Capstone Project

THE BATTLE OF NEIGHBORHOODS

KIRTAN MANKAD

Introduction

- **Background:** Safety is a top concern when moving to a new area. If you do not feel safe in your own home, you are not going to be able to enjoy living there. Hence, prediction of neighborhoods before moving anywhere is a top concern area.
- **Problem:** This project aims to predict safest borough based on crime rates, explore the neighborhoods for venue preferences and do clustering using K-Means Clustering.
- **Interest:** Who are targeting to relocate to London will be interested to find and explore neighborhoods according to there venue preferences.

Data Acquisition and Cleaning

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

- The first data sources of the project is the London Crime Data that shows the crime per borough in London.
- The second source of data is derived from Wikipedia page that contains list of London's boroughs. This page also contains additional information abought these boroughs.
- The third data source is the list of neighborhood in the Royal Borough of Kingston upon Thames as found on the Wikipedia page.

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

- From London crime data, crimes during the most recent year are considered. The major categories of crimes are pivoted with a view to get the total crimes per boroughs for each major category.

- The second data set is derived from Wikipedia page using Beautiful Soup library in python.

- Two datasets are merged on the borough names to form a new data set. The purpose of this dataset s to visualize the crime rates in each borough and identify the borough where least crime recorded during the year.

- After visualizing the crime in each borough we can derive the borough with lowest crime rate. The third data set is created with the names of the neighborhoods and the names of the borough with the latitude longitude obtained using Google Maps API.

- The new dataset is used to generate the10 most common venues for each neighborhood using the Foresquare API. Finally used J Means Clustering algorithm to cluster similar neighborhiids together.

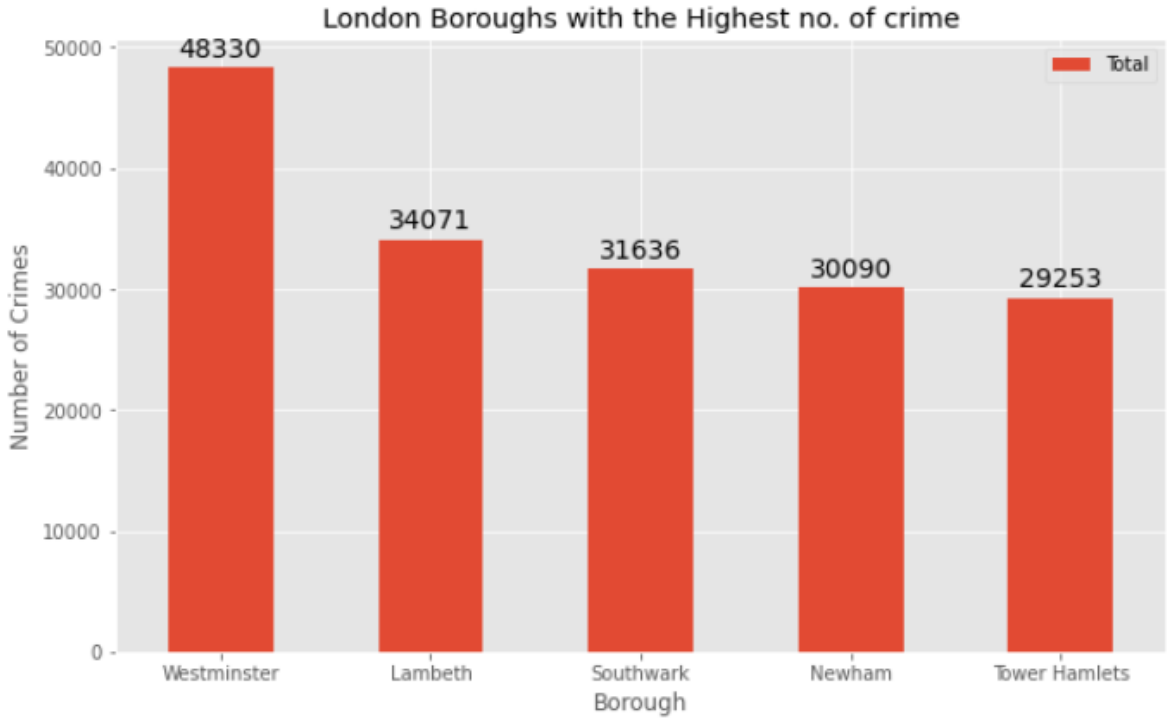
Methodology

Exploratory Data Analysis:

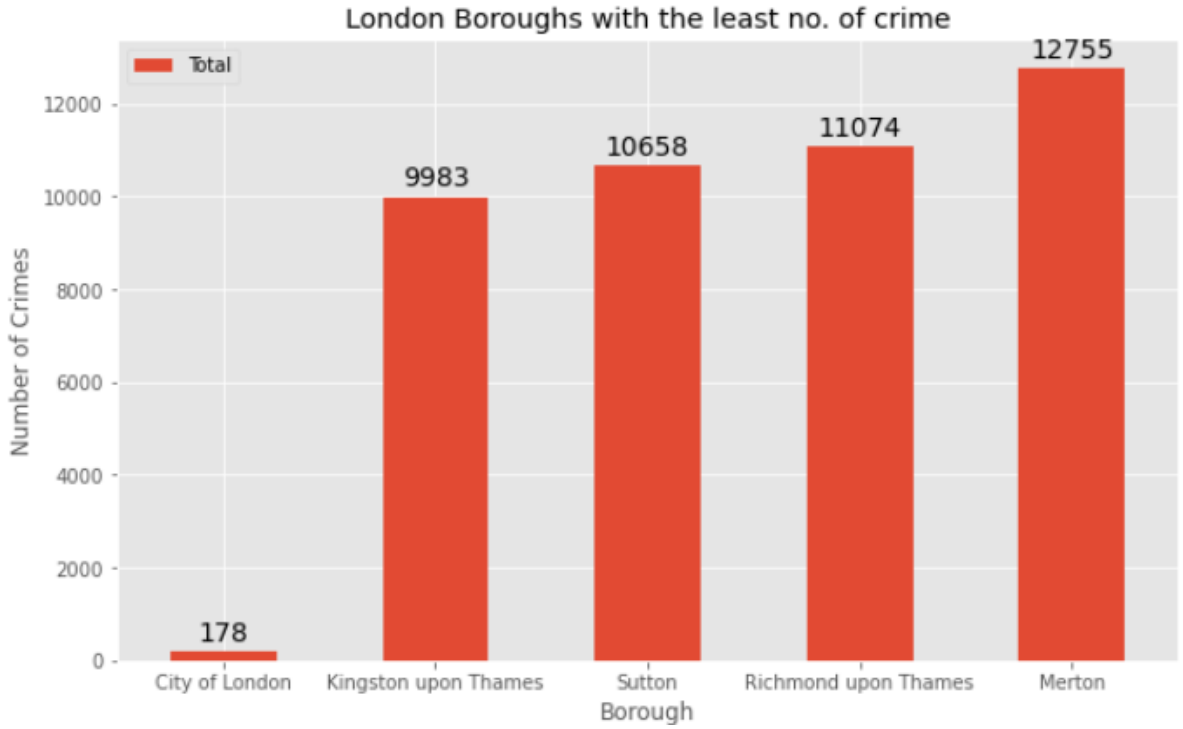
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
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

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.

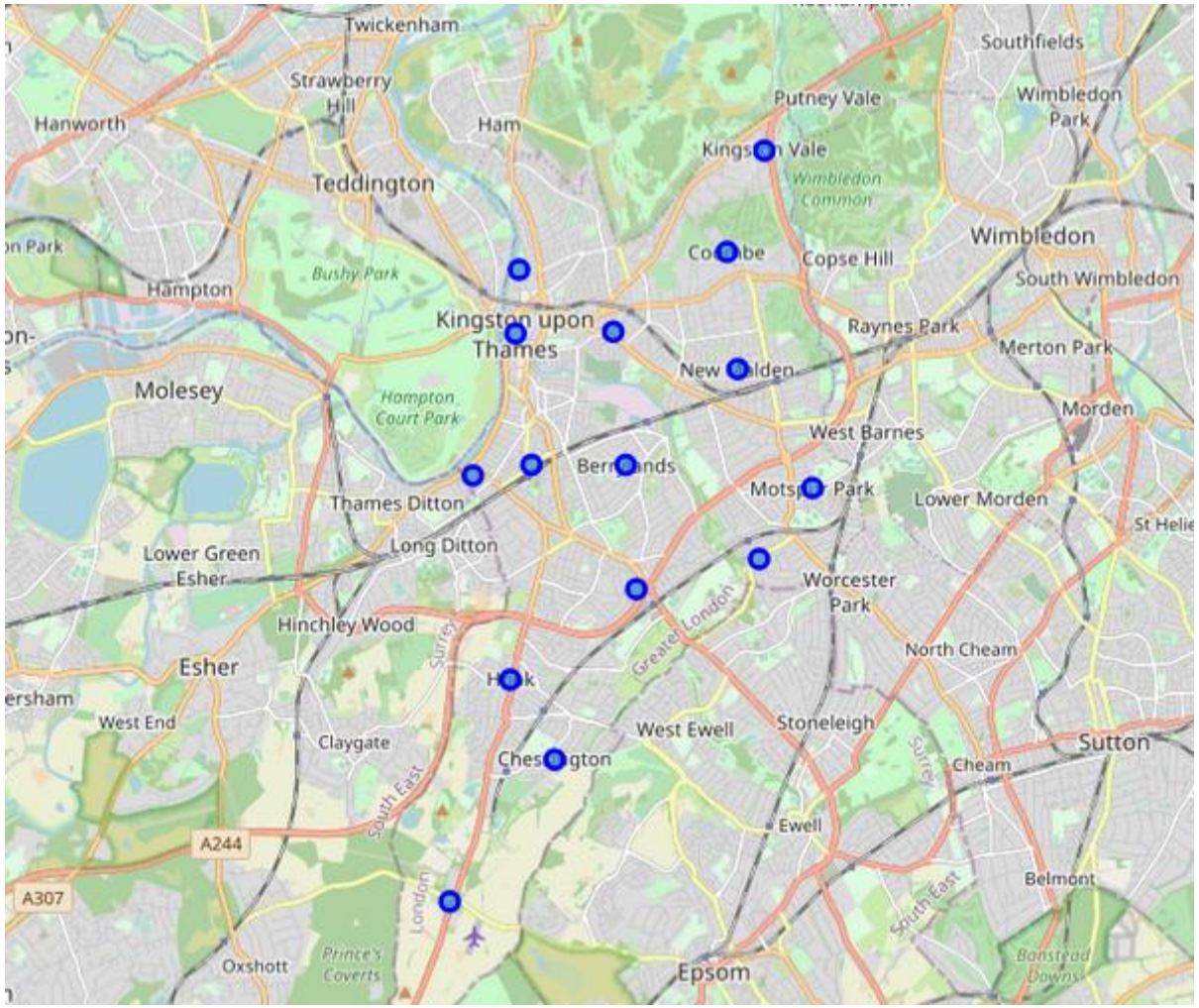
Comparing five boroughs, it is evident that Westminister has the highest crime recorded followed by Labmeth, Southwark ad Newham.



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 (see fig 3.1.3.1). Hence, we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.



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



Modelling

- 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 Foursquare API.

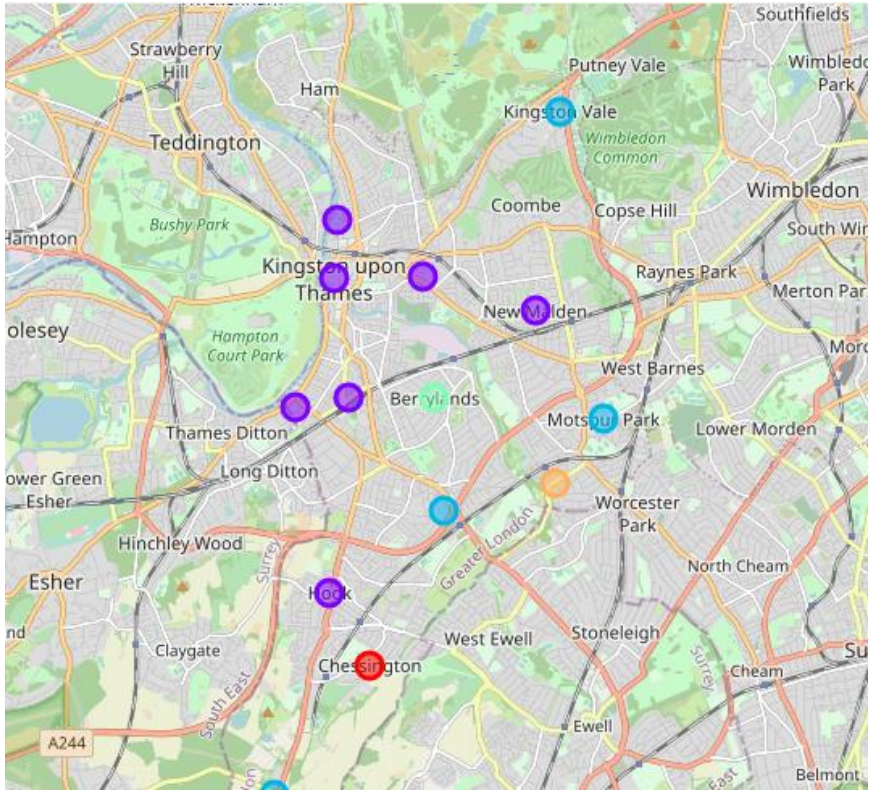
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
2	Berrylands	51.393781	-0.284802	Cafe Rosa	51.390175	-0.282490	Café
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

- To hel[illegible] eans clustering which is the form of unsupervised machince learning.

- I have used the cluster size of 5. Reason of using K-means clustering is to cluster neighborhoods with similar venues.

Result

After performing K-means clustering, we can access each cluster created to see which neighorhoods were assigned to each of the five clusters. I have used folium to visualize clustered neighborhood on map.



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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	0	Park	Wine Shop	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Fast Food Restaurant	Construction & Landscaping

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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Café	Shop & Service	Plaza	Indian Restaurant	Hotel	Park	Gym / Fitness Center	Spa	Supermarket
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Indian Restaurant	Pub	Bakery	Supermarket	Fish & Chips Shop	Wine Shop	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Pub	Sushi Restaurant	Burger Joint	Café	Gift Shop	Department Store	Furniture / Home Store	French Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gastropub	Korean Restaurant	Sushi Restaurant	Supermarket	Bar	Indian Restaurant	Wine Shop	Electronics Store	Deli / Bodega	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Food	Italian Restaurant	Pub	Breakfast Spot	Fried Chicken Joint	Dry Cleaner	Grocery Store	Convenience Store	Coffee Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Park	Café	Pet Café	Fish & Chips Shop	Fast Food Restaurant	Golf Course	Gym
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Bistro	Gym / Fitness Center	Platform	Deli / Bodega	Breakfast Spot

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	9th Most Common Venue	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Garden Center	Grocery Store	Pub	Restaurant	Electronics Store	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Park	Gym	Bus Stop	Soccer Field	Wine Shop	Electronics Store	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Pharmacy	Restaurant	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Coffee Shop	Pizza Place	Café

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	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Wine Shop	Gym / Fitness Center	Café	Bus Stop	Farmers Market	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Fast Food Restaurant

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	10th Most Common Venue
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	4	Train Station	Pub	Food	Construction & Landscaping	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner	Electronics 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 4 and 5 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 proximity, then the neighborhoods in the second cluster is suitable.
- For a family I feel that the neighborhoods in Cluster 3 is 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 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.