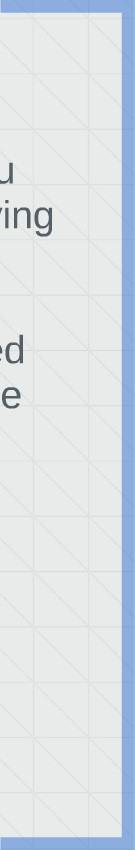
# Presentation by Sujith Krishna Apparasu

# 1. Introduction

* Background: Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.
* Problem: This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.
* Interest: Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.

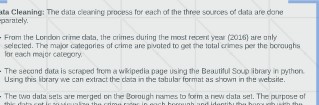
|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | Modelling  • Using | |

Boroughs with the lowest crime rates  th

# 2. Data Acquisition and Cleaning

Data Acquisition: The data acquired for this project is a combination of data from three sources:

* The first data source of the project uses a London crime data that shows the crime per borough in London.
* The second source of data is scraped from a wikipedia page that contains the list of London boroughs. This page contains additional information about the boroughs.
* The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on the wikipedia page.

 3. Methodology

Data Cleaning: The data cleaning process for each of the three sources of data are done separately.

* From the London crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per the boroughs for each major category.
* The second data is scraped from a wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website.
* The two data sets are merged on the Borough names to form a new data set. The purpose of this data set is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.
* After visualizing the crime in each borough we can find the borough with the lowest crime rate. The third data set is created, with the names of the neighborhoods and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.
* The new data set is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

3. Methodology

Exploratory Data Analysis

Statistical summary of crimes



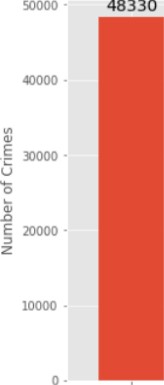
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| count | 33.000000 | 33.000000 | aamoooo | 33.000000 | 33.000000 | 33.000000 | 33.000000 | 33.000000 |
| mean | 2069.242424 | 1941.545455 | 1179.212121 | 479.060606 | 682.666667 | 8913.121212 | 7041.848485 | 22306.696970 |
| std | 737.448644 | 625.207070 | 586.406416 | 223.298698 | 441.425366 | 4620.565054 | 2513.601551 | 8828.228749 |
| min | 2.000000 | 2.000000 | 10.000000 | 6.000000 | 4.000000 | 129.000000 | 25.000000 | 178.000000 |
| 25% | 1531.000000 | 1650.000000 | 743.000000 | 378.000000 | 377.000000 | 5919.000000 | 5936.000000 | 16903.000000 |
| 50% | 2071.000000 | 1989.000000 | 1063.000000 | 490.000000 | 599.000000 | 8925.000000 | 7409.000000 | 22730.000000 |
|  | 2631 .oooooo | 2351.000000 | 1617.000000 | 551 .oooooo | 936.000000 | 10789.000000 | 8832.000000 | 27174.000000 |
| max | 3402.000000 | 3219.000000 | 2738.000000 | 1305.000000 | 1822.000000 | 27520.000000 | 10834.000000 | 48330.000000 |

The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'

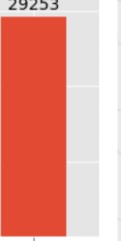
Boroughs with the highest crime rates

London Boroughs with the Highest no. of crime

Total



48330



29253

34071

31636

 30090

VkstmtnsterSuthwark Newham Tower Hamlets

Borough

Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.

Boroughs with the lowest crime rates

London Boroughs with the least no. of crime

Total 12755

12000 

10658 11074

9983

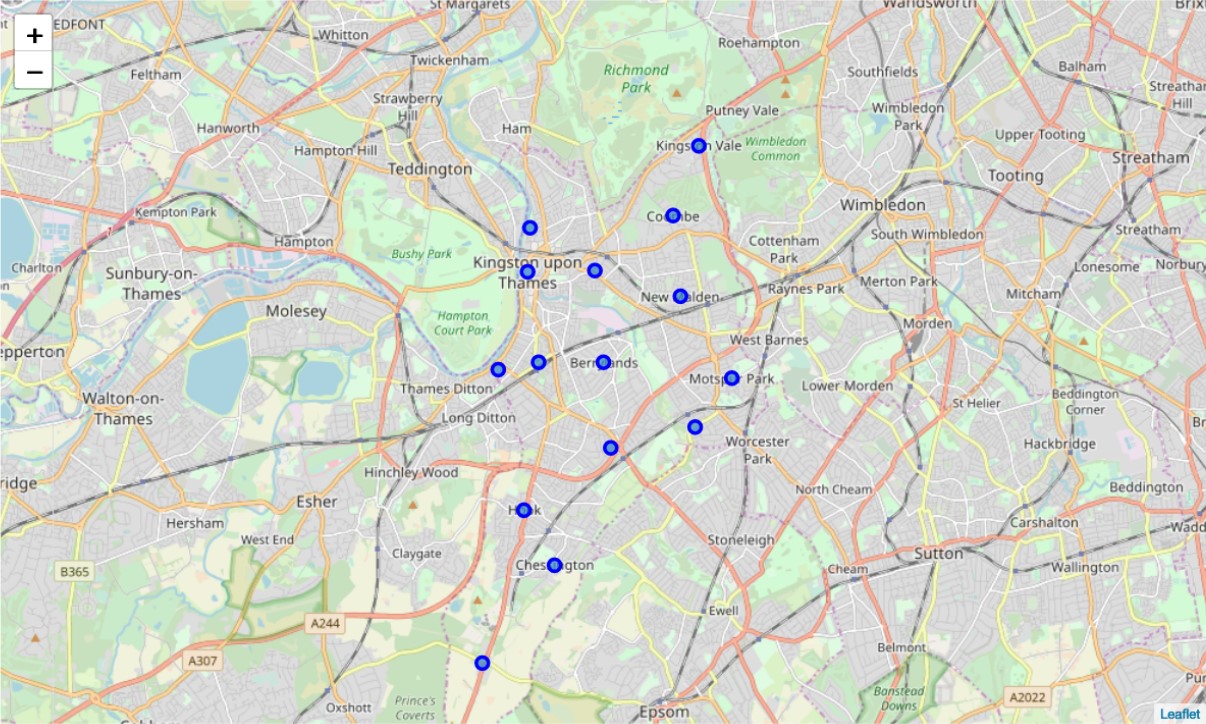
178 o- 

Clty 0t London KJnqston upon Thames 9Jtton Richmond upon Thames Merton Borough

Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.

* City of London has a significantly lower crime rate because it i is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area.
* We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

Neighborhoods in Kingston upon Thames



There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python.

Modelling

* Using the final data set containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API.

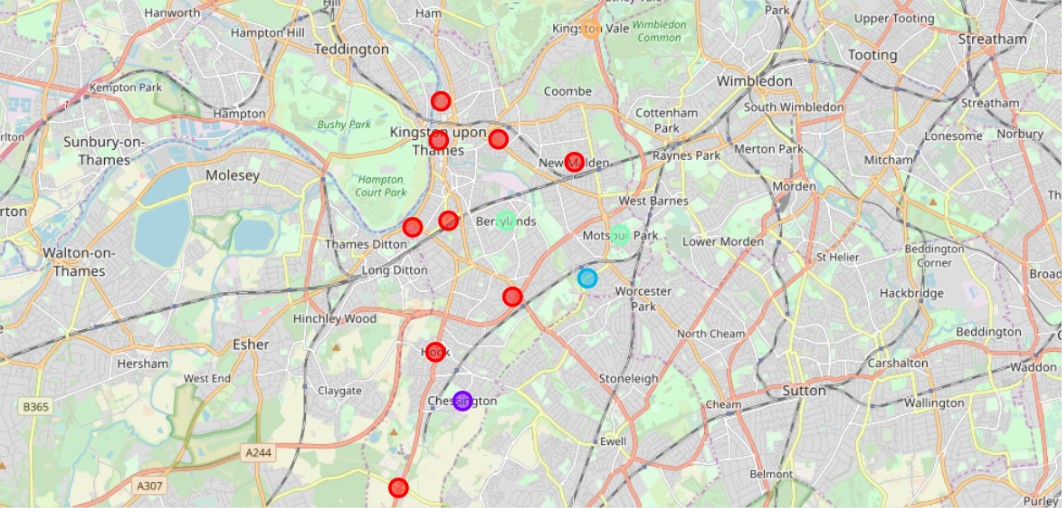
Neighborhood Neighborhood Latitude Neighborhood Longitude Venue Venue Latitude Venue Longitude Venue Category

|  |  |  |  |
| --- | --- | --- | --- |
| O Berrylands | 51.393781 -0.284802 Surbiton Racket & Fitness Club | 51.392676 | -0.290224 Gym / Fitness Center |
| Berrylands | 51.393781 -0.284802 Alexandra Park | 51.394230 | -0.281206 |
| 2 Berrylands | 51.393781 -0.284802 K2 Bus stop | 51.392302 | -0.281534 Bus Stop |
| 3 Berrylands | 51.393781 -0.284802 Cafe Rosa | 51.390175 | -0.282490 |
| 4 | 51.417499 -0.305553 The Boaters Inn | 51.418546 | -0.305915 Pub |

* One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.
* To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.
* We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

## 4. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.



Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster (Blue, Purple and Yellow), these are clusters two three and five. The green cluster consists of two neighborhoods which is the 4th cluster.

1: first cluster

1st Most 2nd Most 3rd Most 4th Most 5th Most 6th Most 7th Most 8th Most

Cluster

Neighborhood Borough Latitude Longitude Common Common Common Common Common Common Common Common Venue Venue Venue

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Hook upon | 51.367898 | -0.307145 | o |  | Convenience  Store | Indian Fish & Chips  Wine Shop Restaurant | Food | Electronics  Store | Farmers Market |
| 5 | Kingston Kingston upon |  |  |  |  |  | Sushi | Record | Cosmetics |  |
| 7 | upon  Kingston | 51.409627 | -0.306262 | O | Coffee Shop | Café | Burger Joint  Restaurant | Shop | Shop |  |

4 Kingston

MaldenConvenienceDiscount Dry Electronics

upon 51041052 -0.319076Pub

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9 |  | Kingston |  |  |  |  |  |  |  | Indian | Fish & | Dry |
|  | New Malden | upon | 51.405335 | -0.263407 | O | Gastropub | Gym | Supermarket | Restaurant | Restaurant | Chips Shop | Cleaner |
| 10 |  | Kingston |  |  |  | Indian |  | Italian |  | Grocery | Farmers | Dry |
|  | Norbiton | upon | 51.409999 | -0.287396 | o | Restaurant | Pub |  | Platform | Store | Market | Cleaner |
| 12 |  | Kingston |  |  |  | Indian |  | Italian |  | Wine | Fast Food | Chinese |
| 13 | Seething Wells | upon  Kingston | 51.392642 | -0.314366 |  | Restaurant | Coffee Shop | Pub | Café  Grocery | Shop | Restaurant | Restaurant  Train |
|  | Surbiton | upon | 51.393756 | -0.303310 | O | Coffee Shop | Pub | Supermarket  Spot | Store | Gastropub | Restaurant | Station |
| 14 | TOIw0rth | Kingston upon | 51078876 | -0.282860 |  | Grocery | Pharmacy | Furniture /  Train Station  Home Store | Pizza Place | Discount | Shop | Bus Stop |

RushettStore Cleaner Store

The cluster one is the biggest cluster with 9 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores

2: second cluster.

1st Most 3rd Most 4th Most 5th Most 6th Most 7th Most



8th

Most

9th

Most

Store

Shop

Chessington

upon

Gastropub

Common Common Common Common Common

Fast Food

Thames Restaurant Course Restaurant Center Joint Restaurant

The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wine shops.

Cluster 3: Looking into the neighborhoods in the third cluster.

1st Most 2nd 3rd Most

Common



4th

Most

5th

Most

6th

Most

7th

Most

8th

Most

9th

Most

The third cluster has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

4: fourth cluster.

 Kingston Gym / Fish &

Wine Electronics Farmers Fast Food

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Berrylands | upon 51.393781  Kingston | -0.284802 | 3 | Fitness Center | Park Café | Bus Stop  Soccer | Shop | Chips  Store Shop  Wine Fast Food | Market Restaurant  Dry Electronics |
| Motspur Park | upon 51090985 | -0.248898 | 3 | Park | Gym Restaurant | field | Bus Stop | Shop Restaurant | Cleaner |

The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

Cluster 5: Looking into the neighborhoods in the fourth cluster.



The fifth cluster has one neighborhood which consists of Venues such as

Grocery shops, Bars, Restaurants, Furniture shops, and Department stores.

# 5. Discussion

* The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighborhoods to which they want to relocate based on the most common venues in it.
* For example if a person is looking for a neighborhood with good connectivity and public transportation we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues.
* If a person is looking for a neighborhood with stores and restaurants in a close proximity then the neighborhoods in the first cluster is suitable.
* For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.
* The preference of venues may vary from person to person, they can select a neighborhood based on ones priorities.

# 6.Conclusion

* This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood.
* We have just taken safety as a primary concern to shortlist the safest borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.