

# **London's New Hospital**

Alexander Shynkarenko

April 16, 2019

## **1. Introduction**

### **Background**

Given the rapidly growing and aging population, the need in hospitals is as high as it ever was. This notion is exceptionally prevalent in the developed countries, such as United Kingdom. While designing a new location for a hospital, one must consider multiple factors to determine whether it is reasonable to put a new establishment in a given location.

### **Problem**

United Kingdom's National Health Service (NHS) has decided to put a new hospital in London. However, it wants to choose a location of this new hospital in a way that maximizes the amount of patients' needs satisfied in the city.

**What are the best-suited boroughs in London to host the new hospital?**

### **Interest**

Although this is just a simulation scenario, in a real-world example NHS-like entities are indeed interested in the skills of a data scientist to decide the optimal location for their venues.

## 2. Data

There are two primary data sources for this project.

- A dataset of the London boroughs and corresponding information, retrieved from Wikipedia. It was scrapped using *beautiful soup 4* library[1]
- Geospatial data of hospitals in London areas fetched using Foursquare API. [2]

The primary dataset [1] consists of **32** rows and **10** columns. Each row represent 1 **borough** of London. Given that the hospitals are rare when looked at a **neighborhood** level, I have chosen to work with **boroughs** instead.

Each borough in the dataset has the following properties, represented by columns:

1. Borough – name of a borough
2. Inner – whether it is considered part of Inner London
3. Royal – is status of royal borough/city applicable to this borough
4. Local Authority – name of a local authority
5. Political Control – what party maintains control over the local authority
6. HQ - address of the local authority, considered to be a center of a borough
7. Area(sq. miles) – area of borough in square miles
8. Population – population of a borough
9. Location – coordinates of the borough's center in DMS format
10. Position – position within the map on the website

	Borough	Inner	Royal	Local Authority	Political Control	HQ	Area(sq. miles)	Population	Location	Position
0	Barking and Dagenham			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51°33'39"N 0°09'21"E	25
1	Barnet			Barnet London Borough Council	Conservative	North London Business Park, Oakleigh Road South	33.49	369,088	51°37'31"N 0°09'06"W	31
2	Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51°27'18"N 0°09'02"E	23
3	Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51°33'32"N 0°16'54"W	12
4	Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51°24'14"N 0°01'11"E	20

Figure 1 Initial representation of primary dataset

### 3. Methodology

In order to predict the best neighborhoods to put a new hospital, we must segment them into **clusters** of similar boroughs and pick the most suited cluster.

What makes a borough well suited for a new hospital? It must satisfy several conditions:

1. Provide **immediate** service to **as many new people** as possible
2. Be **as far from existing establishments** as possible
3. Contain **the least amount of hospitals within**

In other words, we must maximize **population**, **distance between nearest hospitals** while minimizing **amount of hospitals within borough**.

At the first step of the analysis, I visualized the map of London and put markers signifying boroughs over them. This was done using *folium* library (DMS coordinates were transformed to decimal coordinates)

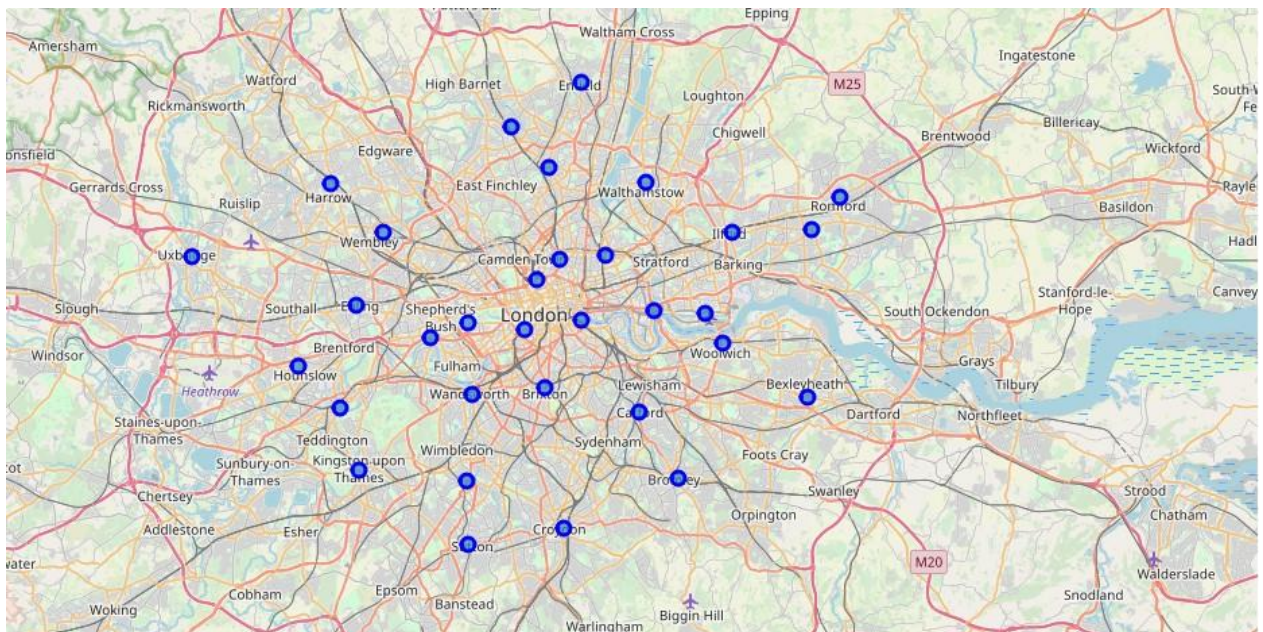


Figure 2 Map of London. Blue dots represent borough centers

This visualization does not provide enough information *per se*. Although now we know the general location of boroughs, that is not enough to make any assumptions about which boroughs are better suited as places for new hospital.

Using Foursquare API I have fetched information on hospital within the limits of a borough. To simplify this process I've calculated the distance from a center of each borough using its area as a perfect circle and finding a radius of that circle. This fetch request has net me a list of **141** hospitals. Given slight inaccuracies of this method, **6** of them were duplicates, but that was not a problem as even if the hospital is technically in other borough, but still sufficiently close to the center of another, then it can efficiently provide service to the patients.

Then I superimposed locations of those hospitals on the same London map.

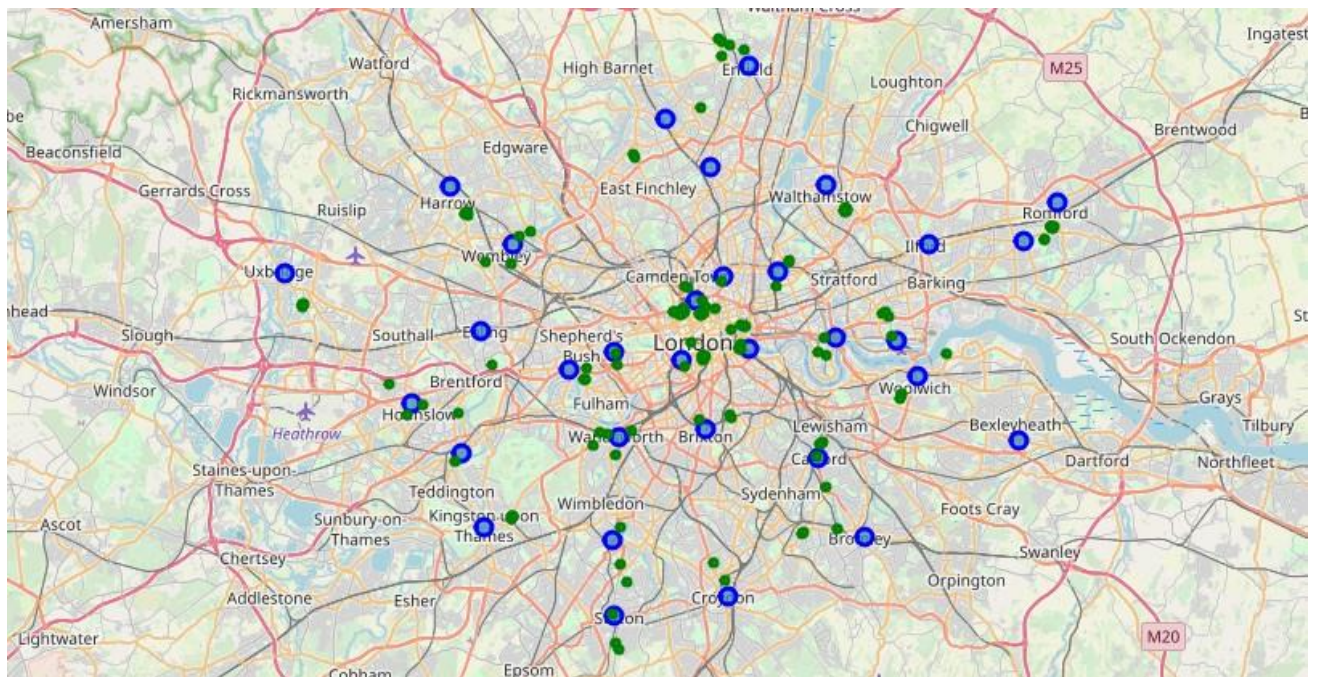


Figure 3 Map of London. Blue dots represent borough centers. Green dots represent hospitals



Now we can see much more. As expected, many hospitals are located within Inner London, most notably Camden borough. On the other hand, it looks that there are boroughs, especially on the outskirts, that are fully devoid of hospitals. Those boroughs may be primary targets for new hospitals.

Looking closer into the dataset fetched from Foursquare we can confirm that boroughs of **Bexley**, **Haringey**, and **Redbridge** lack any hospitals and therefore are the prime targets for a new one.

This may have concluded this analysis, but let's say that maybe the contractor is more interested in locations that already have **some** hospitals, but generally not enough. To understand which boroughs among those that have hospitals are the best for a new one, there must be additional pre-processing of data.

First, we add the overall amount of hospitals for each borough, adding 'Venue' column.

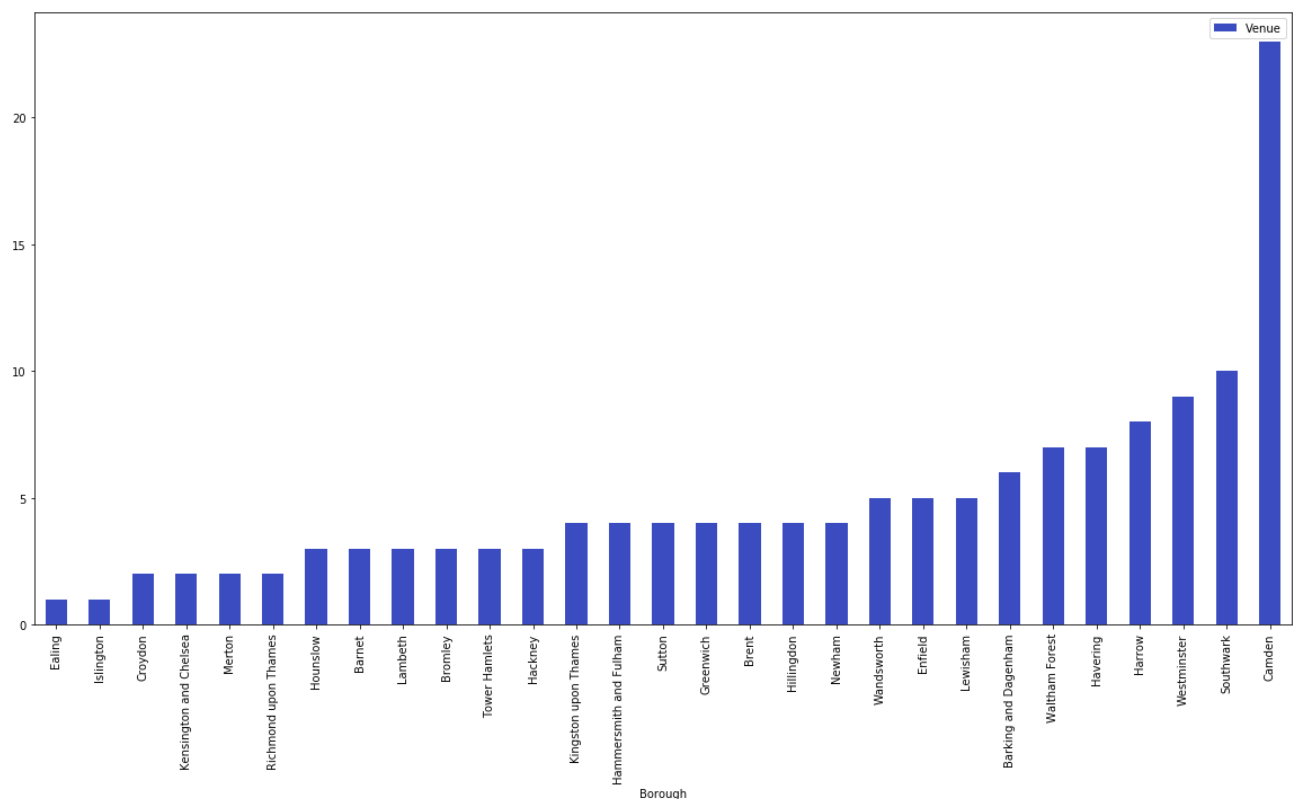


Figure 4 Bar Chart of venues for each of the borough. Note the disproportionate number of hospitals in Camden borough

Then, we get a mean **distance** to a hospital for each borough by calculating *vincenty* distance between the borough's center and each hospitals location and finding the mean. This becomes the column 'Distance'. Now we have a dataset that is ready to be put through clustering analysis. But before that, I checked the correlations between main features of a dataset.

	Population	Area(km^2)	Population Density	Venue	Distance
Population	1	0.442095	-0.175389	-0.113649	0.415593
Area(km^2)	0.442095	1	-0.794263	-0.156613	0.774874
Population Density	-0.175389	-0.794263	1	0.106996	-0.751027
Venue	-0.113649	-0.156613	0.106996	1	-0.103462
Distance	0.415593	0.774874	-0.751027	-0.103462	1

Figure 5 Correlation matrix of the main features of a dataset

The most interesting feature that in general, amount of hospitals is not strongly correlated with any other value on its own. Other findings are as expected, notably distance is strongly negatively correlated with population density and strongly positively correlated with area of the borough.

The next step is clustering analysis. I decided to use k-means clustering, as it is a regional clustering model that works well with the task of segmenting data into general, non-overlapping clusters. The first step, data normalization, was done using `StandardScaler()` of *skit-learn* library. **k** of 4 was chosen after testing model with different values. The model `KMeans()` has clustered dataset into 4 clusters with distinct characteristics.

	Population	Venue	Distance
Label			
0	251610.333333	7.444444	1351.802184
1	307791.333333	4.777778	1551.043282
2	187699.125000	3.125000	1176.745913
3	361444.666667	2.000000	2109.650791

Figure 6 Mean values for our target parameters between clusters

Cluster 3 is clearly the most suitable for a new hospital, as boroughs in it has the largest population, the least venues and the longest distance between hospitals. Three boroughs constitute cluster 3: **Croydon, Ealing, and Barnet**. We would recommend those boroughs as locations for a new hospital.

As a final step, we color our marks on the map in different colors, according to their clusters. Cluster 3 is colored red.

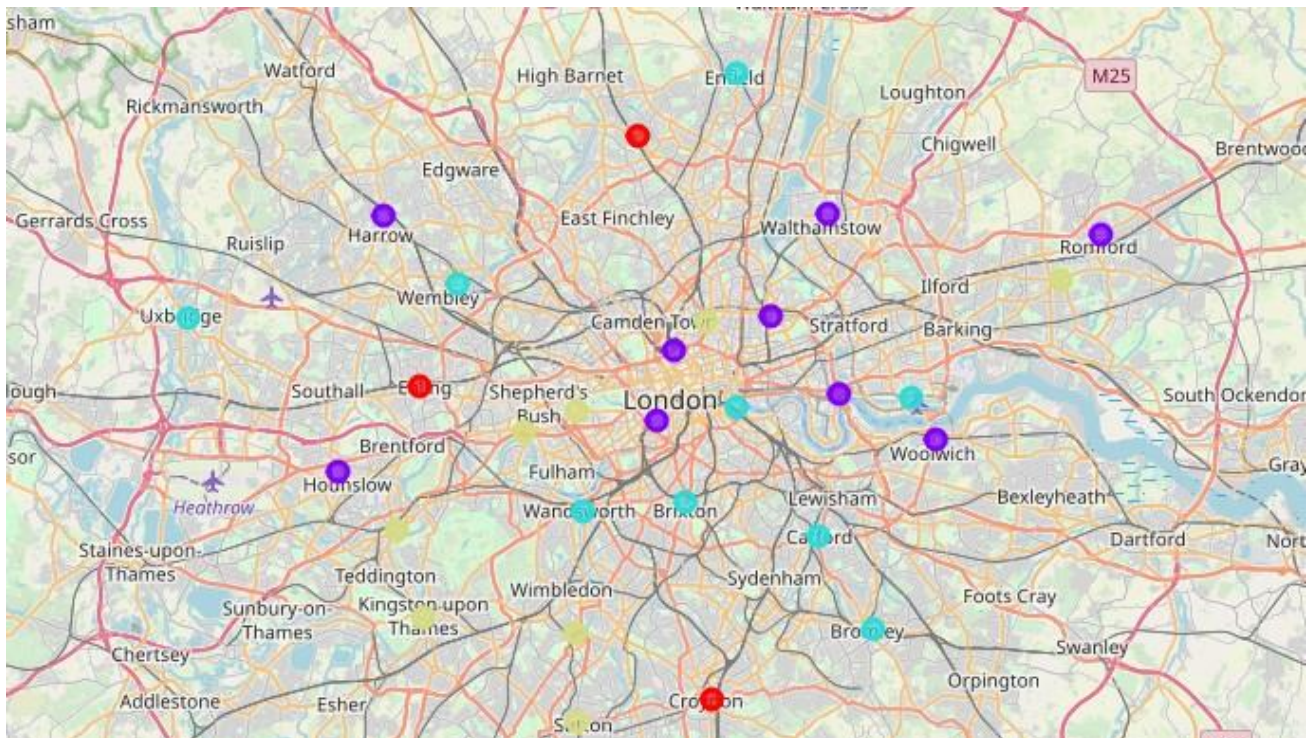


Figure 7 Map of London. Red dots represent boroughs of cluster 3

#### 4. Results

The question posed before this analysis was **what are the best-suited boroughs in London to host the new hospital?**

After conducting data analysis, I have concluded that the most suitable boroughs are **Bexley, Haringey, and Redbridge** as they completely lack any hospital venues. However, if it is more desired to put a new hospital at a location that already hosts a hospital, then **Croydon, Ealing, and Barnet** are the optimal boroughs for that. **Camden** is the most overpopulated borough in terms of hospitals.

## **5. Discussion**

It should be noted that all of the boroughs that lack sufficient hospitals are located far from the center of London, which may hint at the fact that more attention should be averted to the suburban areas.

Another point of concern is disproportionate number of hospitals in Camden borough. Even compared to other Inner London boroughs it has twice more the amount of hospitals, which may lead to potential logistics issues within the neighborhoods of that borough.

More data would be necessary to make this analysis more accurate. In particular, given the demographics (age and social), rate of accidents and crime rate of each borough may give additional insight towards what locations require additional hospitals.

## **6. Conclusions**

In conclusion I would like to thank my instructors and fellow learners that have helped me immensely on this journey of getting familiar with the data science. Data science has really grown onto me over the time it took me to finish this specializations and I am eager to apply my knowledge on practice and learn more.

## **References**

1. [https://en.wikipedia.org/wiki/List\\_of\\_London\\_boroughs](https://en.wikipedia.org/wiki/List_of_London_boroughs)
2. <https://foursquare.com/>
3. [https://nbviewer.jupyter.org/github/wreport/Coursera\\_Capstone/blob/master/London%27s%20New%20Hospital.ipynb](https://nbviewer.jupyter.org/github/wreport/Coursera_Capstone/blob/master/London%27s%20New%20Hospital.ipynb)