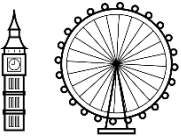# The Battle of Neighborhoods

- Finding the best area to build a student hall in London -
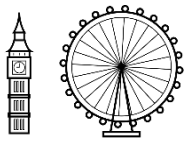
1. Problem

110,000+
international students
every year

High demand for
student halls

- 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 2016-2017 London welcomed **112,200 international students** to its higher education institutions, which make up **29 percent of students at higher education institutions**.

- One of the big concerns for international students when moving to a new city would be finding an accommodation. Student halls are the most reliable means of housing for students, but it is not easy to secure a place in one as they are in high demand. 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. 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 focuses 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.

2. Data

- In this project, three different datasets will be used to solve the problem - London Recorded Crime, List of London Boroughs, and Foursquare API.

(1) London Recorded Crime

- Description: London crime records classified by boroughs and crime type in the last 24 months

- Size: 1594 * 27

- Source: London Datastore, "London Recorded Crime : Geographic Breakdown"

- URL: https://data.london.gov.uk/dataset/recorded_crime_summary

|	MajorText	MinorText	BoroughName	201703	201704	201705	201706	201707	201708	201709	...	201805	201806	201807	201808	201809	201810
0	Arson and Criminal Damage	Arson	Barking and Dagenham	2	13	6	14	2	5	8	...	4	12	6	5	3	8
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	139	139	147	150	143	169	134	...	126	123	127	101	107	131
2	Burglary	Burglary - Business and Community	Barking and Dagenham	44	32	29	19	42	30	25	...	24	33	30	18	33	32
3	Burglary	Burglary - Residential	Barking and Dagenham	93	101	129	71	95	83	81	...	93	77	94	84	99	94
4	Drug Offences	Drug Trafficking	Barking and Dagenham	9	4	4	6	7	1	6	...	7	6	8	7	9	5

2. Data
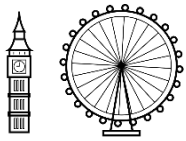
(2) List of London Boroughs

- Description: Information on 32 boroughs of London. 'Population' and 'Coordinates' will be used for the project.

- Size: 33 * 10

- Source: Wikipedia, "List of London Boroughs"

- 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 ⇕	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
Croydon			Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	372,752	51.3714°N 0.0977°W	19

2. Data
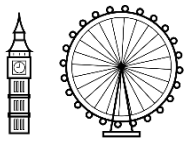
(3) Foursquare API

- Description: List of top 50 popular places in each borough. However, some suburban areas may have less than 50 places registered on Foursquare.

- Size: 1112 * 7

- Source: Foursquare API

- URL: https://api.foursquare.com

|	Borough	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
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	Beaconsfield Road (BF)	51.622827	-0.151466	Bus Stop
9	Barnet	51.6252	-0.1517	Oakleigh Cafe	51.623412	-0.154899	Café
10	Bexley	51.4549	0.1505	Zizzi	51.455834	0.150347	Italian Restaurant

3. Methodology
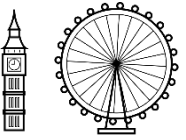
(1) Exploratory Analysis: Which borough is more dangerous?

- To observe the datasets better, each datasets were wrangled and cleansed into useful forms.

- Combining the London Recorded Crime data and the List of London Boroughs data, the number of average monthly crime records per 1000 people was calculated.

- When we visualize the boroughs by total number of crimes and by the ratio of crime to population, it is noticeable that the top two dangerous boroughs, Westminster and Camden, still remain the top two dangerous places.
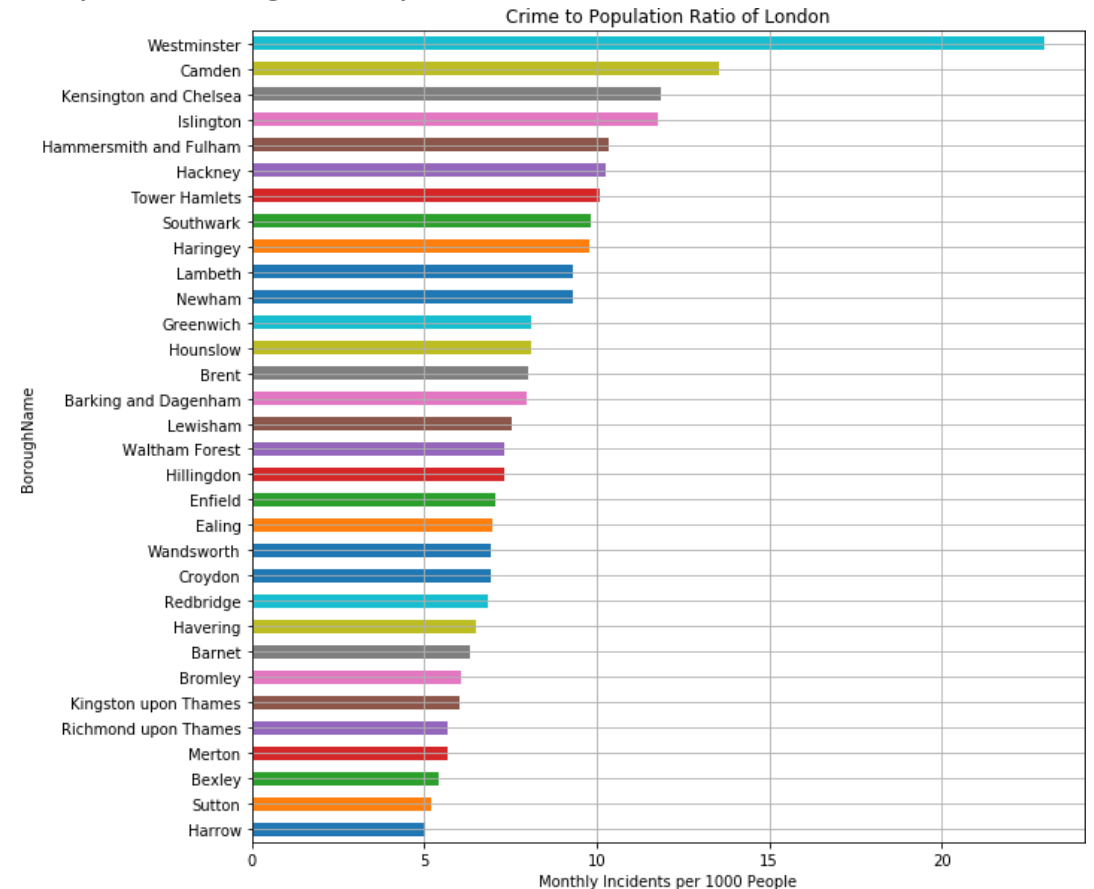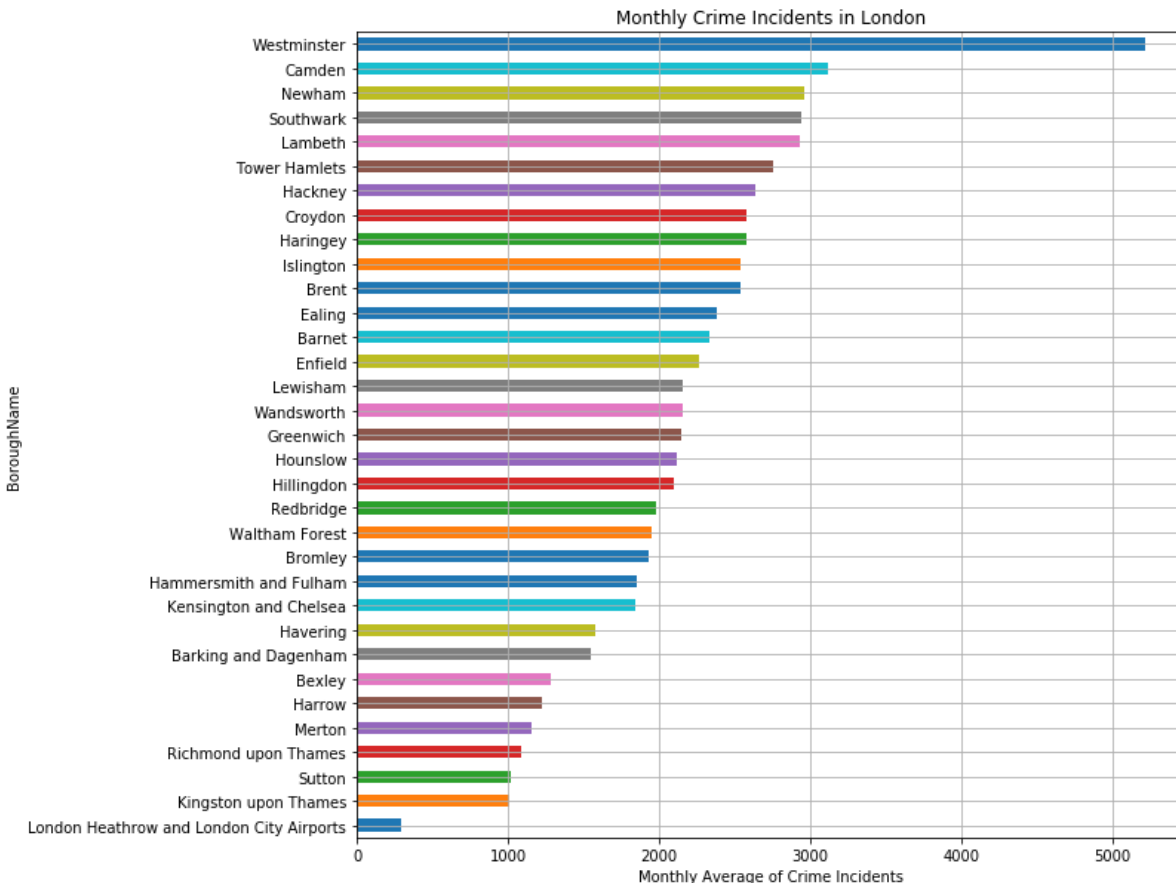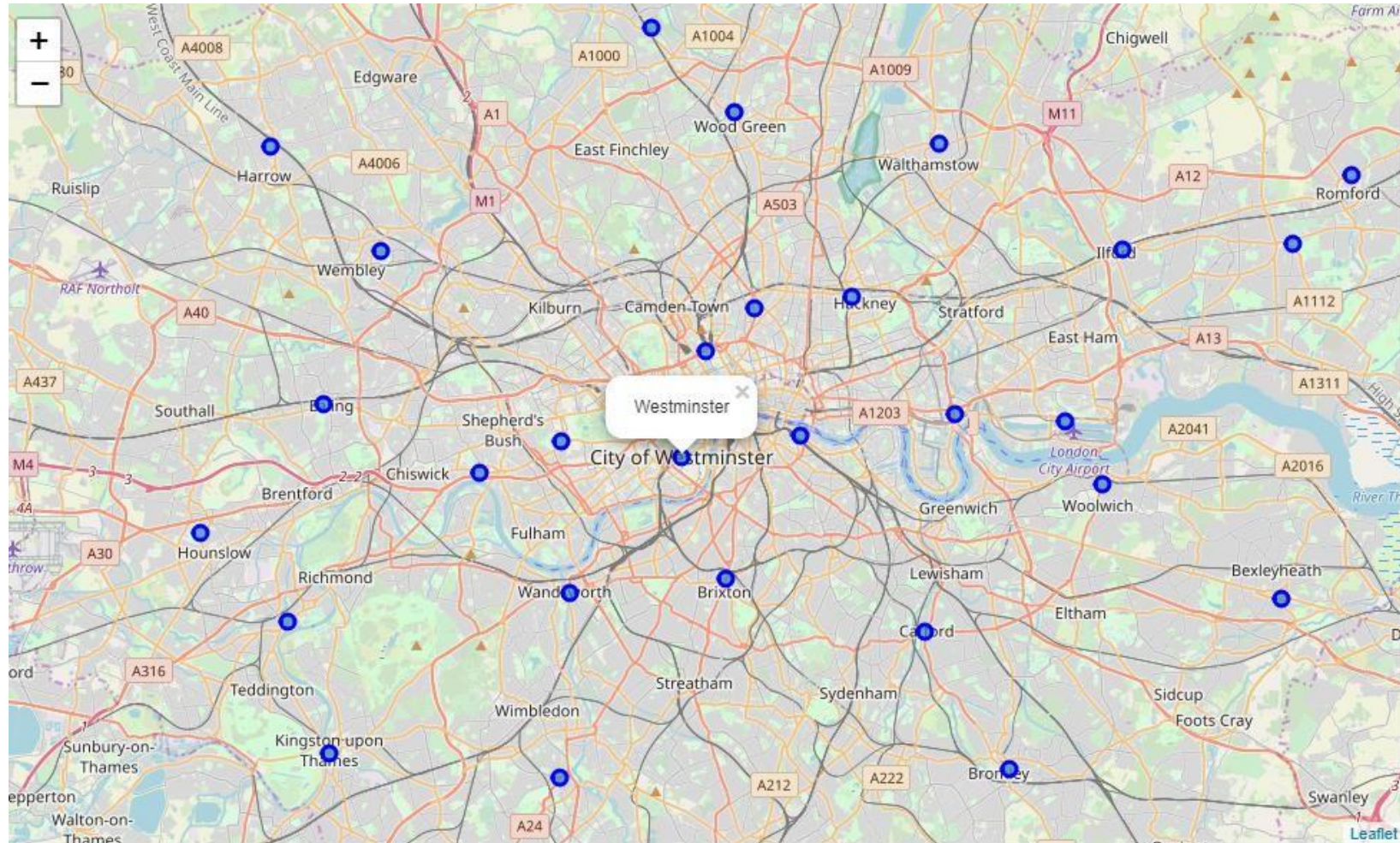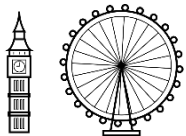
3. Methodology
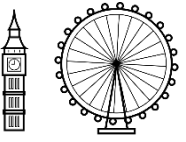
(2) Exploratory Analysis: Map of London

- Then with coordinates of each borough from List of London Boroughs data, we could visualize each borough on the map to get familiar with the Greater London area.

3. Methodology

(3) Cluster Analysis: Popular venues

- From the top 50 venue data we have acquired from Foursquare, we used one hot encoding method to find out what venue categories are most popular. Venues from the same boroughs were grouped by borough names and popular categories were discovered by frequency.

```
----Havering----                          ----Hillingdon----
              venue  freq                              venue  freq
0     Clothing Store  0.11             0        Coffee Shop  0.14
1        Coffee Shop  0.09             1     Clothing Store  0.08     ...
2      Shopping Mall  0.09             2  Italian Restaurant  0.06
3               Pub   0.06             3           Pharmacy  0.04
4  Fast Food Restaurant 0.06          4        Burger Joint  0.04
```

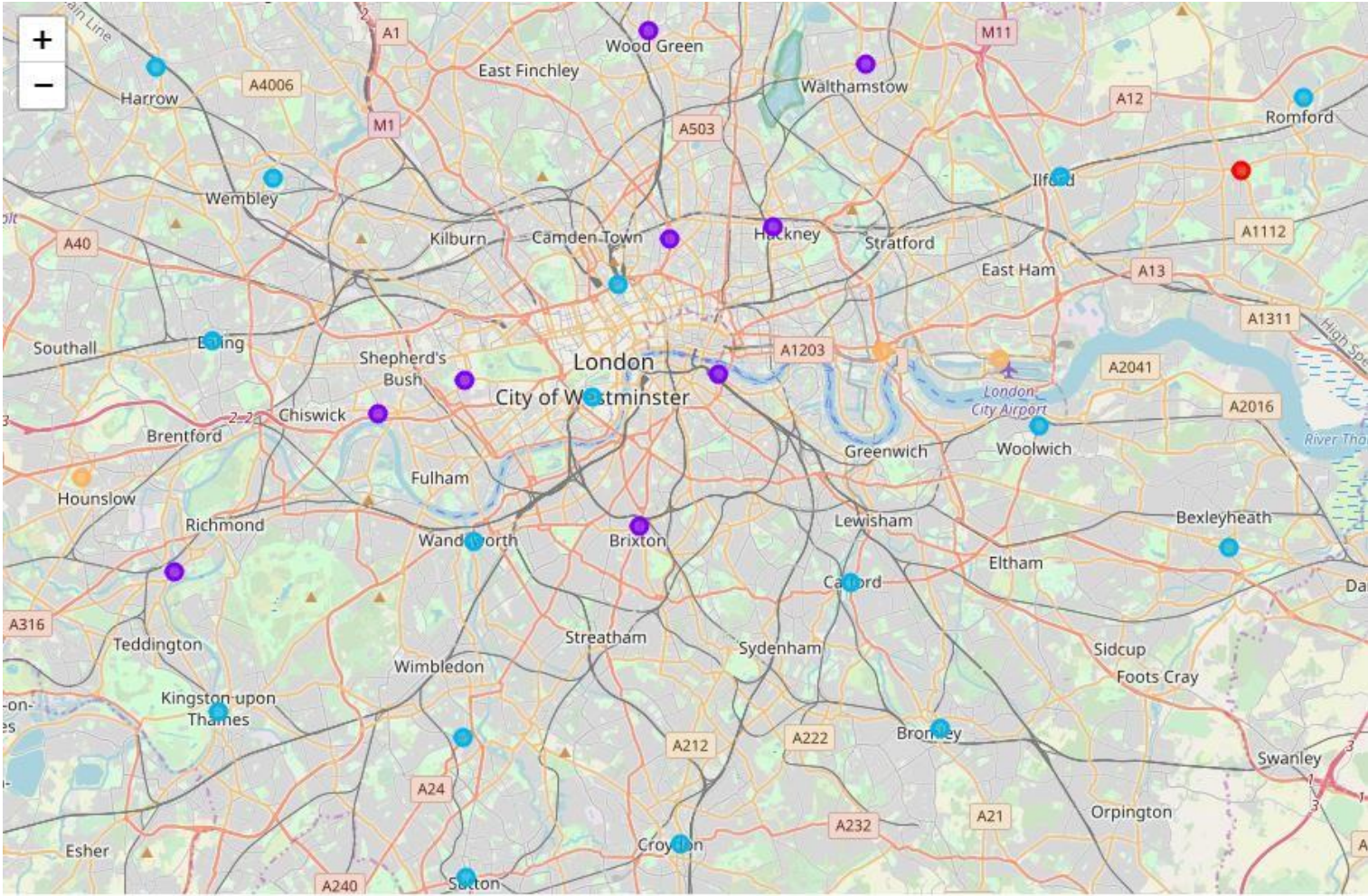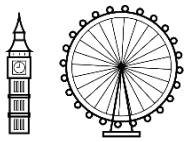|	BoroughName	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking and Dagenham	Gym / Fitness Center	Park	Pool	Supermarket	Golf Course	Martial Arts Dojo	Bus Station	Electronics Store	English Restaurant	Dessert Shop
1	Barnet	Café	Bus Stop	Yoga Studio	Gastropub	Garden	Gaming Cafe	Furniture / Home Store	French Restaurant	Food Court	Food
2	Bexley	Pub	Fast Food Restaurant	Italian Restaurant	Supermarket	Clothing Store	Coffee Shop	Grocery Store	Furniture / Home Store	Pharmacy	Sandwich Place
3	Brent	Coffee Shop	Hotel	Clothing Store	Sporting Goods Shop	American Restaurant	Italian Restaurant	Sandwich Place	Bar	Grocery Store	Electronics Store
4	Bromley	Coffee Shop	Clothing Store	Gym / Fitness Center	Bar	Burger Joint	Pizza Place	English Restaurant	Cosmetics Shop	Irish Pub	Pub

3. Methodology
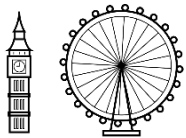
(4) Cluster Analysis: K-Means Clustering

- 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	0	Gym / Fitness Center	Park	Pool
1	Barnet	3	Café	Bus Stop	Yoga Studio
2	Bexley	2	Pub	Fast Food Restaurant	Italian Restaurant
3	Brent	2	Coffee Shop	Hotel	Clothing Store
4	Bromley	2	Coffee Shop	Clothing Store	Gym / Fitness Center

3. Methodology

(4) Cluster Analysis: K-Means Clustering

- Each clusters were then given a name based on the characteristics they display.

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Barking and Dagenham	7.981223	0	Gym / Fitness Center	Park	Pool

Cluster 0 : Healthy area (gym, park, pool, ...)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
10	Hackney	10.248829	1	Pub	Coffee Shop	Cocktail Bar
11	Hammersmith and Fulham	10.339890	1	Pub	Italian Restaurant	Indian Restaurant
12	Haringey	9.791554	1	Fast Food Restaurant	Pub	Grocery Store

Cluster 1 : Lively area (Pub, restaurant, bar, ...)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Bexley	5.416788	2	Pub	Fast Food Restaurant	Italian Restaurant
3	Brent	8.006345	2	Coffee Shop	Hotel	Clothing Store
4	Bromley	6.070980	2	Coffee Shop	Clothing Store	Gym / Fitness Center

Cluster 2 : Busy area (Coffee shop, clothing store...)

3. Methodology
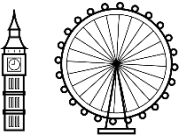
(4) Cluster Analysis: K-Means Clustering

- Each clusters were then given a name based on the characteristics they display.

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Barnet	6.328888	3	Café	Bus Stop	Yoga Studio

Cluster 3 : Quiet area (Bus stop, yoga studio, garden …)

	BoroughName	CrimeToPop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
16	Hounslow	8.081911	4	Bed & Breakfast	Café	Hotel
23	Newham	9.295644	4	Hotel	Coffee Shop	Airport
28	Tower Hamlets	10.107552	4	Italian Restaurant	Hotel	Coffee Shop

Cluster 4 : Traveler area (B&B, hotel, airport…)

4. Results

- Upon different analysis, we were able to discover the best neighborhoods based on our criteria of safety and atmosphere.

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

- 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 2) which I prefer, and lowest score was given to Traveler area (Cluster 4).

- Then by adding the two scores, we were able to find the best neighborhood scoreing full 2.0 points, Harrow.

	BoroughName	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	7.981223	0.8	0.836389	0.0
1	Barnet	6.328888	3.0	0.928624	0.7
2	Bexley	5.416788	2.0	0.979538	1.0
3	Brent	8.006345	2.0	0.834987	1.0
4	Bromley	6.070980	2.0	0.943021	1.0

|	BoroughName	Safety	Atmosphere	Score
13	Harrow	1.000000	1.0	2.000000
27	Sutton	0.990619	1.0	1.990619
2	Bexley	0.979538	1.0	1.979538
22	Merton	0.965216	1.0	1.965216
19	Kingston upon Thames	0.944754	1.0	1.944754
4	Bromley	0.943021	1.0	1.943021
14	Havering	0.918693	1.0	1.918693
24	Redbridge	0.899268	1.0	1.899268
6	Croydon	0.895563	1.0	1.895563
30	Wandsworth	0.894957	1.0	1.894957
7	Ealing	0.893158	1.0	1.893158
8	Enfield	0.887868	1.0	1.887868
15	Hillingdon	0.873787	1.0	1.873787
25	Richmond upon Thames	0.964540	0.9	1.864540

5. Conclusion

- 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.
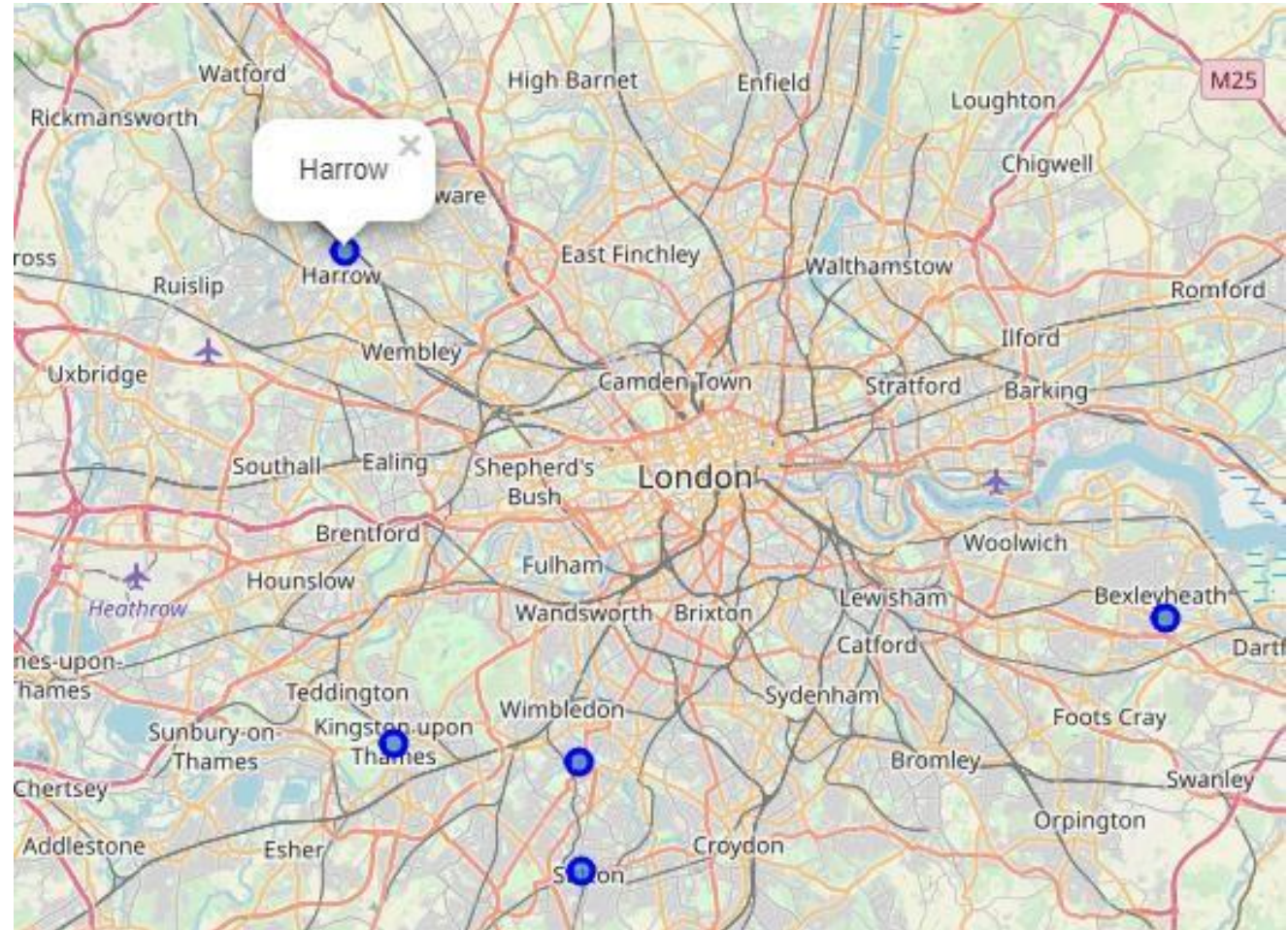
References
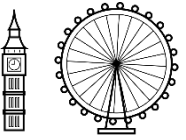
Images:

- Tower Bridge by arloenl evinniev from the Noun Project
- Big Ben by arloenl evinniev from the Noun Project
- The London Eye by arloenl evinniev from the Noun Project
- student by Mello from the Noun Project
- House by Dilla Chee from the Noun Project

Data:

- "London Recorded Crime : Geographic Breakdown", London Datastore
- "List of London Boroughs", Wikipedia
- Foursquare API

Document:

- "The Economic Impact of London's International Studnets", London & Partners (2018)
- Lecture notes from IBM Professional Data Science Specialization, Coursera