

THE BATTLE OF NEIGHBORHOODS

An IBM Coursera Applied Data Science Capstone Project

By Sidharth Sankar Sabat
15th Aug 2020



OBJECTIVE

- The main objective of this Coursera Applied Data Science Capstone project is to find out best place in London to build a student accommodation.

PROJECT BACKGROUND

- London is a popular destination for higher education. Students in large numbers from across the globe come 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.

DATA ACQUISITION

1. London Recorded Crime:

- Following datasets are used in this project:
 1. London Recorded Crime,
 2. List of London Boroughs,
 3. Foursquare API.
- London crime records by boroughs dataset contains data of most recent 24 months (It consists of 1568 observations and 27 columns)
- Source:** London Datastore
- URL:** https://data.london.gov.uk/dataset/recorded_crime_summary
- CSV:** 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	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	

DATA ACQUISITION

2. List of London Boroughs

- **Description:** Information on 32 boroughs of London. 'Population' and 'Coordinates' are used for the project.
- It contains 32 observations with their coordinates and population data.
- **Source:** Wikipedia
- **URL:**
https://en.wikipedia.org/wiki/List_of_London_boroughs

List of boroughs and local authorities [edit]

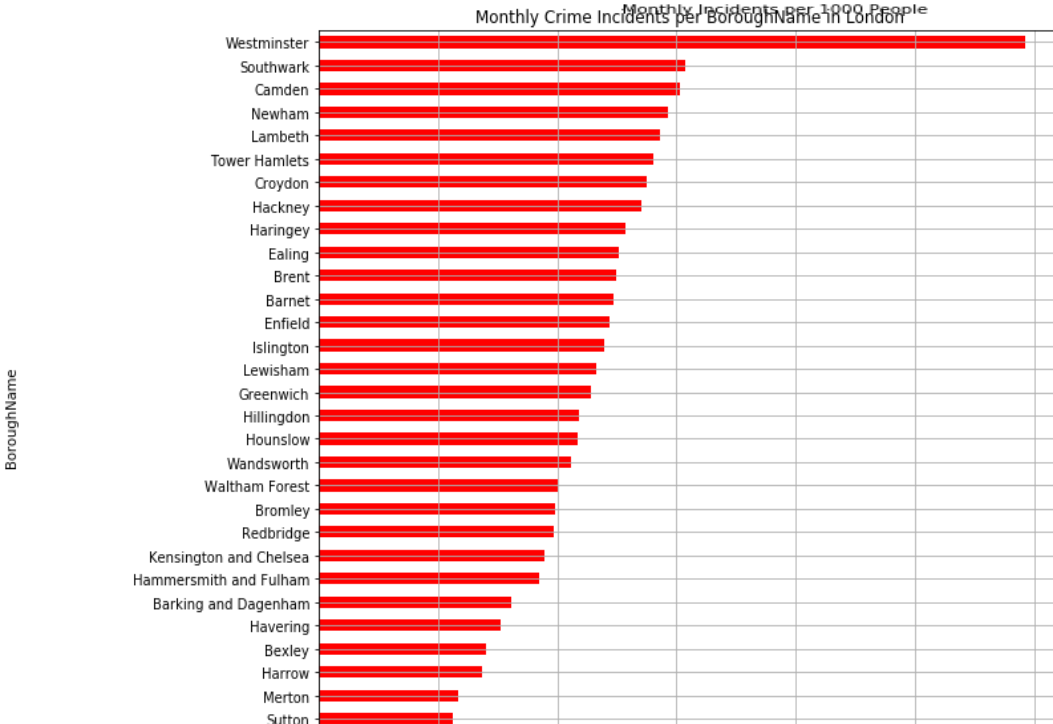
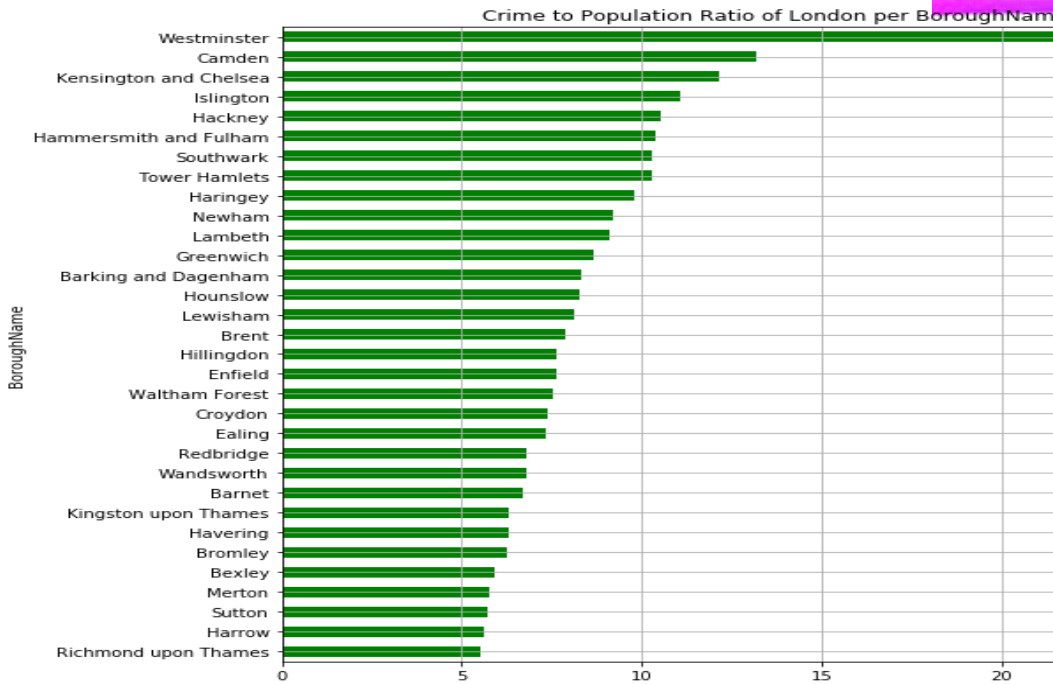
Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) ^[1]	Co-ordinates	Nr. in map
Barking and Dagenham <small>[note 1]</small>			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51.5607°N 0.1557°E﻿ / ﻿	25
Barnet			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

DATA ACQUISITION

3. Foursquare API

- **Description:** List of top 50 popular places in each borough. However, some suburban areas may have less than 50 places registered on Foursquare.
- **Shape and Size:** 1142 observations * 7 Columns
- **Source:** Foursquare API
- **URL:** <https://developer.foursquare.com/>

	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 School
3	Barking and Dagenham	51.5607	0.1557	Morrisons	51.559774	0.148752	Supermarket
4	Barking and Dagenham	51.5607	0.1557	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center
5	Barking and Dagenham	51.5607	0.1557	Becontree Heath Bus Station	51.561065	0.150998	Bus Station
6	Barking and Dagenham	51.5607	0.1557	Dagenham Swimming Pool	51.560946	0.150054	Pool
7	Barnet	51.6252	-0.1517	The Atrium	51.624726	-0.151933	Café
8	Barnet	51.6252	-0.1517	Made Curtains	51.623485	-0.153565	Home Service
9	Barnet	51.6252	-0.1517	Beaconsfield Road (BF)	51.622827	-0.151466	Bus Stop



DATA ANALYSIS AND CLUSTERING

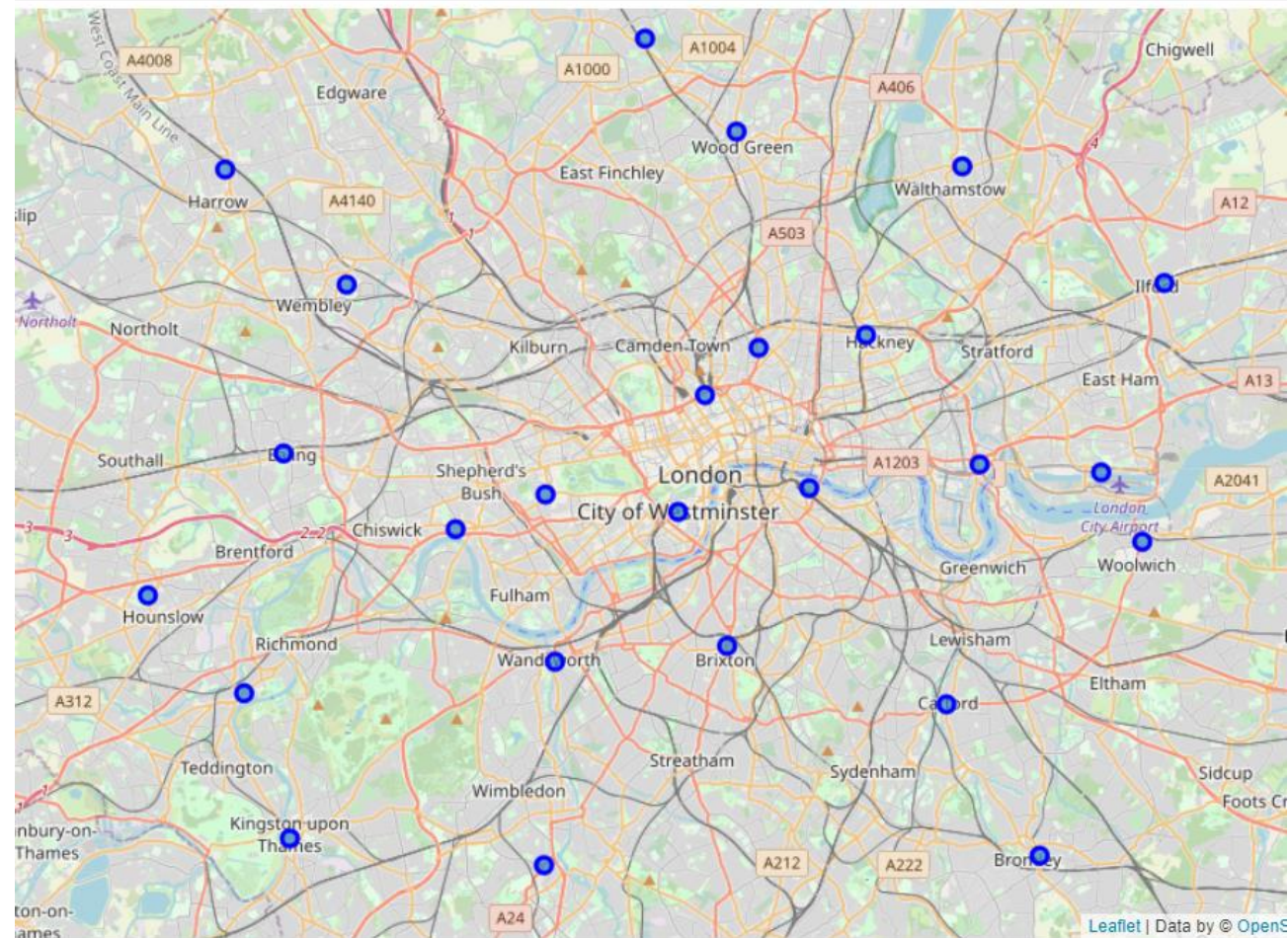
1. Exploratory Analysis: Which borough is more dangerous?

- Westminster has the biggest number of reported crimes, followed by Newham, Camden and Southwark based on Monthly crime incidents per Boroughs.
- Westminster and Camden remain the top two most dangerous places in terms of recorded crime ratio to population.

DATA ANALYSIS AND CLUSTERING

2. Exploratory Analysis: Map of London

- Visualization of London Boroughs through a Map using coordinates data of each borough to get familiar with the Greater London area.



DATA ANALYSIS AND CLUSTERING

----Bexley----

	venue	freq
0	Clothing Store	0.11
1	Coffee Shop	0.11
2	Pub	0.11
3	Supermarket	0.07
4	Fast Food Restaurant	0.07

----Camden----

	venue	freq
0	Café	0.08
1	Hotel	0.08
2	Coffee Shop	0.06
3	Burger Joint	0.04
4	Pizza Place	0.04

3. Cluster Analysis: Popular Venues

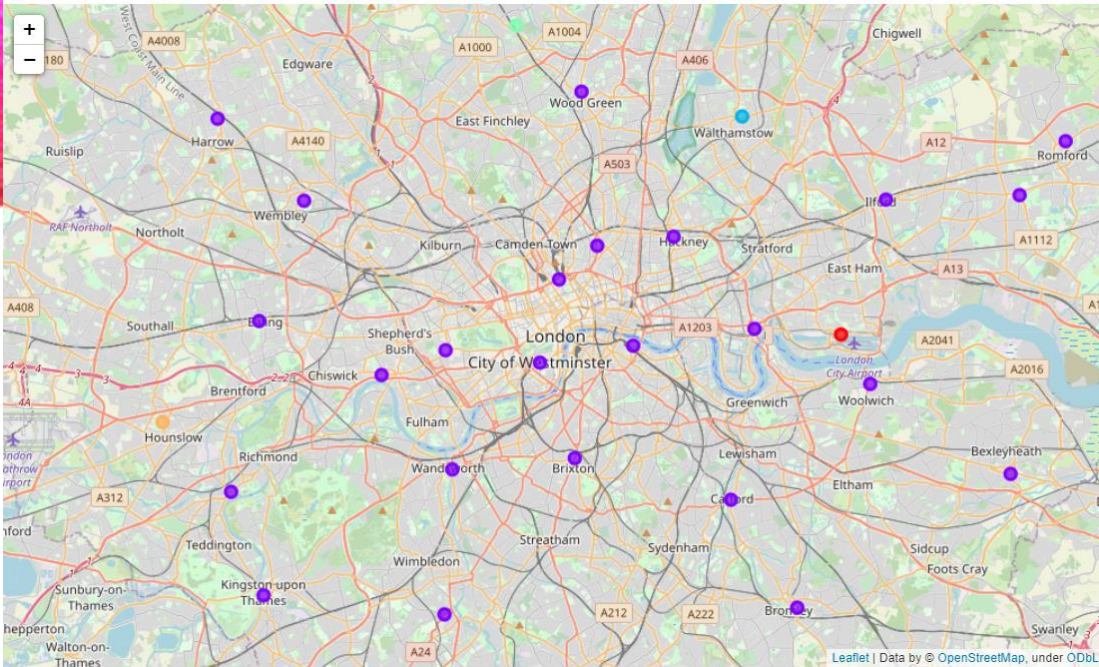
- One hot encoding method used to find out what venue categories are most popular (using the top 50 venue acquired from Foursquare)
- Venues from the same boroughs were grouped by borough names and popular categories were discovered by frequency.

	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

DATA ANALYSIS AND CLUSTERING

4. Cluster Analysis: K-Means Clustering

- Based on the common venue categories, boroughs were grouped into five clusters as displayed on the map by colors.



	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Barking and Dagenham	1	Pool	Bus Station	Supermarket
1	Barnet	3	Café	Home Service	Bus Stop
2	Bexley	1	Coffee Shop	Clothing Store	Pub
3	Brent	1	Coffee Shop	Hotel	Grocery Store
4	Bromley	1	Clothing Store	Coffee Shop	Pizza Place
5	Camden	1	Hotel	Café	Coffee Shop

DATA ANALYSIS AND CLUSTERING

4. Cluster Analysis: K-Means Clustering (Contd.)

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
23	Newham	0	Hotel	Chinese Restaurant	Airport

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

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Barking and Dagenham	1	Pool	Bus Station	Supermarket
2	Bexley	1	Coffee Shop	Clothing Store	Pub
3	Brent	1	Coffee Shop	Hotel	Grocery Store
4	Bromley	1	Clothing Store	Coffee Shop	Pizza Place

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

DATA ANALYSIS AND CLUSTERING

4. Cluster Analysis: K-Means Clustering (Contd.)

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
29	Waltham Forest	2	Pub	Pool	Gym / Fitness Center

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Barnet	3	Café	Home Service	Bus Stop

	BoroughName	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
16	Hounslow	4	Metro Station	Park	Café

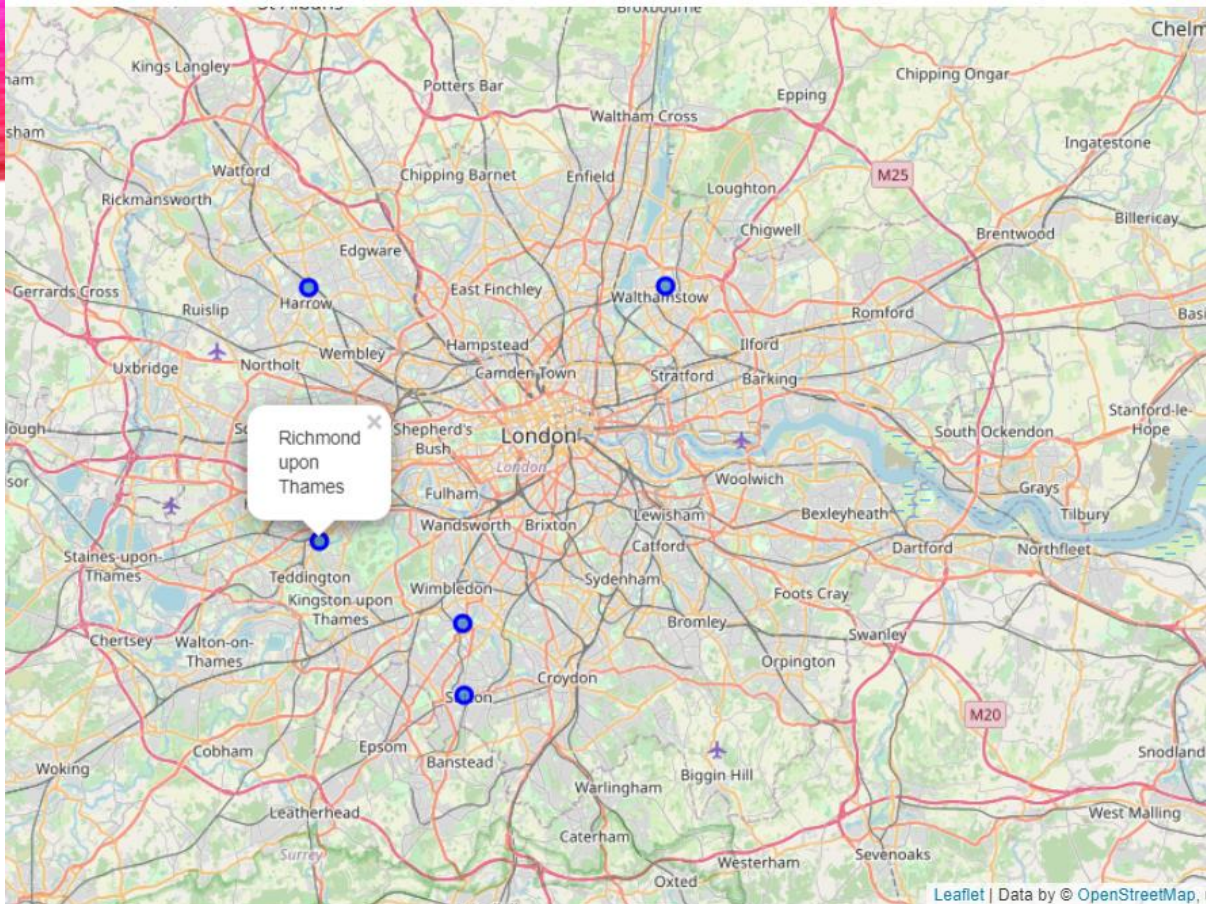
Cluster 2: Fitness Area (Pool, Gym, Pub)

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

Cluster 4: Refreshing area (Park, Cafe)

RESULTS AND CONCLUSION

- By taking safety and atmosphere scores into consideration, the best neighborhood to build a student accommodation is **Richmond upon Thames** with score as 1.9 points.
- The top five boroughs all belong to the lively Area and fitness area cluster with many pubs, restaurants, pub, pool, Gym/Fitness Centre and coffee shops around.
- However, based on the Map visualization the top five neighborhoods to live in, 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, the factors taken into consideration are based on simplified factors, safety, atmosphere, what type of venues are popular and how many crime incidents are recorded.



	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



REFERENCES

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