



# Capstone Project - The Battle of Neighbourhoods -Final

# Introduction

- **Background:** Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy in your new home.
- **Problem:** The 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 interest to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.

# Data Acquisition

**Data Acquisition:** The data acquire for this project is 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 data source is scraped from Wikipedia page that contains the list of London boroughs.
- The third source of data is the list of neighborhoods in Royal borough of Kingston upon Thames

# Data Cleaning

**Data Cleaning:** The data cleaning process for each of the three sources of data re done separately.

- From 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 boroughs for each major category.
- The second data is scraped from Wikipedia page using beautiful soup library of python.
- The two data sets are merged on the borough names. The purpose of this data set is to visualize the crime rates in each borough and identify the borough with the latest crime rate
- The third data set is created with the names of the neighborhoods and the name of the borough with the latitude and longitude using Google API geocoding.
- The new data set is used to generate most common venues for each neighborhood using the foursquare API, finally using K means algorithm to cluster the similar neighborhoods.

# Methodology

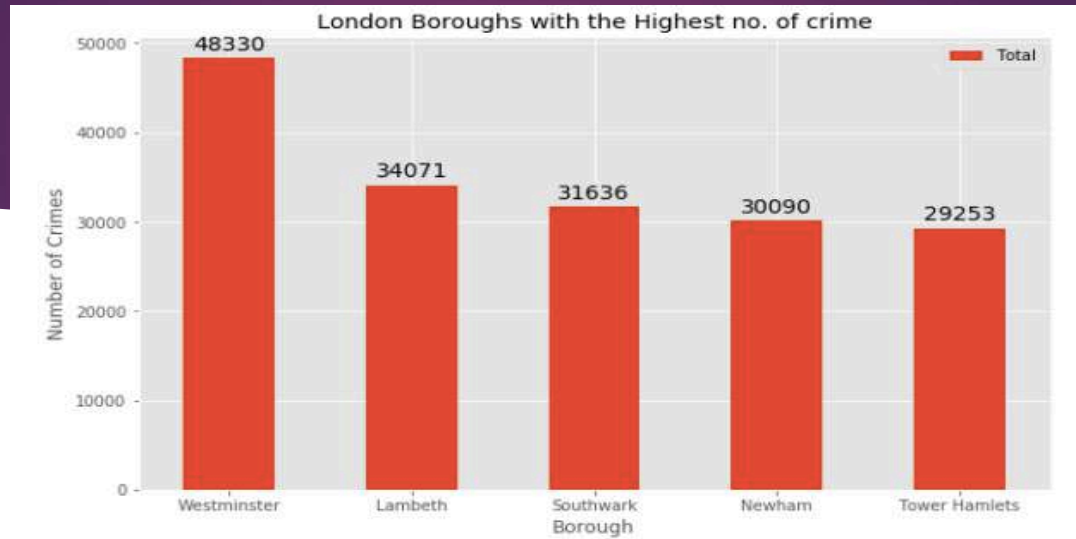
## Exploratory Data Analysis:

	Isao_code	borough	major_category	minor_category	value	year	month
0	E01004177	Sutton	Theft and Handling	Theft/Taking of Pedal Cycle	1	2016	8
1	E01000733	Bromley	Criminal Damage	Criminal Damage To Motor Vehicle	1	2016	4
2	E01003989	Southwark	Theft and Handling	Theft From Shops	4	2016	8
3	E01002276	Havering	Burglary	Burglary in a Dwelling	1	2016	8
4	E01003674	Redbridge	Drugs	Possession Of Drugs	2	2016	11

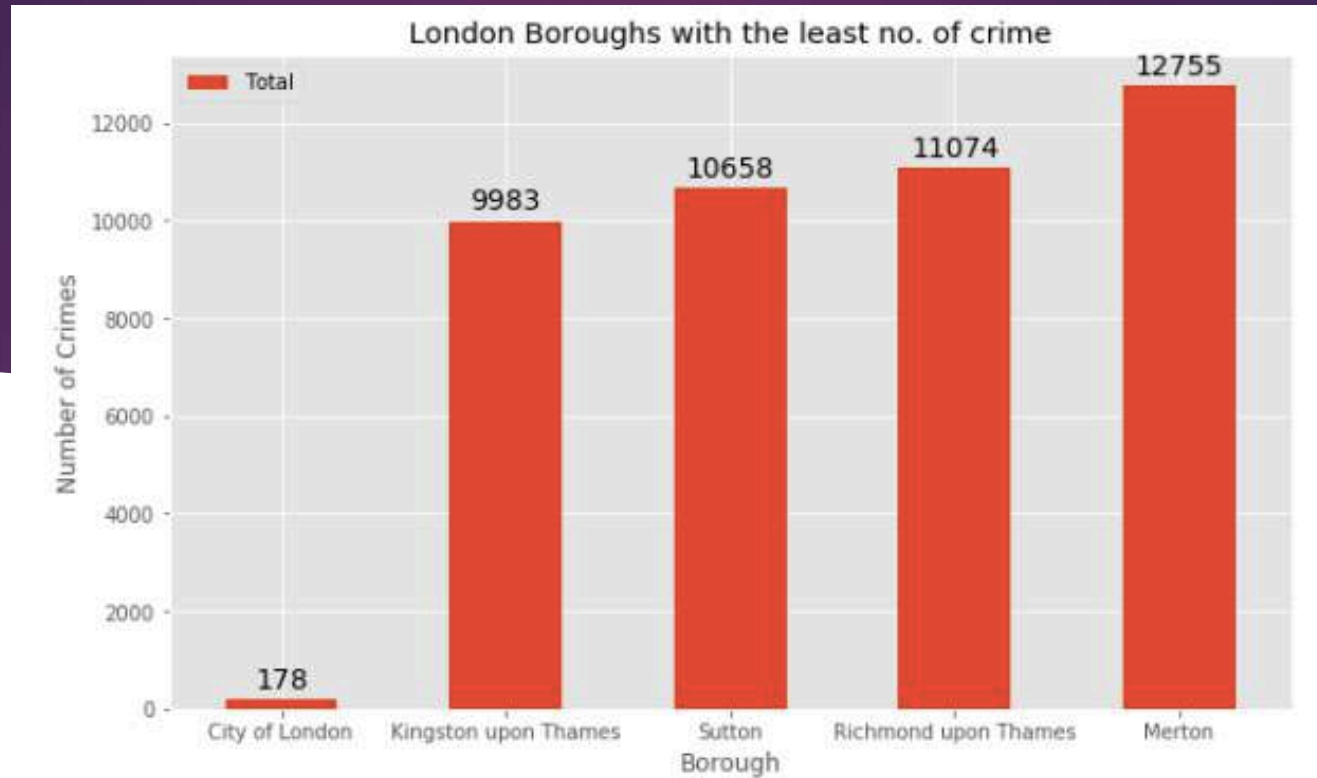
## Statistical Summary of crimes

The count for each of the major categories of crime returns value 33 which is the number of London boroughs. 'Theft and handling's the highest reported crime during year 2016 followed "Violence against the person" ,"Criminal Damage"

## Boroughs with the highest crime rates

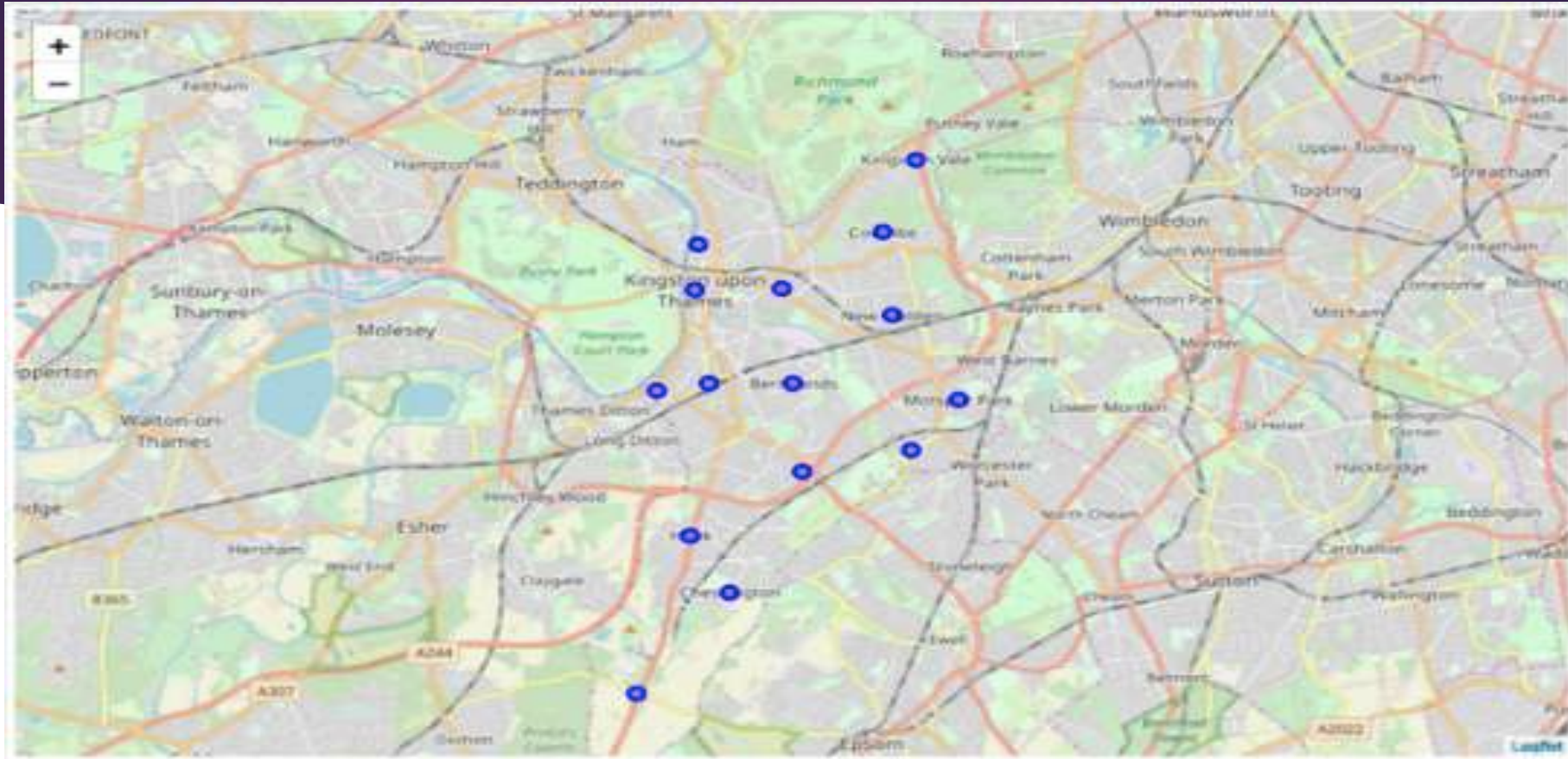


Comparing five boroughs with the highest crime rate during the year 2016, its evident that Westminster has the highest crime recorded followed by Lambeth, Southwark, Newnham and Town Hamlets.



Comparing five boroughs with lowest crime rate during 2016, City of London has the lowest record crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.

## Neighborhoods in Kingston upon Thames



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



## Modelling

- Using the final data set containing the neighborhoods in the 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 the Foursquare API

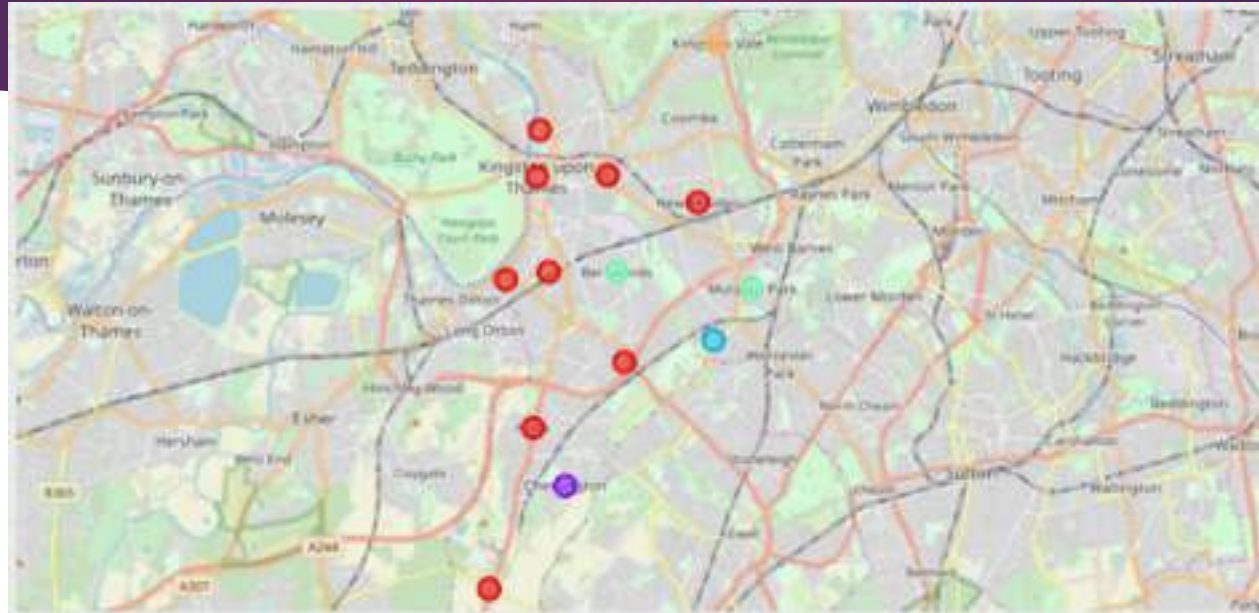
1:

	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.
- To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K-means clustering.

# Result

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



Each cluster is color coded, we can see that majority of the neighborhood falls in the red cluster which is the first cluster, three neighborhoods have their own cluster(Blue, Purple and yellow), these are clusters two, three and five. The green cluster consists of the two neighborhoods which is the fourth cluster.

## Cluster1 Looking into the neighborhoods in the first 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place

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.

## Cluster2 : Looking into the neighborhoods in the 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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture Store

Capstone Project - The Big Data Analytics  
Neighborhoods - London  
Clustering - K-Means

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

## Cluster3 Looking into the neighborhoods in the 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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Car Wash

The third cluster has one neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

## Cluster4 : Looking into the neighborhoods in 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Park	Gym	Restaurant	Soccer Field	Bus Stop	Wine Shop	Fast Food Restaurant

The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

## Cluster5 : Looking into the neighborhoods in 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Italian Restaurant	Soccer Field	Garden Center	Furniture / Home Store	Fashion Store

The fifth cluster has one neighborhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores.

# Discussion

- The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose 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 Clusters 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 that the neighbourhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, 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 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. 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.



Thank You !!