CAPSTONE PROJECT - The Battle of the Neighborhoods

Erhan KORAL

April, 01, 2010

1. Introduction

1.1. Background

Istanbul is a very large city with a population of nearly 16 million. It has a population density of 2844 people per square kilometer. Its population increased over 1 million in last 5 years. The city divided into 39 Boroughs and 782 Neighborhoods. Being such a crowded city leads some health problems in aspects of citizens, especially nowadays when COVID-19 seriously threaten humanity. With a high population, health service planning becomes one of the major responsibilities for government and local managements.

1.2. Problem

Authorities should effectively determine areas with weak health services and medical establishments and orient the funds for developments of health services in those areas. This project aims to classify areas according to their health and medical capabilities and find optimal locations for enhancing health service points and medical establishments. So the target audience is clearly authorities, who is liable for planning and serving health and medical services in Istanbul.

2. Data Description

2.1. Data Sources

- Based on definition of problem, lots of factors that will influence the decision should be considered as a parameter. For example number of existing medical establishments, number of beds, capabilities of emergency service or ambulance, number of intensive care units, number and branches of medical service personnel, number and type of diagnosis devices are some of them. But much of this information is not public. So only open data will be used in this project. These are :
- number of existing medical establishments in the neighborhood (specific type of health care services)
- number of beds, if any
- capabilities of emergency service or ambulance

Following data sources will be needed to extract/generate the required information:

- Centers of Boroughs will be generated by using **Google Maps**
- Coordinate of Istanbul center obtained using **geopy geolocator**
- Number of medical establishments and their type and location in every neighborhood will be obtained using **Istanbul Municipality Open Data Platform**
- The Data is in the form of CSV. It is provided by Istanbul Municipality IT Department at their Open Data Platform. This is the link for the data.

https://data.ibb.gov.tr/dataset/bd3b9489-c7d5-4ff3-897c-8667f57c70bb/resource/f2154883-68e3-41dc-b2be-a6c2eb721c9e/download/salk-kurum-ve-kurulularna-ait-bilgiler.csv

2.2. Data Cleaning and Preparing

Data set will be acquired from Istanbul Municipality Open Data Platform in CSV format. It includes names and types of those establishments, Boroughs, Neighborhoods, communication info, their capabilities such as emergency service, ambulance, number of beds and geolocal information. First we will add "First Aid Station" counts per Borough as a column into our data frame. Unnecessary columns will be dropped so we can focus our main purposes. Column names and records are in Turkish, they are causing some problems with utf-8 encoding so we will translate them to English and manipulate some of them. We will drop some records that include only side health services (for example dialysis centers, veterinaries, physical rehabilitation centers) We will convert 'Ambulance' and 'Emergency_Service' column records to numerical data so we can use them. 1 will refer to existence and 0 will refer to non-existence. We will also drop rows that 'Ambulance','Emergency_Service' and 'Bed' records includes NaN values.

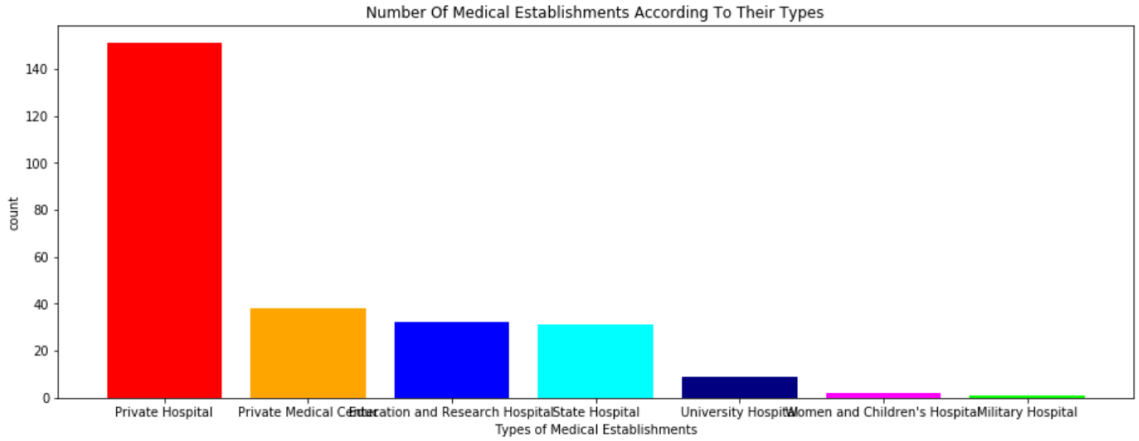
	Borough	Name	Sub_Category	Emergency_Service	Bed	Ambulance	Neighborhood	Latitude	Longitude
4	KADIKÖY	Medipol Üniversitesi Hastanesi Kadýköy	University Hospital	1	72.0	1	KOÞUYOLU	41.004663	29.034348
10	ADALAR	Adalar 1 Nolu Acil Yardým Ýstasyonu	Emergency Station	0	3.0	1	MADEN	40.874674	29.132746
21	FATIH	Ýstanbul Eðitim ve Araþtýrma Hastanesi	Education and Research Hospital	1	507.0	0	CERRAHPAÞA	41.003074	28.938300
43	KARTAL	Kartal Koþuyolu Yüksek Ýhtisas Eðitim ve Araþt...	Education and Research Hospital	1	465.0	0	CEVÝZLÝ	40.915824	29.171704
44	SISLI	Özel Ýac Ýstanbul Aesthetic Týp Merkezi	Private Medical Center	1	8.0	0	ESENTEPE	41.069186	29.006582
54	SILIVRI	Silivri Devlet Hastanesi Selimpaþa Ek Hizmet B...	State Hospital	1	35.0	1	SELÝMPAÞA	41.054333	28.379361
57	FATIH	Ýstanbul Haseki Eðitim ve Araþtýrma Hastanesi ...	Education and Research Hospital	1	125.0	0	HIRKA-Ý ÞERÝF	41.020465	28.937514

3. Data Analysis

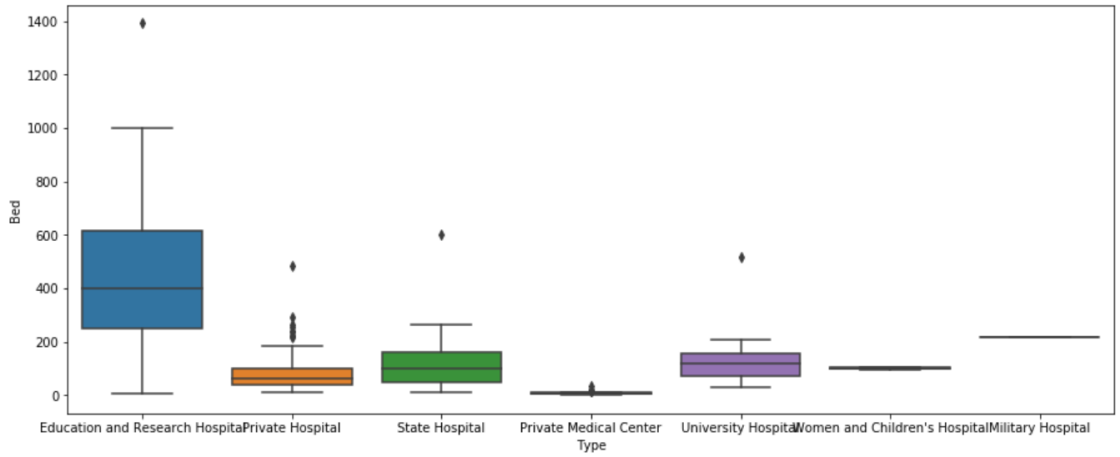
Second step in our analysis we will deep dive into data and try to resolve what kind of information can we use. We will explore **First Aid Counts per Boroughs, Types of Establishments and Their Counts across the city, Distribution of Establishments according to their types in each borough, Distribution of Beds according to Establishment Type** and try to present them by visualization charts. We will also use **boxplot** to identify relation between Establishment Type and Bed Counts and also detect outliers according to number of beds.

These are the numbers of medical establishments according to their type across the Istanbul.

	Facility Type	Facility_counts
0	Private Hospital	151
1	Private Medical Center	38
2	Education and Research Hospital	32
3	State Hospital	31
4	University Hospital	9
5	Women and Children's Hospital	2
6	Military Hospital	1

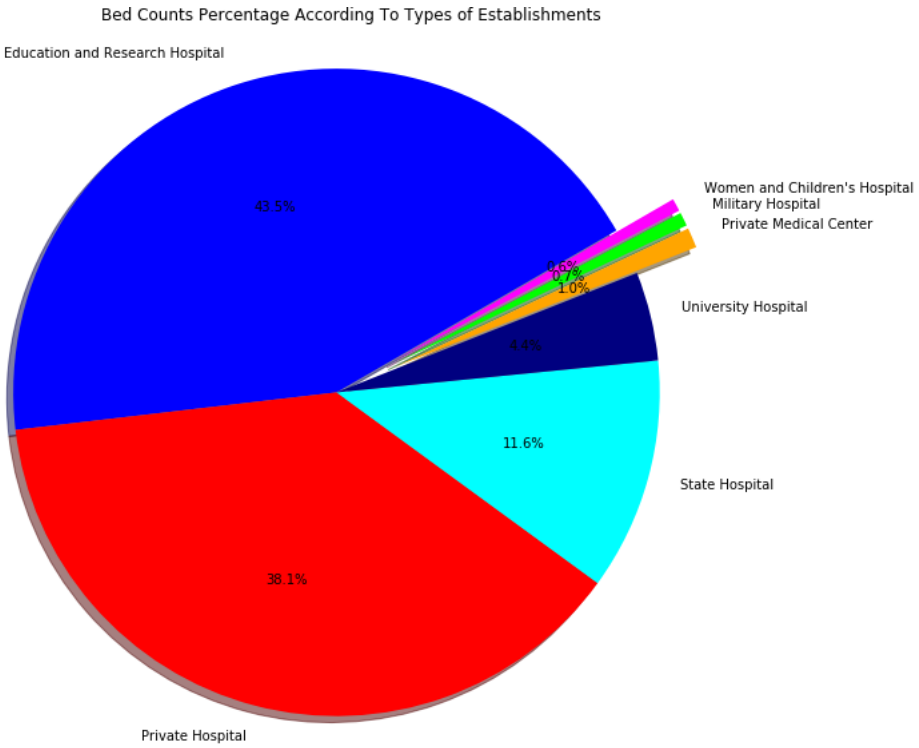


Let's look at the relationship between "Establishment Type" and "Bed counts". We can easily detect outliers with boxplot visualization.

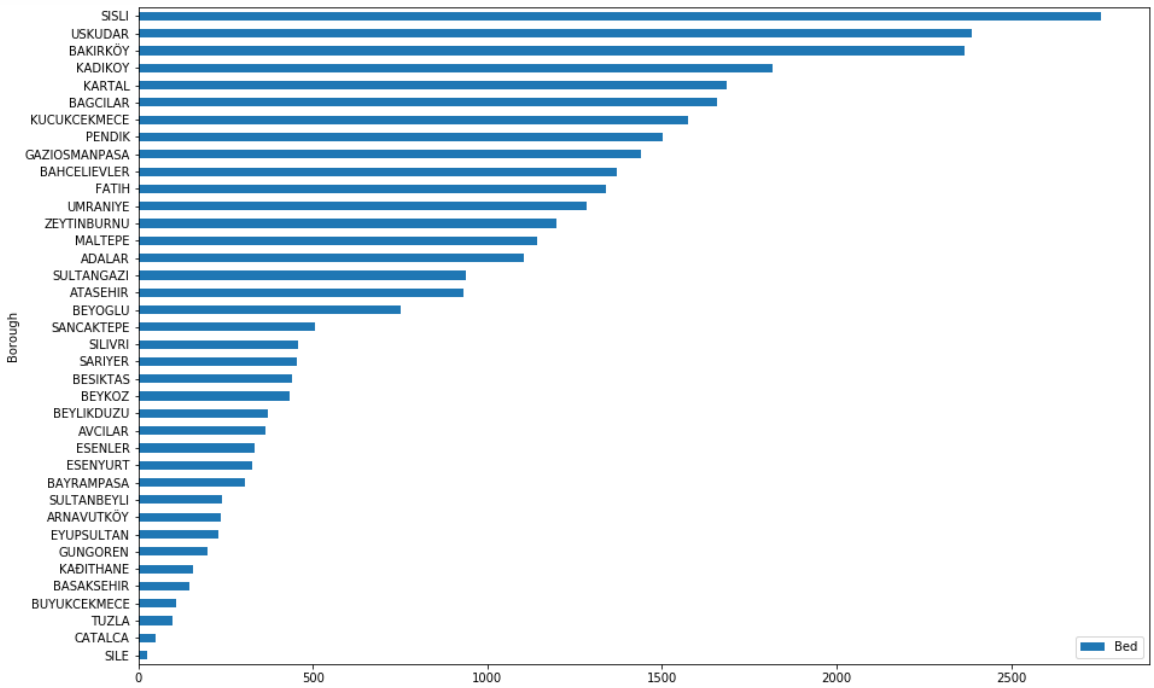


It is important to identify distribution of beds according to type of establishments and create a pie chart for better understanding the situation. It will be so meaningful info for discriminating the private's and state's capability.

	Type	Bed
0	Education and Research Hospital	13763.0
2	Private Hospital	12057.0
4	State Hospital	3670.0
5	University Hospital	1403.0
3	Private Medical Center	317.0
1	Military Hospital	216.0
6	Women and Children's Hospital	201.0



And also it will much meaningful seeing them by means of boroughs.



This chart nicely covers density of beds by means of boroughs.
Then we can list Boroughs by their average Emergency Service, Bed and Ambulance numbers. These are average numbers with respect to numbers of total medical establishment in that Borough.
Now we are approaching a meaningful table for clustering. We are trying to collect all kind of info about boroughs in form of numbers about health and medical service capability. We are merging

two frames and adding First Aid Station number. After, we are trying to find aid capability radio by multiplying number of first aid stations and ambulance ratio.

	Borough	Emergency_Service_Ratio	Ambulance_Ratio	Bed_Counts	First Aid Station	Latitude	Longitude	Firs Aid Ratio
0	ADALAR	1.000000	1.000000	1105.0	4	40.870022	29.128099	4.00
1	ARNAVUTKÖY	1.000000	0.500000	235.0	2	41.179635	28.745557	1.00
2	ATASEHIR	1.000000	1.000000	930.0	3	40.977382	29.107801	3.00
3	AVCILAR	1.000000	0.500000	366.0	0	40.991604	28.717076	0.00
4	BAGCILAR	0.916667	0.458333	1658.0	6	41.042299	28.844779	2.75

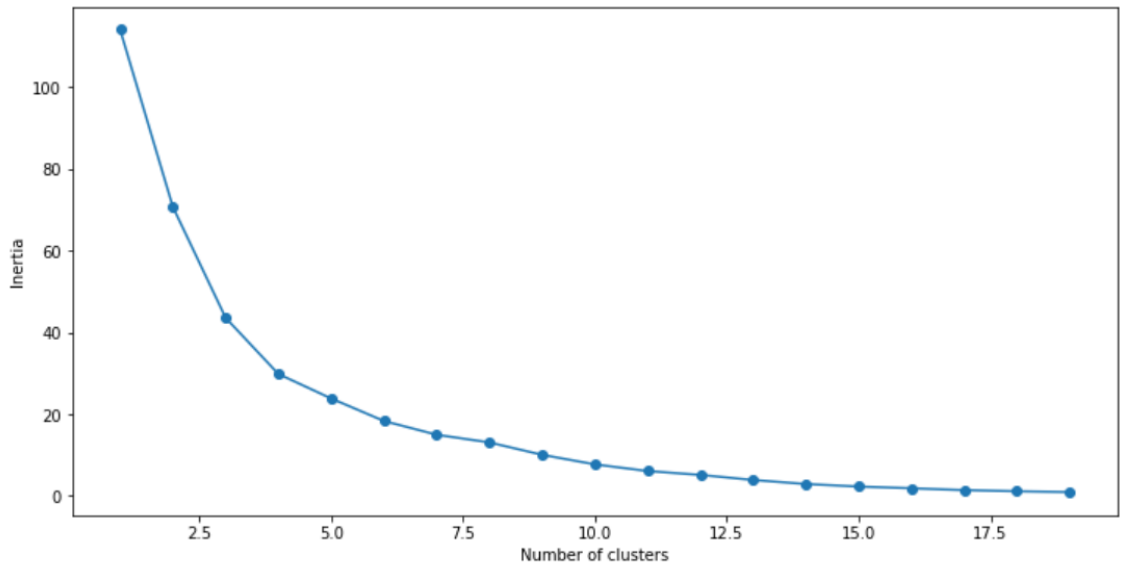
4. Modeling

In clustering, we are looking at the data and then try to club similar observations and form different groups, hence it is an unsupervised learning problem. Clustering is a widely used technique in data driven sectors. It is actually being used in almost every domain, ranging from banking to health, image recognition to business.

We will first normalize the dataset.

Using the K-Means++ algorithm, we are optimizing the step where we randomly pick the cluster centroid. We are more likely to find a solution that is competitive to the optimal K-Means solution while using the K-Means++ initialization.

One of the most common doubts everyone has while working with K-Means is selecting the right number of clusters. One thing we can do is plot a graph, also known as an elbow curve, where the x-axis will represent the number of clusters and the y-axis will be an evaluation metric.

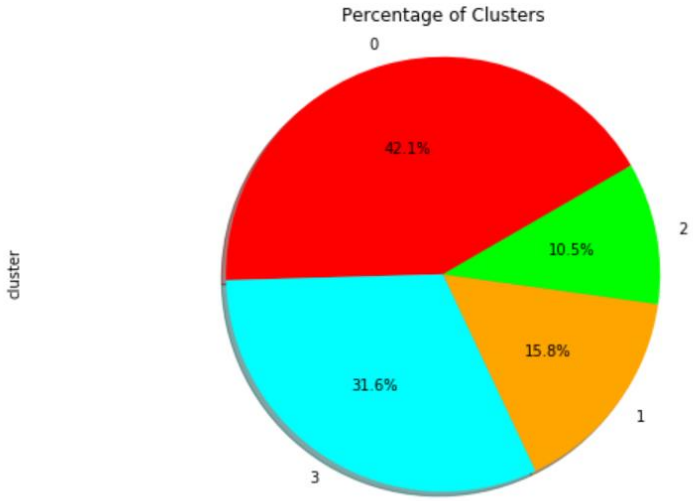


We can choose 4 for number of clusters since we can easily se 4 as a breaking point. Let's set the number of clusters as 4 and fit the model.

5. Discussion and Results

So, we have 4 different cluster produced by K-means. We can assess them as:
Very High Capacity, no precedence,

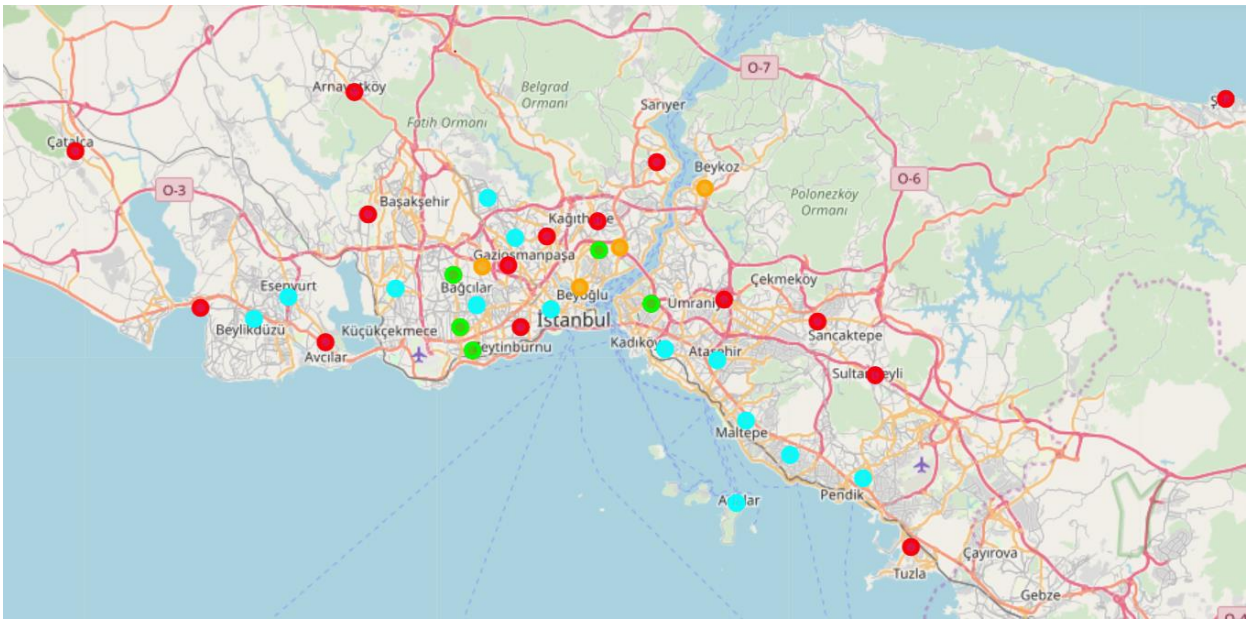
High Capacity, no precedence,
Normal Capacity, to be considered,
Low Capacity, have precedence.



After now we can label boroughs according to their clusters by means of their health and medical capability. At the frame below, we can see detailed list of boroughs with their cluster labels.

	Borough	Cluster Labels	Emergency_Service_Ratio	Ambulance_Ratio	Bed_Counts	First Aid Station	Latitude	Longitude
0	ADALAR	3	1.000000	1.000000	1105.0	4	40.870022	29.128099
1	ARNAVUTKÖY	0	1.000000	0.500000	235.0	2	41.179635	28.745557
2	ATASEHIR	3	1.000000	1.000000	930.0	3	40.977382	29.107801
3	AVCILAR	0	1.000000	0.500000	366.0	0	40.991604	28.717076
4	BAGCILAR	1	0.916667	0.458333	1658.0	6	41.042299	28.844779

At last we can Show classified boroughs according to their heath and medical capacity in form of different colors according to their cluster.



5. Conclusion

In this project we tried to find a way to classify boroughs and neighborhoods of Istanbul by analyzing data and using classification algorithms.

As we know in metropolis such as Istanbul, one of the challenging problems is planning and serving medical and health services. It especially gains much more importance in nowadays as humanity continuing to suffer from COVID-19. And we see most affected part is millions who lives in big cities.

We used dataset which is provided by Istanbul Municipality Open Data Platform. This data set is providing limited information about medical capabilities. It should be considered health and medical service planning depends on much more amount and complexity of data. This project is just a case study example and may not yield high quality results for real life. For more accurate results, the data set must be expanded.

We used the K-means algorithm as classification tool. We determined optimum k value by Elbow method. We classified boroughs into four clusters by means of their health and medical capabilities. At last we visualized clusters on Istanbul map by different colors according to their clusters.