The Battle of Neighborhoods

Finding the best area to build a student dormitory in London

1. Introduction

1.1 Background

London is a leading global educational centre, having one of the largest populations of overseas students of any city in the world. In 2015/16 in the UK as a whole, nearly 2.3 million students were registered at 163 higher education (HE) providers, with about 372,000 or 16% of the UK total studying at 39 HE providers in London.

According to the data published by the Higher Education Statistical Agency (HESA), in the academic year 2016-2017 London welcomed 112,200 international students to its higher education institutions, which make up 29 percent of students at higher education institutions. This means that at least 110,000 students are looking for a new home in London every year, even if domestic students from outside of London are not considered.

The low quality of the dormitories offered by schools themselves, the shortage of student housing, and the relatively good economic resources of the international students who attend schools in London offered an economic opportunity to firms such as Unite Students which have built thousands of units of student housing in London.

1.2 Problem Description:

Student dormitories are the most reliable means of housing for students, especially for one who is completely new to the city and is not familiar with how rental contract works. However, as they are in high demand, it is not easy to secure a place in one. Therefore this project aims to explore different neighborhoods of London and find the best area to build a new dormitory so that international students could solve this persistent problem and real-estate company may find potential business opportunities.

From the student perspective, a lot of factors come into play while searching for the best accommodation, including location, safety and rent fee. However, in this project, we will only focus on the safety and the general atmosphere of the neighborhood for simplicity purpose.

1.3 Target Audience:

This research is expected to benefit real-estate investors looking for a profitable location or international students looking for a place to live in London.

This would interest anyone who would to pursue a high education in London in the future.

2. Data Acquisition and Preprocessing

One city will be analyzed in this project: London.

We will be using the below datasets for analyzing London.
In this project, three different datasets will be used to solve the problem:

- London Recorded Crime,
- List of London Boroughs,
- Foursquare API.

After acquiring them from original and reliable sources, they will be wrangled and cleansed into more useful forms for further analysis.

2.1 London Recorded Crime :

London crime records by boroughs in the last 24 months

source: London Datastore

website url: https://data.london.gov.uk/dataset/recorded_crime_summary

data url: https://data.london.gov.uk/download/recorded_crime_summary/d2e9ccfc-a054-41e3-89fb-53c2bc3ed87a/MPS%20Borough%20Level%20Crime%20%28most%20recent%2024%20months%29.csv'

	MajorText	MinorText	BoroughName	201711	201712	201801	201802	201803	201804	201805	...	201901	201902
0	Arson and Criminal Damage	Arson	Barking and Dagenham	7	4	2	3	6	3	4	...	5	2
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	108	119	135	111	115	122	126	...	97	127
2	Burglary	Burglary - Business and Community	Barking and Dagenham	27	21	38	33	38	36	24	...	45	24
3	Burglary	Burglary - Residential	Barking and Dagenham	88	124	143	134	122	75	93	...	114	107
4	Drug Offences	Drug Trafficking	Barking and Dagenham	5	6	4	5	7	3	8	...	5	1

The table above is the London crime records classified by boroughs and crime type in the recent 24 months. It contains 1575 observations and 27 columns. It was acquired directly from London Datastore.

For further analysis, the number of crimes were calculated into monthly averages, and crime categories were not considered in this research for the purpose of simplicity. This process cleaned the dataset above and makes it more comparable.

	BoroughName	MonthlyAverage
0	Barking and Dagenham	1586.250000
1	Barnet	2445.541667
2	Bexley	1370.333333
3	Brent	2550.083333
4	Bromley	1986.500000

2.2 List of London Boroughs

Information on boroughs and their population & coordinates:

- Population can be used to calculate the ratio of reported crime to population for better comparison.
- Coordinates can be used to get neighborhood data from Foursquare.

source: Wikipedia;

url: https://en.wikipedia.org/wiki/List_of_London_boroughs

Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates
Barking and Dagenham [note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51.5607°N 0.1557°E
Barnet			Barnet London Borough Council	Conservative	North London Business Park, Oakleigh Road South	33.49	369,088	51.6252°N 0.1517°W
Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51.4549°N 0.1505°E
Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51.5588°N 0.2817°W
Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51.4039°N 0.0198°E
Camden	✓		Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	229,719	51.5290°N 0.1255°W
Croydon			Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	372,752	51.3714°N 0.0977°W

From the original data, we will only use on population and coordinates. Population can be used to calculate the ratio of reported crime to population for better comparison, while coordinates can be used to get neighborhood data from Foursquare. So the simplified data for our analysis looks as following.

	BoroughName	Population	Latitude	Longitude
0	Barking and Dagenham	194352	51.5607	0.1557
1	Barnet	369088	51.6252	-0.1517
2	Bexley	236687	51.4549	0.1505
3	Brent	317264	51.5588	-0.2817
4	Bromley	317899	51.4039	0.0198

2.3 Foursquare API:

Listoftop50popularplacesintheneighborhood

source: Foursquare
url: https://api.foursquare.com

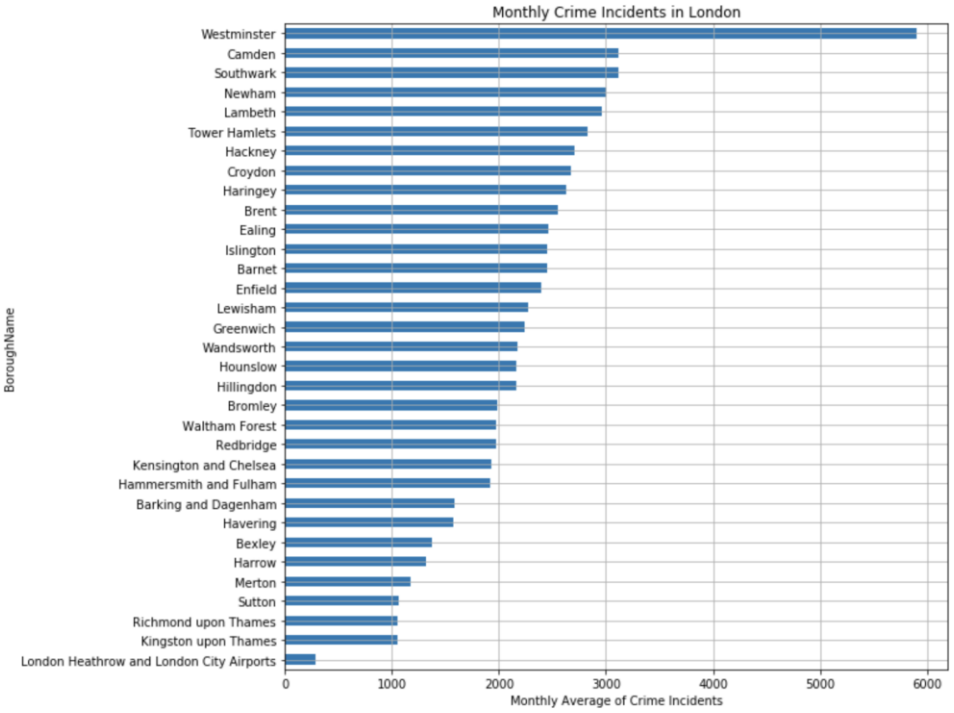
Finally, Foursquare API was used to filter out the top 50 most popular venues in each neighborhood. By using the 'explore' function of requesting URL, we can request data through Foursquare platform. Here is the clean version of data we get.

	BoroughName	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.5607	0.1557	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.5607	0.1557	Crowlands Heath Golf Course	51.562457	0.155818	Golf Course
2	Barking and Dagenham	51.5607	0.1557	Robert Clack Leisure Centre	51.560808	0.152704	Martial Arts Dojo
3	Barking and Dagenham	51.5607	0.1557	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center
4	Barking and Dagenham	51.5607	0.1557	Becontree Heath Bus Station	51.561065	0.150998	Bus Station

3. Methodology

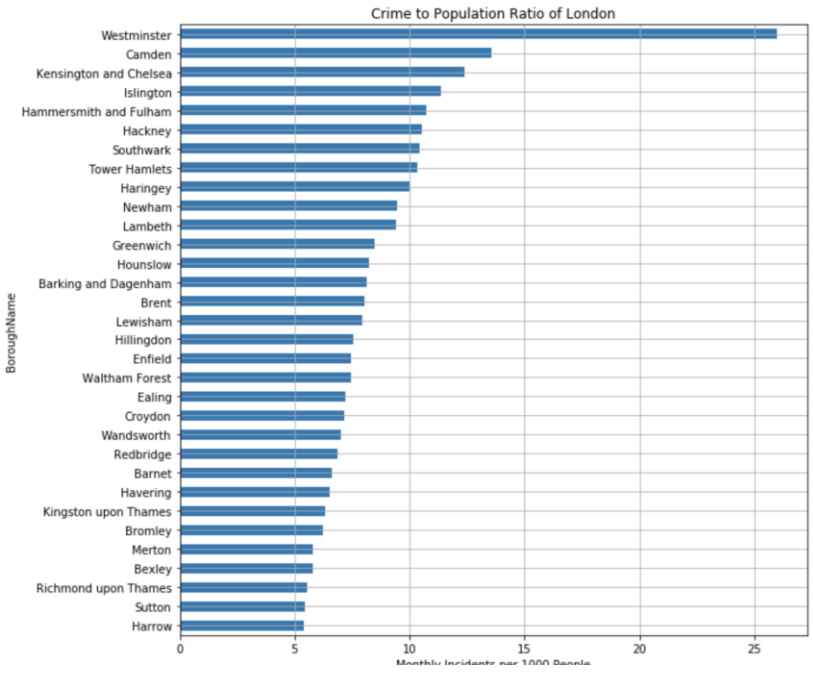
3.1 Exploratory Analysis

After cleansing datasets to more useful forms, I used visualizational tool to create some graphs to interpret the data I have.



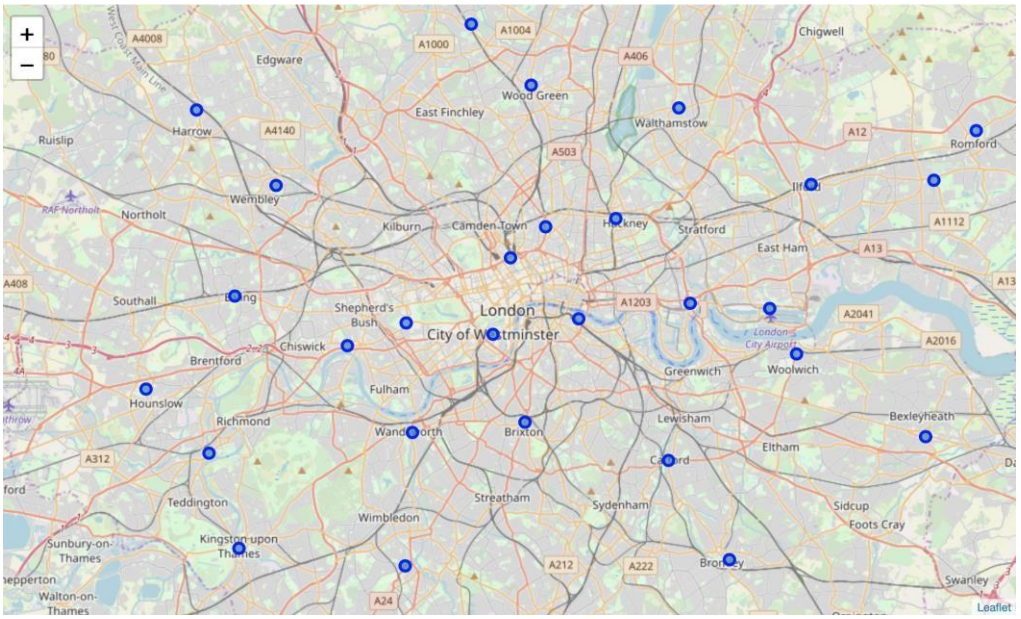
This is a bar chart displaying boroughs in descending order of monthly crime incidents. Westminster has the biggest number of reported crime, followed by Camden, Southwark, Newham and Lambeth.

However, the population of different boroughs varies, it is not wise to directly compare the absolute number of incidents. Instead, we should consider the ratio of crime incidents to people. Thus, I have used the population to calculate the number of recorded crimes per 1000 people in each borough.



It is obvious that Westminster and Camden still remain the top two of the most dangerous places in terms of recorded crime ratio to population. However, from the rank has been changed since the third borough.

And before conducting further with the analysis, I have observed the geographical distribution of boroughs to get an idea of the Greater London area.



3.2 Cluster Analysis

Afterwards, K-means clustering was conducted in order to group by the boroughs according to what venues they have using Foursquare data, in order to evaluate the atmosphere of each borough.

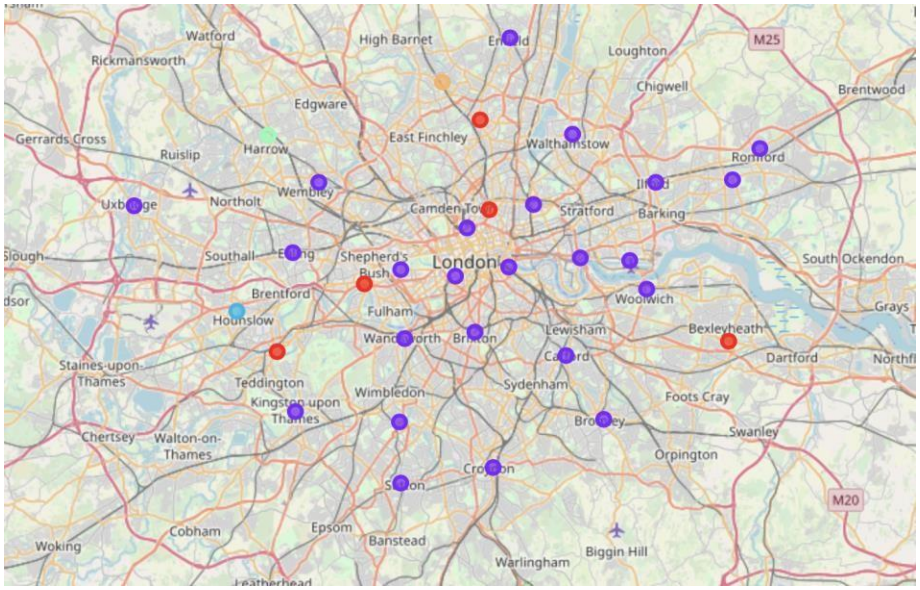
As the first step of cluster analysis, one hot encoding was conducted to assign binary values to each venue categories.

	BoroughName	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	...	Turkish Restaurant	Vegetarian / Vegan Restaurant	Video Game Store
0	Barking and Dagenham	0	0	0	0	0	0	0	0	0	...	0	0	0
1	Barking and Dagenham	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Barking and Dagenham	0	0	0	0	0	0	0	0	0	...	0	0	0
3	Barking and Dagenham	0	0	0	0	0	0	0	0	0	...	0	0	0
4	Barking and Dagenham	0	0	0	0	0	0	0	0	0	...	0	0	0

Then, the data was grouped by borough names to find out how many venues of each category exist in the boroughs within the top 50 venues. However, as some boroughs display less than 50 venues due to lack of Foursquare data, the category counts were altered to frequency of how often the category appears among others. Based on the frequency, we could get a list of most common venue categories in each borough as follows.

	BoroughName	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking and Dagenham	Pool	Park	Bus Station	Supermarket	Golf Course	Martial Arts Dojo	Gym / Fitness Center	Discount Store	Fast Food Restaurant	Food Court
1	Barnet	Café	Bus Stop	Yoga Studio	Farmers Market	Furniture / Home Store	French Restaurant	Food Court	Flea Market	Fish Market	Fish & Chips Shop
2	Bexley	Pub	Clothing Store	Pharmacy	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Supermarket	Portuguese Restaurant	Bakery	Plaza
3	Brent	Coffee Shop	Hotel	Grocery Store	Clothing Store	Sporting Goods Shop	Sandwich Place	Bar	American Restaurant	Italian Restaurant	Pizza Place
4	Bromley	Clothing Store	Coffee Shop	Pizza Place	Bar	Burger Joint	Gym / Fitness Center	Women's Store	Pub	Burrito Place	Café

Based on the venue categories, K-means clustering was conducted to allocate the boroughs into 5 different clusters based on their similarity. The color dots below represent different clusters.



After observing each clusters and the characteristics they possess, we have given names for each clusters that best depicts their characteristics.

- Cluster 1: Business area (Coffee shops, Hotels, Markets)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
3	Brent	8.037733	0	Coffee Shop	Hotel	Grocery Store	Clothing Store	Sporting Goods Shop	Sandwich Place	Bar
5	Camden	13.579816	0	Hotel	Coffee Shop	Café	Breakfast Spot	Pizza Place	Hotel Bar	Train Station
6	Croydon	7.160145	0	Coffee Shop	Pub	Gym / Fitness Center	Asian Restaurant	Hotel Bar	Brewery	Burger Joint
7	Ealing	7.202438	0	Coffee Shop	Vietnamese Restaurant	Park	Clothing Store	Italian Restaurant	Pizza Place	Pub
13	Harrow	5.405648	0	Indian Restaurant	Coffee Shop	Thai Restaurant	Convenience Store	Platform	Supermarket	Yoga Studio
18	Kensington and Chelsea	12.386939	0	Café	Garden	Gym / Fitness Center	French Restaurant	Modern European Restaurant	Restaurant	Juice Ba

- Cluster 2: Recreation area (Parks, Cafe)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Mo Comm Venue
16	Hounslow	8.239109	1	Park	Café	Bed & Breakfast	Chinese Restaurant	Yoga Studio	Furniture / Home Store	French Restaurant	Food Court

- Cluster 3 : Transportation area (Café, Bus stop…)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Mos Commo Venue
1	Barnet	6.625904	2	Café	Bus Stop	Yoga Studio	Farmers Market	Furniture / Home Store	French Restaurant	Food Court	Flea Market

- Cluster 4 : Catering area (Pubs, Coffee shops, Restaurants …)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Con Ven
2	Bexley	5.789643	3	Pub	Clothing Store	Pharmacy	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Sup
4	Bromley	6.248840	3	Clothing Store	Coffee Shop	Pizza Place	Bar	Burger Joint	Gym / Fitness Center	Wor Stor
8	Enfield	7.457190	3	Coffee Shop	Clothing Store	Optical Shop	Sandwich Place	Shopping Mall	Pub	Boo
9	Greenwich	8.483013	3	Coffee Shop	Supermarket	Fast Food Restaurant	Clothing Store	Pub	Plaza	Hote
10	Hackney	10.514164	3	Pub	Coffee Shop	Cocktail Bar	Café	Bakery	Clothing Store	Groc Stor
11	Hammersmith and Fulham	10.742135	3	Pub	Italian Restaurant	Indian Restaurant	Gastropub	Café	Clothing Store	Japa Rest

- Cluster 5 : Travelling area (Pool,Parks…)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Mo Comm Venue
0	Barking and Dagenham	8.161737	4	Pool	Park	Bus Station	Supermarket	Golf Course	Martial Arts Dojo	Gym / Fitness Center	Discou Store

4. Results:

Upon different analysis, we were able to discover the best neighborhoods based on our criteria of safety and atmosphere. Now we will review all the analysis made in this project before we make a conclusion on which area to live as an international student or invest as a student accommodation builder.

Like mentioned in the beginning, our key criteria of location decision will base on safety and atmosphere.

4.1 Safety

For safety, I normalized crime to population ratio and reversed the score so that 1 is the highest safety score which represents the neighborhood with no crimer per person.

	BoroughName	CrimeToPop	Cluster Labels	Safety
0	Barking and Dagenham	8.161737	4	0.866111
1	Barnet	6.625904	2	0.940721
2	Bexley	5.789643	3	0.981346
3	Brent	8.037733	0	0.872135
4	Bromley	6.248840	3	0.959038

4.2 Atmosphere

For atmosphere, I gave an arbitrary score to each cluster based on personal preference, as preference is not easy to quantify without subjectivity. Highest score was given to Catering area (Cluster 4) which I prefer, and lowest score was given to Travelling area (Cluster 5).

	BoroughName	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	8.161737	4.0	0.866111	0.6
1	Barnet	6.625904	2.0	0.940721	0.7
2	Bexley	5.789643	3.0	0.981346	1.0
3	Brent	8.037733	0.0	0.872135	0.9
4	Bromley	6.248840	3.0	0.959038	1.0
5	Camden	13.579816	0.0	0.602904	0.9
6	Croydon	7.160145	0.0	0.914768	0.9
7	Ealing	7.202438	0.0	0.912713	0.9
8	Enfield	7.457190	3.0	0.900337	1.0
9	Greenwich	8.483013	3.0	0.850503	1.0
10	Hackney	10.514164	3.0	0.751831	1.0

4.3 Final Score

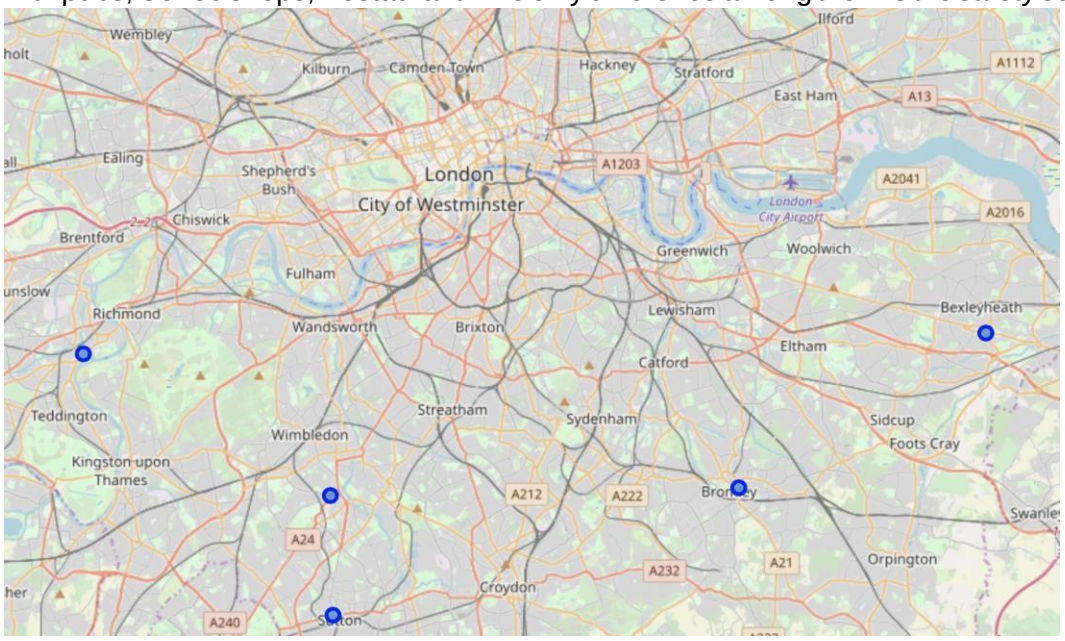
Then by adding up the two scores, we were able to find the top 5 neighborhood. Among them, Sutton has the best score of 1.998.

	BoroughName	Safety	Atmosphere	Score
27	Sutton	0.998688	1.0	1.998688
25	Richmond upon Thames	0.993863	1.0	1.993863
2	Bexley	0.981346	1.0	1.981346
22	Merton	0.981277	1.0	1.981277
4	Bromley	0.959038	1.0	1.959038

5. Conclusion

5.1 Final report of analysis

Given this analysis, we have found that the top five boroughs below to build a student dormitory, based on safety and atmosphere of the neighborhood. The top five boroughs all belong to the catering area with pubs, Coffee shops, Restaurant. The only difference among them is the safety score.



5.2 Limitations and recommendation for future study

However, when we map the top five neighborhoods to live in, it is easily noticeable that they are all located in far out suburbs. This is due to many limitations this research holds.

In this case, we only take crime rate and atmosphere into consideration, while there are also a lot of other factors determining the living quality of dormitory. In addition, we ignored the severity of crime for the purpose of simplicity, which means serious crime like homicide is treated the same as the lit crime like theft.

To get rid of the limitation, we need to acquire further data such as the housing price, the population density, and the availability of public transportation tools. In addition, we may assign different weight to the crime in terms of its severity. Nevertheless, this research is still enlightening and provide us a way to evaluate the factors of dormitory and explore its neighborhoods.

Despite some limitations, this research was still enjoyable in that we were to explore the neighborhoods in depth.

References:

- “London Recorded Crime : Geographic Breakdown”, London Datastore
- “List of London Boroughs”, Wikipedia
- Foursquare API
- “The Economic Impact of London’s International Studnets”, London & Partners (2018)
- Lecture notes from IBM Professional Data Science Specialization, Course