Coursera Capstone Project

By, Uma Subbiah (August 2019)

Finding the Optimal Location to Establish a New Hospital in London

IBM Applied Data Science Capstone

Table of Contents

Coursera Capstone Project

IBM Applied Data Science Capstone

Finding the Optimal Location to Establish a New Hospital in London

Project by
Uma Subbiah
(August 2019)

---

1. Introduction:

London, the capital of England, is an extremely diverse metropolitan city. Healthcare in the UK is largely covered by the NHS (National Health Services) that aims to provide free and universal medical services to the residents of the UK. The city of London consumes as estimate one fifth of the NHS budget in England [1]. The NHS provides all kinds of medical services like check-ups, surgeries and mental health care. Supposing the NHS approaches me with a proposal to build a new hospital within the city limits of London, I aim to be able to tell them which location in London would be best suited for this establishment. Using data science and machine learning, I will predict the borough that will benefit the most from these services, in terms of number of people who will be benefitted, security and accessibility.

1.1 Motivation

Hospitals are essential establishments that need to be present in each and every locality. The need for easily accessible, safe and reliable hospitals is evident and access to healthcare is a human right. Though there are already a large number of hospitals functioning to provide services to the people of London, I aim to find the most optimal location for the construction of a new hospital based on certain pre-defined criteria.

1.2 Problem Statement

To use data science approaches to analyse and select the most optimal location in the capital city of London to establish a new hospital that aims to serve and provide

medical assistance to citizens from all walks of life and of various age groups. Further, I will adopt clustering techniques to help find the most suitable borough in London to establish a new hospital.

1.3 Stakeholders

The stakeholders in my project are: the hospital management, the citizens of the borough and the doctors working in the new hospital.

1. **The hospital management:** In order to be useful to citizens, the hospital must be located in a locality that is safe, easily accessible and has a large number of elderly people and patients.
2. **The citizens of the borough:** The residents of the borough that the hospital is going to be built in will be directly benefited from the establishment of the same.
3. **The doctors working in the new hospital:** The doctors will prefer the location if it is safe to work in. It will be better if the hospital is located in a borough where it will benefit more patients.

2. Data Required

To solve the problem, I will be using the following data:

- **Foursquare location data API**
  - Used to find the lists boroughs in London.
  - This API provides location specific details like nearby venues, close neighbourhoods and is easily integrated into geospatial data analysis projects.
  - I will be using the

- **Latitude and longitude data:**
  - To obtain the coordinates of the boroughs of interest to my project.
  - This is required to plot the map and also to get the venue data.
  - This API is also available as a csv file that can be joined with the Foursquare location data to provide a way to plot the locations on a folium map.

- **Demographic Data of the population of London**
  - Age distribution data of the population of London, available at https://www.ageuk.org.uk/london/about-us/media-centre/facts-and-figures/ .

- This data will be scraped using a web-scraping tool like Beautiful Soup that can help decide which parts of London have higher proportions of elderly people and will benefit from the construction of a hospital there.
- This can be viewed as density clusters on a folium map.

- **Crime rate data in London:**
  - Available at : https://en.wikipedia.org/wiki/Crime_in_London.
  - Since we also aim to consider the crime rate details of the location of our hospital, I will be using Wikipedia's crime in London data to determine the safest area to establish a hospital.

3. Methodology

Approach

In order to determine the most suitable location for the establishment of a new hospital, I've considered the factors that best determine the success of a hospital as detailed in []. The main highlights of this article show that hospitals run better when they are close to the patients' residences, and when the doctors feel comfortable working there. In this aspect, I've considered the proportion of elderly population and the crime rate of cities as factors for determining the most suitable location.

Exploratory Data Analysis (EDA)

In order to carry out this project, I've used 3 csv files and 2 json/geojson files. The age-distribution dataset has been analysed, and clustered based on the proportion of elderly people in a particular location. From this, I've extracted the boroughs with the highest number of elderly people. Another data file contains crime rate records, recorded borough wise in London. From here, I've extracted the safest cluster of boroughs. The intersection (pandas merge command/SQL join command) gives us the safest boroughs with a high proportion of elderly people.

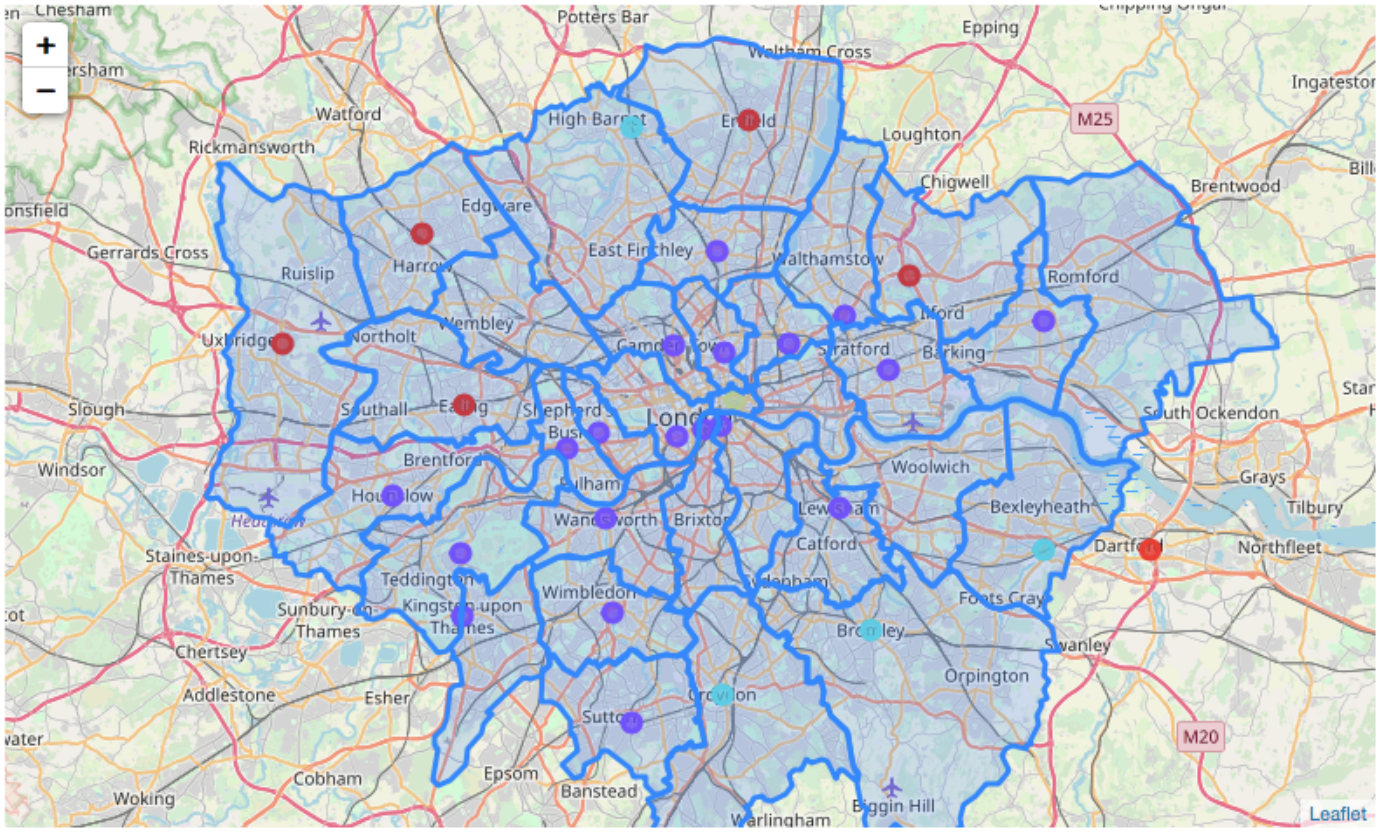

Figure 1: Clusters of boroughs in London, according to the number of elderly patients. The turquoise markers indicate boroughs with the highest number of elderly.

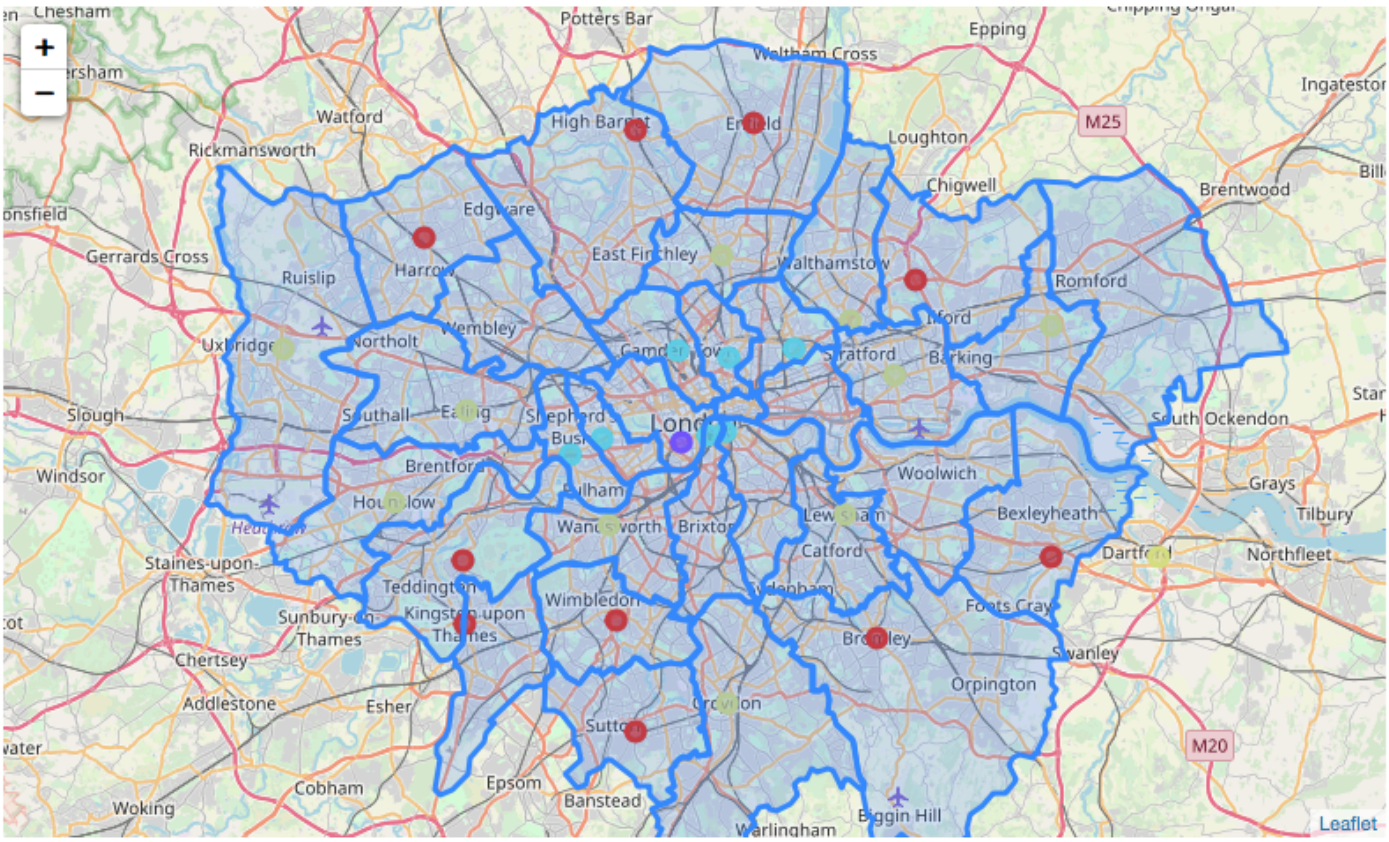

Figure 2: Clusters of boroughs in London, according to the crime rate of each. The red markers indicate the safest boroughs in terms of crime rate.

With the three clusters obtained from the stage described above, I decided to check whether any of these boroughs already have well-established hospitals present. So, for each borough in London, I used a FourSquare API call to find the nearest hospitals within a 1000km radius. The hospitals were plotted on a borough-map of London, as seen below.

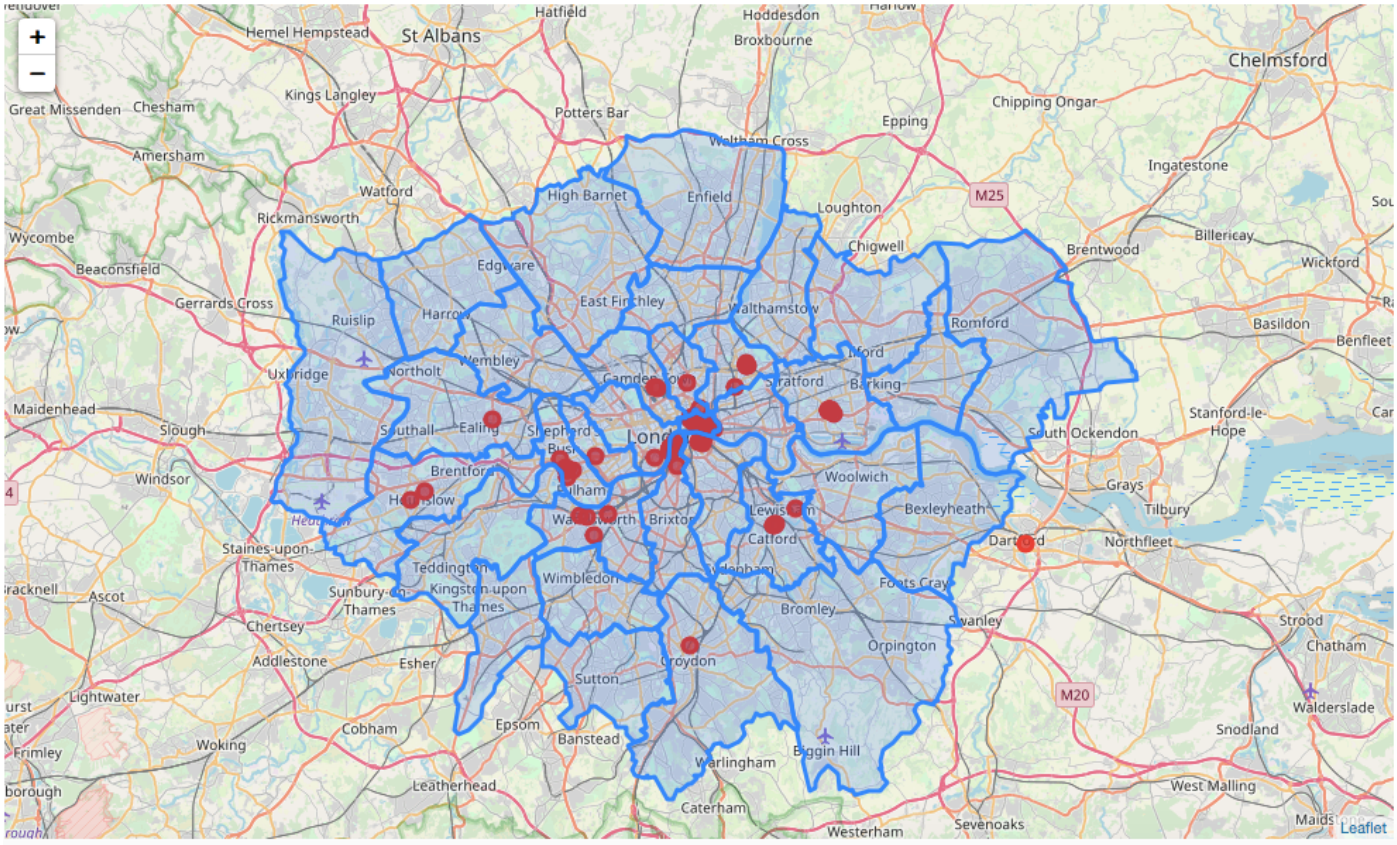

Figure 3 A plot of the nearest hospitals within a range of 1000km from each borough. We observe a concentration of hospitals and medical centers towards the center of London.

From the above map, we can see that the three boroughs we had obtained as a result of the low crime rate-high elderly population combination do not have established hospitals close to them. This shows that the establishment of a new hospital in these areas has a high chance of success.

Application of Machine Learning

In order to determine clusters of cities that are similar in terms of elderly population and crime prone areas, I've used a machine learning approach called clustering. Clustering is an unsupervised machine learning technique that is used to group similar data points together. The similarity is usually measured by distance, in particular, Euclidean distance. I've used sklearn's inbuilt clustering function to perform this operation (ref: [3]). After grouping the data points into clusters, I've extracted the cluster with the highest proportion of elderly people and taken the intersection with the cluster having the lowest crime rate. The resultant three boroughs were considered for further analysis.

4. Results

Tabulate results

The three safest boroughs with a high proportion of elderly population is obtained by joining the two data frames containing the respective information. From this, we see that the following three boroughs are the best locations to establish a new hospital:

	borough	Latitude_x	Longitude_x	Rate	Population 65+
0	Barnet	51.648784	-0.172913	16.923889	47400
1	Bexley	51.441679	0.150488	14.442778	47900
2	Bromley	51.402805	0.014814	16.421667	51900

Figure 4 The three most suitable boroughs - A summary

These boroughs have a fairly low crime rate and a large population of elderly (> 65) people. Establishing a hospital here seems to be a rational management decision.

Visualisation of the 3 locations most suitable for building a new hospital (the three red dots):

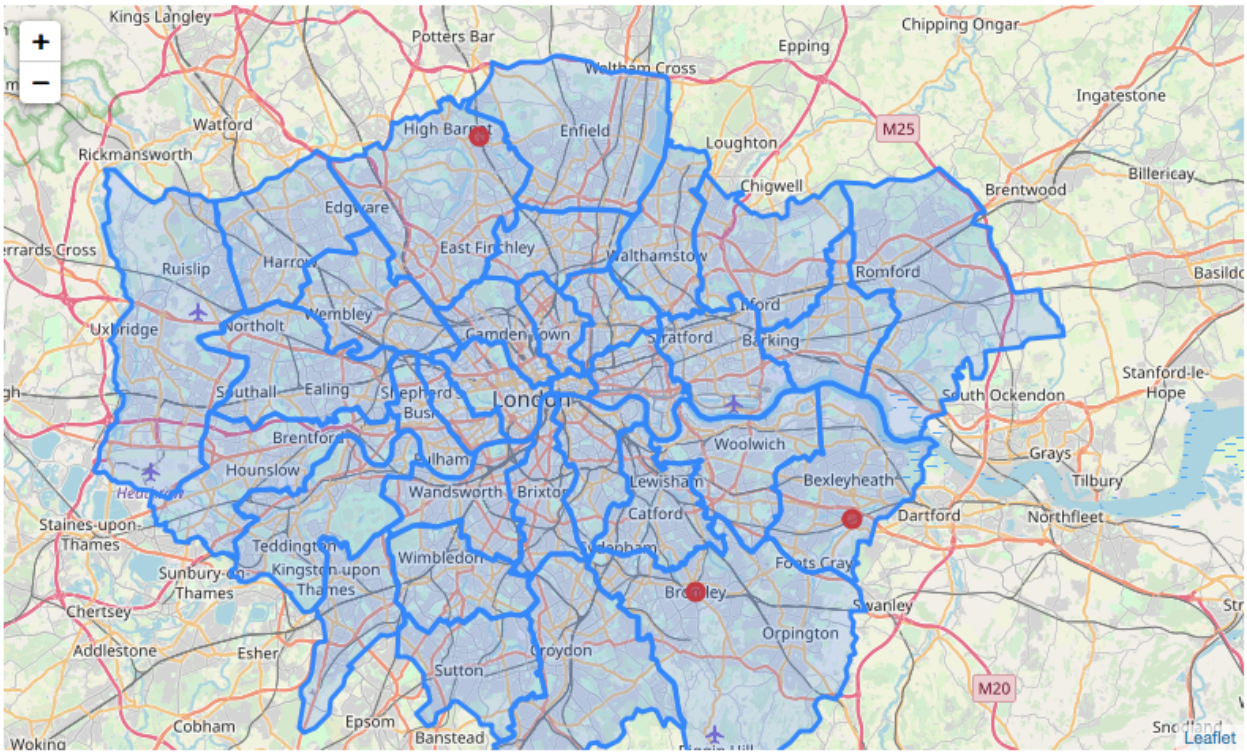

Figure 5: Visualisation of the three most suitable locations for the establishment of a new hospital in London

5. Discussion

The results of the clustering algorithm on both the datasets shows that there are a large number of feasible locations to establish hospitals, if we consider each factor individually, but only 3 main possibilities if we consider both the criteria together. For a new hospital to be successful, the absence of an existing hospital may be a good point.
From the plot of the nearest hospitals from each borough, we can see that the existing hospitals are clustered toward the center, concentrated around the borough of the City of London. So, that eliminates the City of London and neighboring boroughs as potential locations. In addition, we consider the initial two criteria : high proportion of elderly patients and low crime rate.

Thus, I've considered a combination of all three criteria (large elderly population, low crime rate, absence of existing hospitals), to draw my conclusions.

6. Conclusion

In this project, I've determined the most suitable locations in London for the establishment of a new hospital. I considered three major criteria to help decide the location. Using APIs like FourSquare and GeoCoder facilities, I was able to plot the locations and clusters of towns on a folium map and decide the most suitable location for the establishment of a new hospital in London. Based on the three criteria deemed most relevant, I've concluded that there are three boroughs, where a new hospital will prove useful: Barnet, Bexley and Bromley. These three locations have been plotted in Figure 5 and the reasons behind choosing these three places has been discussed in this report.

7. Links

Notebook:
https://github.com/TheGoldenAurora/Coursera_Capstone/blob/master/Final%20Project/Coursera%20Capstone.ipynb

Blog Post:
https://thegoldencode.wordpress.com/2019/08/31/finding-the-optimal-location-to-establish-a-new-hospital-in-london/

8. References

[1] Healthcare in London (https://en.wikipedia.org/wiki/Healthcare_in_London)

[2] 8 Important Things to Consider when Choosing a Location for the Hospital – Hospital Planning, Rahul Choudary (http://www.shareyouressays.com/knowledge/8-important-things-to-consider-when-choosing-a-location-for-the-hospital-hospital-planning/115957)

[3] The 5 Clustering Algorithms Data Scientists Need to Know, George Seif (https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68)