IBM Professional Data Science Specialization

Coursera Applied Data Science Capstone

The Battle of Neighborhoods

Finding the best place in London to build a student accommodation

By Sidharth Sankar Sabat

Date: 15th Aug 2020

Table of Contents

1. Introduction 3
1.2 Problem Description 3
2. Data Acquisition and Data Preprocessing 3
2.1 London Recorded Crime Dataset 4
2.2 List of London Boroughs 5
2.3 Foursquare API 5
3. Data Analysis and Clustering 6
3.1 Exploratory Analysis 6
3.2 Cluster Analysis 8
4. Results 11
4.1 Safety 11
4.2 Atmosphere 11
4.3 Final score 11
5. Conclusion 12
5.1 Final result of analysis 12
5.2 Limitations and recommendation for future study 13
6. References 13

1. Introduction

London is a popular destination for higher education. Students in large numbers from across the globe comes every year for higher education to London.

One of the major concerns for international students when moving to a new city like London is to find an optimal and best accommodation. Student halls are the most reliable means of housing for students, but it is not very easy to find a place in one of the big cities like London as they are in high demand. Hence the goal of the project is to explore different neighborhoods of London and find the best area to build a new student hall for international students in London to solve this persistent problem and to find a new business opportunity. 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.

From the student perspective, a lot of factors come into play when finding the best accommodation, including location and rent. However, this study will focus only on the safety and the general atmosphere of the neighborhood. Distance to universities are also an important factor in choosing a student hall, but as student halls accept students from different universities, it will be disregarded in this project.

1.2 Problem Description

Student halls are the most reliable means of housing for students, especially if one is completely new to the city and is not familiar with how rental contracts work. However, as they are in high demand in the international cities like in London, it is not easy to secure a place in one of those. Therefore, this project aims to explore different neighborhoods of London and find the best area to build a new student hall for international students to solve this persistent problem and to find a new business opportunity as well.

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.

From the student perspective, a lot of factors come into play when finding the best accommodation, including location and rent. In this project, however, the study will only focus on the safety and the general atmosphere of the neighborhood for simplification. Distance to universities are also an important factor in choosing a student hall, but as student halls accept students from different universities, it will be disregarded in this project.

2. Data Acquisition and Data Preprocessing

In this project, three different datasets will be used to solve the problem - London Recorded Crime, List of London Boroughs, and Foursquare API. After acquiring those from respective sources, they will be wrangled and cleansed into more useful forms for further analysis.

2.1 London Recorded Crime Dataset

Shown below is the London crime records classified by boroughs and crime type in the most recent 24 months. It consists of 1568 observations and 27 columns. It was acquired directly from London Datastore.

	MajorText	MinorText	LookUp_BoroughName	201807	201808	201809	201810	201811	201812	201901	...	201909	201910	201911	201912	202001	
0	Arson and Criminal Damage	Arson	Barking and Dagenham	6	5	3	8	5	1	5	...	6	9	8	6	4	
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	127	101	107	132	105	88	97	...	109	109	97	121	97	
2	Burglary	Burglary - Business and Community	Barking and Dagenham	30	18	33	32	39	33	45	...	37	30	30	25	31	
3	Burglary	Burglary - Residential	Barking and Dagenham	94	84	99	94	106	164	114	...	80	97	114	130	116	
4	Drug Offences	Drug Trafficking	Barking and Dagenham	8	7	10	7	7	4	5	...	7	8	13	3	14	

For further analysis, the number of crimes were calculated into monthly averages and Sum of all incidents. Crime categories are not considered in this research for simplification. The above dataset is simplified to a cleaned data frame shown as below. The below table is the values of Borough Names with the most number of incidents (In descending order).

	BoroughName	SumOfAllIncidents	MonthlyAverageSum
32	Westminster	142096	5920.666667
27	Southwark	73726	3071.916667
5	Camden	72764	3031.833333
24	Newham	70182	2924.250000
20	Lambeth	68701	2862.541667

2.2 List of London Boroughs

The second dataset used in this project is the information of boroughs in London, web scrapped from Wikipedia.

List of boroughs and local authorities [edit]

Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
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	25
Barnet			Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369,088	51.6252°N 0.1517°W	31
Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51.4549°N 0.1505°E	23
Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51.5588°N 0.2817°W	12
Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51.4039°N 0.0198°E	20
Camden	✓		Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	229,719	51.5290°N 0.1255°W	11
Croydon			Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	372,752	51.3714°N 0.0977°W	19
Ealing			Ealing London Borough Council	Labour	Perceval House, 14-16 Uxbridge Road	21.44	342,494	51.5130°N 0.3089°W	13
Enfield			Enfield London Borough Council	Labour	Civic Centre, Silver Street	31.74	320,524	51.6538°N 0.0799°W	30
Greenwich [note 2]	✓ [note 3]	Royal	Greenwich London Borough Council	Labour	Woolwich Town Hall, Wellington Street	18.28	264,008	51.4892°N 0.0648°E	22
Hackney	✓		Hackney London Borough Council	Labour	Hackney Town Hall, Mare Street	7.36	257,379	51.5450°N 0.0553°W	9

From the original data, population and coordinates data are being used. Population data are used to calculate the ratio of reported crime to population for better comparison and coordinates data are 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

Finally, Foursquare API is used to call the top 50 popular venues in each neighborhood. This is done using the 'explore' function of requesting URL. Below is a sample of the dataset. It has 1142 observations and 7 columns.

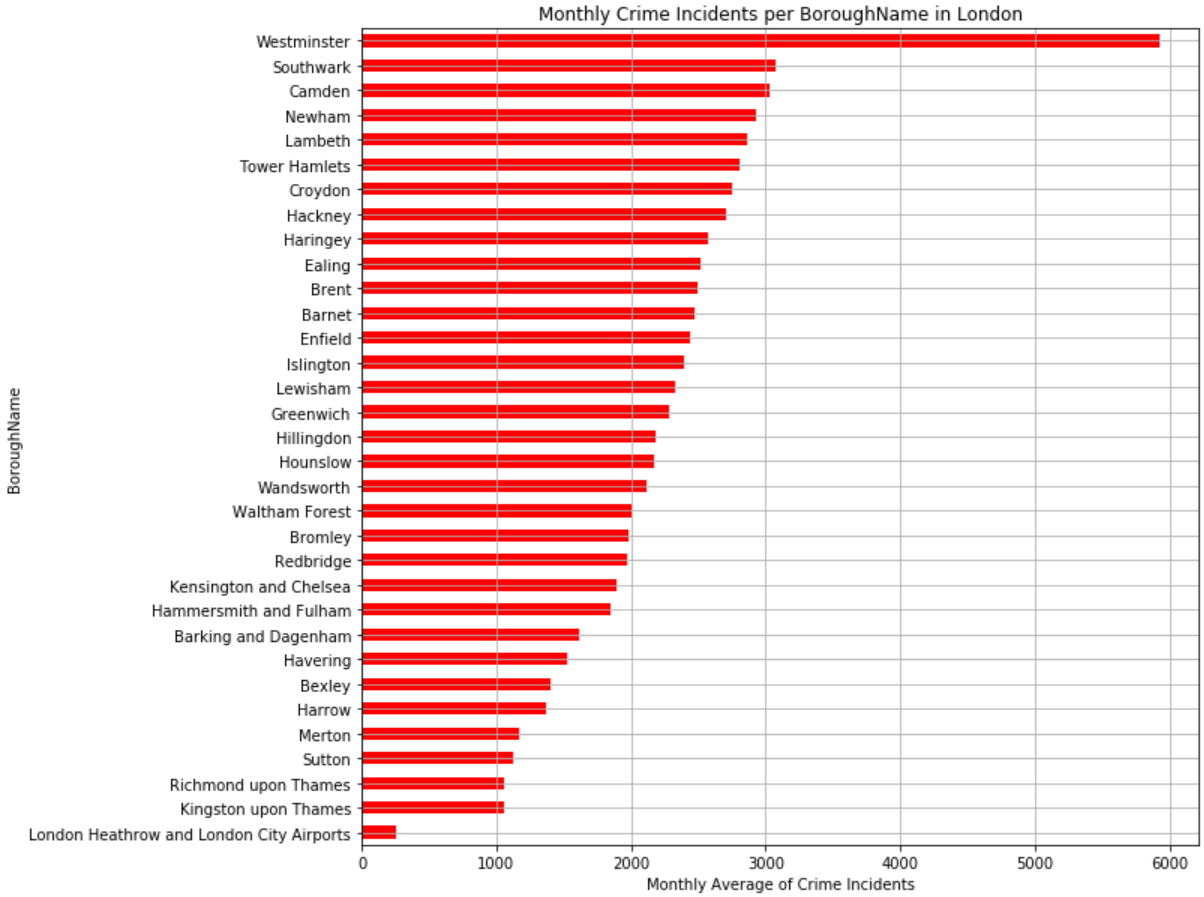
	BoroughName	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.5607	0.1557	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.5607	0.1557	Crowlands Heath Golf Course	51.562457	0.155818	Golf Course
2	Barking and Dagenham	51.5607	0.1557	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center
3	Barking and Dagenham	51.5607	0.1557	Robert Clack Leisure Centre	51.560808	0.152704	Martial Arts Dojo
4	Barking and Dagenham	51.5607	0.1557	Morrisons Becontree Heath	51.559774	0.148752	Supermarket

3. Data Analysis and Clustering

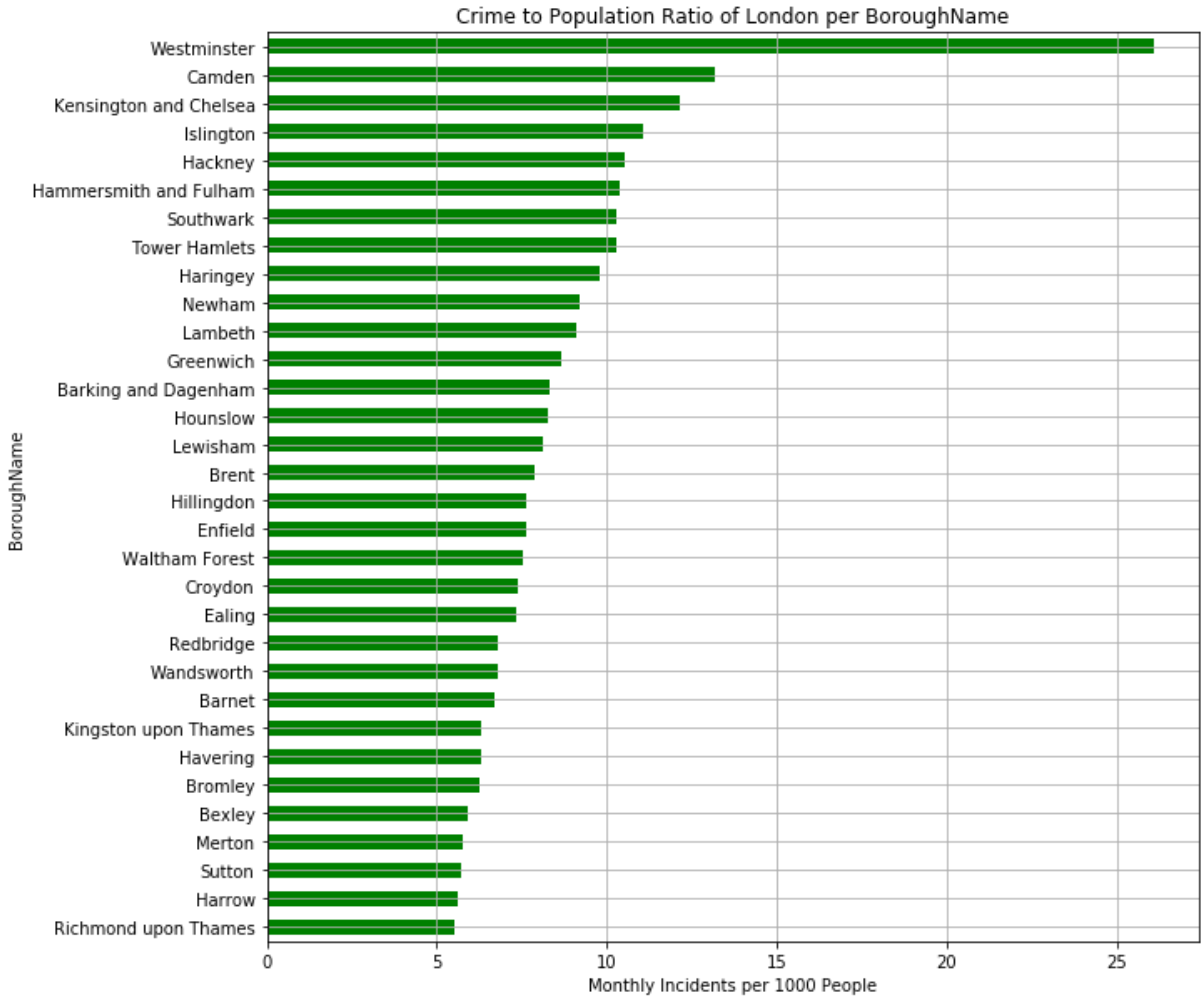
3.1 Exploratory Analysis

After cleansing datasets, some of visualizations created as below to demonstrate the data in more meaningful way.

The below bar chart displays Monthly crime incidents per Boroughs in London. Westminster has the biggest number of reported crimes, followed by Newham, Camden and Southwark.

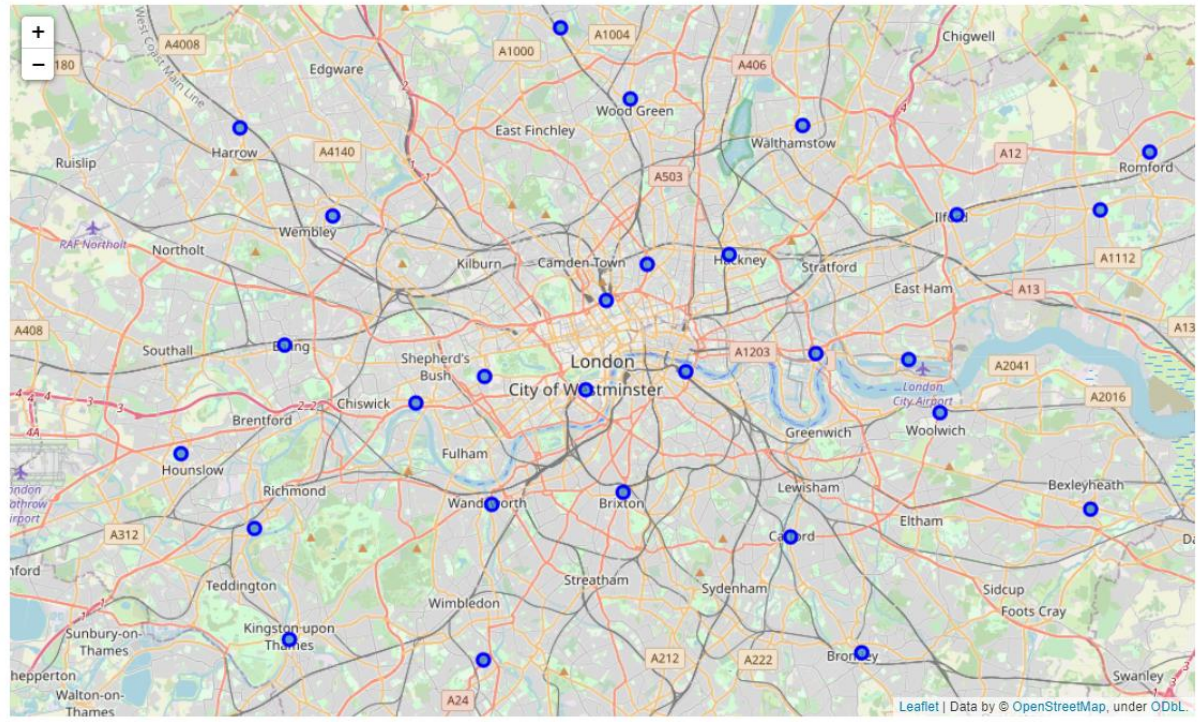


However, as different boroughs have different sizes of population, it is not wise to directly compare the absolute number of incidents. Instead, we should consider the ratio of crime incidents to people. Thus, the population data is used to calculate the number of recorded crimes per 1000 people in each borrow as shown in the below bar chart.



It is noticeable that Westminster and Camden remain the top two most dangerous places in terms of recorded crime ratio to population. However, ranks of Boroughs have been changed from the third place.

And before commencing further with the analysis, it would be wise to analyze the locations of each borough to get an idea of the Greater London area. Below map is giving an idea of the locations of each Borough.



3.2 Cluster Analysis

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

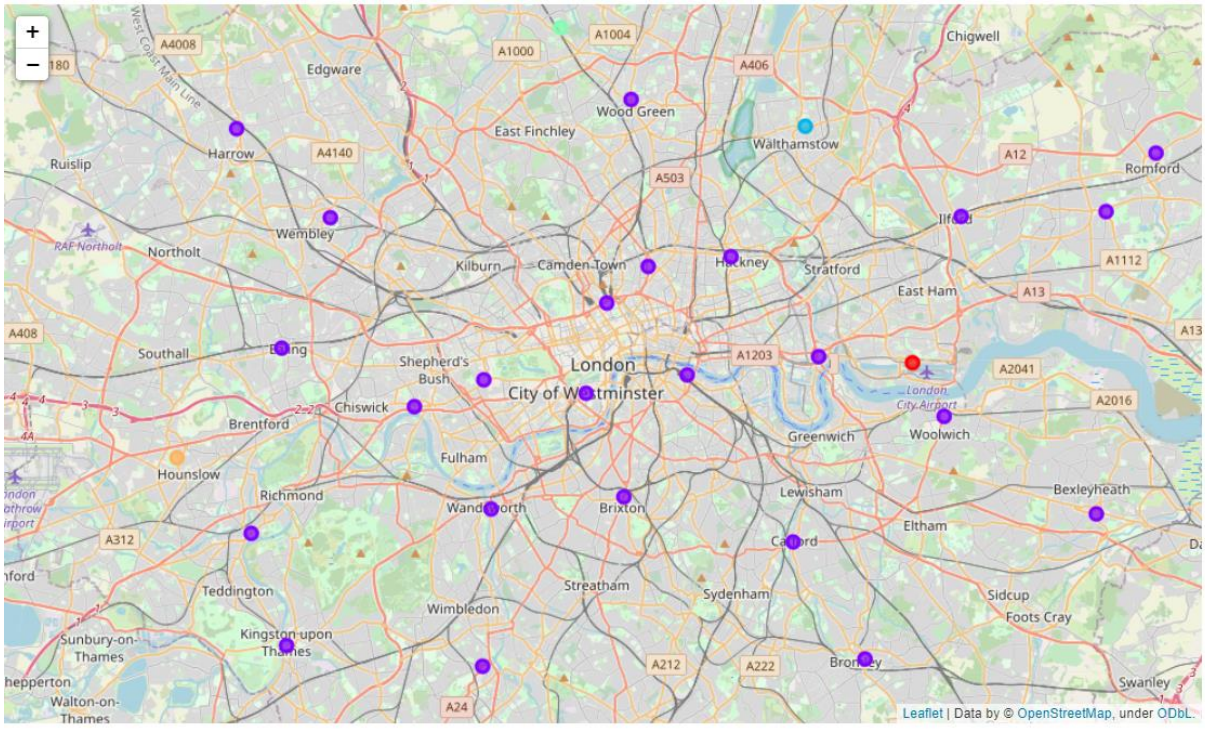
As the first step of cluster analysis, one hot encoding was conducted to give binary values to each venue categories.

	BoroughName	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	...	Turkish Restaurant	Used Bookstore	Vegetarian / Vegan Restaurant	Video Game Store	Viet Re
0	Barking and Dagenham	0	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	0	
2	Barking and Dagenham	0	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	0	
4	Barking and Dagenham	0	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 attain 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	Gym / Fitness Center	Bus Station	Supermarket	Martial Arts School	Park	Golf Course	Dim Sum Restaurant	English Restaurant	Fast Food Restaurant
1	Barnet	Café	Home Service	Bus Stop	Yoga Studio	English Restaurant	Flea Market	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market
2	Bexley	Coffee Shop	Pub	Clothing Store	Supermarket	Fast Food Restaurant	Pharmacy	Bakery	Video Game Store	American Restaurant	Plaza
3	Brent	Coffee Shop	Hotel	Grocery Store	Clothing Store	Bar	Sandwich Place	Italian Restaurant	Sporting Goods Shop	American Restaurant	Ice Cream Shop
4	Bromley	Clothing Store	Coffee Shop	Pizza Place	Burger Joint	Gym / Fitness Center	Bar	Fast Food Restaurant	Furniture / Home Store	Burrito Place	Café

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



Below is the sample of the cluster data after labeled with their cluster labels.

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	1	Pool	Bus Station	Supermarket	Gym / Fitness Center	Park
1	Barnet	3	Café	Home Service	Bus Stop	Yoga Studio	Electronics Store
2	Bexley	1	Coffee Shop	Clothing Store	Pub	Pharmacy	Supermarket
3	Brent	1	Coffee Shop	Hotel	Grocery Store	Clothing Store	Sandwich Place
4	Bromley	1	Clothing Store	Coffee Shop	Pizza Place	Burger Joint	Gym / Fitness Center
5	Camden	1	Hotel	Café	Coffee Shop	Breakfast Spot	Hotel Bar
6	Croydon	1	Pub	Coffee Shop	Portuguese Restaurant	Asian Restaurant	Park
7	Ealing	1	Coffee Shop	Italian Restaurant	Pub	Vietnamese Restaurant	Clothing Store
8	Enfield	1	Clothing Store	Coffee Shop	Supermarket	Pub	Optical Shop
9	Greenwich	1	Coffee Shop	Pub	Clothing Store	Fast Food Restaurant	Plaza

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

- Cluster 0: Traveler area (Hotel, Restaurant, Airport)

	BoroughName	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
23	Newham	0	Hotel	Chinese Restaurant	Airport	Airport Lounge	Airport Service	Pharmacy	Sandwich Place	Duty-free Shop	Rafting	Escape Room

- Cluster 1: Lively area (Pool, Bus Station, Pub, Coffee Shop, Supermarket, Clothing Store)

	BoroughName	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking and Dagenham	1	Pool	Gym / Fitness Center	Bus Station	Supermarket	Martial Arts School	Park	Golf Course	Dim Sum Restaurant	English Restaurant	Fast Food Restaurant
2	Bexley	1	Coffee Shop	Pub	Clothing Store	Supermarket	Fast Food Restaurant	Pharmacy	Bakery	Video Game Store	American Restaurant	Plaza
3	Brent	1	Coffee Shop	Hotel	Grocery Store	Clothing Store	Bar	Sandwich Place	Italian Restaurant	Sporting Goods Shop	American Restaurant	Ice Cream Shop
4	Bromley	1	Clothing Store	Coffee Shop	Pizza Place	Burger Joint	Gym / Fitness Center	Bar	Fast Food Restaurant	Furniture / Home Store	Burrito Place	Café
5	Camden	1	Hotel	Café	Coffee Shop	Garden	Hotel Bar	Breakfast Spot	Pizza Place	Italian Restaurant	Train Station	Burger Joint

- Cluster 2: Fitness Area (Pool, Gym, Pub)

	BoroughName	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Waltham Forest	2	Pub	Pool	Gym / Fitness Center	Construction & Landscaping	Concert Hall	Coffee Shop	Garden Center	Beer Store	Tea Room	Grocery Store

- Cluster 3: Quiet area (Cafe, Bus Stop, Home Service)

	BoroughName	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Barnet	3	Café	Home Service	Bus Stop	Yoga Studio	Electronics Store	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market	Event Space

- Cluster 4: Refreshing area (Park, Cafe)

	BoroughName	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
16	Hounslow	4	Metro Station	Park	Café	Bed & Breakfast	Yoga Studio	English Restaurant	Fish Market	Film Studio	Fast Food Restaurant	Farmers Market

4. Results

Based on different analysis, we have analyzed the best neighborhoods based on criteria of safety and atmosphere. Now let's review all the analysis made in this project before making a conclusion on which area to live as an international student or invest as a student accommodation builder.

Like mentioned in the beginning, the key criteria for deciding a location will be based on safety and atmosphere.

4.1 Safety

For safety, we normalized crime to population ratio and reversed the score so that 1 represents the neighborhood with least crime per person.

	BoroughName	CrimeToPop	Cluster Labels	Safety
0	Barking and Dagenham	8.328317	1	0.861862
1	Barnet	6.684607	3	0.942858
2	Bexley	5.945264	1	0.979291
3	Brent	7.844676	1	0.885694
4	Bromley	6.229573	1	0.965281

4.2 Atmosphere

For atmosphere, we gave an arbitrary score to each cluster based on personal preference.

	BoroughName	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	8.328317	1.0	0.861862	0.9
1	Barnet	6.684607	3.0	0.942858	0.7
2	Bexley	5.945264	1.0	0.979291	0.9
3	Brent	7.844676	1.0	0.885694	0.9
4	Bromley	6.229573	1.0	0.965281	0.9

4.3 Final score

Then by taking both the above two scores into consideration, we found the best neighborhood scoring 1.9 points, i.e. **Richmond upon Thames.**

	BoroughName	Safety	Atmosphere	Score
25	Richmond upon Thames	1.000000	0.9	1.900000
29	Waltham Forest	0.899685	1.0	1.899685
13	Harrow	0.995243	0.9	1.895243
27	Sutton	0.990822	0.9	1.890822
22	Merton	0.987051	0.9	1.887051

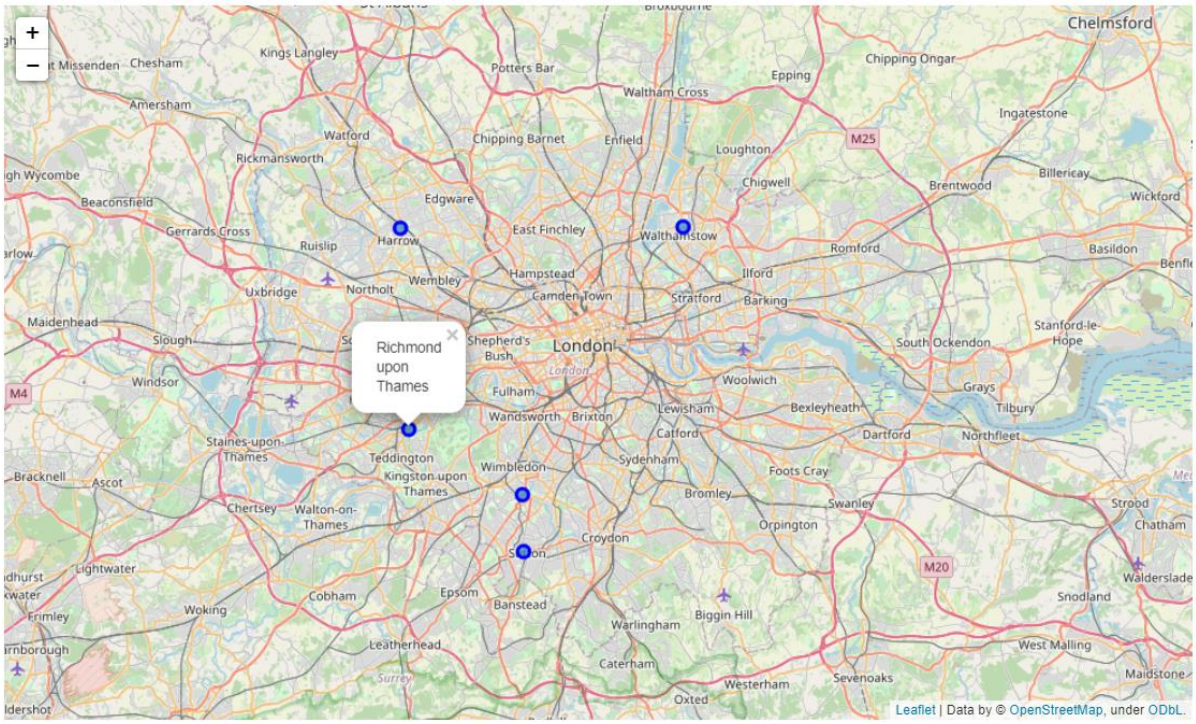
5. Conclusion

5.1 Final result of analysis

Based on the analysis so far and based on safety and atmosphere of the neighborhood, it is clear that below five boroughs are the best places to build a student hall in London.

	BoroughName	SumOfAllIncidents	MonthlyAverageSum	Population	Latitude	Longitude	CrimeToPop
25	Richmond upon Thames	25375	1057.291667	191365	51.4479	-0.3260	5.525000
29	Waltham Forest	48231	2009.625000	265797	51.5908	-0.0134	7.560751
13	Harrow	32835	1368.125000	243372	51.5898	-0.3346	5.621538
27	Sutton	26854	1118.916667	195914	51.3618	-0.1945	5.711264
22	Merton	28229	1176.208333	203223	51.4014	-0.1958	5.787772

The top five boroughs all belong to the lively Area and fitness cluster area with many pubs, restaurants, pub, pool, Gym/Fitness Centre and coffee shops. Therefore, what differentiates them is the safety score, which was calculated from monthly recorded crimes per 1000 people.



5.2 Limitations and recommendation for future study

However, it obviously seems strange that the best places to live in London are all far out suburbs. This is due to limitations this research holds. Among numerous factors that determine a good neighborhood, the factors taken into consideration are based on simplified factors, what type of venues are popular and how many crime incidents are recorded. This means that serious crimes like homicide was treated the same as a comparatively petty crime like shoplifting. Moreover, the number of stores in the neighborhood may be as important as what type of stores there are.

To overcome the limitations of this study, some more data would be required such as distance to city center and housing prices etc. Also, taking crime categories into a consideration and weighing them differently may be helpful. However, this research is still meaningful in a way to explore the neighborhoods in depth and to decide the best place to build a student accommodation.

6. References

- London Recorded Crime: Geographic Breakdown", London Datastore
- List of London Boroughs, Wikipedia
- Foursquare API
- Notes from IBM Professional Data Science Specialization, Coursera