**Capstone project – The Battle of Neighbourhoods**

**London’s Crime and Safest Areas**

**Introduction/Business Problem**

Before moving to a new area we don’t really know, to a new city in our own country, or even to a foreign country, we need to know a bit about the place we’re going to live in beforehand. Especially when it comes to very big cities like New York or London, it’s crucial to know where we can rent/buy a house or a business, namely which areas are the ones with the venues we like and are also safe. Therefore, safety will be our main concern, as nobody wants to spend time and money just to risk being robbed, or worse. No place in the world is crime-free, but making an effort to explore and understand crime stats for the place we want to move to is never a waste of time.

In this report, the crime statistics for London will be considered. The goal is to identify the safest borough of London, explore its neighbourhoods, and detect the 10 most common venues in each neighbourhood. These data might help any person who wishes to move to a safe and interesting area in London.

**Data**

*Data acquisition*

The first source of data will be the crime statistics dataset for London, 2008-2016, found on Kaggle at the following link <https://www.kaggle.com/jboysen/london-crime> . Since we are in 2020, only the most recent data from 2016 will be considered. The dataset is divided into the following columns:

* Lsoa code (Lower Super Output Area)
* Borough
* Major category (high level categorisation of crime)
* Minor category (low level categorisation of crime)
* Value (count of monthly reported crime)
* Year
* Month.

The second and third sources of data are Wikipedia pages, from which the list of London boroughs and the list of neighbourhoods in the borough of Kingston upon Thames are scraped, respectively. The page with the list of London boroughs, <https://en.wikipedia.org/wiki/List_of_London_boroughs> , contains the following sections:

* Borough (names)
* Inner (to identify the borough as in inner or outer London)
* Status (Royal, City, or other)
* Local authority
* Political control (which party is in control)
* Headquarters
* Area (in square miles)
* Population (as recorded in 2013)
* Coordinates
* Nr. In map (number assigned to boroughs on a map).

The page with the list of neighbourhoods of Kingston upon Thames <https://en.wikipedia.org/wiki/List_of_districts_in_the_Royal_Borough_of_Kingston_upon_Thames> was consulted to create a dataset with latitude and longitude coordinates for each neighbourhood.

*Data cleaning*

The dataset imported from Kaggle included the years from 2008 to 2016. Since this report is being written in 2020, only the most recent data from 2016 were kept. All rows with value = 0 were also removed. A pivot table with the numbers of major categories of crime per borough was created (fig. 1).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Borough** | **Burglary** | **Criminal Damage** | **Drugs** | **Other Notifiable Offences** | **Robbery** | **Theft and Handling** | **Violence Against the Person** | **Total** |
| **0** | Barking and Dagenham | 11 | 9 | 7 | 5 | 8 | 68 | 40 | 148 |
| **1** | Barnet | 34 | 23 | 9 | 6 | 4 | 82 | 66 | 224 |
| **2** | Bexley | 9 | 24 | 13 | 5 | 0 | 58 | 53 | 162 |
| **3** | Brent | 30 | 22 | 12 | 7 | 6 | 98 | 86 | 261 |
| **4** | Bromley | 21 | 15 | 6 | 8 | 0 | 59 | 62 | 171 |

*Fig.1.* Pivot table with crimes per borough

The Wikipedia page with the data about London’s boroughs was scraped using Beautiful Soup library. The table built from the webpage was cleaned and merged with the dataset containing the crime data. The merged table allows us to see the crime rates in each borough and consequently the ones with the highest and lowest number of crimes recorded in 2016.

The merged table shows Kingston upon Thames as the safest borough, because of its lowest crime rates. The Wikipedia page with the list of neighbourhoods of Kingston upon Thames was consulted in order to create a new dataset with the neighbourhoods and corresponding coordinates (fig. 2). The coordinates were obtained using Google Maps API geocoding.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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. 2.* Pandas data frame with Neighbourhoods of Kingston upon Thames and coordinates

This new dataset will then be used to explore the 10 most common venues for each neighbourhood using Foursquare API; after that, k-means clustering will be used to cluster neighbourhoods with similar characteristics.

**Methodology**

*Exploratory data analysis*

After sorting the boroughs of London in ascending order from the one with the lowest rate of crime to the one with the highest, is clear again that Kingston upon Thames is the safest borough, or at least the one with the lowest number of reported crimes (fig. 3). To be precise, City of London appears to be the safest place, however it is not a London Borough, but only one of its 33 local authority districts. On the other hand, Westminster doesn’t look good, given its high number of crimes reported (fig. 4).

A screenshot of a cell phone

Description automatically generated

*Fig. 3.* The 5 Boroughs with the lowest crime rate. Kingston upon Thames appears to be the safest borough

A picture containing drawing

Description automatically generated

*Fig. 4.* The 5 Boroughs with the highest crime rate, with Westminster leading the infamous competition

A dataset with the 15 neighbourhoods of Kingston upon Thames and respective coordinates was created. A corresponding map was created using Folium (fig. 5).



*Fig. 5.* Kingston upon Thames’ neighbourhoods

*Modelling*

Using Foursquare API, all the venues within a 500 meters radius of each neighbourhood were found, and a json file was produced. After converting the json file to a pandas data frame, all the venues, their characteristics (name and category) and their coordinates can be observed (fig. 6).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Neighborhood Latitude** | **Neighborhood Longitude** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Venue Category** |
| **0** | Berrylands | 51.393781 | -0.284802 | Surbiton Racket & Fitness Club | 51.392676 | -0.290224 | Gym / Fitness Center |
| **1** | Berrylands | 51.393781 | -0.284802 | Alexandra Park | 51.394230 | -0.281206 | Park |
| **2** | Berrylands | 51.393781 | -0.284802 | K2 Bus Stop | 51.392302 | -0.281534 | Bus Stop |
| **3** | Berrylands | 51.393781 | -0.284802 | Kamala Food and Wine | 51.397810 | -0.284045 | Wine Shop |
| **4** | Canbury | 51.417499 | -0.305553 | Canbury Gardens | 51.417409 | -0.305300 | Park |

*Fig. 6.* Neighbourhoods and their venues

One hot encoding, a representation of categorical variables as binary vectors, was done on the venues data, which were then grouped by neighbourhood. By calculating the frequency of each venue in each neighbourhood, the 10 most common venues per neighbourhood were found.

**Results**

K-means clustering was used to cluster the neighbourhoods with similar venues. The 15 neighbourhoods were clustered into 5 clusters.

The following map was produced (fig. 7):



*Fig. 7.* Clustered neighbourhoods. Red = cluster 0; Purple = cluster 1; Blue = cluster 2; Green = cluster 3; Orange = cluster 4

The following tables (fig. 8-12) show the neighbourhoods of each cluster, with their coordinates and their 10 most common venues (the complete tables can be seen in the notebook).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Borough** | **Latitude** | **Longitude** | **Cluster Labels** | **1st Most Common Venue** | **2nd Most Common Venue** |
| **6** | Kingston Vale | Kingston upon Thames | 51.431850 | -0.258138 | 0 | Grocery Store | Sandwich Place |
| **7** | Malden Rushett | Kingston upon Thames | 51.341052 | -0.319076 | 0 | Grocery Store | Garden Center |
| **14** | Tolworth | Kingston upon Thames | 51.378876 | -0.282860 | 0 | Grocery Store | Restaurant |

*Fig. 8.* Cluster 1 (0)

The three neighbourhoods in cluster 1 seem to be ideal for whomever wants to live near a grocery store.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Borough** | **Latitude** | **Longitude** | **Cluster Labels** | **1st Most Common Venue** | **2nd Most Common Venue** |
| **3** | Coombe | Kingston upon Thames | 51.41945 | -0.265398 | 1 | Tea Room | Wine Shop |

*Fig. 9.* Cluster 2 (1)

The neighbourhood in cluster 2 could not be clustered with any other neighbourhood, since tea rooms and wine shops seem to be its commonest venues.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Borough** | **Latitude** | **Longitude** | **Cluster Labels** | **1st Most Common Venue** | **2nd Most Common Venue** |
| **1** | Canbury | Kingston upon Thames | 51.417499 | -0.305553 | 2 | Pub | Park |
| **4** | Hook | Kingston upon Thames | 51.367898 | -0.307145 | 2 | Fish & Chips Shop | Bakery |
| **5** | Kingston upon Thames | Kingston upon Thames | 51.409627 | -0.306262 | 2 | Café | Coffee Shop |
| **9** | New Malden | Kingston upon Thames | 51.405335 | -0.263407 | 2 | Gastropub | Sushi Restaurant |
| **10** | Norbiton | Kingston upon Thames | 51.409999 | -0.287396 | 2 | Pub | Food |
| **12** | Seething Wells | Kingston upon Thames | 51.392642 | -0.314366 | 2 | Indian Restaurant | Coffee Shop |
| **13** | Surbiton | Kingston upon Thames | 51.393756 | -0.303310 | 2 | Coffee Shop | Pub |

*Fig. 10.* Cluster 3 (2)

The neighbourhoods in cluster 3 are certainly the ones I would recommend to someone who really likes going out for a drink and some exotic food.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Borough** | **Latitude** | **Longitude** | **Cluster Labels** | **1st Most Common Venue** | **2nd Most Common Venue** |
| **11** | Old Malden | Kingston upon Thames | 51.382484 | -0.25909 | 3 | Train Station | Pub |

*Fig. 11.* Cluster 4 (3)

The neighbourhood in cluster 4, as the one in cluster 2, doesn’t seem to be similar to others, since its main venue is a train station. But this information would certainly be useful to someone who commutes to central London for work.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Borough** | **Latitude** | **Longitude** | **Cluster Labels** | **1st Most Common Venue** | **2nd Most Common Venue** |
| **0** | Berrylands | Kingston upon Thames | 51.393781 | -0.284802 | 4 | Wine Shop | Gym / Fitness Center |
| **8** | Motspur Park | Kingston upon Thames | 51.390985 | -0.248898 | 4 | Gym | Park |

*Fig. 12.* Cluster 5 (4)

The two neighbourhoods in cluster 5 are recommended for those who love going to the gym and take care of themselves, unlike the ones living in cluster 3 who are eating and drinking all the time ☺

**Discussion**

As we could see from the data presented in this report and in the related notebook, among London’s boroughs, the one that appears to be the safest (at least in 2016) is Kingston upon Thames. This borough has 15 neighbourhoods which could be grouped (except for two of them) in specific groups, according to the type of venues they offer.

**Conclusion**

Data science has a great potential, since among many other things, it can help a person making such an important decision as is buying a house, opening a business, or renting a new apartment. With code, we could learn which borough of London is the safest (and which one is definitely not safe); also, we could virtually explore the safest borough by dividing it in specific areas that have similar venues in common, so whoever has specific needs like living near a train station or a gym can just have a look at these tables and maps and make a decision.