IBM Professional Data Scientist Specialization
The Battle of Neighborhoods

Finding the best area for a student in London

1. Introduction

1.1 Background

London is a popular destination for higher education where diverse students from all around the world gather to study.
According to the data published by the Higher Education Statistical Agency (HESA), in the academic year 2016-17, there were 2317880 students at UK higher education institutions. As of 2016/17, according to official international enrollment statistics, 442,750 international students were attending university in the UK. Of which 112205 students were in London, which make up 24 percent of immigrant 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.

1.2 Problem

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, 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 student hall for international students to solve this persistent problem as well as 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. In this project, however, the study will only focus on the safety and the general atmosphere of the neighborhood for simplification. Distances 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 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 them from original and reliable sources, they will be wrangled and cleansed into more useful forms for our further analysis.

2.1 London Recorded Crime

	MajorText	MinorText	LookUp_BoroughName	201707	201708	201709	201710	201711	201712	201801	...	201809	201810	201811
0	Arson and Criminal Damage	Arson	Barking and Dagenham	2	5	8	7	7	4	2	...	3	8	5
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	143	169	134	132	108	119	135	...	107	131	105
2	Burglary	Burglary - Business and Community	Barking and Dagenham	42	30	25	23	27	21	38	...	33	32	39
3	Burglary	Burglary - Residential	Barking and Dagenham	95	83	81	122	88	124	143	...	99	94	106
4	Drug Offences	Drug Trafficking	Barking and Dagenham	7	1	6	7	5	6	4	...	9	6	7

Shown above is London crime records classified by boroughs and crime type in the recent 24 months. It consists of 1584 Observations and 27 columns. It was acquired directly from London Data Store.

For further analysis, the numbers of crimes were calculated into monthly averages, and crime categories were not considered in this research for simplification. This process turned the above dataset to a simple one as below

	LookUp_BoroughName	MonthlyAverage
0	Barking and Dagenham	1574.916667
1	Barnet	2395.375000
2	Bexley	1329.708333
3	Brent	2553.958333
4	Bromley	1938.791667

.

2.2 List of London Boroughs

The second dataset used was information on boroughs in London, scrapped from Wikipedia.

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	North London Business Park, Oakleigh Road South	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

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, and 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

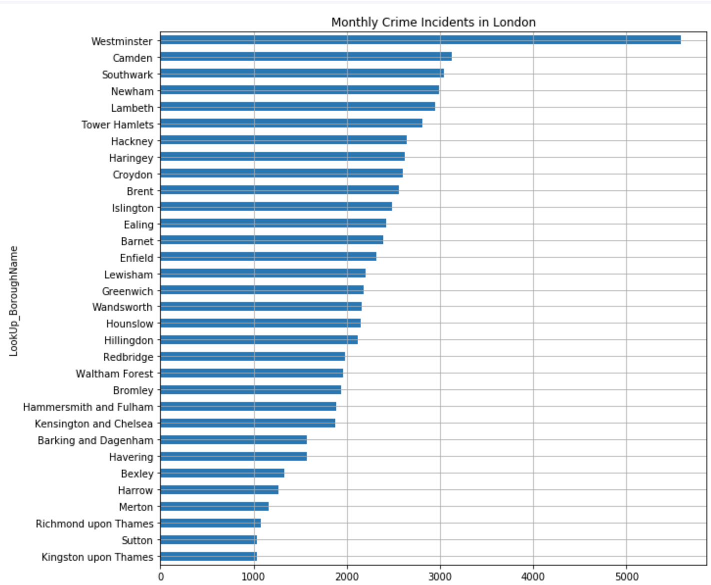
Finally, Foursquare API was used to call the top 50 popular venues in each neighborhood. This was done using the 'explore' function of requesting URL. We were able to acquire data looking like this.

	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	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center
2	Barking and Dagenham	51.5607	0.1557	Crowlands Heath Golf Course	51.562457	0.155818	Golf Course
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. Methodology

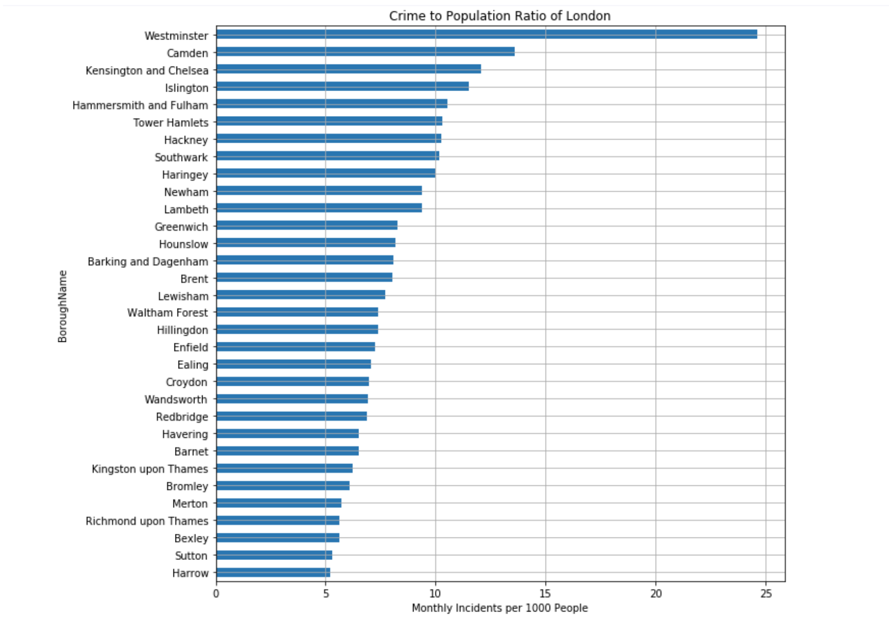
3.1 Exploratory Analysis

After cleansing datasets to more useful forms, we created some visualizations to interpret the data we have better.



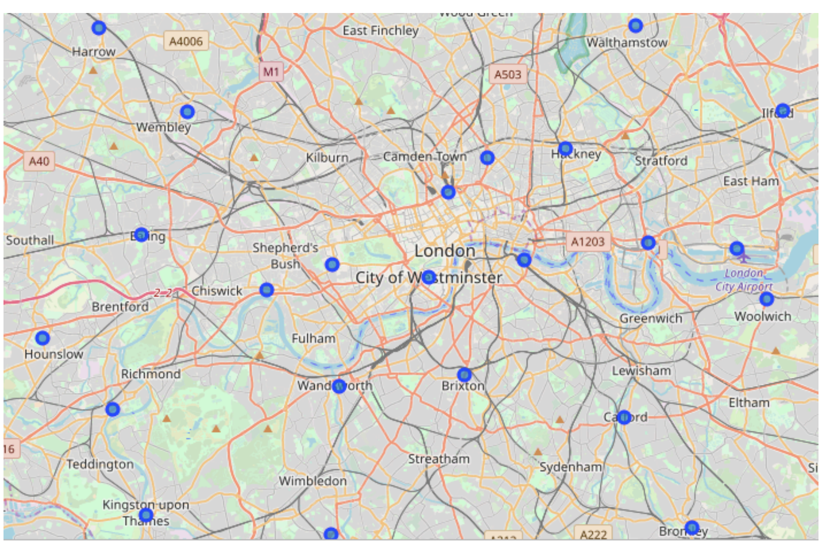
This is a bar chart displaying boroughs in descending order of monthly crime incidents. Westminster has the biggest number of reported crime, followed by Camden, Southwark and Newham.

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, I have used the population to calculate the number of recorded crimes per 1000 people in each borough.



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

And before commencing further with the analysis, I have observed the locations of each borough to get an idea of the Greater London area.



3.2 Cluster Analysis

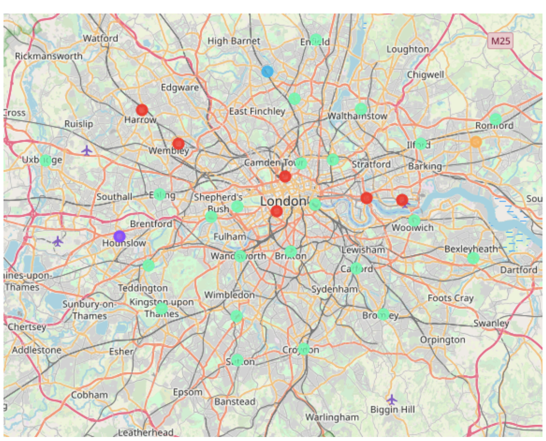
Afterwards, K-means clustering was conducted in order to group the boroughs according to what venues they have using Foursquare data, in order to feel the atmosphere of each borough.
As the first step of cluster analysis, one hot encoding was conducted to give binary values to each venue categories.

BoroughName	African Restaurant	Airport	Airport Lounge	Airport Service	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	...	Turkish Restaurant	Vegetarian / Vegan Restaurant	
Barking and Dagenham	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.00	0.000000	0.0
Barnet	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.00	0.000000	0.0
Bexley	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.00	0.000000	0.0
Brent	0.000	0.000000	0.000000	0.000000	0.04	0.00	0.000000	0.00	0.020000	...	0.00	0.000000	0.0
Bromley	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.025000	...	0.00	0.000000	0.0
Camden	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.02	0.000000	0.0
Croydon	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.054054	...	0.00	0.000000	0.0
Ealing	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.020000	0.00	0.020000	...	0.00	0.000000	0.0
Enfield	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.04	0.000000	0.0
Greenwich	0.025	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.025000	...	0.00	0.000000	0.0
Hackney	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.02	0.020000	0.0
Hammersmith and Fulham	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000	...	0.00	0.000000	0.0
Haringey	0.000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.00	0.000000		0.00	0.000000	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	Martial Arts Dojo	Pool	Bus Station	Supermarket	Park	Gym / Fitness Center	Golf Course	Yoga Studio	Flea Market	Fish Market
1	Barnet	Café	Bus Stop	Yoga Studio	Farmers Market	French Restaurant	Food Court	Food	Flea Market	Fish Market	Fish & Chips Shop
2	Bexley	Coffee Shop	Pub	Italian Restaurant	Supermarket	Fast Food Restaurant	Clothing Store	Hotel	Grocery Store	Furniture / Home Store	Multiplex
3	Brent	Coffee Shop	Hotel	Clothing Store	Sporting Goods Shop	Bar	Grocery Store	Sandwich Place	Italian Restaurant	American Restaurant	Café
4	Bromley	Coffee Shop	Clothing Store	Pizza Place	Burger Joint	Bar	Gym / Fitness Center	Noodle House	Stationery 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 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 0 : Traveler area (B&B, hotel, airport...)

	MonthlyAverage	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 Most Common Venue
3	2553.958333	8.049947	0	Coffee Shop	Hotel	Clothing Store	Sporting Goods Shop	Bar	Grocery Store	Sandwich Place	Italian Restaurant
5	3127.916667	13.616273	0	Hotel	Coffee Shop	Café	Train Station	Hotel Bar	Pizza Place	Breakfast Spot	Bookstore
13	1266.500000	5.203968	0	Coffee Shop	Indie Movie Theater	Indian Restaurant	Convenience Store	Platform	Supermarket	Exhibit	Food
23	2989.250000	9.393452	0	Hotel	Airport	Airport Lounge	Airport Service	Chinese Restaurant	Pharmacy	Sandwich Place	Rafting
28	2810.458333	10.298869	0	Italian Restaurant	Hotel	Coffee Shop	Café	Gym / Fitness Center	Grocery Store	Outdoor Sculpture	Pizza Place
31	5584.916667	24.620402	0	Hotel	Sandwich Place	Coffee Shop	Theater	Sushi Restaurant	Hotel Bar	Juice Bar	Fast Food Restaurant

Cluster 1 : Lively area (B&B, Cafe, ...)

	MonthlyAverage	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 Most Common Venue	9th Most Common Venue
16	2147.958333	8.185598	1	Bed & Breakfast	Pizza Place	Café	Park	Yoga Studio	Falafel Restaurant	Food Court	Food	Flea Market

Cluster 2 : Quiet area (Bus stop, yoga studio..)

	MonthlyAverage	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	2395.375	6.489983	2	Café	Bus Stop	Yoga Studio	Farmers Market	French Restaurant	Food Court	Food	Flea Market	Fish Market	Fish & Chips Shop

Cluster 3 : Busy area (Coffee shop, clothing store...)

	MonthlyAverage	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 Most Common Venue
2	1329.708333	5.618003	3	Coffee Shop	Pub	Italian Restaurant	Supermarket	Fast Food Restaurant	Clothing Store	Hotel	Grocery Store
4	1938.791667	6.098766	3	Coffee Shop	Clothing Store	Pizza Place	Burger Joint	Bar	Gym / Fitness Center	Noodle House	Stationery Store
6	2604.333333	6.986772	3	Coffee Shop	Pub	Asian Restaurant	Gym / Fitness Center	Spanish Restaurant	Malay Restaurant	Bookstore	Clothing Store
7	2425.333333	7.081389	3	Coffee Shop	Italian Restaurant	Burger Joint	Park	Bar	Bakery	Pizza Place	Vietnamese Restaurant
8	2322.166667	7.244907	3	Coffee Shop	Sandwich Place	Clothing Store	Pub	Bookstore	Supermarket	Mobile Phone Shop	Café
9	2185.250000	8.277211	3	Coffee Shop	Fast Food Restaurant	Sandwich Place	Pub	Hotel	Supermarket	Grocery Store	Clothing Store

Cluster 4 : Healthy area (gym, park, pool, ...)

	MonthlyAverage	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 Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1574.916667	8.103424	4	Martial Arts Dojo	Pool	Bus Station	Supermarket	Park	Gym / Fitness Center	Golf Course	Yoga Studio	Flea Market	Fish Market

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 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	8.103424	4	0.850670
1	Barnet	6.489983	2	0.933767
2	Bexley	5.618003	3	0.978676
3	Brent	8.049947	0	0.853424
4	Bromley	6.098766	3	0.953915

4.2 Atmosphere

For atmosphere, we 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 Busy area (Cluster 3), which I prefer, and lowest score was given to Traveler area (Cluster 0).

	BoroughName	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	8.103424	4	0.850670	0
1	Barnet	6.489983	2	0.933767	0
2	Bexley	5.618003	3	0.978676	0
3	Brent	8.049947	0	0.853424	0
4	Bromley	6.098766	3	0.953915	0

4.3 Final score

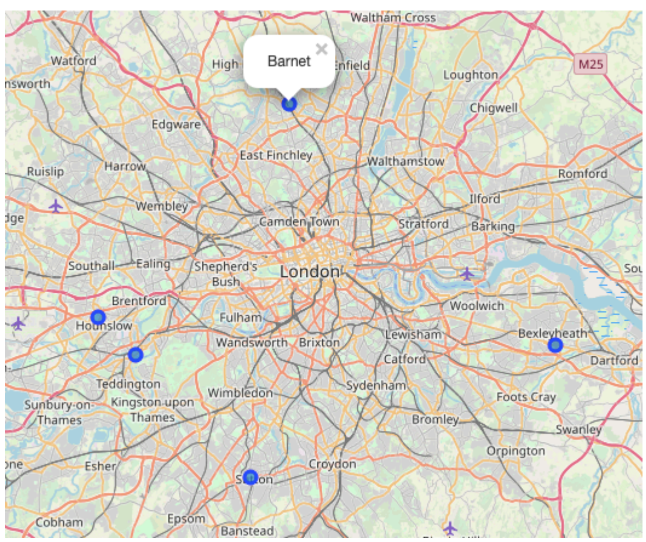
By adding the two scores, the best neighborhood scoring highest 1.93 points, Barnet.

	BoroughName	Safety	Atmosphere	Score
1	Barnet	0.933767	1.0	1.933767
16	Hounslow	0.846438	0.9	1.746438
27	Sutton	0.995057	0.7	1.695057
2	Bexley	0.978676	0.7	1.678676
25	Richmond upon Thames	0.977837	0.7	1.677837
22	Merton	0.974568	0.7	1.674568
4	Bromley	0.953915	0.7	1.653915
19	Kingston upon Thames	0.947670	0.7	1.647670
14	Havering	0.933193	0.7	1.633193
24	Redbridge	0.913862	0.7	1.613862
30	Wandsworth	0.910427	0.7	1.610427
6	Croydon	0.908181	0.7	1.608181
7	Ealing	0.903308	0.7	1.603308

5. Conclusion

5.1 Final result of analysis

From this analysis, we have found that the five boroughs below are the best places to build a student hall, based on safety and atmosphere of the neighborhood. The top five boroughs all belong to the Busy Area cluster, with many coffee shops and clothing stores. 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, 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.
Among numerous factors that determine a good neighborhood, we only took into consideration

what type of venues are popular and how many crime incidents are recorded for the sake of simplification. 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, we will need further data such as distance to city center, housing prices or ratio of the number of stores to population. Also, taking crime categories into factor and weighting them differently may be helpful.
Despite some limitations, this research was still enjoyable in that we were able 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, Coursera