

# Capstone Project : The Battle of Neighborhoods

Presentation by:  
Eeshan Nandkumar Chaudhari

# 1. Introduction

- **Background:** Safety is a top concern when moving out to a new area. If you don't feel safe home, you're not going to enjoy the living there
- **Problem:** This project aims to select the safest borough in London based on total crimes, explore the neighbourhoods of that borough to find 10 most common venues in each neighbourhood and finally clustering them with K-Means Clustering
- **Interest:** People who are considering to relocate in London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around it

## 2. Data Acquisition and Cleaning

**Data Acquisition:** The data acquired for this project is a combination of 3 sources :

- The first source of the data uses a London Crime Data that shows crimes per borough in London.
- The second source of data is scraped from a Wikipedia Page that contains the list of London Boroughs.
- The third source of data is the list of neighbourhoods in the Royal Borough of Kingston upon Thames as found on Wikipedia

**Data Cleaning:** The data cleaning process for each data is done separately:

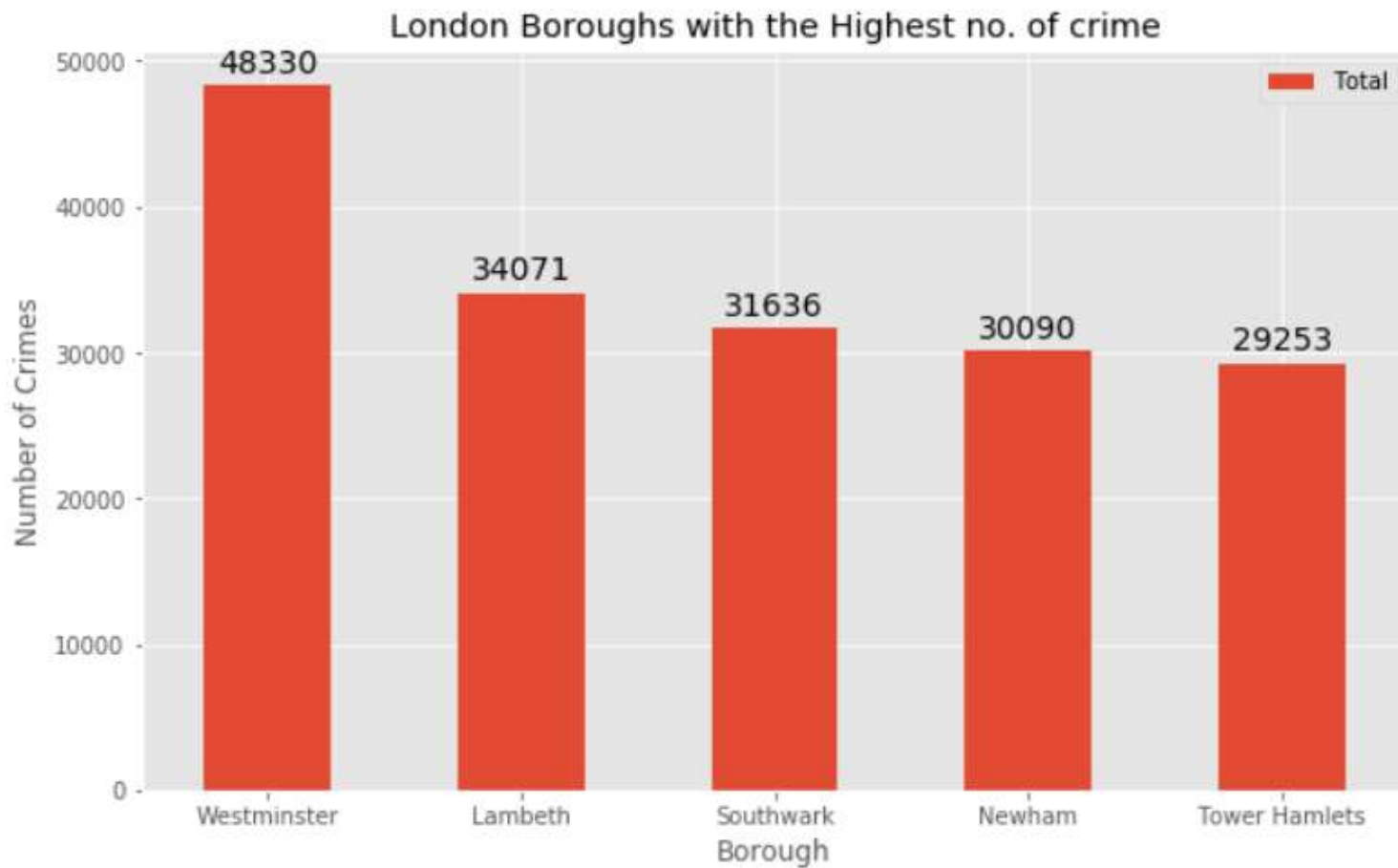
- From London Crime Data, the crimes are selected and the major categories are pivoted to get the total crimes per borough for each major category
- Second data is scraped from Wikipedia page using the Beautiful Soup library in Python. Data is extracted in tabular format
- Two datasets are merged on the Borough names to form a new dataset for visualizing the crime rates and details of the borough
- After visualizing, the borough with the lowest crime rate is found. The third dataset is created with names of neighbourhoods and the name of the borough with latitude and longitude by *Google Maps API Geocoding*
- New dataset is used to generate the 10 most common venues for each neighbourhood using the *Foursquare API*, and finally clustered by K-Means Clustering

# 3. Methodology

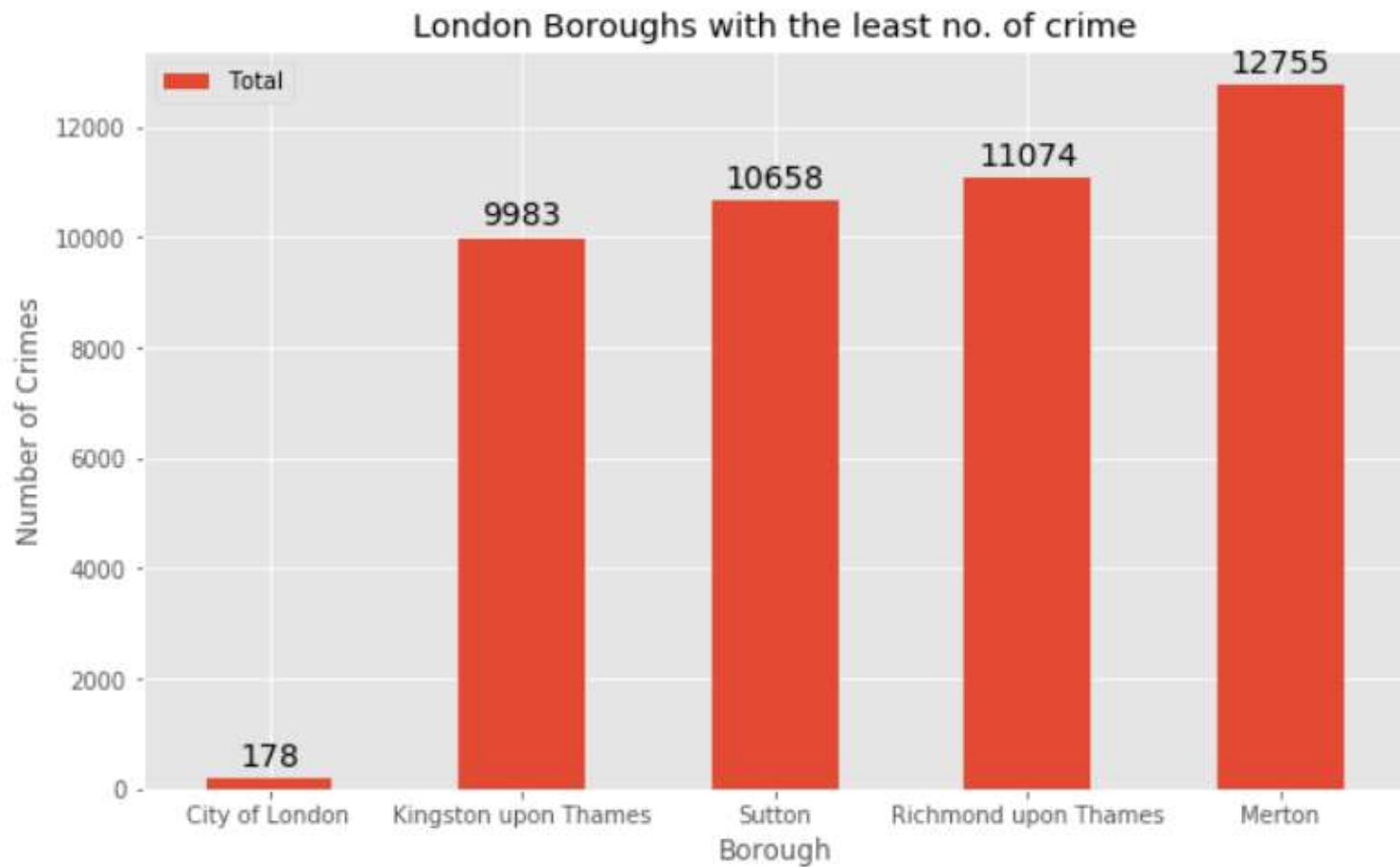
## Exploratory Data Analysis

	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

## Statistical Summary of Crimes



Boroughs with the Highest Crime Rates

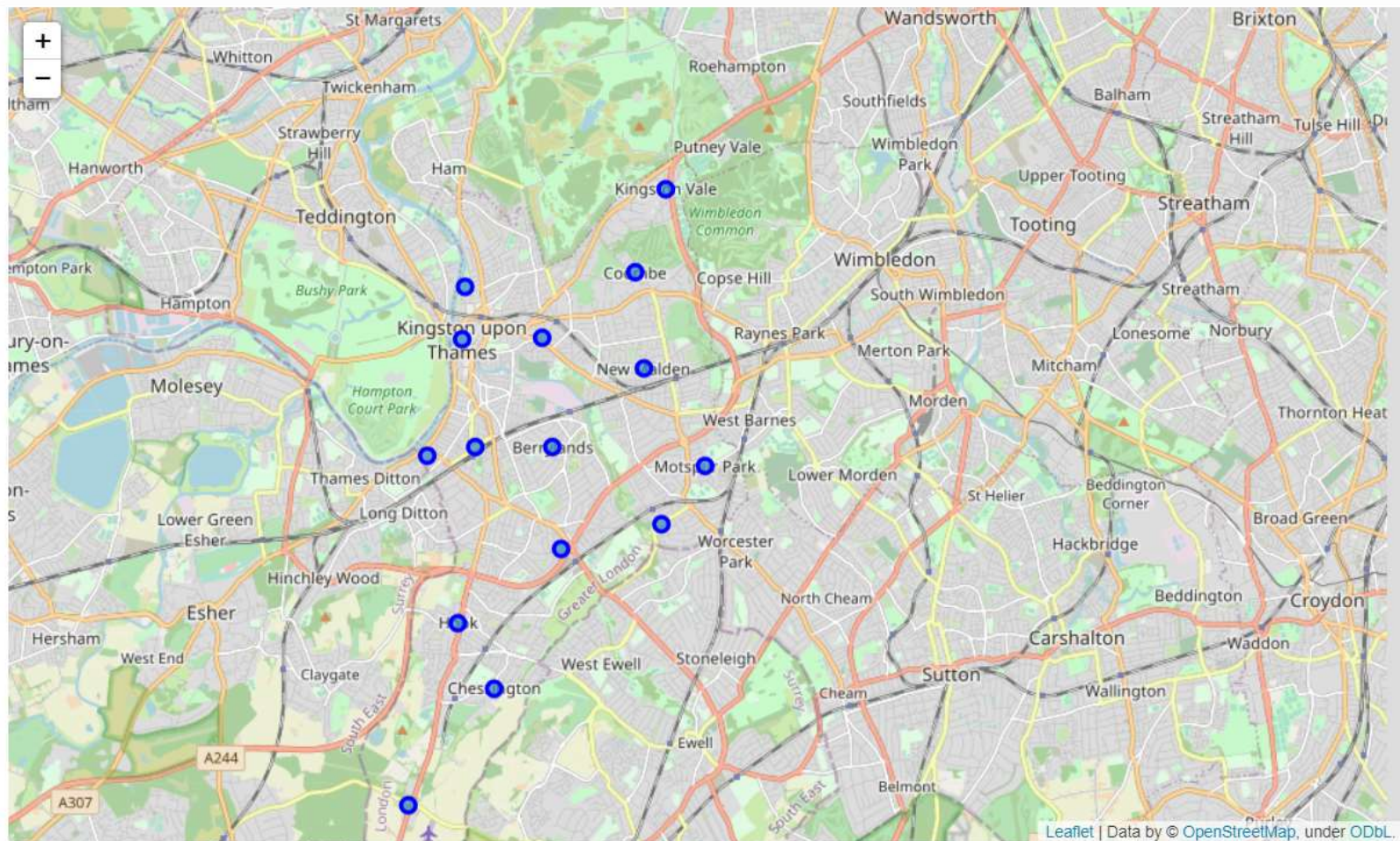


Boroughs with the Lowest Crime Rates

- From the above plots, it is evident that City of London has the lowest number of crimes. But it is not a London Borough as it has an area of 1.12 square miles and population of 7000 as of 2013. Hence it is a very small area.
- The next borough with the lowest crimes is Kingston upon Thames. Hence this is considered



# Neighbourhoods in Kingston upon Thames



# Modelling:

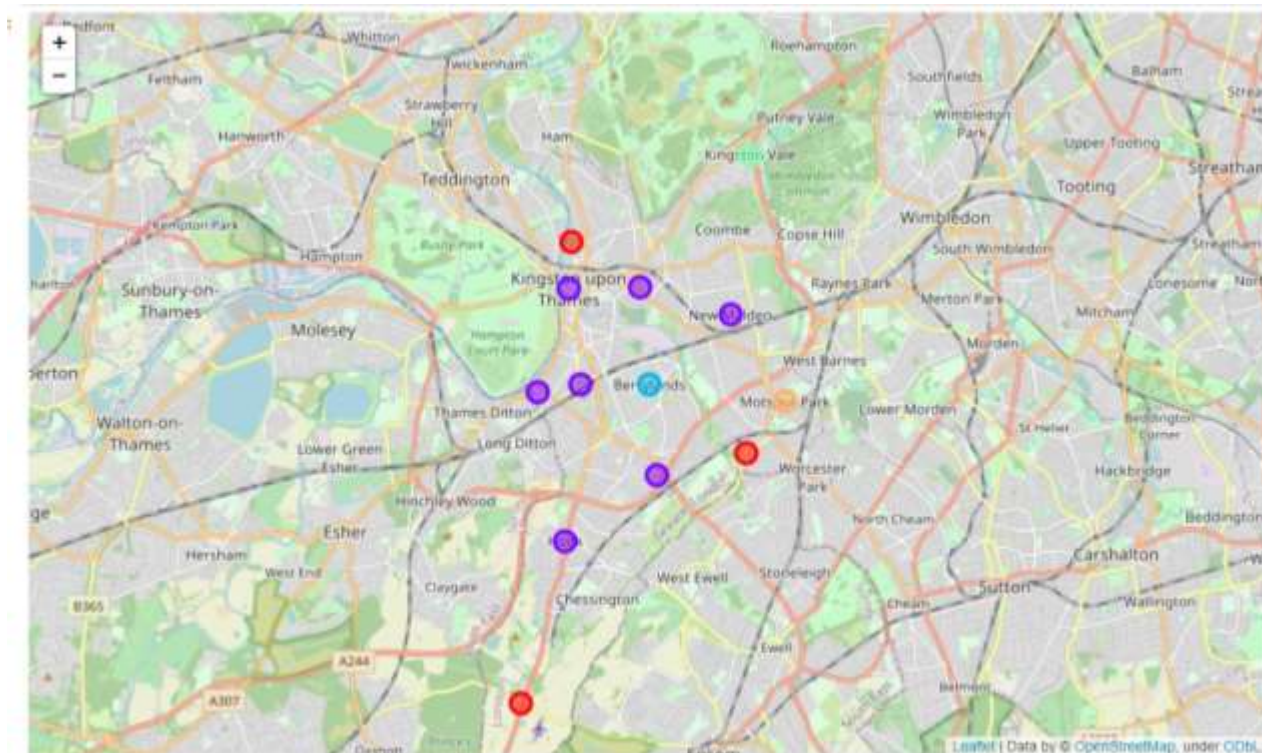
- Using the final dataset containing the neighbourhoods Kingston upon Thames (KUT) along with latitude and longitude, all neighbourhood within 500 meter radius of each neighbourhood are found by connecting to Foursquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Sutton 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	102 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	51.418548	-0.305915	Pub

- One hot encoding is done on the venues data . Venue data is then grouped by the neighbourhood and mean is calculated. Finally the 10 common venues are calculated
- To help people find similar neighbourhoods in the safest borough, clustering similar neighbourhoods using K-Means Clustering is done
- 5 clusters are used. The reason for using K-Means clustering is to cluster neighbourhoods with similar venues together so that people can shortlist the area of interests based on venues around each neighbourhood

## 4. Results

- After K-Means clustering, each cluster can be accessed to see which neighbourhoods were assigned to each of the 5 clusters. The clustered neighbourhoods are visualised by *Folium* Library





- Each cluster is colour coded for the ease of presentation. Here, majority of the neighbourhood falls in Violet Cluster which is 1<sup>st</sup> Cluster
- Looking into the neighbourhoods of the 5 clusters in the next slide

# 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Café	Shop & Service	Gym / Fitness Center	
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Grocery Store	Garden Center	Pub	Restaurant	Electronics Store	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Depot
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Train Station	Food	Pub	Construction & Landscaping	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Garden Center	

The first cluster contains 3 neighbourhoods with Pubs, Grocery Store and Train Station as the most common venue

# 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
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Indian Restaurant	Fish & Chips Shop	Bakery	Supermarket	Cosmetics Shop	Deli / Bodega	Department Store	Electronics Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Coffee Shop	Burger Joint	Sushi Restaurant	Pub	French Restaurant	German Restaurant	Theater
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Indian Restaurant	Gastropub	Chinese Restaurant	Sushi Restaurant	Supermarket	Gym	Bar	Korean Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Italian Restaurant	Food	Platform	Pub	Rental Car Location	Pharmacy	Pizza Place
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Harbor / Marina	Chinese Restaurant	Café	Italian Restaurant	Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Pharmacy	Italian Restaurant	Bakery	Bar	Betting Shop
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Restaurant	Bowling Alley	Bus Stop	Café	Coffee Shop	Furniture / Home Store	Hotel

The second cluster contains 7 neighbourhoods with Restaurants, Grocery Store and Coffee Shops as the most common venue

# 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	2	Gym / Fitness Center	Park	Bus Stop	Train Station	Fast Food Restaurant	Deli / Bodega	Department Store	Electronics Store	Farmers Market

The third cluster contains 1 neighbourhood with Gym as the most common venue

# 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	3	Grocery Store	Sandwich Place	Bar	Soccer Field	Cosmetics Shop	Deli / Bodega	Department Store	Electronics Store	Farmers Market

The fourth cluster contains 1 neighbourhood with Grocery Store as the most common venue



# 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
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Construction & Landscaping	Gym	Park	Soccer Field	Fast Food Restaurant	Cosmetics Shop	Deli / Bodega	Department Store	Electronics

The fifth cluster contains 1 neighbourhood with Construction and Landscaping as the most common venue

## 5. Discussion

- The aim of this project is to help people who want to relocate to the safest borough in London, people can choose the neighbourhoods to which they want to relocate based on the most common venues
- For example if someone wants to relocate to a neighbourhood with many Restaurants, he can relocate to the second cluster
- For a family, I feel cluster 1 is more suitable as it has multiple venues nearby which will be suitable

## 6. Conclusion

- This project helps a person get a better understanding of the neighbourhoods with respect to common venues before moving out there
- This project just considers safety as the primary concern to shortlist the safest borough. The future scope of this project can involve other factors like cost of living, job opportunities, etc.