The Battle of Neighborhoods in London

Project of Applied Data Science Capstone

Coursera – IBM Data Science

Pedro Pereira

2020

1. Introduction

1.1. Background

Nowadays we are living in an exponentially globalized world where people are increasingly moving from one city to another, between different countries (or even continents) chasing their dreams or just to find a better place to live and call home. Happiness and comfort are two very important pylars in human being's lives, existing several factors which may affect them either positive and/or negativelly. One factor is safety, a fundamental concept that sometimes is underappreciated in highly developed countries (mainly because criminality rate is decreasing) but it is essential to choose the place where we would live and mostly when we are moving to a new/ unknown place.

1.2. Problem

London is the capital and largest city of England and United Kingdom (UK), as well as one of most important cities in western part of Europe, with a total population of around 9 million people, distributed by their 33 boroughs (including the City of London). Although it is well known as a crucial metropolis with a inumerous offer, quality of life, a diverse range of people and cultures, etc., London is also a city with a vast criminal record and therefore this variable plays an important role, or being even decisive, at the time of choosing a place when we decide to move and to settle a new life there.

1.3. Interest

Based on this premise, the project main goal can be posed by the following question: “what is the most suitable/ safest area to live within London city?” Puting the hands on the data will help to discover the answer in order to make a recommendation for the target audience. In fact, this project is very interesting to everybody who is coming to London from abroad (or even within London area) and want to rent/ buy a new house or also for any businessman who wants to find the best area to move its company’s office or even to implement the headquarters of its start-up company.

2. Data

2.1. Data Sources

To answer the introductory question, and to achieve the objective previously described, available free data was used and analised throughout the project, according to the London criminal records downloaded from Kaggle website (source: *https://www.kaggle.com/jboysen/london-crime*). Besides this data, consisting into the main dataset of the project, an additional dataset was also used, obtained from wikipedia website (source: *https://en.wikipedia.org/wiki/List_of_London_boroughs*), consisting in a list and features of London’s boroughs.

2.2. Data Description

Regarding the main dataset of London criminality, the data matrix (i.e. table) is composed by 3419099 rows and 7 columns as shown in Figure 1. In more detail, the variables presented in the columns correspond to:

- **lsoa_code** (code for Lower Super Output Area in Greater London (lsoa)).
- **borough** (Name of London boroughs).
- **major_category** (Major categorization of crimes).
- **minor_category** (Minor categorization of crimes according to major category).
- **value** (Number of crimes monthly reported in given borough).
- **year** (Year of reported crimes, Jan/2008 - Dez/2016).
- **month** (Month of reported crimes, Jan - Dez (1-12)).

	lsoa_code	borough	major_category	minor_category	value	year	month
0	E01004177	Sutton	Theft and Handling	Theft/Taking of Pedal Cycle	1	2016	8
1	E01000086	Barking and Dagenham	Theft and Handling	Other Theft Person	1	2009	5
2	E01001301	Ealing	Theft and Handling	Other Theft Person	2	2012	1
3	E01001794	Hackney	Violence Against the Person	Harassment	1	2013	2
4	E01000733	Bromley	Criminal Damage	Criminal Damage To Motor Vehicle	1	2016	4

Fig. 1 – First five rows of main dataset.

The additional dataset with respect to London's boroughs is composed by 33 rows and 10 columns as shown in Figure 2. The variables presented in the columns correspond to:

- Borough
- Inner
- Status Local Authority
- Political Control
- Headquarters
- Area
- Population
- Coordinates

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	Barking and Dagenham [note 1]	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E	25
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W	31
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E	23
3	Brent	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W	12
4	Bromley	NaN	NaN	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E	20

Fig. 2 – First five rows of additional dataset.

This additional dataset has several variables available, however only Borough and Population were the variables to be used for this project. Although the coordinates will also be necessary for this work, the variable Coordinates of this table was discarded since they are obtained afterwards applying geocoder to retrieve them and to start the battle of neighborhoods.

For the final part of this work, a list of neighborhoods was used, regarding the selected borough, and obtained from wikipedia website (source: https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames).

3. Methodology

The methodology for the development of this project was three-folded: 1) the exploratory data analysis, 2) the segmentation of neighborhoods and 3) the clustering neighborhoods.

Before the first method being applied, the data needed to be scrapped and manipulated to obtain the structured format and generate the dataframes. Regarding the main dataset, the rows with variable "value" equal to zero needed to be removed from the dataframe, i.e., the crimes were reported in the given boroughs, however they were classified as zero since they probably did not trully happened and therefore these rows were discarded from the original dataset. Then, the merge of both datasets was done, grouping the rows by boroughs.

3.1. Exploratory Data Analysis of London Criminality

The exploratory analysis of the data was an important section of this work because this type of analysis is very useful for a more statistical comprehensive interpretation through the combination of several variables of London criminal records. In this section, this interpretation took into account the number and type of crimes comitted in the boroughs of London, the evolution of criminality over time, the comparison of crimes between boroughs, etc. Different visualization tools were used such as line plots, scatter plot and histograms to understand and to determine the safest borough.

3.2. Segmenting Neighborhoods of the safest borough in London

After choosing the most suitable/ safest borough in London, a new dataframe was created with the list of 15 neighborhoods. At this point, the geopy library was used to get the values of geographical coordinates (latitude and longitude) associated to each neighborhood. Then, location data from Foursquare API was used to explore neighborhoods within the selected borough, retrieving the most common venue categories in each neighborhood. This part of the project is the so-called Segmentation of Neighborhoods, where the 10 most common venues with a radius of 750 meters, around each neighborhood coordinates center, were retrieved. Finally, the folium library was also used to visualize the location of neighborhoods in the maps of London's borough.

3.3. Clustering Neighborhoods of the safest borough in London

The final section of the methodology describes the clustering of neighborhoods, consisting in grouping the neighborhoods into clusters. For this, firstly an one hot encoder (source: *https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f*) was applied to perform the "binarization" of each category, i.e. the top 10 most common venues associated to each neighborhood. Then, and before applying the clustering algorithm, the several rows were grouped by neighborhood and the mean of the frequency of venues was taken. The K-means clustering algorithm was then applied to generate the several clusters of neighborhoods, using lastly the folium library to visualize the neighborhoods in London and their emerging clusters. Before the decision/ recommendation making of most suitable neighborhood, the analysis of each cluster and respective most common venues was performed.

4. Results

In this report section the several figures, obtained throughout the methods applied in this project, are presented using the matplot library.

4.1. Exploratory Data Analysis

Figure 3 illustrate the bi-distribution, traduced by a scatter plot, between the total number of crimes reported in 2013 and the population number in the same year for all the boroughs of Great London. It is clearly visible that a good positive correlation may be achieved between the data. Although the application of a supervised machine learning algotihm (e.g. linear regression) might be used to fit the data (the dashed blue line is one possible fitting solution), this was considered out of the scope regarding the main goal of this project. Nevertheless, it is possible to interpret that the higher is the population number for a given borough, the higher is the number of crimes reported. Moreover, the "Westminster" and "City of London" boroughs are the ouliers of the data, being also the dangerous and safest boroughs in London, respectively. However, we can consider that "Westminster" is the clear outlier because it does not even fall into the data behavior and the error to the predicted line would be bigger; the contrary is observed when "City of London" is considered, since even being an outlier of the data, the error of predicted values for this range in the plot would be smaller and it follows the linear relationship of the dataset.

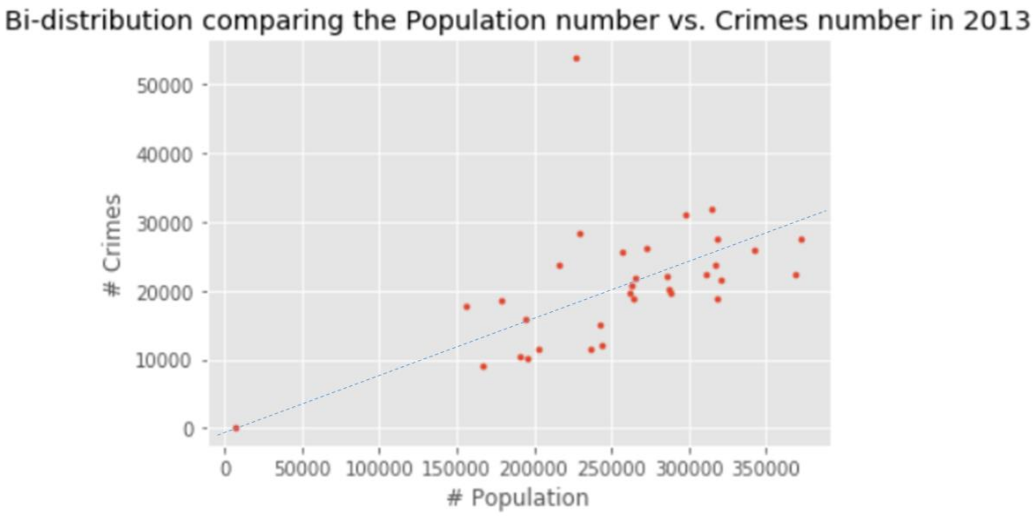

Fig. 3 – Scatter plot between population and crimes number of boroughs in London (2013).

major_category	
Burglary	754293
Criminal Damage	630938
Drugs	470765
Fraud or Forgery	5325
Other Notifiable Offences	106349
Robbery	258873
Sexual Offences	1273
Theft and Handling	2661861
Violence Against the Person	1558081
Name: value, dtype: int64	

minor_category	
Assault with Injury	451001
Burglary in Other Buildings	263011
Burglary in a Dwelling	491282
Business Property	21295
Common Assault	413690
Counted per Victim	3840
Criminal Damage To Dwelling	154116
Criminal Damage To Motor Vehicle	265463
Criminal Damage To Other Building	66003
Drug Trafficking	35819
Going Equipped	5530
Handling Stolen Goods	16100
Harassment	458124
Motor Vehicle Interference & Tampering	56224
Murder	949
Offensive Weapon	37983
Other Criminal Damage	145356
Other Drugs	2998
Other Fraud & Forgery	1485
Other Notifiable	100819
Other Sexual	1005
Other Theft	980085
Other Theft Person	308842
Other violence	70778
Personal Property	237578
Possession Of Drugs	431948
Rape	268
Theft From Motor Vehicle	569956
Theft From Shops	345142
Theft/Taking Of Motor Vehicle	216538
Theft/Taking of Pedal Cycle	168974
Wounding/GBH	125556
Name: value, dtype: int64	

Fig. 4 – Total number of crimes grouped by major (left) and minor (right) crime categories. The red rectangles are the maximum value of each category.

Due to extension of crime categories (especially the minor category) these values were preferred to be presented as described in Figure 4 and not as histogram, for instance. It is clear that “Theft and Handling” major category is the most often crime happened in London, followed by “Violence Against the Person”, both with more than 1 million crimes reported between 2008 and 2016.

Figures 5 and 6 are line plots illustrating the evolution of top2 types of crimes in London over the years and months, respectively. Starting from Figure 5, the general behavior is quite different between both crime types over the years. The highest number of reports happened in 2012 for “Theft and Handling, in spite of “Violence Against the Person” crime type that is the second most often in London, its behavior was stable between 2008 and 2013, significantly increasing afterwards.

Interpreting monthly both crimes in Figure 6, the general evolution behavior between “Theft and Handling” and “Violence Against the Person” crimes is similar over the months. Moreover, between 2008 and 2016 the most critical months were March, July and October for both crime types, being February and December the “safest” months with less reported cases.

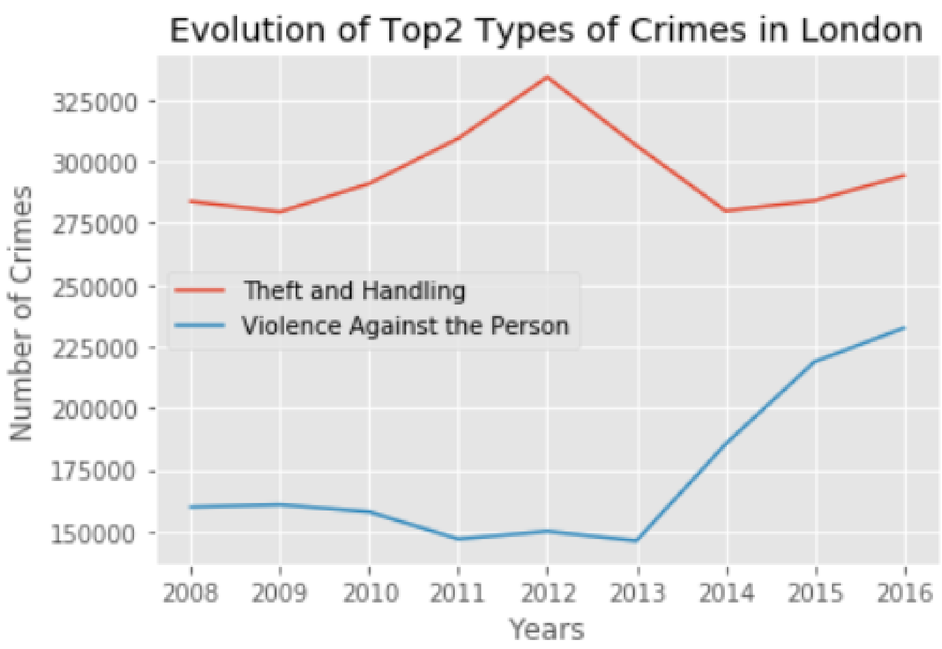

Fig. 5 – Line plot of evolution of top2 types of crimes in London over the years.

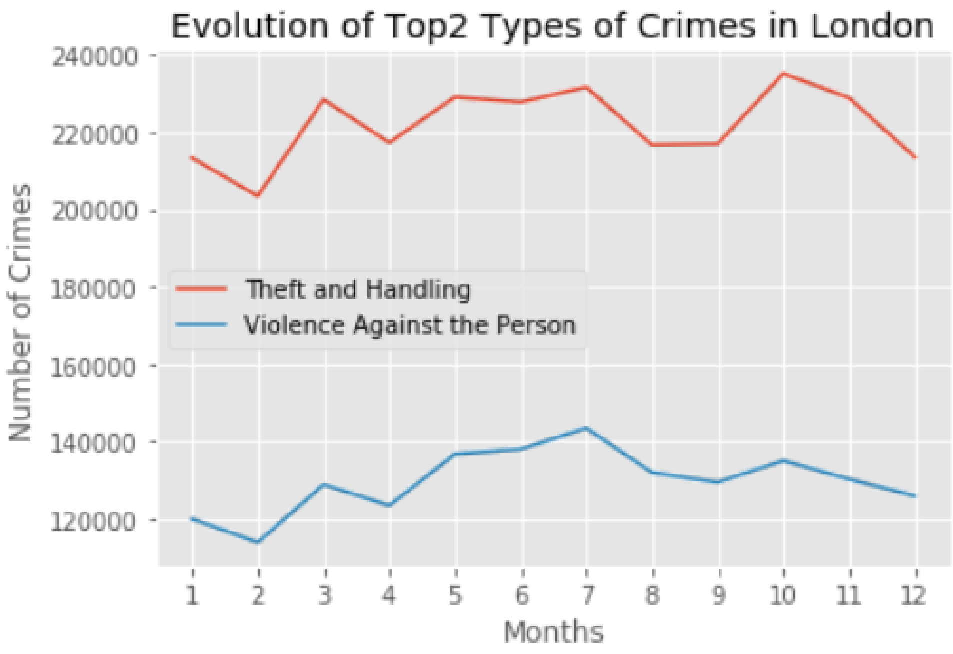

Fig. 6 – Line plot of evolution of top2 types of crimes in London over the months.

Figures 7 and 8 are the histograms of top5 boroughs with highest and lowest number of crimes, respectively, in Great London area. Clearly between 2008-2016 "Westminster" borough presents the higher number of reported crimes (Figure 7), therefore it's the most dangerous (almost half a million crimes); "City of London" is the safest borough in London with 780 reported crimes.

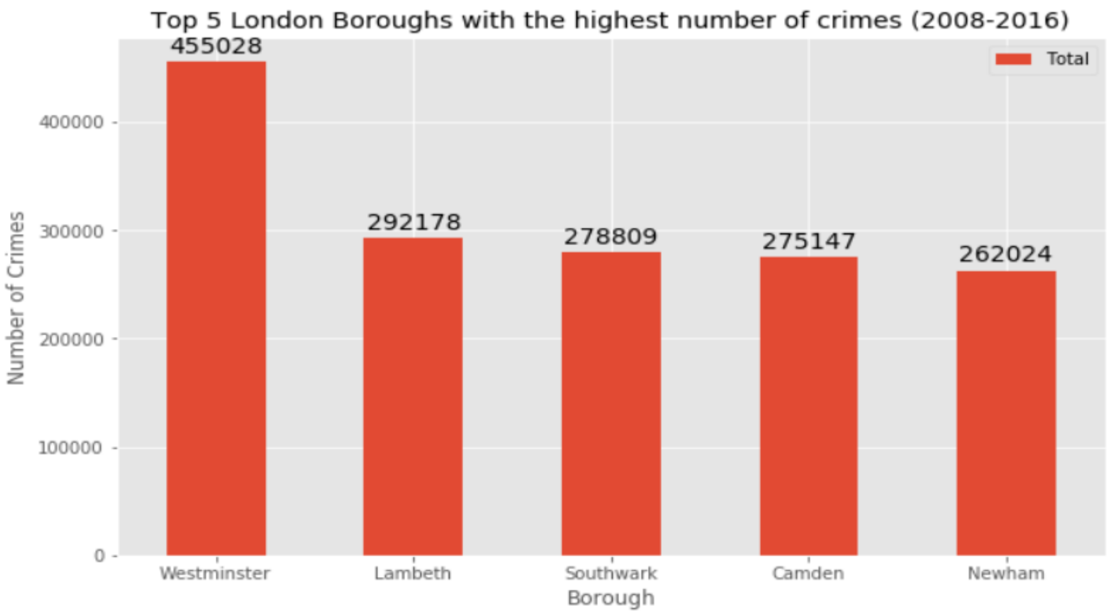

Fig. 7 – Histogram of top5 (higher values) of total crimes in London boroughs.

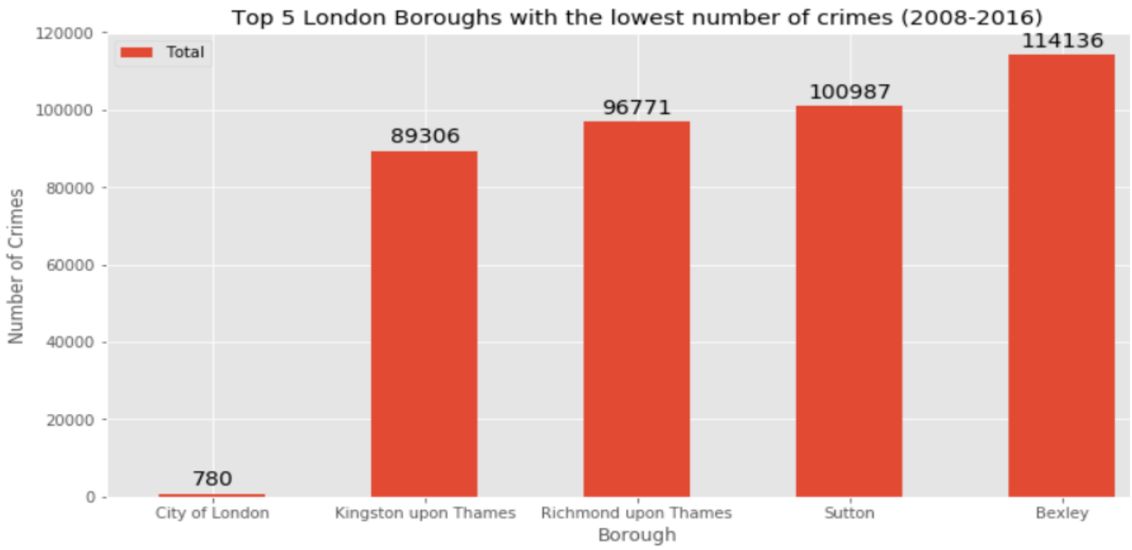

Fig. 8 – Histogram of top5 (lower values) of total crimes in London boroughs.

According to wikipedia webpage, “City of London” doesn’t consist trully a borough of Great London, although it was considered for the previous exploratory data analysis. For this reason, next results illustrated in Figures 9, 10 and 11 encompass only the interpretation between “Kingston upon Thames” and “Richmond upon Thames” to select the safest Borough in London.

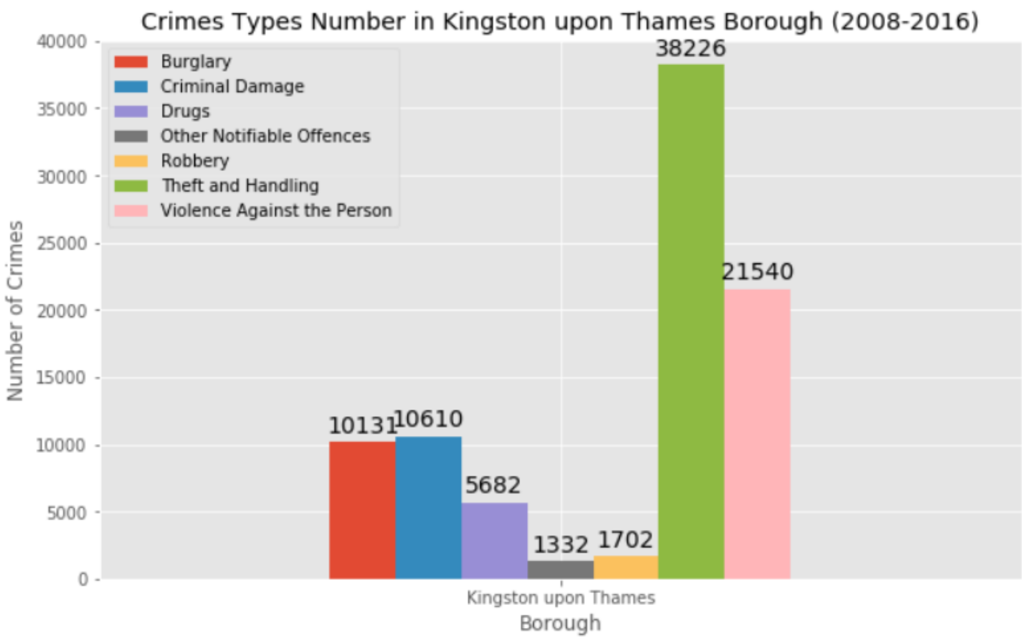

Fig. 9 – Histogram of major crimes types in Kingston upon Thames borough.

Figure 9 and 10 show a good similarity in the general behavior of Crime Types, within Major Category, between both analysed boroughs (although frequency number is different). Clearly the crime types “Theft and Handling” and “Violence Against the Person” happen more often. Similarly between both Boroughs, crime types classified as “Robbery” and “Other Notifiable Offences” are the less reported.

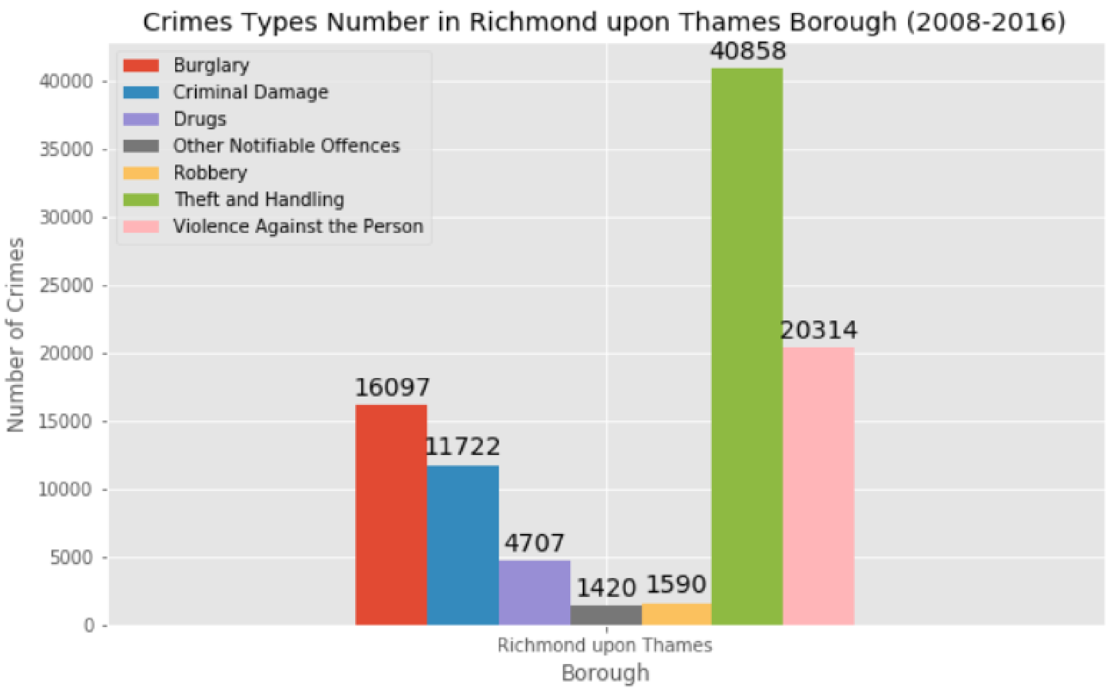

Fig. 10 – Histogram of major crimes types in Richmond upon Thames borough.

In Figure 11, between time intervals of 2008-2009 and 2015-2016 the evolution behavior is quite similar. Critical years of criminality maximums were 2008, 2012 and 2016, for “Richmond upon Thames”, and mostly 2008 for “Kingston upon Thames”. “Richmond upon Thames” registered a peak in crimes between 2010 and 2013. “Kingston upon Thames” shows a significant decrease in crimes between 2012 and 2013.

Based on this final exploratory data analysis between the two safest boroughs (discarding at this time the “City of London”), “Richmond upon Thames” “Kingston upon Thames”, the selected borough to conduct this project was “Kingston upon Thames” and will be used from now on.

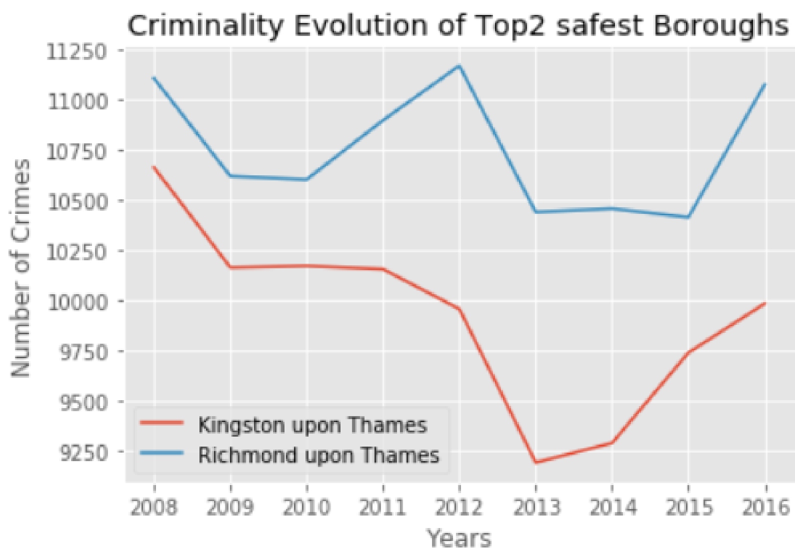

Fig. 11 – Line plot of evolution of total crimes over the years in Kingston upon Thames and Richmond upon Thames.

4.2. Segmenting Neighborhoods of safest borough in London

In this section, the results of segmentation of neighborhoods from "Kingston upon Thames" borough are presented. Figure 12 shows a table with the results after using the geolocator (with a user_agent = "London agent") to find the latitude and longitude of each neighborhood.

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames	51.393781	-0.284802
1	Canbury	Kingston upon Thames	51.417499	-0.305553
2	Chessington	Kingston upon Thames	51.358336	-0.298622
3	Coombe	Kingston upon Thames	51.419450	-0.265398
4	Hook	Kingston upon Thames	51.367898	-0.307145
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898
9	New Malden	Kingston upon Thames	51.405335	-0.263407
10	Norbiton	Kingston upon Thames	51.409999	-0.287396
11	Old Malden	Kingston upon Thames	51.382484	-0.259090
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366
13	Surbiton	Kingston upon Thames	51.393756	-0.303310
14	Tolworth	Kingston upon Thames	51.378876	-0.282860

Fig. 12 – Neighborhoods from Kingston upon Thames borough and respective coordinates retrived using geopy library.

The geolocator was also used to determine the central coordinates of the borough, i.e. the central point of “Kingston upon Thames”, which is “Berrylands” with values of 51.3937811; -0.2848024. It is worth to mention that the coordinates of the center for the entire borough are very similar with those of “Berrylands” neighborhood center (Figure 12).

Using the elements of Figure 12, the spatial distribution of neighborhoods was ploted and shown in the map of Figure 13. Now that we have all the neighborhoods’ location let’s move to the battle of neighborhoods!

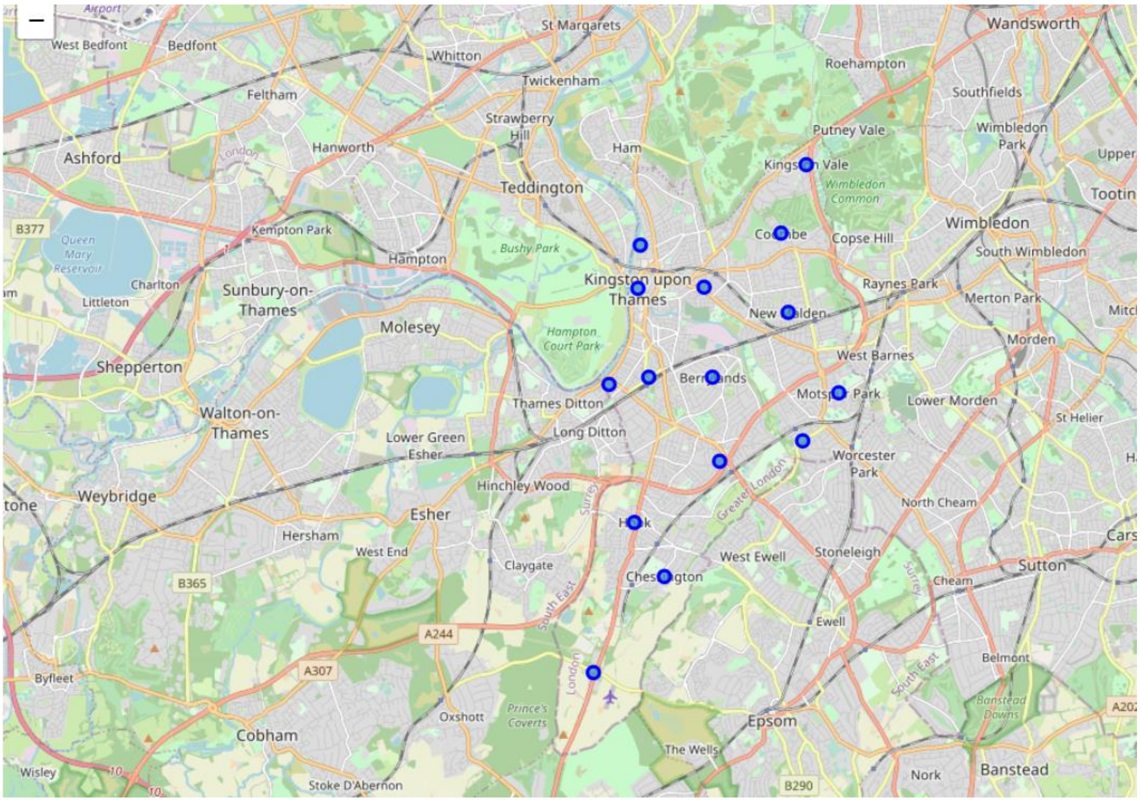

Fig. 13 – Map of neighborhoods (blue dots) from Kingston upon Thames borough.

4.3. Clustering Neighborhoods of safest borough in London

The final results shown in this section are related with the modeling of the data using firstly the Foursquare API to get the venues to every neighborhood and applying the K-means clustering algorithm to find the similar groups of neighborhoods based on similarity of ranked common venues, i.e. the mean frequency of venues within each neighborhood.

Figure 14 shows the final map of “Kingston upon Thames” borough with the five clusters of neighborhoods. The most important parameters used to obtain the final clustering were the definition of the radius to explore the venues of each neighborhood (radius =

750 meters) and the pre-defined number of clusters for the K-means algorithm. Clearly cluster 2 is located in the western part of the borough and cluster 3 regards the the most eastern part of the borough. Clusters 1 and 4 are clearly distinct in what common venues concern.

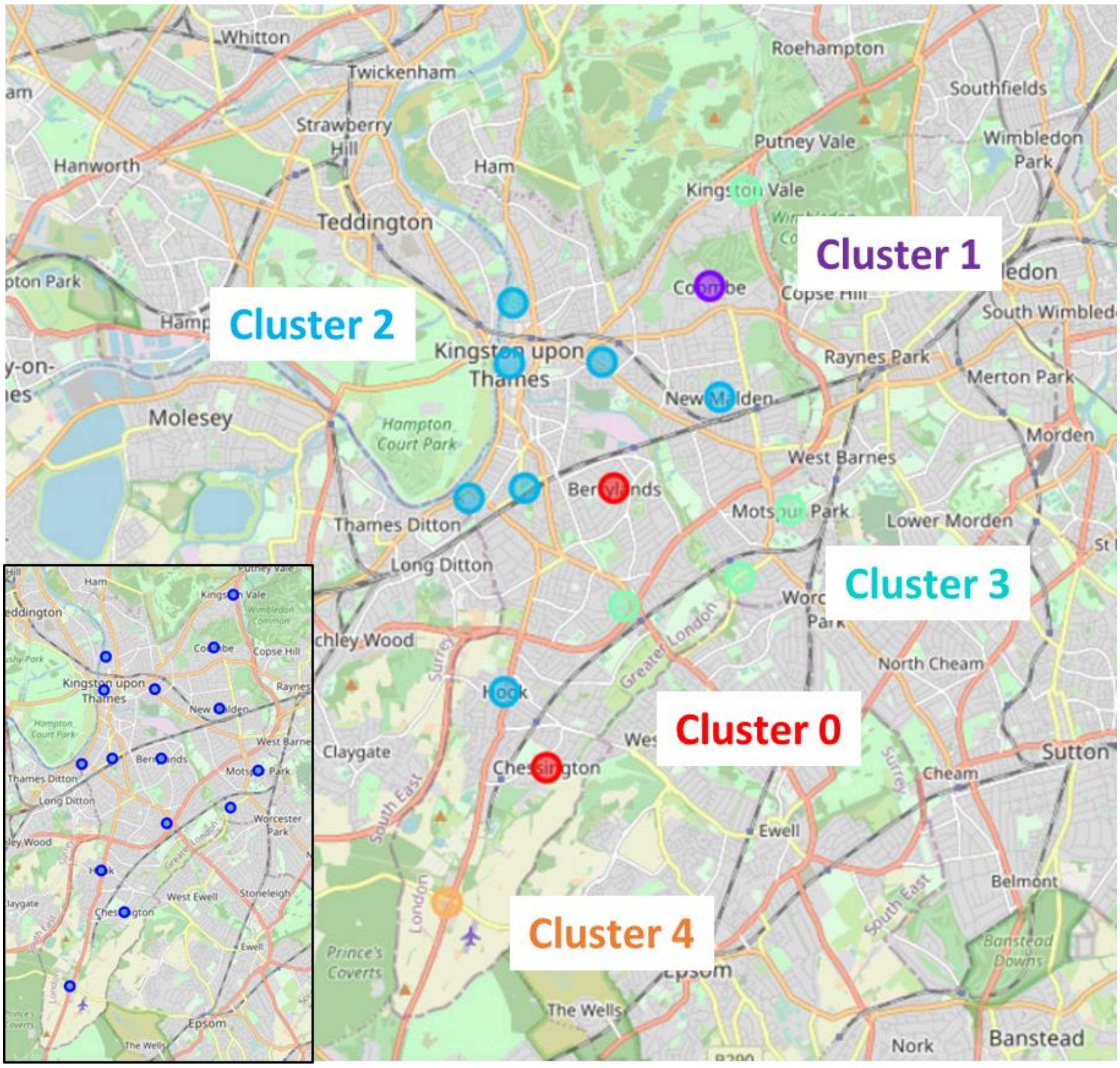

Fig. 14 – Map of five clusters of neighborhoods from Kingston upon Thames borough.

Finally, Figure 15 presents the five final tables as the output of clustering, showing the content of each cluster and respective common venues.

A more comprehensive analysis of each cluster, individually and the comparison between each other, is acomplished in the next section of discussion.

	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	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	0	Pub	Bus Stop	Coffee Shop	Park	Platform	Gym / Fitness Center	Train Station	Farmers Market	Department Store	Discount Store
2	Chessington	Kingston upon Thames	51.358336	-0.298622	0	Train Station	Golf Course	Breakfast Spot	Convenience Store	Fast Food Restaurant	Platform	Garden	Furniture / Home Store	Fried Chicken Joint	French Restaurant

	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	10th Most Common Venue
3	Coombe	Kingston upon Thames	51.41945	-0.265398	1	Golf Course	Hotel	Garden	Spa	Food	Discount Store	Donut Shop	Electronics Store	Farmers Market	Fast Food Restaurant

	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	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	2	Café	Coffee Shop	Pub	Hotel	Italian Restaurant	Supermarket	Plaza	Turkish Restaurant	Fish & Chips Shop	Indian Restaurant
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Breakfast Spot	Indian Restaurant	Convenience Store	Park	Fast Food Restaurant	Fish & Chips Shop	Supermarket	Platform	Bakery	Grocery Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	2	Café	Coffee Shop	Pub	Burger Joint	Sushi Restaurant	Mexican Restaurant	Department Store	Cosmetics Shop	Clothing Store	Portuguese Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Korean Restaurant	Supermarket	Indian Restaurant	Fast Food Restaurant	Clothing Store	Café	Used Auto Dealership	Karaoke Bar	Department Store	Gastropub
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	2	Italian Restaurant	Gastropub	Indian Restaurant	Pub	Fried Chicken Joint	Gym / Fitness Center	Japanese Restaurant	Pharmacy	Pizza Place	Coffee Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	2	Coffee Shop	Pub	Indian Restaurant	Restaurant	Gastropub	French Restaurant	Golf Course	Grocery Store	Gym / Fitness Center	Fish & Chips Shop
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	2	Coffee Shop	Pub	Grocery Store	Restaurant	French Restaurant	Fish & Chips Shop	Indian Restaurant	Italian Restaurant	Farmers Market	Deli / Bodega

	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	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	3	Coffee Shop	Stables	Outdoors & Recreation	Sandwich Place	Soccer Field	Bus Stop	Grocery Store	Bar	Bistro	Food Truck
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Tennis Court	Park	Soccer Field	Bus Stop	Japanese Restaurant	Steakhouse	Bakery	Grocery Store	Furniture / Home Store
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	3	Grocery Store	Gym / Fitness Center	Japanese Restaurant	Bakery	Steakhouse	Train Station	Gastropub	Food	Discount Store	Donut Shop
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	3	Grocery Store	Soccer Field	Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Discount Store	Garden Center	Sandwich Place	Café

	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	10th Most Common Venue
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	4	Theme Park Ride / Attraction	Pub	Restaurant	Garden Center	Hotel	Grocery Store	Bar	Food Truck	Donut Shop	Electronics Store

Fig. 15 – Tables of each cluster and respective group of neighborhoods and top10 most common venues for each one.

5. Discussion

Recalling the results shown in Figure 15, the analysis of each cluster was performed and the top5 common venues were selected to a facilitate a better understanding the differences and the similarities of neighborhoods. It is important to notice that different k values were tested in the K-means algorithm, however there were always two distinct neighborhoods when compared to the remaining ones, i.e. Coombe (cluster 1) and Malden Rushett (cluster 4). When using the radius equal to 500 meters or lesser, there were neighborhoods were no venues were found.

Keeping the clustering k-value equal to 5 and the exploration radius of 750 meters, the following list of common venues, corresponding to each cluster, are the following:

Cluster 0:

- Pub
- Train Station
- Bus Stop
- Coffee Shop
- Golf Course

Cluster 1:

- Golf Course
- Hotel
- Garden
- Spa
- Food

Cluster 2:

- Café
- Breakfast Spot
- Coffee Shop
- Supermarket
- Restaurants (Korean, Italian, Indian)

Cluster 3:

- Coffee Shop
- Gym/ Fitness Center
- Grocery Store
- Tennis Court
- Soccer Field

Cluster 4:

- Theme Park Ride/ Attraction
- Pub
- Restaurant
- Garden Center
- Hotel

After analysing the clusters, the neighborhood recommendation would be Berrylands, neighborhood represented as a star in Figure 16.

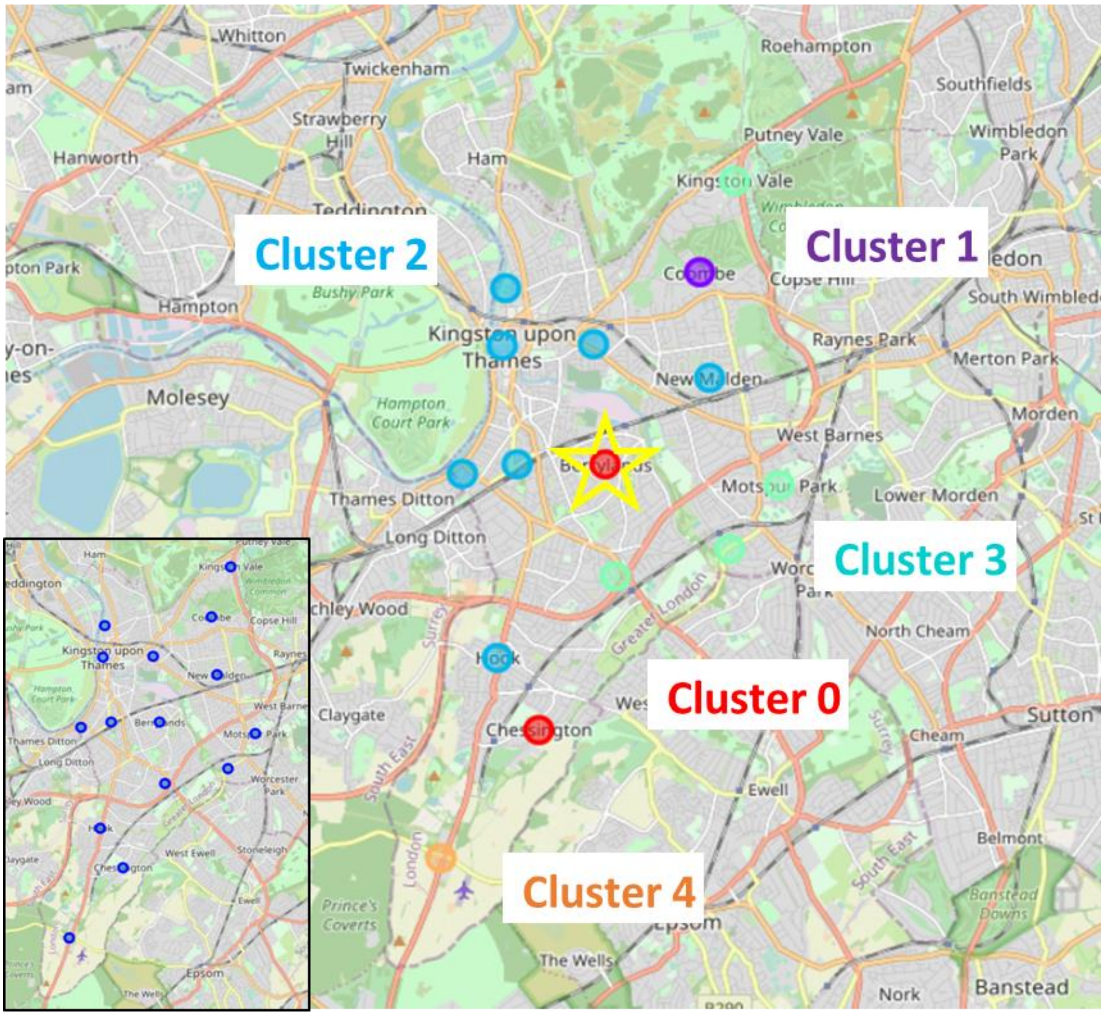

Fig. 16 – Map of five clusters of neighborhoods from Kingston upon Thames borough. The star represents the recommended neighborhood.

Independently if the interested audience is just a citizen who wants to find an house or a businessman to open a new office, neighborhood Berrylands would be the most suitable area where just a walk of a few minutes we can find social venues as pubs or cofee shops, or public transport points as train station and bus stops, or either if one is interested in taking a golf course is also very closed to the neighbohood center.

Moving a bit further, i.e. 750 meters or more, food shops restaurants, grocery, breakfast spots) and supermarkt restaurant stores can also be found as wells as gym places, cafes and sport and leisure zones (e.g. soccer, tenis, garden).

Common Venues (max <=750 meters)

- Pub
- Train Station
- Bus Stop
- Coffee Shop
- Golf Course

Common Venues from adjacent Clusters 2 and 3 (max > 750 meters)

- Café
- Gym/ Fitness Center
- Breakfast Spot
- Supermarket
- Grocery Store
- Restaurants (Korean, Italian, Indian)
- Tennis Court
- Soccer Field
- Golf Course
- Hotel/ Spa
- Garden

6. Conclusions

In this project the exploratory analysis of London crime data was done to better understand how criminality components evolved during last years and to find the safest borough (Kingston upon Thames).

Segmenting and clustering of neighborhoods belonging to Kingston upon Thames borough was achieved using Foursquare API and K-means algorithm, allowing to explore and retrieve the most common venues around each neighborhood and cluster them based on venues similarity; After analysing the several clusters and respective venues, the final recomendation to everybody who is interested to find a new house, or a new work office for a start-up company, would be Berrylands neighborhood.

As future work of this project, the extension of this method to other London boroughs would be valuable but also the incorporation of property/ offices costs to complement this work, since only topics of safety and common venues were taken in consideration.

7. References

London Crime records:

https://www.kaggle.com/jboysen/london-crime

List of London boroughs:
https://en.wikipedia.org/wiki/List_of_London_boroughs

One hot encoding:

https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f

List of neighborhoods in Kingston upon Thames borough:

https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames