The Battle of the Neighborhoods

Identifying the safest borough in London

I. Introduction

Background of the Problem:

London is both the capital and the largest city of England. It is also a financial hub full of booming businesses as well as a popular tourist destination. Due to the competitive environment of the city, it is therefore also known to have a fairly high cost of living. Those choosing to relocate there, whether for work purposes or otherwise, are therefore encouraged to undertake extensive research on the various boroughs and neighborhoods which are available to them before settling permanently. As with any other decision, there are numerous factors that would need to be considered before coming to a final decision. One of these factors is the safety of the borough - which can be quantified by looking at the different crime rates. The goal of this project is to therefore provide insights into the 'environment' of each of the 32 boroughs of London with a particular focus on identifying the safest borough. Once the safest borough has been identified, I will then proceed to explore the different neighborhoods within that borough with regards to the 10 most common venues in each neighborhood.

Description of the Approach:

To achieve the goal I set out above, I will carry out the following plan:

1. To begin with, I will need to collect data on the crime statistics of each London Borough. The dataset I will be using to accomplish this can be found at data.london.gov.uk and it includes the most recent crimes recorded from 2018 to 2020. More on this in the Data section.
2. Having determined the safest borough from the analysis in step 1, I will proceed to characterize the neighborhoods within that particular borough. The list of neighborhoods can be found on wikipedia. Following that, I will explore the borough to identify the 10 most common venues in each neighborhood. This will be achieved using Foursquare location data.
3. Finally, each neighborhood will be placed into one of five clusters. I will then carry out an analysis on the characteristics of each cluster and which individuals would be best suited/attracted to each (based on the common venues).

Who is the target audience?

Those who might be interested in this project are expatriates (or British citizens in other cities) looking to relocate to London and who value safety. It will be of particular interest to those that have families with young children, or otherwise vulnerable family members, who would greatly benefit from choosing the safest borough. From there, one can further tailor their options by selecting the neighborhood that offers venues which cater to their needs the most.

II. Data

Data Acquisition

The data for this project will be acquired from three sources (links to all three can be found at the end of this section). Firstly, the project will utilise a crime dataset which counts the number of crimes in each London borough, per month, (from May 2018 - April 2020), according to crime type. The dataset contains the following columns:

- **MajorText**: The higher-level/general categorisation of the crime
- **MinorText**: The lower-level/specific categorisation of the crime (within the MajorText category)
- **LookUp_BoroughName**: The common name for the London borough
- **Year and month (multiple columns)**: Monthly reported count of each crime type in given borough

The second source of data will be scraped from a wikipedia page that contains the list of London boroughs. It contains the following columns:

- **Borough**: The name of the London borough
- **Inner**: Categorization of either an inner or outer London borough
- **Status**: Categorization of a Royal, City, or other borough
- **Local authority**: The local authority assigned to the borough
- **Political control**: The political party that controls the borough
- **Headquarters**: Location of the borough's headquarters
- **Area (sq mi)**: Area of the borough in square miles
- **Population (2013 est)**: The population of the borough recorded in 2013
- **Coordinates**: The latitude and longitude of the borough
- **Nr. in map**: The number assigned to each borough to represent visually on a map

The third data source is the list of nieghborhoods in the borough of 'Richmond upon Thames' as found on wikipedia. The dataset will be manually created from scratch and will include the following columns:

- **Neighborhood**: Name of the neighborhood in the borough
- **Borough**: Name of the borough
- **Latitude**: Latitude of the neighborhood
- **Longitude**: Longitude of the neighborhood

Links to the data sources:

- Crime data (source 1)
- List of London boroughs (source 2)
- List of neighborhoods in Richmond upon Thames (source 3)

Data Cleaning

The data sources will each be prepared separately. For the first source, only the London crime data from the most recent period (April 2019 to April 2020) will be used. We will produce two dataframes here: one that shows both the major and minor crimes, as well as one that shows the total crimes for each borough:

	Major Category	Minor Category	Borough	Total Crimes
0	Arson and Criminal Damage	Arson	Barking and Dagenham	73
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	1460
2	Burglary	Burglary - Business and Community	Barking and Dagenham	368
3	Burglary	Burglary - Residential	Barking and Dagenham	1257
4	Drug Offences	Drug Trafficking	Barking and Dagenham	99

Fig 1.1 - Major and minor crimes in each London borough (2019-04 to 2020-04)

	Borough	Total Crimes
0	Richmond upon Thames	13608
1	Kingston upon Thames	13727
2	Sutton	15149
3	Merton	15376
4	Harrow	18485
5	Bexley	18973
6	Havering	20114
7	Barking and Dagenham	21617
8	Hammersmith and Fulham	24724
9	Kensington and Chelsea	25740
10	Redbridge	25831

Fig 1.2 - Total number of crimes in each London borough (2019-04 to 2020-04)

Next, we will use the Beautiful Soup package to scrape data on each London borough from a table in a wikipedia page. The initial table is quite cluttered as it includes columns we are not necessarily interested in (e.g. the political party in power) and it also contains tags and notes that we do not require. As such, we will drop the unnecessary columns and remove any unwanted extra text.

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)	Nr. in map
0	Barking and Dagenham [note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	25
1	Barnet			Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369,088	31
2	Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	23
3	Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	12
4	Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	20
5	Camden	Y		Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	229,719	11
6	Croydon			Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	372,752	19
7	Ealing			Ealing London Borough Council	Labour	Perceval House, 14-16 Uxbridge Road	21.44	342,494	13
8	Enfield			Enfield London Borough Council	Labour	Civic Centre, Silver Street	31.74	320,524	30
9	Greenwich [note 2]	Y [note 3]	Royal	Greenwich London Borough Council	Labour	Woolwich Town Hall, Wellington Street	18.28	264,008	22

Fig 2.1 - The initial dataframe contains unwanted columns and tags/notes

	Borough	Area (sq mi)	Population (2013 est)	Total Crimes
0	Barking and Dagenham	13.93	194,352	21617
1	Barnet	33.49	369,088	32925
2	Bexley	23.38	236,687	18973
3	Brent	16.70	317,264	32039
4	Bromley	57.97	317,899	26507
5	Camden	8.40	229,719	41109
6	Croydon	33.41	372,752	36965
7	Ealing	21.44	342,494	33317
8	Enfield	31.74	320,524	32348
9	Greenwich	18.28	264,008	30513
10	Hackney	7.36	257,379	35647

Fig 2.2 - We will only keep these four columns of interest

Furthermore, because we are interested in the **crime rate**, a new column will be added to calculate this value by first dividing the number of reported crimes by the population, then multiplying the answer by 1,000. This will get us the **crime rate per 1,000 residents**.

	Borough	Area (sq mi)	Population (2013 est)	Total Crimes	Crime Rate
25	Richmond upon Thames	22.17	191365	13608	71.11
22	Merton	14.52	203223	15376	75.66
13	Harrow	19.49	243372	18485	75.95
27	Sutton	16.93	195914	15149	77.32
2	Bexley	23.38	236687	18973	80.16
19	Kingston upon Thames	14.38	166793	13727	82.30
14	Havering	43.35	242080	20114	83.09
4	Bromley	57.97	317899	26507	83.38
30	Wandsworth	13.23	310516	27611	88.92
1	Barnet	33.49	369088	32925	89.21

Fig 2.3 - London boroughs sorted by crime rate

Finally, the third data source (a wikipedia page containing a list of nieghborhoods in the borough of 'Richmond upon Thames') will be used alongside geopy to produce the dataframe in fig 3.1 (on the next page) which contains the coordinates of each neighborhood. The result will then be used with foursquare location data to explore the venues within each neighborhood.

	Neighborhood	Borough	Latitude	Longitude
0	Barnes	Richmond upon Thames	51.471896	-0.238744
1	East Sheen	Richmond upon Thames	51.462371	-0.267094
2	Ham and Petersham	Richmond upon Thames	51.444235	-0.314040
3	Hampton	Richmond upon Thames	51.415027	-0.369141
4	Hampton Hill	Richmond upon Thames	51.427844	-0.355081
5	Hampton Wick	Richmond upon Thames	51.414452	-0.312674
6	Kew	Richmond upon Thames	51.480663	-0.291929
7	Mortlake	Richmond upon Thames	51.469887	-0.268523
8	Whitton	Richmond upon Thames	51.451169	-0.357976
9	Richmond and Richmond Hill	Richmond upon Thames	51.461353	-0.303277
10	Strawberry Hill	Richmond upon Thames	51.438592	-0.339937
11	St Margarets	Richmond upon Thames	51.456709	-0.322412
12	Teddington	Richmond upon Thames	51.427784	-0.333653
13	Twickenham	Richmond upon Thames	51.446744	-0.328189
14	Heathfield North	Richmond upon Thames	51.452399	-0.337211

Fig 3.1 - Coordinates of each neighborhood in Richmond upon Thames

III. Methodology

Exploratory Data Analysis

First, the descriptive statistics of the London crime data will be produced. This returns the mean, standard deviation, minimum, maximum, 1st quartile, 2nd quartile, and 3rd quartile for the population, total number of crimes and crime rate.

	Population (2013 est)	Total Crimes	Crime Rate
count	32.000000	32.000000	32.000000
mean	262777.718750	29913.906250	115.808750
std	57346.611409	12411.117231	52.630362
min	155594.000000	13608.000000	71.110000
25%	224047.500000	23947.250000	87.535000
50%	263697.000000	29668.500000	100.955000
75%	311447.500000	34280.500000	129.797500
max	372752.000000	82479.000000	363.600000

Fig 4.1 - Descriptive statistics of London crime data

The count for each column is 32, which indicates the number of London boroughs. The mean population of the London boroughs is 262,777 (estimated in 2013) and the mean total crime and crime rates are 29,913 crimes per borough and 115 crimes per 1,000 residents, respectively. The difference between the lowest and highest crime rate is also quite stark (71 vs. 363 reported crimes per 1,000 residents).

Comparing the five boroughs with the highest crime rates, we can see that the borough of Westminster has a disproportionately higher crime rate than the next four boroughs (double the crime rate of Camden!):

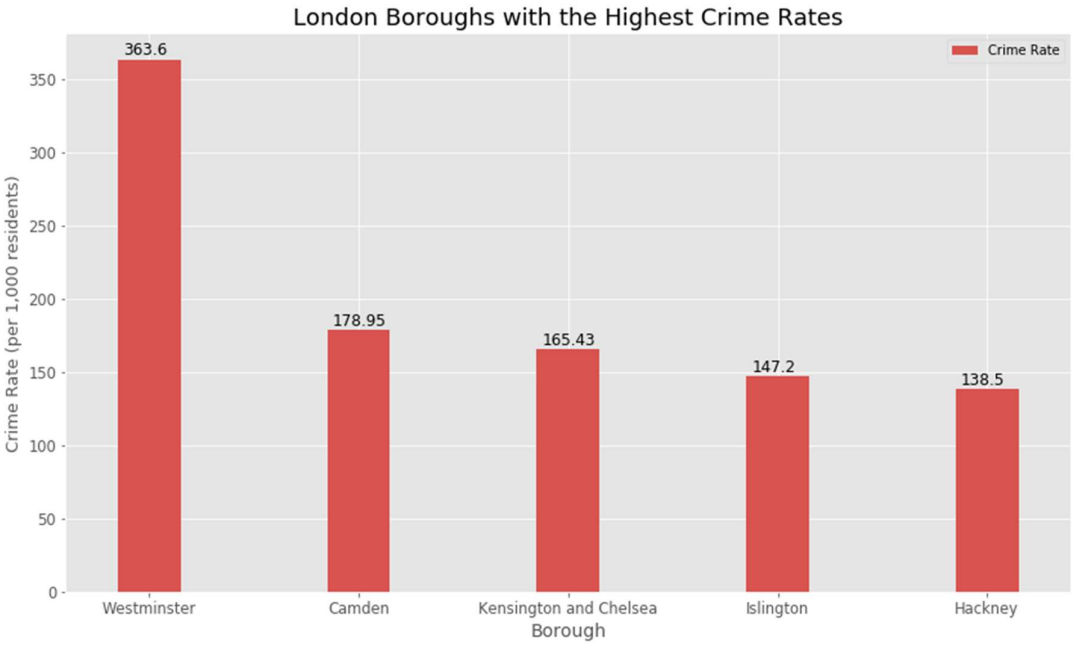

Fig 4.2 - London boroughs with the highest crime rates

Comparing the five boroughs with the lowest crime rates, we can see that there is very little difference between the numbers (they all lie between 70 and 80 crimes per 1,000 residents):

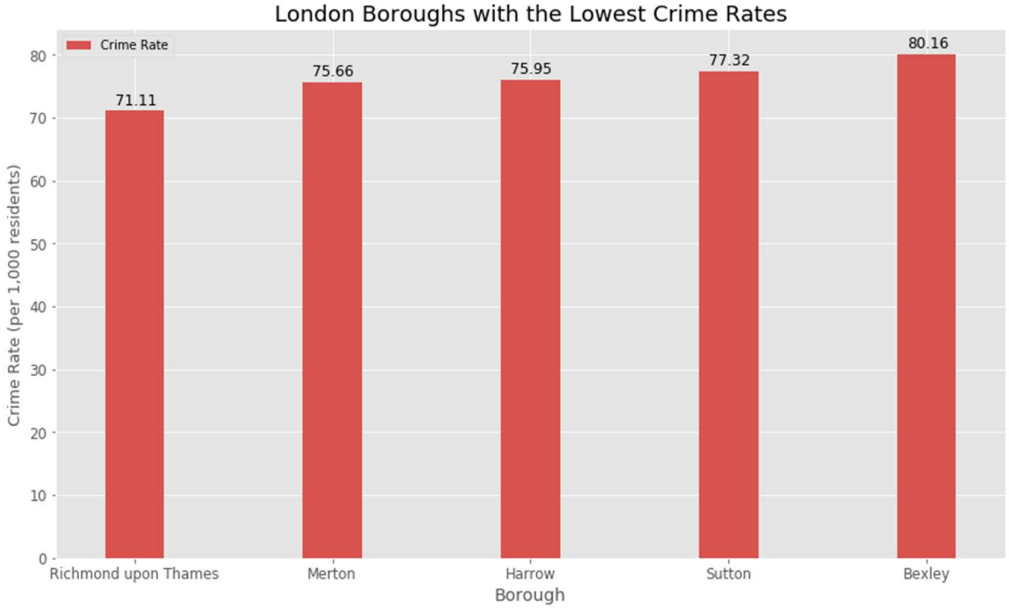

Fig 4.3 - London boroughs with the lowest crime rates

Now that we have determined the safest borough based on the crime rate (Richmond upon Thames), we will look into the types of crimes that have occurred in that time period (April 2019 to April 2020).

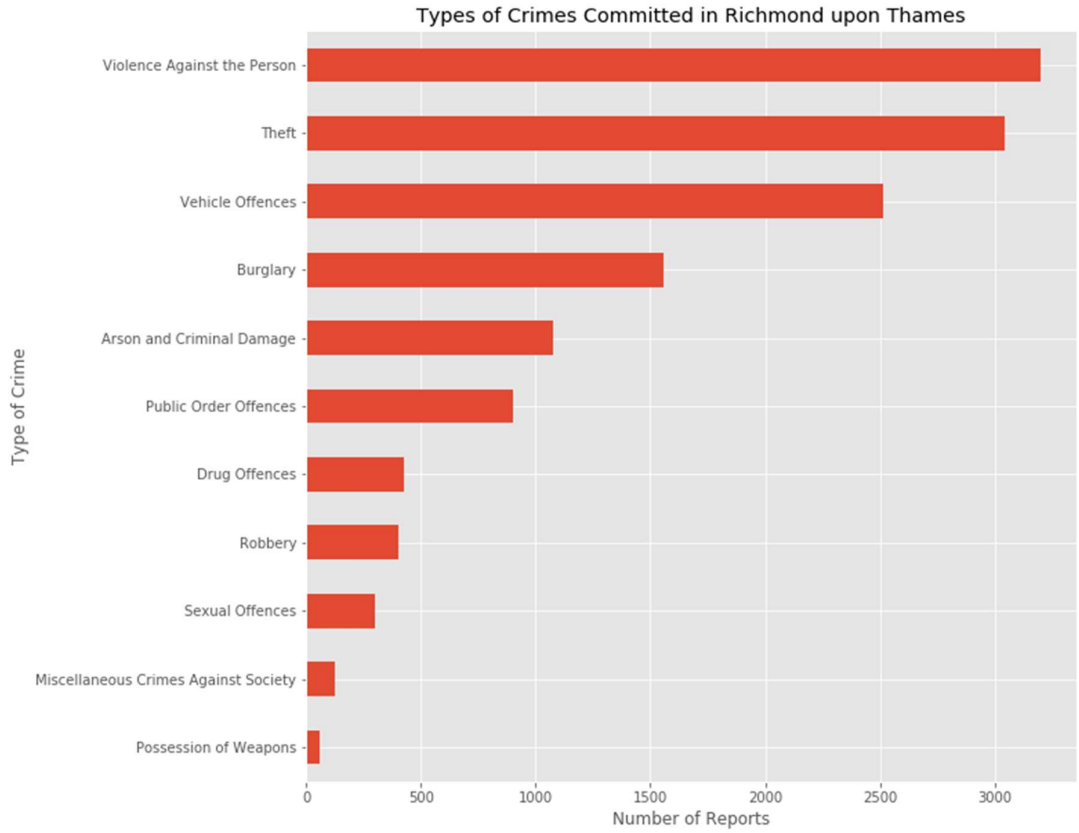

Fig 4.4 - Types of crimes committed in Richmond upon Thames

As can be seen in the above graph, the two most common types of crime committed in Richmond upon Thames (RuT) are violence against another person and theft, whilst the least common types of crime are possession of weapons, sexual offences, and miscellaneous crimes. Whilst RuT is the safest borough, and crimes are not all that common, knowing the most common types of crimes that are committed will help individuals better prepare for such cases. For example, because violence and theft are the most common types of crime, one might keep safe by reducing interaction with people they don't know, as well as being vigilant of their personal belongings when in public.

Moving on, the next step is to look at the different neighborhoods within the borough of RuT. There are 15 boroughs in RuT as is visualised in the Folium map below (next page):

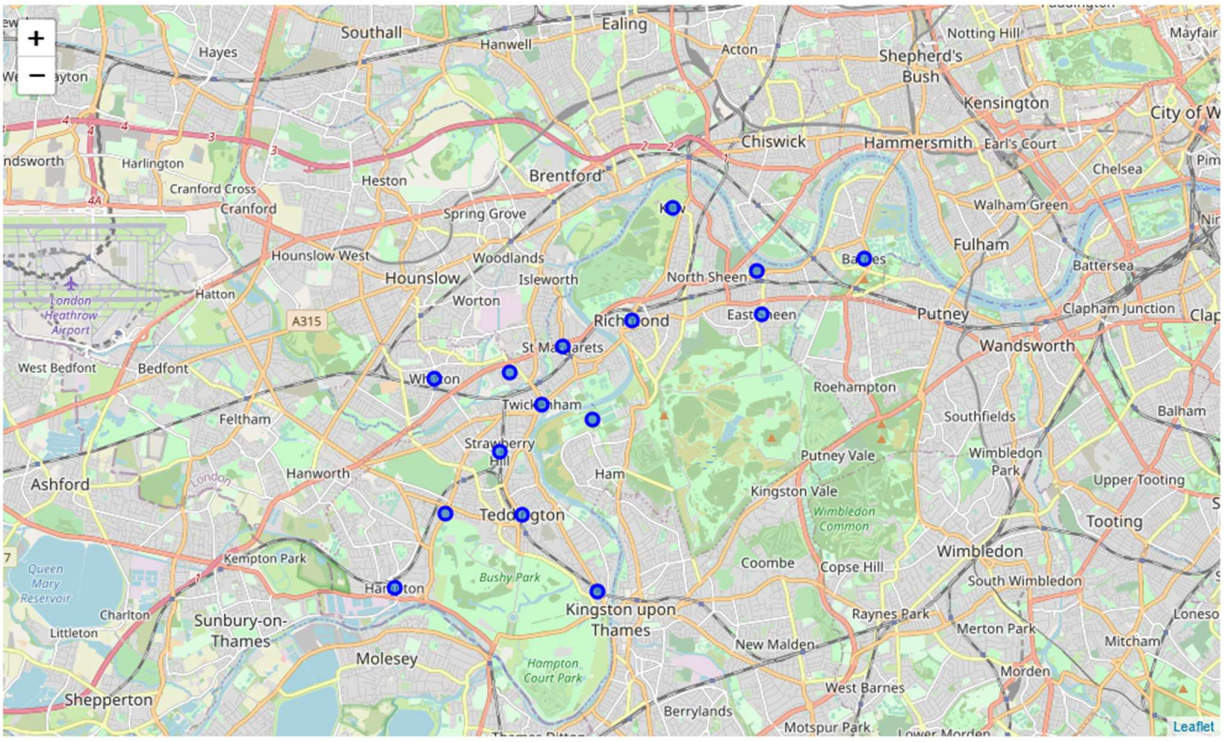

Fig 4.5 - Folium map visualising the 15 neighborhoods in RuT

Modelling

Using the dataframe from fig 3.1 and foursquare location data, a new merged dataframe will be created which displays the venue data for each neighborhood. The additional columns added will include the likes of the venue name, venue latitude/longitude, and venue category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barnes	51.471896	-0.238744	Olympic Studios Cafe + Dining Room	51.475158	-0.240333	Indie Movie Theater
1	Barnes	51.471896	-0.238744	Barnes Common	51.469260	-0.238903	Park
2	Barnes	51.471896	-0.238744	ArteChef	51.474705	-0.241282	Pizza Place
3	Barnes	51.471896	-0.238744	Alma Café	51.474880	-0.239207	Breakfast Spot
4	Barnes	51.471896	-0.238744	Awesome Thai Cuisine	51.474905	-0.240909	Thai Restaurant

Fig 5.1 - Neighborhood and venue data using Foursquare

Next, one hot encoding will be performed on the data (which is essentially the representation of categorical variables as binary vectors). The data is then grouped according to the neighborhood, and the mean of the frequency of each venue category is calculated. Finally, the top 10 most common venues will be identified for each neighborhood. The result is a dataframe of the following form:

	Neighborhood	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	Barnes	Park	Food & Drink Shop	Café	Indie Movie Theater	Pizza Place	Community Center	Movie Theater	Pub	Restaurant	Breakfast Spot
1	East Sheen	Coffee Shop	Pub	Pizza Place	Café	Creperie	Indian Restaurant	Sporting Goods Shop	Stationery Store	Beer Store	Supermarket
2	Ham and Petersham	Sports Club	Boat or Ferry	Historic Site	Park	Café	Garden	Playground	Trail	Garden Center	Art Gallery
3	Hampton	Grocery Store	Café	Pizza Place	Train Station	Park	Seafood Restaurant	Soccer Stadium	Department Store	Coffee Shop	Comedy Club
4	Hampton Hill	Grocery Store	Wine Shop	Pub	Butcher	Italian Restaurant	Fast Food Restaurant	Coffee Shop	Discount Store	Comic Shop	Community Center

Fig 5.2 - The top 10 most common venues in each neighborhood

Following this, we will perform k-means clustering to cluster the neighborhoods into **five clusters** depending on the types of venues we can find in each and their similarities. This is particularly useful as it allows people the ability to effectively shortlist the neighborhoods they are interested in without looking too much into the specifics.

IV. Results

To visualise the five clusters, a Folium map is created with color-coded circle markers for each neighborhood. We can see that the purple cluster is the largest cluster indicating that those neighborhoods are very similar to each other. Three neighborhoods have their own clusters indicating they have a unique set of venues.

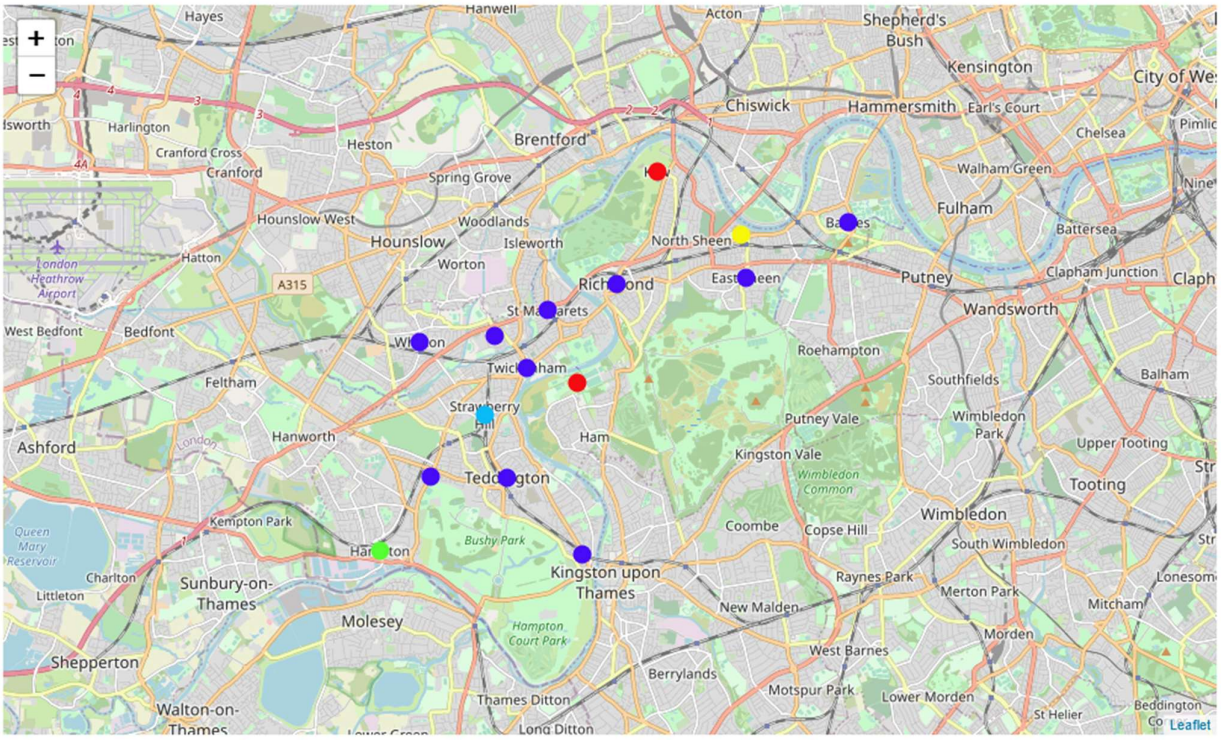

Fig 6.1 - The five clusters of neighborhoods in Richmond upon Thames

Here, we will look deeper into each cluster individually. Depicted below is the first cluster (cluster 0) which is depicted by the purple markers. This cluster is the biggest cluster as it contains 10 of the 15 neighborhoods in the borough of Richmond upon Thames. Looking at the data, we can see that the most common venues in these neighborhoods are pubs, coffee shops/cafés and restaurants. This cluster is very much oriented towards individuals that regularly eat out and/or meet friends at cafés and pubs. It is thus recommended for those that enjoy socialising often, and also perhaps for those that dislike cooking meals at home.

	Neighborhood	Borough	Latitude	Longitude	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
0	Barnes	Richmond upon Thames	51.471896	-0.238744	0	Park	Food & Drink Shop	Café	Indie Movie Theater	Pizza Place	Community Center	Movie Theater	Pub
1	East Sheen	Richmond upon Thames	51.462371	-0.267094	0	Coffee Shop	Pub	Pizza Place	Café	Creperie	Indian Restaurant	Sporting Goods Shop	Stationery Store
4	Hampton Hill	Richmond upon Thames	51.427844	-0.355081	0	Grocery Store	Wine Shop	Pub	Butcher	Italian Restaurant	Fast Food Restaurant	Coffee Shop	Discount Store
5	Hampton Wick	Richmond upon Thames	51.414452	-0.312674	0	Pub	Plaza	Indian Restaurant	Coffee Shop	Train Station	Park	Restaurant	Sports Club
8	Whitton	Richmond upon Thames	51.451169	-0.357976	0	Coffee Shop	Indian Restaurant	Pub	Pizza Place	Grocery Store	Fish & Chips Shop	Bakery	Supermarket
9	Richmond and Richmond Hill	Richmond upon Thames	51.461353	-0.303277	0	Pub	Café	Italian Restaurant	Coffee Shop	Bakery	Restaurant	French Restaurant	Burger Joint
11	St Margarets	Richmond upon Thames	51.456709	-0.322412	0	Pub	French Restaurant	Coffee Shop	Gym	Seafood Restaurant	Deli / Bodega	Park	Comedy Club
12	Teddington	Richmond upon Thames	51.427784	-0.333653	0	Pub	Café	Coffee Shop	Hotel	Mediterranean Restaurant	Italian Restaurant	Gym / Fitness Center	Deli / Bodega
13	Twickenham	Richmond upon Thames	51.446744	-0.328189	0	Pub	Coffee Shop	Italian Restaurant	Café	Platform	Pharmacy	Pizza Place	Grocery Store
14	Heathfield North	Richmond upon Thames	51.452399	-0.337211	0	Rugby Stadium	Gym / Fitness Center	Gym	Hotel Bar	Museum	Restaurant	Bus Station	Breakfast Spot

Fig 6.2 - Cluster 0: recommended for young adults that enjoy socialising

Our second cluster (cluster 1) consists of two neighborhoods. The most common venues in these two neighborhoods includes the likes of gardens, parks, sports clubs, historic sites and cafés. This kind of environment would likely attract those that enjoy the outdoors. It is particularly suitable for older individuals as the abundance of outdoor venues (e.g. gardens/parks) provides a safe space for some mild physical activity.

	Neighborhood	Borough	Latitude	Longitude	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
2	Ham and Petersham	Richmond upon Thames	51.444235	-0.314040	1	Sports Club	Boat or Ferry	Historic Site	Park	Café	Garden	Playground	Trail	Garden Center
6	Kew	Richmond upon Thames	51.480663	-0.291929	1	Garden	Botanical Garden	Café	Playground	Coffee Shop	Wine Shop	Restaurant	Pub	Ice Cream Shop

Fig 6.3 - Cluster 1: quieter environment recommended for older individuals

The third cluster (cluster 2) has just one neighborhood which consists of venues such as train stations, pubs, restaurants and various stores. It is suitable for those that don't own a car/bicycle and would prefer commuting by public transport.

	Neighborhood	Borough	Latitude	Longitude	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
10	Strawberry Hill	Richmond upon Thames	51.438592	-0.339937	2	Train Station	Pub	History Museum	Thai Restaurant	Convenience Store	Women's Store	Discount Store	Coffee Shop	Comedy Club

Fig 6.4 - Cluster 2: recommended for those that regularly commute by public transport

The fourth cluster (cluster 3) also has one neighborhood and has venues such as grocery stores, restaurants, train stations, parks and soccer stadiums. It is suitable for families as it has a good mix of everything. The most common venues are grocery stores, which is good for regularly stocking up on supplies for the family. Restaurants, parks, and the soccer stadium are also good for keeping families occupied during their downtime.

	Neighborhood	Borough	Latitude	Longitude	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
3	Hampton	Richmond upon Thames	51.415027	-0.369141	3	Grocery Store	Café	Pizza Place	Train Station	Park	Seafood Restaurant	Soccer Stadium	Department Store	Coffee Shop

Fig 6.5 - Cluster 3: recommended for families with school-children

The fifth and final cluster (cluster 4) has one neighborhood which consists of restaurants, golf courses, parks, athletics/sports venues and pubs. This is likely to attract and suit those that enjoy a good balance between leisure and exercise. However, the lack of public transport suggests that individuals would need to commute privately (or it could also suggest that a vast majority don't need to commute in the first place).

	Neighborhood	Borough	Latitude	Longitude	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
7	Mortlake	Richmond upon Thames	51.469887	-0.268523	4	American Restaurant	Tapas Restaurant	Golf Course	Park	Pub	Grocery Store	Athletics & Sports	Comedy Club	Community Center

Fig 6.6 - Cluster 4: recommended for active and social individuals

V. Discussion

As initially determined, the aim of this project is to help identify the safest borough in London and characterise the neighborhoods within that borough. We ultimately identified the safest borough to be Richmond upon Thames, a medium-sized borough with a crime rate of just 71 registered crimes per 1,000 residents. This is impressive when compared to boroughs such as Westminster and Camden that have crime rates of 364 and 179 per 1,000 residents. Note however, that although RuT

has the lowest crime rate, the other five boroughs with the lowest crime rates also fall below 80 recorded crimes per 1,000 residents. Those that do not find any of the neighborhoods withing RuT to suit them may therefore look into the next safest boroughs without there being too much of a difference. Nonetheless, the neighborhoods within RuT offer a good variety (as demonstrated by the five clusters). For example, a family might look into the neighborhood of Hampton as there is a very good mix of venues that will suit the needs of both the children and the parents. Single, working adults that are sociable might opt for one of the neighborhoods within cluster 0 as it offers many opportunities to eat out and meet with friends for a drink.

VI. Conclusion

All in all, this project has been successful in i) determining the safest borough in London based on crime rate, ii) identifying and clustering the neighborhoods within that borough, and iii) offering advice/comments about the findings. Although the report is useful as a simple and brief overview about the subject matter, it is advisable to also utilise alternative resources to supplement ones research when choosing the perfect borough/neighborhood for settlement. Furthermore, the findings are fairly restrictive as the report only considers safety as the primary factor for the analysis. It would therefore not be useful to those that prefer to prioritise other characteristics such as cost of living or opportunity for employment. Furthermore, we made no distinction between the 'seriousness' of each crime type (effectively viewing homicides and petty thefts to have the same weight). This doesn't necessarily nullify the findings of this report, although it is worth keeping the idea in mind. That being said, this project could potentially look into these other factors in the future.

VII. References

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

- https://en.wikipedia.org/wiki/London_Borough_of_Richmond_upon_Thames#List_of_neighbourhoods

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