Report

Capstone project: The Battle of the Neighbourhoods

Elena Lestini

1. Introduction

1.1 Background

Education and safety play a big role in family decisions to move from one London borough to another. More space and good schooling attract a growing number of families out of central London into greater London. Greater London has a lot to offer, with shops, cafés, libraries, theatres and parks. There is a lot going on for families. It is also well connected by train, bus and underground, making it easy to commute to work.

1.2 Problem

The problem we want to solve with this data science project is to find good schools in safe boroughs in London. London comprises of 32 boroughs. Finding a good school in a safe borough is an important task for many families with young children and teenagers. Many families decide to move borough to make sure their children attend a good school. Although there is a lot of interest in finding good state schools in London, it is quite difficult to have a real grasp of the situation. Detailed reports on this topic are available, but they are not easy to read, especially for people moving from countries with a different school system and/or from non-English-speaking countries. The information and statistics available are not provided in a user-friendly manner and this can be quite frustrating for families moving to London.

1.3 Interest

There is a lot of interest in this topic, from families with children and teenagers but also from couples who plan to start a family and foreign families moving to London for work.

This is such a popular topic that every year reports on schools and catchment areas make the headlines of the newspapers. Some examples are reported in the links below.
London property prices in catchment area of state good schools,
Top of the class:10 of the best London boroughs to move to for good state

schools,
London's 10 super state schools

There are also many blogs available (e.g. TheGoodSchoolGuide, Mumsnet) where parents discuss, ask, post messages. It is a growing community and an ongoing topic. This is also discussed a lot on social media, such as Facebook.

2. Data acquisition and cleaning

2.1 Data acquisition

This project uses open source data from London local authorities regarding youth crime (see https://www.london.gov.uk/mopac-disclosure-log/london-knife-crime-statistics)

	London Borough	2017/18 Total	2017/18 With Injury	2018/2019 Total	2018/2019 With Injury	% Change in total 2017/18 - 2018/19
0	Westminster	650.0	169.0	985.0	199	0.5154
1	Southwark	866.0	318.0	777.0	239	-0.1028
2	Haringey	794.0	229.0	764.0	179	-0.0378
3	Newham	787.0	235.0	696.0	197	-0.1156
4	Brent	766.0	241.0	680.0	191	-0.1123
5	Tower Hamlets	715.0	217.0	667.0	221	-0.0671

Table 1. Knife crimes per borough in years 2017/18 and 2018/19.

Features used in this project: **London Borough** and **2018/2019 Total** (number of knife crimes per borough in 2018/2019); Table 1. shows the original dataset.

High-school General Certificate of Secondary Education (GCSE) achievements to find the best performing schools in the safest London boroughs (see https://data.london.gov.uk/dataset/gcse-results-by-borough and Table2).

	London Borough	% pupils achieving Math & English strong 9-5 pass	% pupils achieving Math & English standard 9-4 pass
0	Barking and Dagenham	40.2	60.0
1	Barnet	60.6	76.0
2	Bexley	51.0	69.1
3	Brent	50.9	69.5
4	Bromley	49.3	70.0
5	Camden	47.0	66.0
6	Croydon	42.2	63.1
7	Ealing	51.0	69.2

Table 2. Math & English GCSEs strong and standard pass (2015-2018).

Features used in this project: **London Borough** and **% pupils achieving Math & English strong 9-5 pass** (percentage of pupils achieving top grades in Math and English GCSEs, 2015-2018);

The **List of London Boroughs** and the corresponding **coordinates** were obtained by scraping a Wikipedia page (https://en.wikipedia.org/wiki/List_of_London_boroughs) with BeautifulSoup package in Python. The data were stored in Table 3. However, undesired characters and strings were removed (cleaning) as they would have caused later problems in the visualisation with Folium and in the search of popular venues with Foursquare API.

Before cleaning

	Borough	coordinates
1	Barking and Dagenham [note 1]	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /...
2	Barnet	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /...
3	Bexley	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /...
4	Brent	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /...
5	Bromley	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /...
6	Camden	51°31′44″N 0°07′32″W / 51.5290°N 0.1255°W /...

After cleaning

	Borough	coordinates
0	Barking and Dagenham	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /...
1	Barnet	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /...
2	Bexley	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /...
3	Brent	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /...
4	Bromley	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /...
5	Camden	51°31′44″N 0°07′32″W / 51.5290°N 0.1255°W /...

Table 3. Boroughs and coordinates scraped from Wikipedia.

After visualising the pupils GCSEs achievement 2015-2018 per borough and the total number of knife crime offences per borough (2018/2019), the Royal Borough of Sutton was identified as the safest borough with the best performing high school.

The data on the neighbourhoods within Sutton were scraped from Wikipedia and their GPS latitude and longitude was obtained with Google Geocoding API. The final table showing Sutton neighbourhoods and the corresponding latitude and longitude was obtained (Table 4).

	Borough	District	Latitude	Longitude
0	Sutton	Bandon Hill	51.364777	-0.134833
1	Sutton	Beddington	51.371988	-0.132393
2	Sutton	Beddington Corner	51.386942	-0.149532
3	Sutton	Belmont	51.343785	-0.201152
4	Sutton	Benhilton	51.371642	-0.191571
5	Sutton	Carshalton	51.365788	-0.161086
6	Sutton	Carshalton Beeches	51.357196	-0.169351
8	Sutton	Cheam	51.357616	-0.216241
9	Sutton	Hackbridge	51.379613	-0.156754
10	Sutton	Little Woodcote	51.346076	-0.145932

Table 4. Borough of Sutton, neighbourhoods and corresponding latitude and longitude.

In this format (Borough, District (i.e. neighbourhood), Latitude, Longitude) the data were ready to be used for map visualization with Folium and to retrieve the most popular venues with Foursquare API.

3. Methodology

3.1 Exploratory Data Analysis

Bar-charts and maps were useful to determine which boroughs are the safest with the best achieving high schools.

3.2 Knife-crime and high schools

Bar-charts (Fig. 1) were employed for the visualization of both the GCSEs Math & English strong pass and number of knife crimes in the 22 London Boroughs. The bar-charts show that Sutton is the safest borough in London with the lowest number of knife crimes and with the best achieving high schools in London. Kingston upon Thames is the second best.

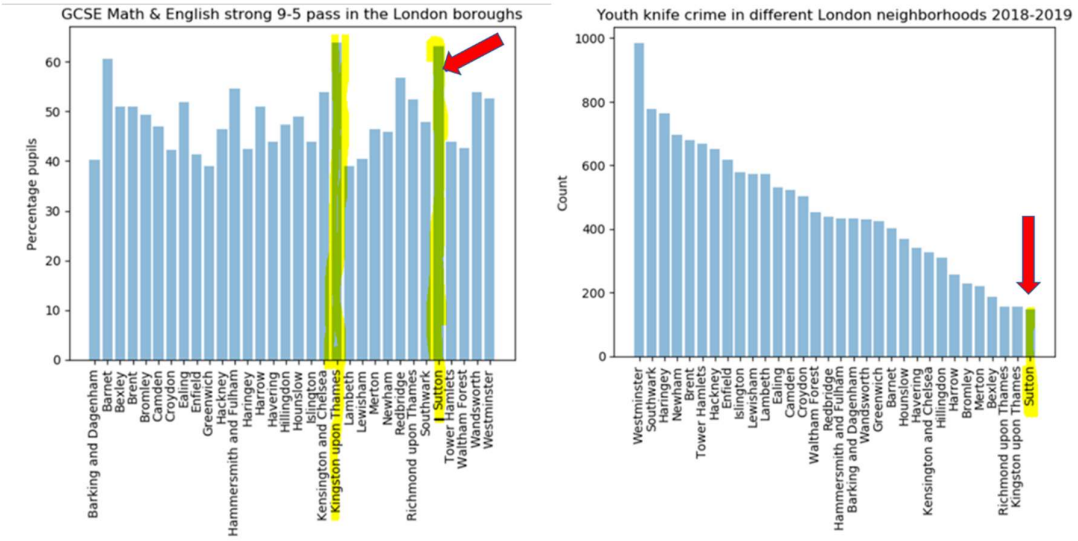

Fig. 1 (Left hand-side) strong achievements in Math & English per borough from 2015-2018; (right hand-side) number of offences (knife crime 2018/2019) per borough.

Due to such results Sutton and its neighbourhoods were further explored by using Google Geocoding API, Foursquare API and k-means clustering modelling.

3.3 Spatial visualisation of Sutton neighbourhoods

The spatial visualisation of Sutton neighbourhoods with Folium was also carried out and the result is shown in Figure 2. Each neighbourhood is represented by a coloured dot on the London map.

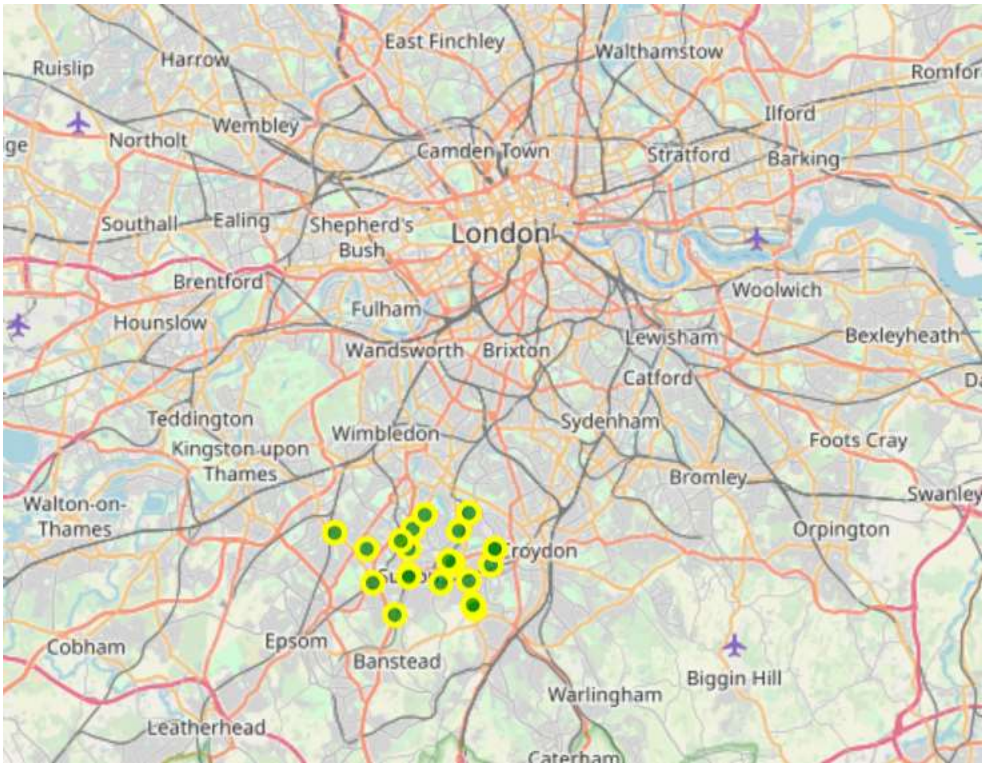

Fig. 2 Geospatial map of the Sutton neighbourhoods.

3.4 Modelling

After explorative analysis Table 4 was created (Sutton neighbourhoods and corresponding coordinates) and latitude and longitude were used to request from Foursquare API the top ten most popular venues per neighbourhood in a radius of 500 m. The response is a json nested file that is finally converted into a pandas dataframe (Table 5).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bandon Hill	51.364777	-0.134833	The Plough	51.367633	-0.132089	Pub
1	Bandon Hill	51.364777	-0.134833	Demesne Rd Allotments	51.362752	-0.140418	Garden
2	Bandon Hill	51.364777	-0.134833	Mellows Park	51.360692	-0.134101	Park
3	Beddington	51.371988	-0.132393	Carew Manor	51.370983	-0.136604	Park

Table 5. Screenshot of the most popular venues in radius of 500 m of each Sutton neighbourhood. These are the first 3 rows out of 220 rows.

The categorical variables in Table 5 were transformed into numerical data by hot encoding. Numerical data are necessary to perform machine learning and k-mean clustering.

An unsupervised K-mean machine learning clustering algorithm was employed on the hot-encoded data. The aim is to create clusters of neighbourhoods. Neighbourhoods that are similar fall in the same cluster. In this case the common denominator for the Sutton neighbourhoods are the facilities and venues. Neighbourhoods with similar facilities and venues belong to the same cluster and families can explore neighbourhoods based on the similarities or dissimilarities in terms of facilities and venues. This would help them have a better idea of the area and possibly make the right choice.

4. Results

K-means clustering provided 5 clusters (n=5 was set). The Sutton neighbourhoods in each cluster are similar for most popular venues.

	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	Belmont	Electronics Store	Asian Restaurant	Train Station	Pub	Women's Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub
1	Benhilton	Indian Restaurant	Park	Coffee Shop	Clothing Store	Pizza Place	Supermarket	Grocery Store	Gym / Fitness Center	Gastropub	Fast Food Restaurant
2	Carshalton	Grocery Store	Pub	Park	Café	Veterinarian	Historic Site	Train Station	Hotel	Tea Room	Platform
3	Carshalton Beeches	Health Food Store	Italian Restaurant	Train Station	Bakery	Grocery Store	Electronics Store	Hardware Store	Gym / Fitness Center	Gym	Gastropub
4	Cheam	Grocery Store	Italian Restaurant	Pub	Coffee Shop	Indian Restaurant	Soccer Field	Creperie	Gastropub	Gym / Fitness Center	Park
5	North Cheam	Coffee Shop	Supermarket	Fast Food Restaurant	Seafood Restaurant	Soccer Field	Social Club	Pub	Gym / Fitness Center	Grocery Store	Turkish Restaurant
6	Sutton	Pub	Clothing Store	Coffee Shop	Italian Restaurant	Bar	Pizza Place	Café	Women's Store	Sandwich Place	Hotel
7	Sutton High Street	Pub	Coffee Shop	Italian Restaurant	Bar	Pizza Place	Café	Electronics Store	Hotel	Irish Pub	Park
8	The Wrythe	Pub	Grocery Store	Café	Gym / Fitness Center	Historic Site	Hotel	Train Station	Platform	Tea Room	Park
9	Wallington	Grocery Store	Bakery	Supermarket	Spanish Restaurant	Pizza Place	Portuguese Restaurant	Pub	Convenience Store	Coffee Shop	Fast Food Restaurant
10	Worcester Park	Grocery Store	Pharmacy	Lebanese Restaurant	Pub	Fish & Chips Shop	Platform	Convenience Store	Coffee Shop	Pizza Place	Pet Store

Table 6. Sutton, Cluster 1.

11 neighbourhoods (i.e. Belmont, Benhilton, Carshalton, Carshalton Beeches, Cheam, North Cheam, Sutton, Sutton High Street, The Wrythe, Wallington, Worcester Park) fall into the biggest cluster (cluster 1, Table 6), indicating that such neighbourhoods are similar.

Neighbourhoods in cluster 1 seem to offer much more amenities than the neighbourhoods belonging to other clusters. In cluster 1 we can find a lot of international restaurants, pubs, coffee shops and grocery stores. Moreover, four neighbourhoods have a train station which can be a plus.

Cluster 2

```
1 sutton_districts_merged.loc[sutton_districts_merged['Cluster Labels'] == 1, sutton_districts_merged.columns[[1] + list(range
```

	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	Little Woodcote	Park	Garden Center	Sports Club	Coffee Shop	Women's Store	Electronics Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store
1	St. Helier	Turkish Restaurant	Coffee Shop	Park	Breakfast Spot	Women's Store	Garden Center	Fish & Chips Shop	French Restaurant	Furniture / Home Store	Garden
2	Woodcote Green	Park	Garden Center	Coffee Shop	Women's Store	Electronics Store	Health Food Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store

Table 7. Sutton, Cluster 2.

Cluster 2 (Table 7) comprises of three Sutton neighbourhoods (Little Woodcote, St. Helier, Woodcote Green). The three neighbourhoods have parks and garden centres as most common venues. There are few restaurants and leisure areas in these three neighbourhoods.

Cluster 3

```
1 sutton_districts_merged.loc[sutton_districts_merged['Cluster Labels'] == 2, sutton_districts_merged.columns[[1] + list(range
```

	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	Hackbridge	Convenience Store	Train Station	Park	River	Women's Store	Garden	Fast Food Restaurant	Fish & Chips Shop	French Restaurant	Furniture / Home Store
1	Rosehill	Park	Supermarket	Gym / Fitness Center	Tennis Court	Fast Food Restaurant	Athletics & Sports	Gym	Grocery Store	Electronics Store	Hardware Store
2	Sutton Common	Grocery Store	Athletics & Sports	Train Station	Gym / Fitness Center	Tennis Court	Park	Garden Center	Fish & Chips Shop	French Restaurant	Furniture / Home Store

Table 8. Sutton, Cluster 3.

Cluster 3 (Table 8) comprises of three Sutton neighbourhoods (Hackbridge, Rosehill and Sutton Common). The three neighbourhoods have convenience stores, supermarkets and parks as most common venues. Similarly to cluster 2, it seems that there are few restaurants and leisure areas in these three neighbourhoods. Conveniently, Hackbridge and Sutton Common both have a train station.

Cluster 4

```
1 sutton_districts_merged.loc[sutton_districts_merged['Cluster Labels'] == 3, sutton_districts_merged.columns[[1] + list(range
```

	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	Beddington Corner	Business Service	Furniture / Home Store	Racetrack	Women's Store	Electronics Store	Health Food Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store

Table 9. Sutton, Cluster 4.

Cluster 4 (Table 9) comprises of only one neighbourhood, Beddington Corner. Such neighbourhood is different to all the others. Indeed, one can see that the most popular venues (Business Service, Furniture Store, Home Store, Racetrack etc.) are quite different to those found for the other clusters.

Cluster 5

```
sutton_districts_merged.loc[sutton_districts_merged['Cluster Labels'] == 4, sutton_districts_merged.columns[[1] + list(range
```

	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	Bandon Hill	Park	Garden	Pub	Women's Store	Electronics Store	Hardware Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub
1	Beddington	Indian Restaurant	Hardware Store	Park	Pub	Electronics Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub	Garden Center
2	South Beddington	Indian Restaurant	Hardware Store	Park	Pub	Electronics Store	Gym / Fitness Center	Gym	Grocery Store	Gastropub	Garden Center

Table 10. Sutton, Cluster 5.

Cluster 5 (Table 10) comprises of three Sutton neighbourhoods (Bardon Hill, Beddington, South Beddington). Parks, gyms, Electronics stores are quite common popular venues. No train station is available for any of these neighbourhoods in a radius of 500 m.

The five clusters were visualised on the London map with Folium (Fig. 3). Each cluster is encoded with a colour. Dots of the same colour represent neighbourhoods belonging to the same cluster.

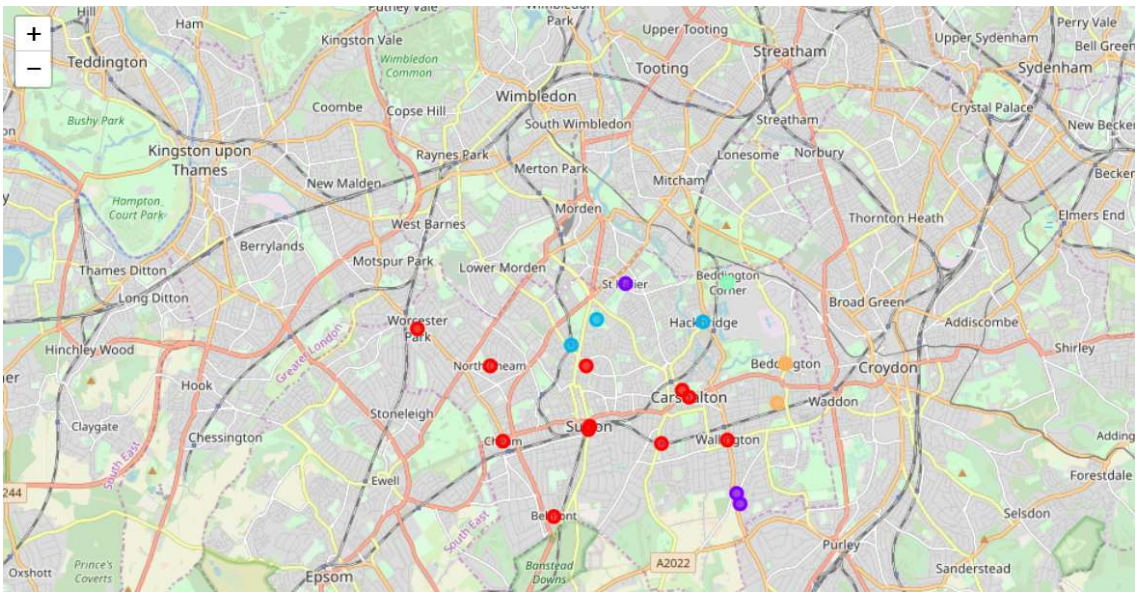

Fig. 3 Geospatial map of the Sutton neighbourhoods clustered in 5 groups (same colour means same cluster).

5. Discussion

The Royal Borough of Sutton comprises of 22 neighbourhoods. This borough is the safest (youth crime / knife crime) and has the best performing high schools in London (official datasets from London local authorities).Therefore, Sutton seems a good place for families with children to move to and live.

The Sutton neighbourhoods have been explored and the ten most popular venues per neighbourhood were found (e.g. cafes, food stores, train stations, international restaurants, etc.)

The Sutton neighbourhoods were also grouped in five clusters by the k-mean algorithm and 11 neighbourhoods (i.e. Belmont, Benhilton, Carshalton, Carshalton Beeches, Cheam, North Cheam, Sutton, Sutton High Street, The Wrythe, Wallington, Worcester Park) fall into the biggest cluster (cluster 1), indicating that such neighbourhoods are similar.

Neighbourhoods in cluster 1 offer more amenities than the other Sutton neighbourhoods belonging to other clusters (cluster 2, 3, 4 and 5) while at the same time offering the same level of safety and schooling.

Therefore, cluster 1 Sutton neighbourhoods seem to be the most suitable for families with children who are looking for good schools, a high level of safety but also for a place with many amenities and a vibrant community.

The remaining clusters are quite small (3 and a single neighbourhood, respectively) and less diverse than cluster 1. In such clusters there is little choice of restaurants, although in cluster 5 gyms and pubs are quite common.

The results show neighbourhoods with a train station (in cluster 1 and 3). This can be quite useful for all those people who need to commute for work and might want to buy a house in the vicinity of a train station.

Finally, it is quite interesting to see that parks are among the most common venues in most of the Sutton neighbourhoods. This can be important for families with young children and who own a dog. Open green areas like parks in quiet neighbourhoods are quite popular in a hectic capital city such as London.

6. Conclusion

This project aimed to explore neighbourhoods in the safest boroughs of London with the best achieving high schools.

This is an important topic not only among people living in London but also for those who are planning to move to London from abroad or from other regions in the United Kingdom.

Datasets available from an open source local government website and obtained by scraping few Wikipedia webpages were used to solve the task. When the datasets were visualised, it was found that the Royal Borough of Sutton has the best performing high schools (Math & English GCSEs) and it is also the safest borough in London with low youth crime (i.e. knife crime). Based on these findings Sutton neighbourhoods were explored and the most popular venues per neighbourhood were found using the Foursquare API. Finally, neighbourhoods with similar venues were clustered together using an unsupervised k-mean machine learning algorithm and the results were visualized on the London map.

The five clusters are quite different. Cluster 1, the biggest one, comprises of 11 neighbourhoods and by analysing the most popular venues it was possible to draw the conclusion that such neighbourhoods have a lot to offer in terms of amenities, while at the same time offering a high level of safety and schooling.

The Sutton neighbourhoods belonging to the remaining clusters are quite small (3 or 1 neighbourhoods per cluster) and offer a lower diversity of amenities. Probably, these are quiet neighbourhoods.

It was also possible to find which Sutton neighbourhoods have a train station; this could be quite important for all the people who need to commute to and from central London for work.

In conclusion, this approach gives a good snapshot of a neighbourhood. This could help families in their choice as to where to move, depending on their needs and/or personal interests.