

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# **Capstone Project - Battle of Neighbourhoods - Presentation G HARISH KUMAR**

# 1. Introduction

- ▶ **Background:** Safety is always the top concern when we move to a new place. If you feel your home is not safe, then you are not going to feel safe anywhere.
- ▶ **Problem:** This project aims to find the safest borough in London based on total crimes, explore the neighbourhoods in that borough to find 10 most common venues in neighbourhoods and finally cluster it using k-means clustering.
- ▶ **Interest:** This is mainly for those who are looking to re-locate or shift to London and see the safest places in the neighbourhoods.

## 2. Data Acquisition and Cleaning

- ▶ **Data Acquisition:** The data acquired for this project is from three different sources:
- ▶ The first data uses the crime data of London that shows the crime per borough in London.
- ▶ The second source of data is scrapped from a Wikipedia page that contains list of all the borough in London.
- ▶ The third data source is the list of Neighbourhoods in Royal Borough of Kingston upon Thames as found on Wikipedia page.

### 3.Methodology

- ▶ **Data cleaning:** The data cleaning process for each of the different sources is done differently:
- ▶ For the London Crime data, only the crime in the most recent year is taken(2016). The major categories of crime are selected and pivoted to get the to get the total crimes per borough through each of the major category.
- ▶ The second data is scrapped from a Wikipedia web page using the Beautiful Soup library in python and is put in a tabular format.
- ▶ The two data sets are merged on the borough names to form the new dataset. The purpose of this data set is to visualize the crime rates in each Borough and identify the borough with least crimes recorded during the year 2016.
- ▶ After visualizing the crime rates in each Borough we identify the borough with least crime rate. The data set is created with the names of the neighbourhoods and the names of the borough with latitude and longitude Obtained using Google maps API geocoding.
- ▶ The new dataset is used to generate the 10 most common venues for each neighbourhood using the Foursquare API, finally using k-means clustering algorithm to cluster similar neighbourhoods together.

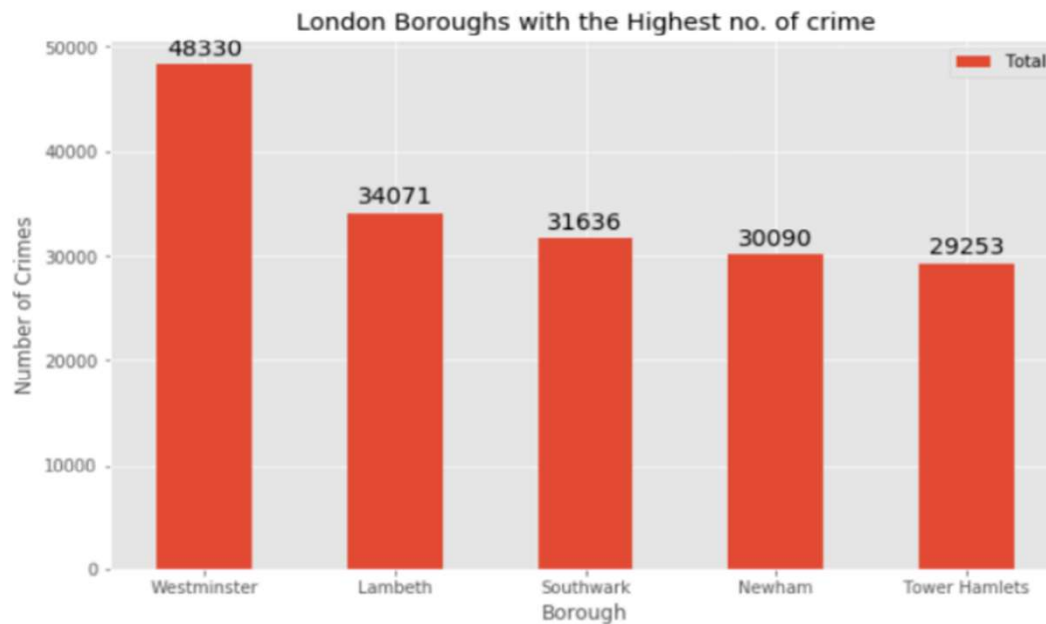
### 3.Methodology

- 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	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

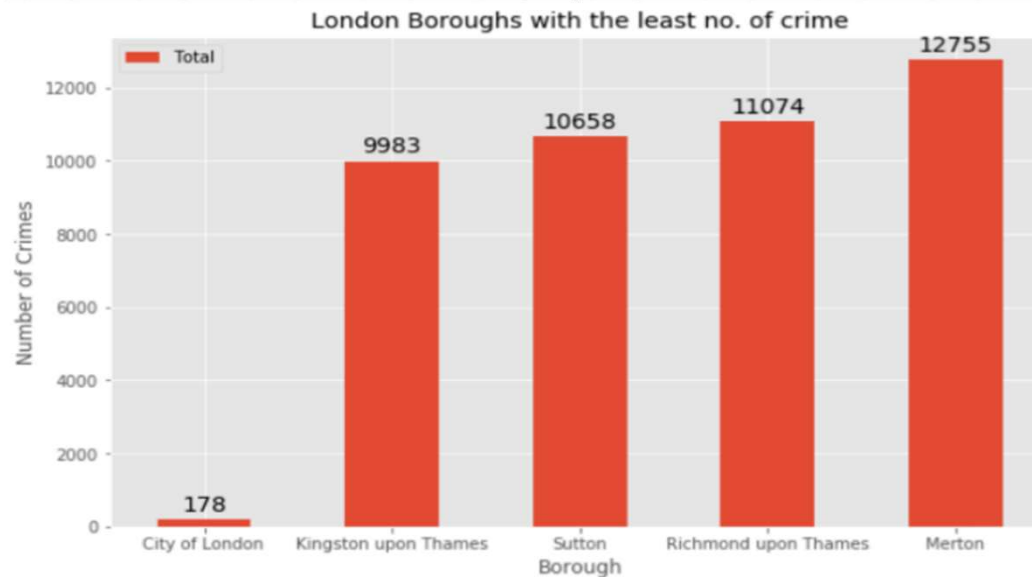
- The count for each the major categories of the crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime in the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and others.

## Borough with highest crimes



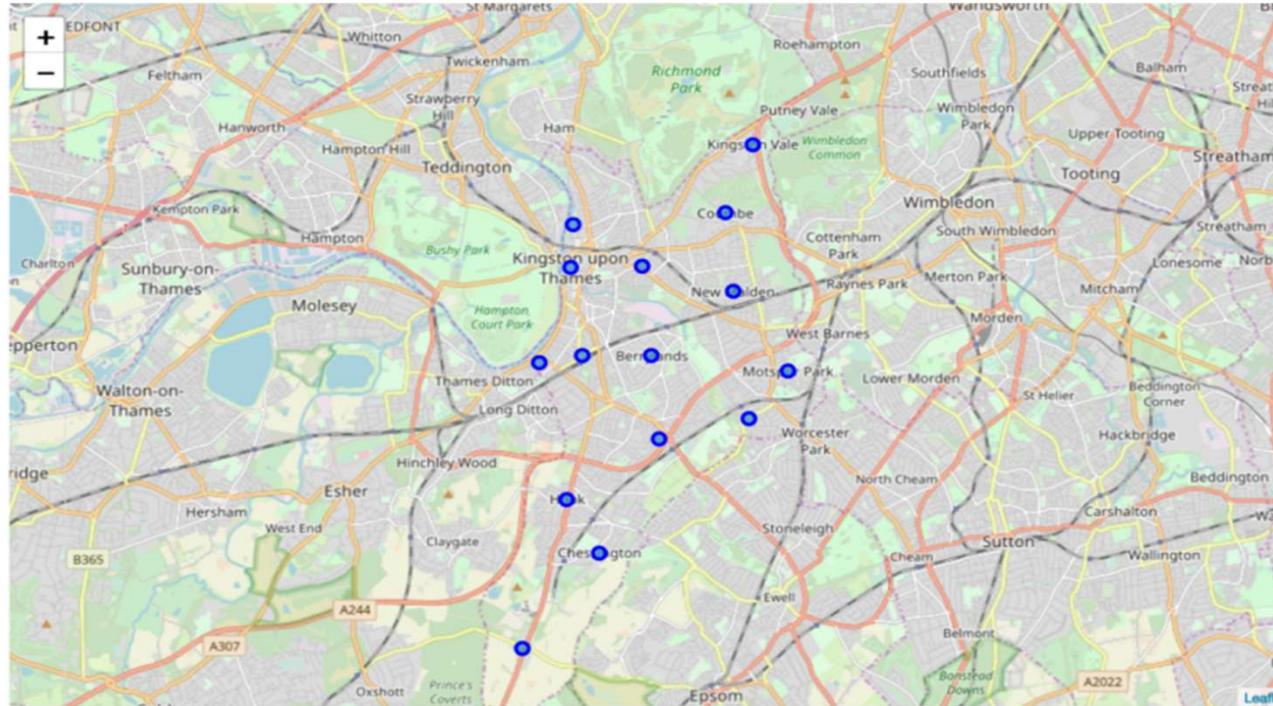
- From the above figure comparing the 5 boroughs it's evident that during the year 2016 Westminster has the highest crime recorded followed by Lambeth, Southwark, Newham, Tower Hamlets. Westminster has the highest crime compared to other 4 places.

## Boroughs with least crime rate



- ▶ Among the 5 boroughs City of London has the least crime rates followed by Kingston upon Thames, Sutton Borough, Richmond upon Thames, Merton.
- ▶ City of London has the least crime rate because it is the 33<sup>rd</sup> division of Greater London but it is not 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.

## Neighbourhoods in Kingston upon Thames



- There are 15 neighbourhoods of Kingston upon Thames, they are visualized using folium using python.



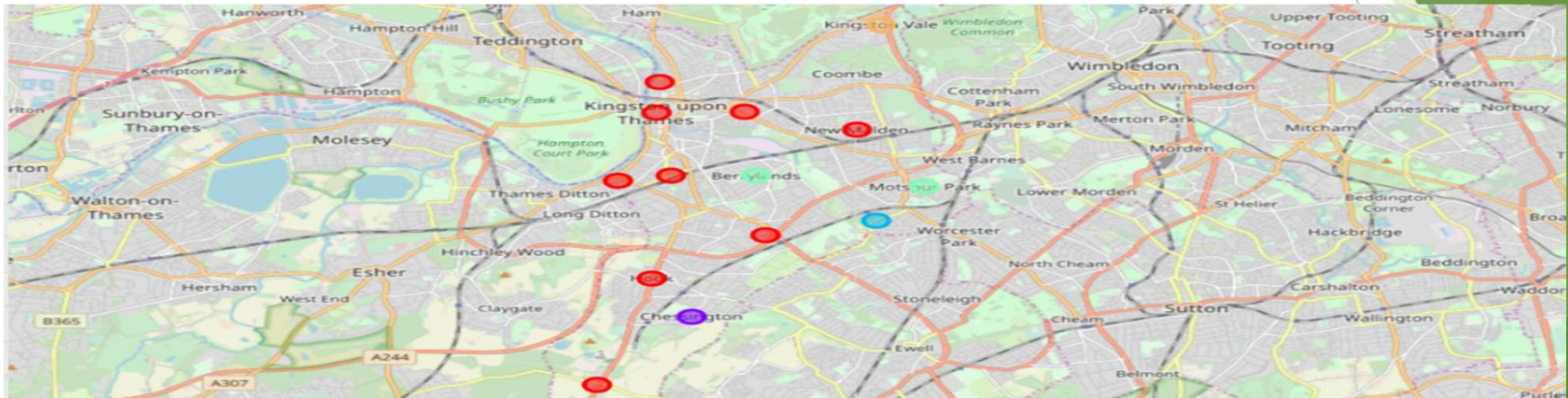
- Modelling
- Using the final dataset containing the neighbourhoods in Kingston upon with latitude and longitude, we can find all the venues within 500 kms radius by connecting to foursquare API.

	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	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 the 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 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 the predefined cluster size.

## 4. Results

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



- Each phase is colour coded for the ease of representation, we can see the majority of the neighbourhood falls in the red cluster which is the first cluster. Three neighbourhoods have their own cluster, these are clusters two, three and five. The green cluster consists of two neighbourhoods which is the fourth cluster.

## Cluster 1: Looking into the neighbourhoods of first clusters

	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

- The cluster one is the biggest cluster with the 9 of the 15 neighbourhoods in the borough Kingston upon Thames. Upon closely examining the neighbourhoods we see that the most common venues in these neighbourhoods are restaurants, pubs, café and super markets.

## Cluster 2: Looking into the neighbourhoods for second 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

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

## Cluster 3: Looking into the neighbourhoods for third 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 Malden	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

The third cluster has one neighbourhood which consists of venues such as Train Stations, Restaurants and Furniture shops.

## Cluster 4 : Looking into the neighbourhoods of the fourth 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

The fourth cluster has two venues in it, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bust stops, Restaurants, Electronic Stores and Soccer fields etc.

## Cluster 5 : Looking into the neighbourhoods of the fifth 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



## 5. Discussion

- ▶ The aim of this project is for people who look to relocate to the safest borough in London, expats can choose the neighbourhoods to which they want to relocate based on the most common venues in it.
- ▶ For example if a person is looking for a neighbourhood with good connectivity and public transportation we can see that Cluster 3 and 4 have train stations and bus stops as the most common venues.
- ▶ If a person is looking for a neighbourhood with stores and restaurants in a close proximity then the neighbourhoods in the first cluster is suitable.
- ▶ For a family I feel the neighbourhoods in cluster four is more suitable due to common venues in cluster, these neighbourhoods have common venues such as Parks, gyms, Bus Stops, Restaurants, Electronic stores and soccer fields which are ideal for a family.

## 6. Conclusion

- ▶ This project helps a person get 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 about more places moving into neighbourhood.
- ▶ We have just taken safety as the primary concern to shortlist the safest places in London.