

The Battle of Neighborhoods

Presentated by - Shashi Dubey





INTRODUCTION

Background: Safety is a top concern when moving to a new area. If you do not feel safe in the area you are planning to move – you will not enjoy living in that area.

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.



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 is a London crime data that shows crime per borough in London.
- The second source of the data is scraped from a Wikipedia page that contains the list of London boroughs. The page contains additional information about the boroughs.
- The third source is the list of the neighborhoods from Foursquare



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

- From the London crime data, the crimes from the most recent year 2016 are only selected. The major categories of crime are pivoted to get the total crimes per borough for each major category.
- The second data is scrapped from a Wikipedia page using the beautifulsoup library in python
- The two data sets are merged on the borough names to form a new data set. The purpose of the data is to visualize the crime rates in each borough and identify the borough with least crime rates during 2016.
- After visualizing the crime in each borough we can find the borough with least crime
- The new data set is used to generate the 10 most common venues of each neighborhood using the Foursquare API and finally the K-means clustering algorithm is used to cluster similar neighborhoods together.

Methodology



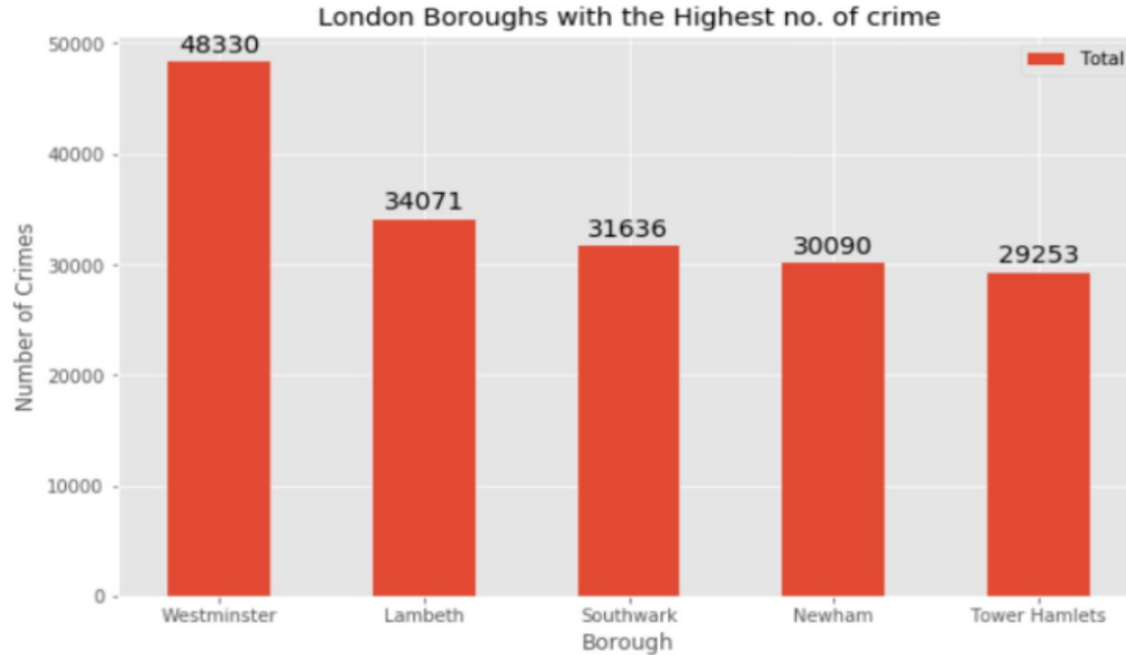
Exploratory Data Analysis

Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

The count of 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.

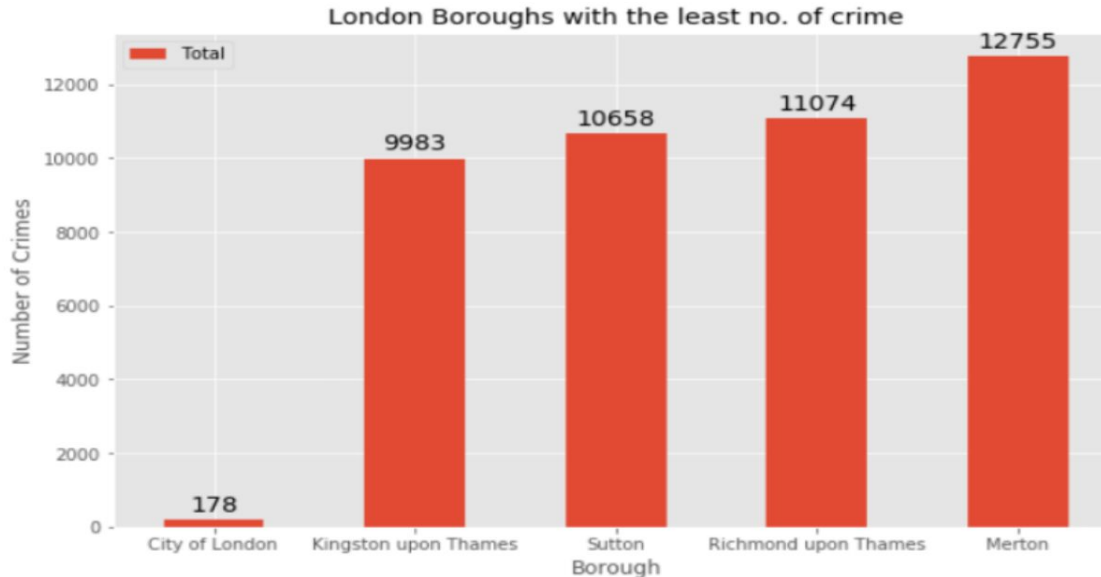
The lowest recorded crimes are drugs, robbery and other notifiable offenses.

Borough with highest crime rates



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

Borough with lowest crime rates



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.

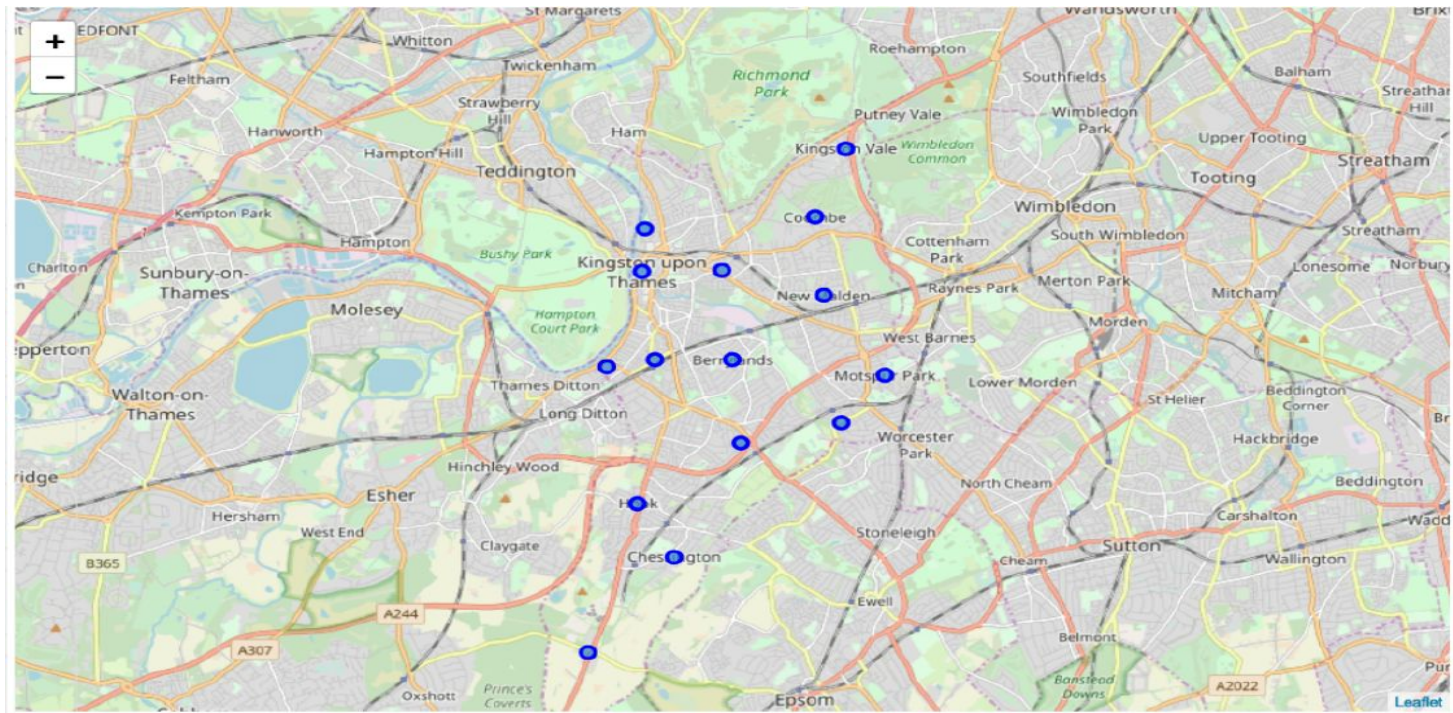
Modelling

- Use 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 meters radius of each neighborhood by connecting Foursquare API

	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	Cafe Rosa	51.390175	-0.282490	Café
4	Canbury	51.417499	-0.305553	The Boater's 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.
- We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reason to conduct K-mean clustering is to find the venue so that people can shortlist the area of their interests based on the venue/amenities around each neighborhood.

Result



After running the k-mean clustering, we can access cluster created to see the neighborhood on the map using folium library.

Cluster 1: Looking the neighborhoods in 1st cluster

	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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

Cluster 2: Looking the neighborhoods in 2nd cluster

	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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant

Cluster 3: Looking the neighborhoods in 3rd cluster

	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
11	Old Maiden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Deli / Bodega

Cluster 4: Looking the neighborhoods in 4th cluster



	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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Park	Gym	Restaurant	Soccer Field	Bus Stop	Wine Shop	Fast Food Restaurant	Dry Cleaner	Electronics Store

Cluster 5: Looking the neighborhoods in 5th cluster

	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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Italian Restaurant	Soccer Field	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Department Store



Discussion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can choose 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 due 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.



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 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 based on safety and a predefined budget