Capstone Project – The Battle of Neighbourhood

Final Presentation
Sanjay Kumar Tiwary

Analysis Objective

- **Background** – One of the important aspect while selecting a locality to shift is Safety. If we don't feel safe in society and home where we live, we cannot sustain there for long.

- **Problem Statement** – This project aims to select the safest borough in London based on the total crimes, explore the neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering

- **Target Audience** – Expats who are considering relocating to London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around each neighbourhood.

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.
- The third data source is the list of Neighbourhoods in the Royal Borough of Kingston upon Thames as found on a Wikipedia page.

Data pre-processing

The data preparation for each of the three sources of data is 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 borough as per the 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. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form.
- The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset
- The third source of data is acquired from the list of neighbourhoods in the safest borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude are left blank.
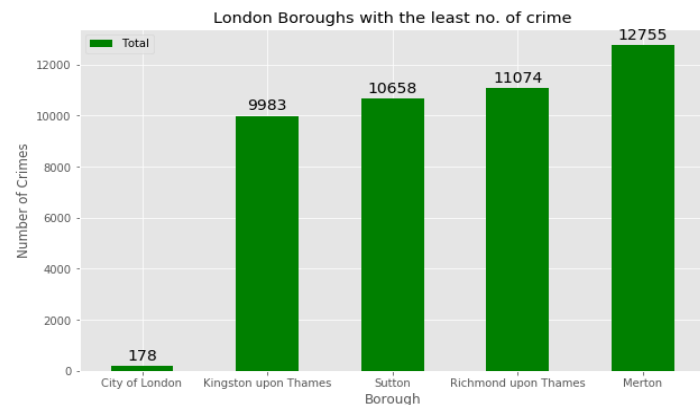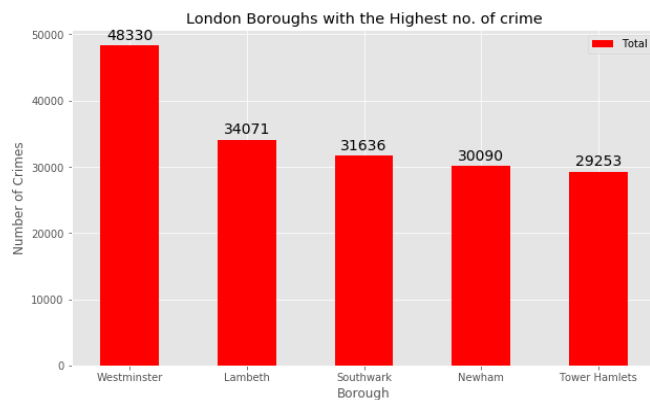
Methodology

	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	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
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

Crime Summary

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'.
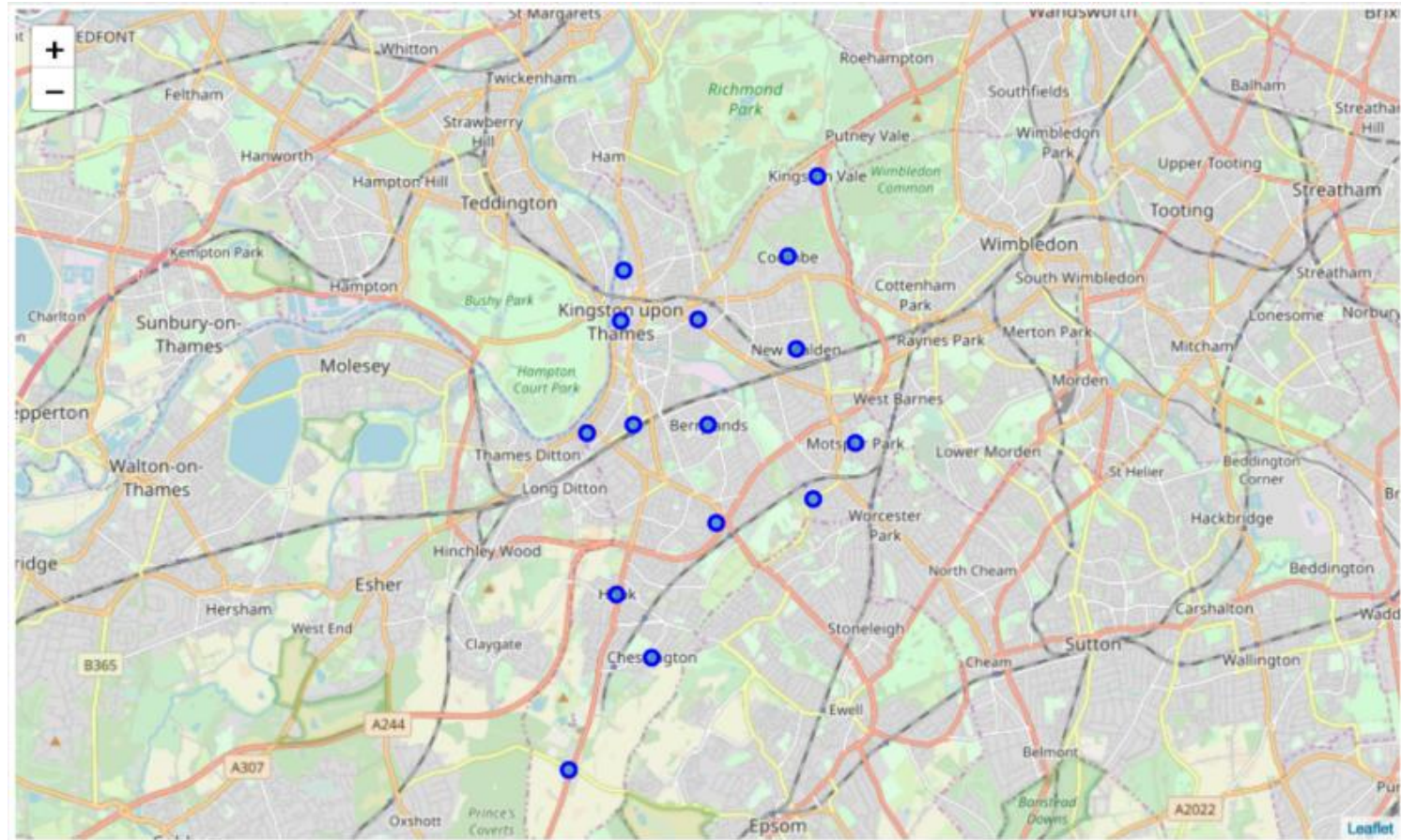
Comparing Crime Records

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 Merto

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.

Hence we considered the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

Neighbourhoods in Kingston upon Thames

There are 15 neighbourhoods in the royal borough of Kingston upon Thames

Modelling

Using the final dataset containing the neighbourhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighbourhood by connecting to the Foursquare API.
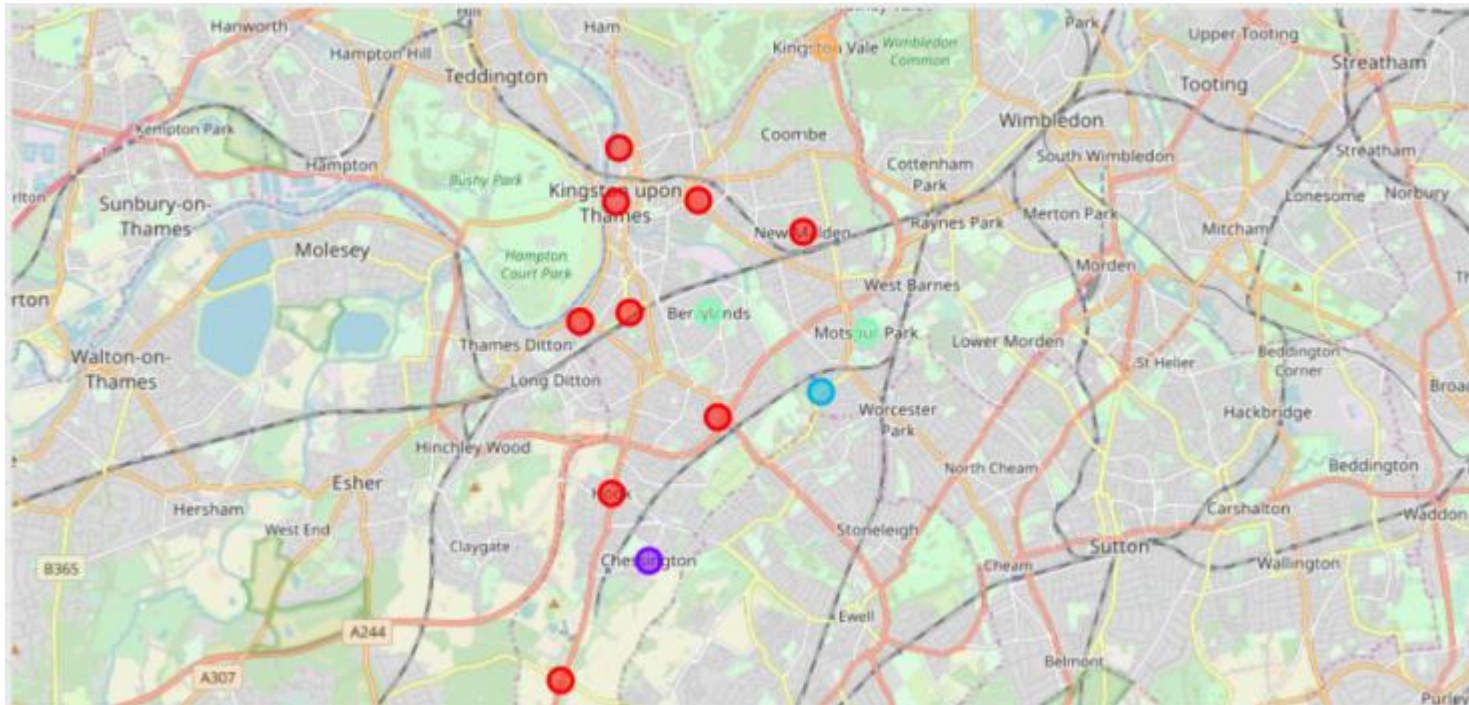
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	Alexandra Park	51.394230	-0.281206	Park
2	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
3	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
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 Neighbourhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighbourhoods.
- To help people find similar neighbourhoods in the safest borough we will be clustering similar neighbourhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.
- We use a cluster size of 5 for this project that will cluster the 15 neighbourhoods into 5 clusters.

Results

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five clusters.

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

Cluster-1

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
Berrylands	Kingston upon Thames	51.393781	-0.284802	0	Bus Stop	Gym / Fitness Center	Park	Wine Shop	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

This cluster has one neighbourhood which consists of Venues such as Bus stop, Park, Dry Cleaner, Fast Food restaurants

Cluster-2

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
Kingston Vale	Kingston upon Thames	51.43185	-0.258138	1	Grocery Store	Italian Restaurant	Bar	Soccer Field	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

This cluster has one neighbourhood which consists of Venues such as Grocery Store, Italian Restaurant, Bar, Soccer Field, Discount Store, Bus stop, Park, Dry Cleaner, Fast Food restaurants

Cluster-3

	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	2	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Gym / Fitness Center	Plaza
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Food	Discount Store	Dry Cleaner	Electronics Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	2	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Asian Restaurant	Cosmetics Shop	Electronics Store
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Convenience Store	Garden Center	Pub	Restaurant	Farmers Market	Deli / Bodega	Department Store	Discount Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Gastropub	Indian Restaurant	Bar	Sushi Restaurant	Supermarket	Gym	Korean Restaurant	Chinese Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	2	Indian Restaurant	Food	Italian Restaurant	Pub	Wine Shop	Platform	Athletics & Sports	Breakfast Spot
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	2	Indian Restaurant	Café	Coffee Shop	Pub	Fast Food Restaurant	Italian Restaurant	Chinese Restaurant	Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	2	Coffee Shop	Pub	Grocery Store	Gastropub	Train Station	Farmers Market	Gym / Fitness Center	Hotel
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Pharmacy	Sandwich Place	Furniture / Home Store	Train Station	Hotel	Discount Store	Coffee Shop

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

Cluster-4

	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	C...
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Construction & Landscaping	Park	Soccer Field	Wine Shop	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store	F...

This cluster has one neighbourhood which consists of Venues such as Soccer Field, Wine Shop, Fast Food restaurants, Electronics Store

Cluster-5

	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	C
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	4	Grocery Store	Construction & Landscaping	Food	Train Station	Bar	Discount Store	Gift Shop	German Restaurant	Gastropub	

The Fifth cluster has one neighbourhood which consists of Venues such as Train stations, Restaurants, and Furniture shops

Conclusion

- This Analysis helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. 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 neighbourhood.
- We have just taken safety as a primary concern to shortlist the borough of London. There can be a list of other parameters e.g. Cost, Pollution index and others.

Cluster-4