

Capstone Project

The Battle of Neighborhoods

By :
Ghazali Akmal Rabbani



1. INTRODUCTION



Background

- Safety is a top concern when moving to a new area. If you don't feel safe in your 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-means clustering

Interest

- Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore in neighborhoods and common venues around each neighborhood

2. DATA ACQUISITION AND CLEANING





Data Acquisition

The data acquired for this project is a combination of data from 3 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 Wikipedia page that contains the list of London boroughs. This page contains additional information about the borough
- The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on the Wikipedia page.

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 (2019) 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 BeautifulSoup library in python. Using this library we can extract the data in the tabular format as shown in the website
- The two datasets are merged on the borough names to form a new dataset. The purpose of this dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2019



Data Cleaning

(continue)

- After visualizing the crime in each borough we can find the borough with the lowest crime rate. The third dataset 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 dataset 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





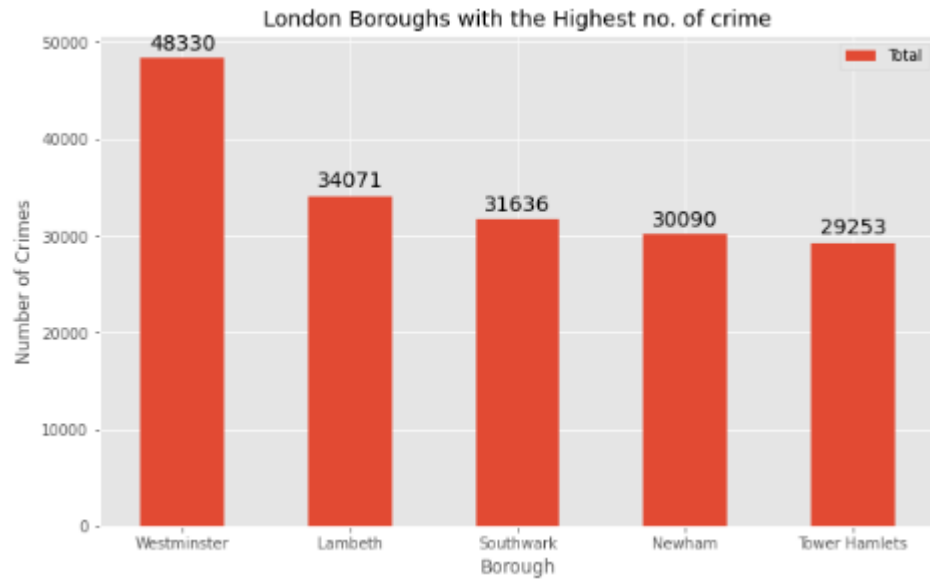
3. METHODOLOGY

a. Exploratory Data Analysis

Statistical Summary of Crimes

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2089.242424	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
75%	2631.000000	2351.000000	1617.000000	551.000000	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

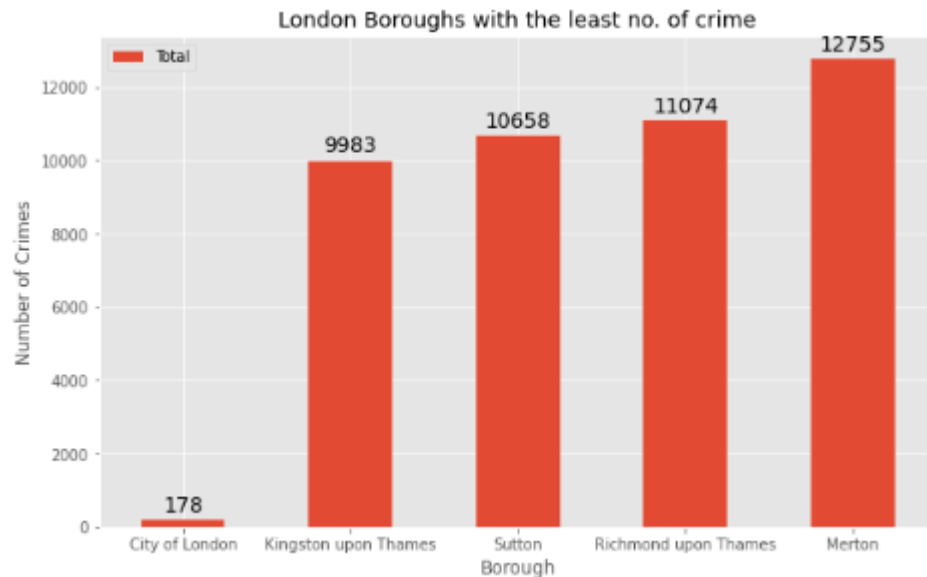
The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2019 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.



Borough with the highest crime rates

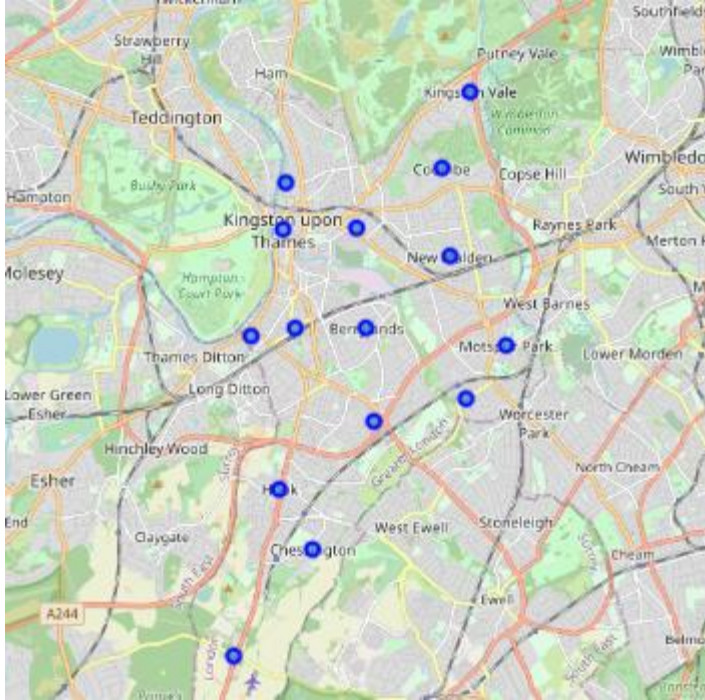
Comparing five boroughs with the highest crime rate during the year 2019 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham, and Tower Hamlets.

Borough with the lowest crime rates



Comparing five boroughs with the lowest crime rate during the year 2019, 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 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 2019 which suggests that it is a small area. Hence we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.



Neighborhood 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

b. Modelling

	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	Canbury	51.417250	-0.305631	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417250	-0.305631	The Boater's Inn	51.418546	-0.305915	Pub

Using the final dataset 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. This returns a json file containing all the venues in each neighborhood which is converted to a pandas dataframe. This data frame contains all the venues along with their coordinates and category.

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). 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. RESULT



Cluster 1

The cluster one is the biggest cluster with 7 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.

	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	51.417250	-0.305631	0	Pub	Fish & Chips Shop	Japanese Restaurant	Café	Supermarket	Spa	Plaza	Shop & Service	Gym / Fitness Center	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Fish & Chips Shop	Pub	Bakery	Supermarket	Indian Restaurant	Fast Food Restaurant	Deli / Bodega	Department Store	Discount Store	Electronics Store
5	Kingston upon Thames	Kingston upon Thames	51.409827	-0.306262	0	Coffee Shop	Café	Pub	Thai Restaurant	Sushi Restaurant	Burger Joint	German Restaurant	Electronics Store	Gym / Fitness Center	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Pub	Italian Restaurant	Indian Restaurant	Food	Platform	Wine Shop	Hardware Store	Pharmacy	Japanese Restaurant	Hotel
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Coffee Shop	Pub	Food	Construction & Landscaping	Train Station	Bakery	Bar	Gastropub	Garden Center	Furniture / Home Store
12	Seething Wells	Kingston upon Thames	51.392642	-0.314386	0	Indian Restaurant	Coffee Shop	Pub	Hotel	Italian Restaurant	Café	Restaurant	Park	Harbor / Marina	Gym / Fitness Center
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Thai Restaurant	Tea Room	Gastropub	Train Station	Farmers Market

Cluster 2

The second cluster has 2 neighborhood which consists of Venues such as Gym, Store and Public places like Park, Bus Stop, and Soccer Field

	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	1	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Farmers Market	Deli / Bodega	Department Store	Discount Store	Electronics Store	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Gym	Park	Bus Stop	Soccer Field	Wine Shop	Farmers Market	Deli / Bodega	Department Store	Discount Store	Electronics Store

Cluster 3

The third cluster has 2 neighbourhood which consists of Venues such as Grocery Store, Bar, Sandwich Place, and 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
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Grocery Store	Sandwich Place	Bar	Soccer Field	Cosmetics Shop	Deli / Bodega	Department Store	Discount Store	Electronics Store	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Restaurant	Train Station	Discount Store	Hotel	Indian Restaurant	Italian Restaurant	Coffee Shop	Pizza Place	Bus Stop

Cluster 4

The forth cluster has 1 neighborhood which consists of Venues such as Store, Bar and Sandwich 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	10th Most Common Venue	
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	3	Grocery Store	Garden Center	Pub	Restaurant	Electronics Store	Cosmetics Shop	Deli / Bodega	Department Store	Discount Store	Farmers Market

Cluster 5

The fifth cluster has 1 neighbourhood which consists of Venues such as Train Station, Coffee Shop, and Pub.

	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
9	New Malden	Kingston upon Thames	51.405335	-0.283407	4	Gastropub	Gym	Supermarket	Bar	Korean Restaurant	Indian Restaurant	Farmers Market	Deli / Bodega	Department Store	Discount Store



5. DISCUSSION

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 public spaces and transportations we can see that **Clusters 2** have Gym, Park and Bus Station 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 **Cluster 1, 3 and 4** is suitable.
- For a family I feel that the neighborhoods in **Cluster 5** are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Gastropub, Gym, Supermarket and Store which is ideal for a family who want to have great life experience from daily life needed to entertainment

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

THANK YOU!

