public or private facility, among others and not just proximity can be responsible for why people visited and rated some facilities and not others. Future work could benefit greatly with more data on the unique features of the various medical facilities. This would help drive more insight into the clusters uncovered by the K-means algorithm. It is highly recommended that more facilities are set up in the areas that need them, as medical facilities are a key way of decongesting the overpopulated areas in the region.

6. CONCLUSIONS

There are areas in the region that appear neglected with respect to the siting of medical facilities. Whereas citing new facilities in these areas may not be the only solution, the challenges provides an opportunity to be explored by entrepreneurs and other stake holders in the health industry.