



# CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS PRESENTATION

---

JAN 2020

MOSES KIPLAGAT



# INTRODUCTION

- Background

When considering a move to a new place, safety is one of the main concerns. If the area you are considering moving to isn't safe, it is probably a place you don't want to move to

- Problem

This project aims to select the safest borough in London based on crimes(total), explore the neighborhoods of that borough to find the 10 most common venues, and finally cluster the neighborhoods using k-means.

- Interest

Migrants who are considering a move to London who would be interested in identifying the safest borough in London and explore its neighborhoods and common venues





# DATA ACQUISITION AND CLEANING

## Data Acquisition

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

- The first data source of this 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. This page contains additional information about the boroughs.
- The third data source is the list of Neighborhoods in the Royal Borough of Kingston Upon Thames as found on the Wiki page.



# DATA CLEANING

## Data Cleaning

- Data cleaning for each of the three sources of data are done separately.
- From the London crime data, only data from year 2016 is selected
- The second data is scraped from Wikipedia using BeautifulSoup library in python
- The two data sets are merged on the borough names to form a new data set. The purpose of this data set is to visualize the crime rates in each borough and identify the borough with the least crimes during the year 2016.
- After visualizing the crime in each borough we can find the borough with the lowest crime rate. The third data set is created, with the names of neighborhoods and name of the borough with the latitude and longitude obtained using Google Maps API Recording.
- The new data set is used to generate the 10 most common venues for each neighborhood using Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.



# METHODOLOGY

## Exploratory Data Analysis

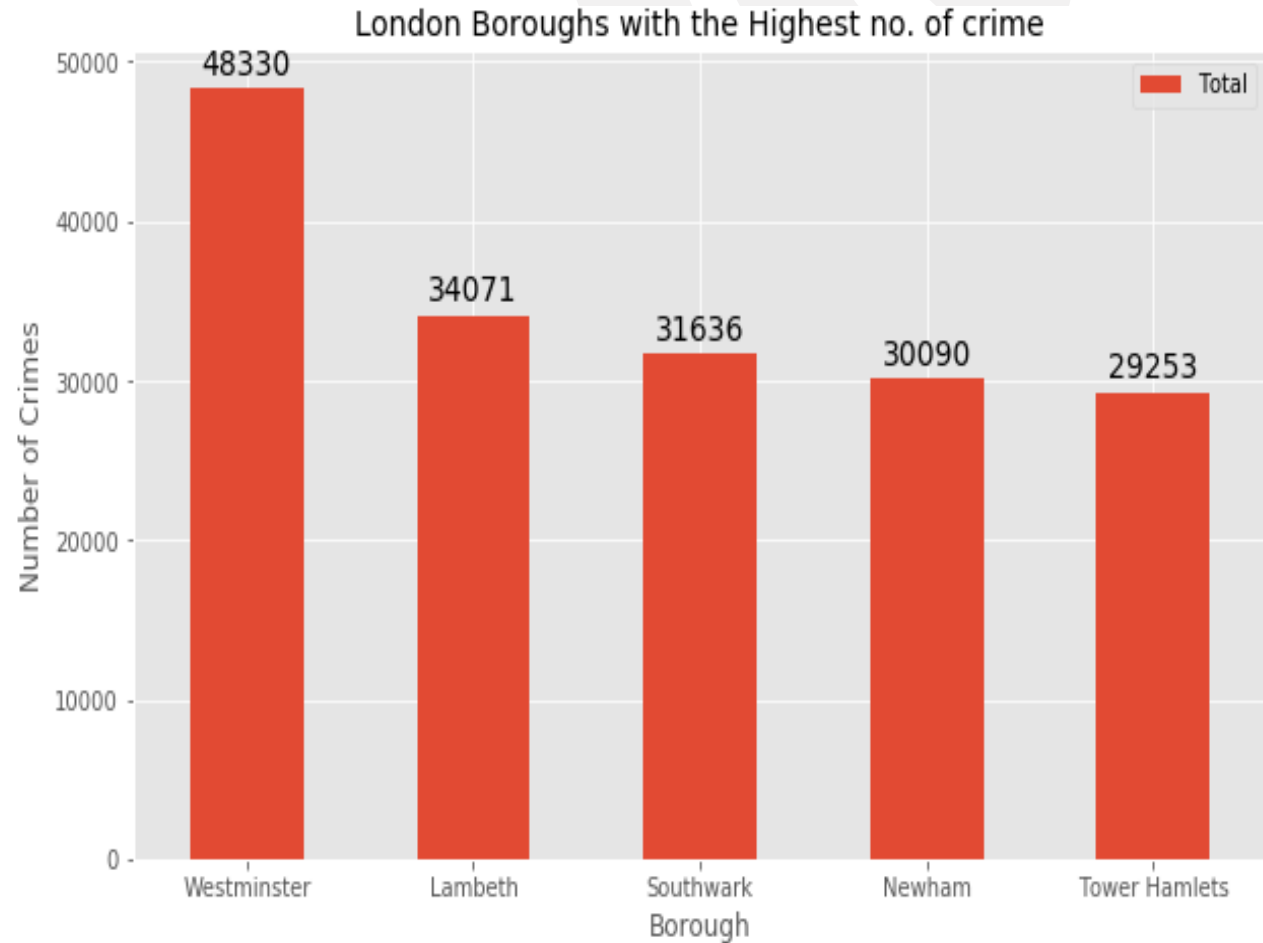
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970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

### *Statistical Summary of Crimes*

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'.

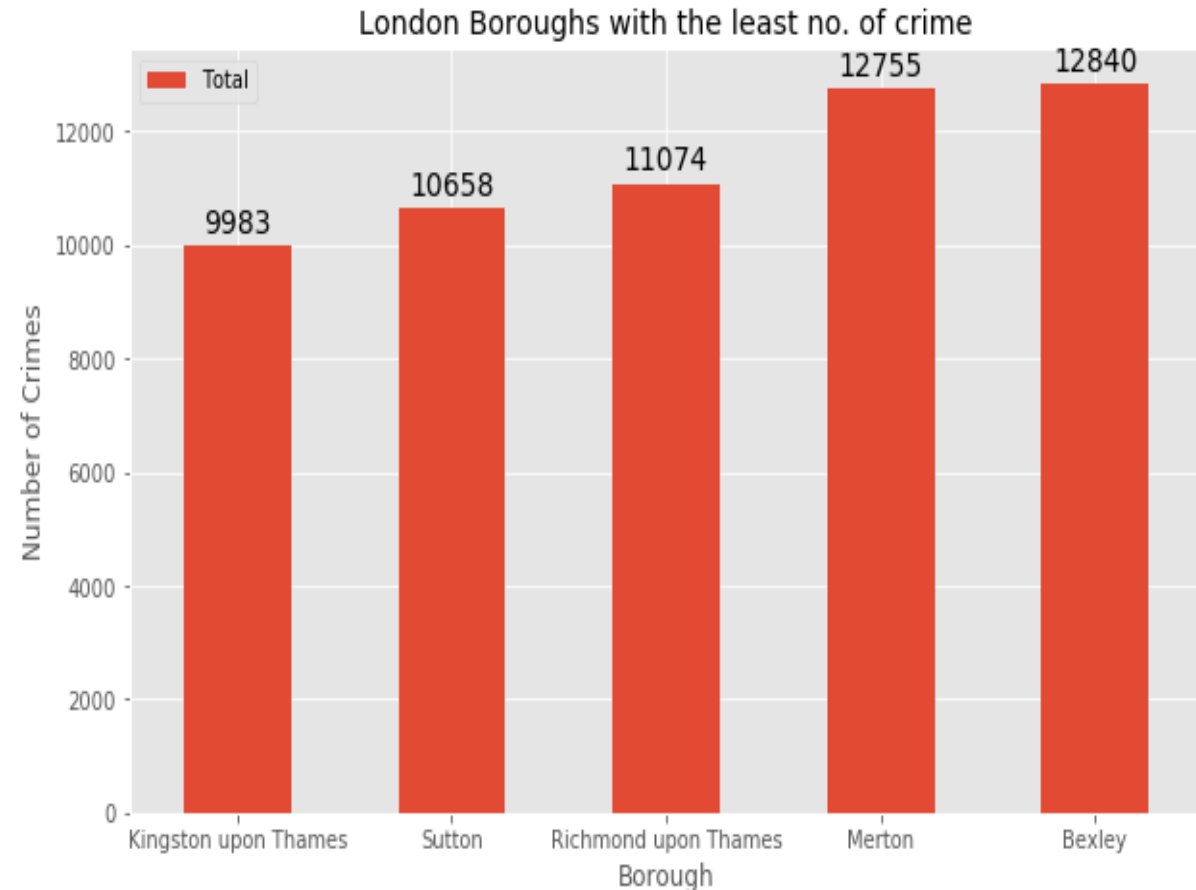
# Boroughs with the highest crime rates

- Comparing 5 boroughs with the highest crime rate during the year 2016 it is evident that Westminster is the highest crimes recorded followed by Lambeth , Southwark, Newham and Tower Hamlets.
- Westminster has significantly higher crime rate than the other 4 boroughs.



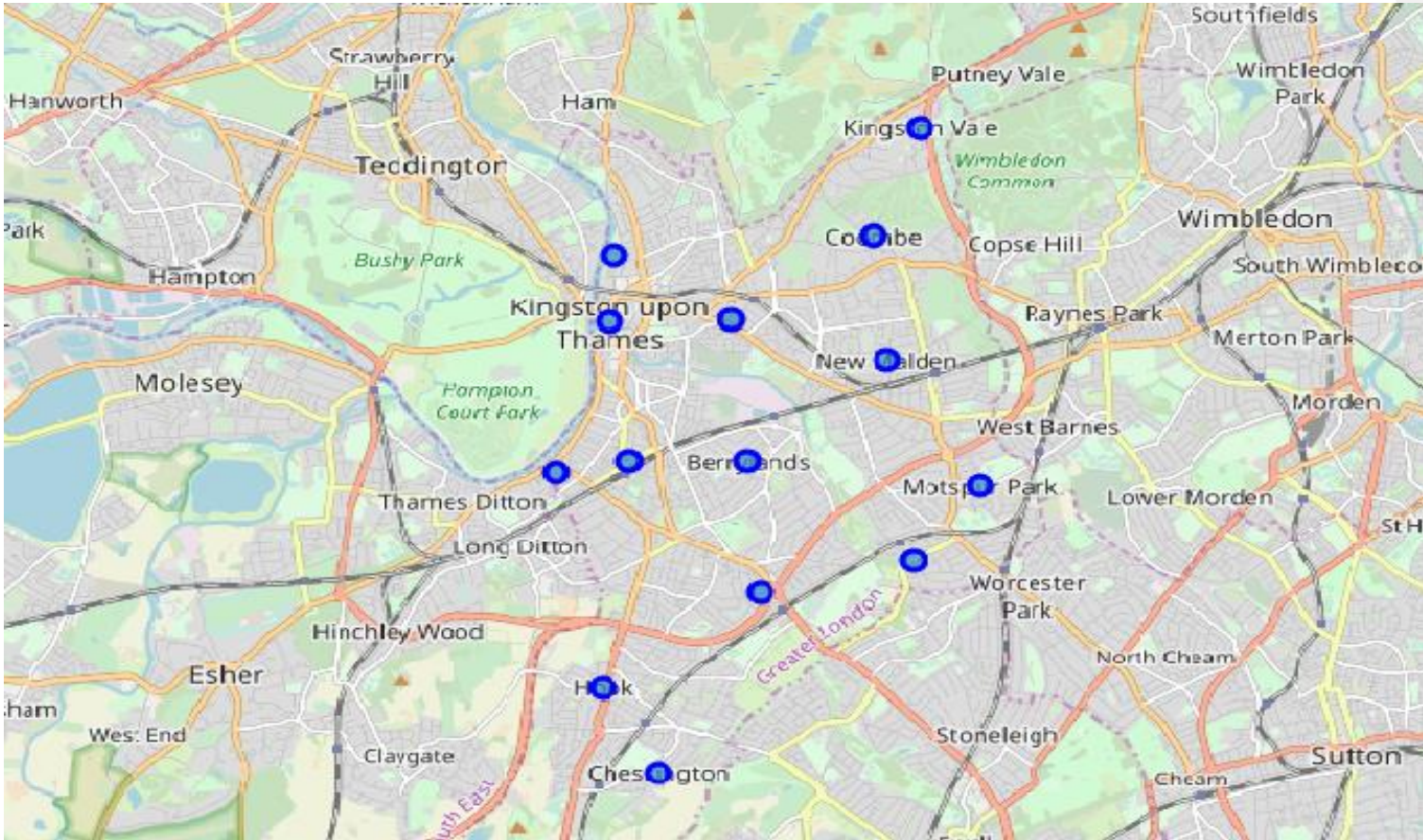
# Boroughs with the lowest crime rates

- Kingston Upon Thames has the lowest recorded crimes followed by Sutton, Richmond upon Thames, Merton and Bexley
- We consider then the next borough with the lowest crime rate as the safest borough in London which is Sutton





# Neighborhoods in Kingston upon Thames



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



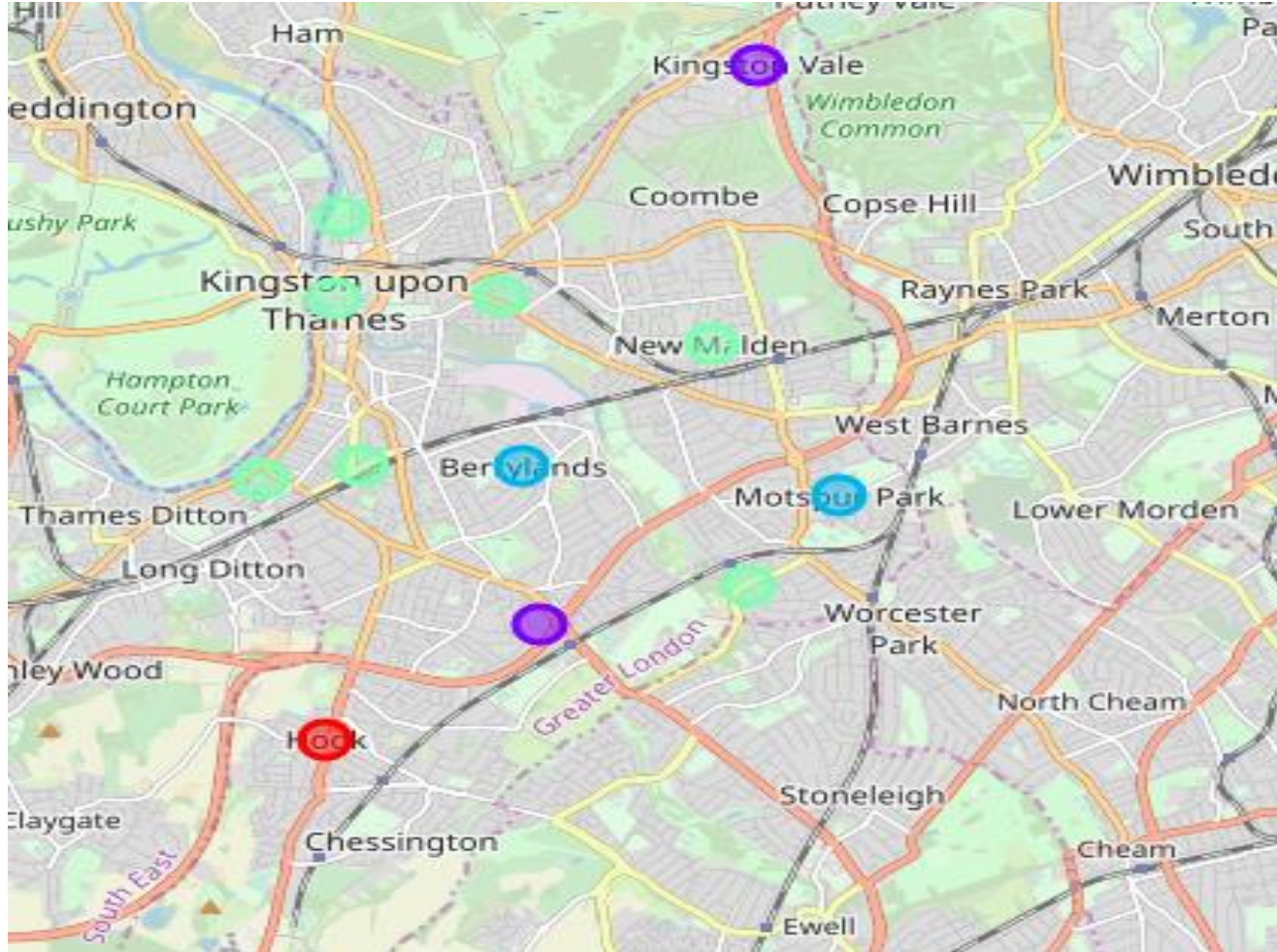
# MODELING

Using the final data set 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 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 to the venues data. 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 which is a form of unsupervised machine learning algorithm that clusters the data based on predefined cluster size.
- We use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reasoning for this is to cluster neighborhoods with similar neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues /amenities around each neighborhood.

# Results



- After running the k-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.
- Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood, we can see the majority falls under the green cluster which is the first cluster. Three neighborhoods have their own cluster(Blue, Purple, Red)

# 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Plaza	Hotel	Shop & Service	Café	Spa	Indian Restaurant	Supermarket	Fish & Chips Shop	Gym / Fitness Center
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Fish & Chips Shop	Bakery	Indian Restaurant	Supermarket	Farmers Market	Deli / Bodega	Department Store	Dessert Shop	Discount Store	Donut Shop
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Italian Restaurant	Café	Clothing Store	Pub	Hotel	Bakery	Sushi Restaurant	Department Store	Bookstore
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Korean Restaurant	Gym	Gastropub	Chinese Restaurant	Bar	Sushi Restaurant	Supermarket	Indian Restaurant	Department Store	Dessert Shop
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Food	Italian Restaurant	Pub	Pizza Place	Breakfast Spot	Coffee Shop	Japanese Restaurant	Pharmacy	Fried Chicken Joint
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Pub	Train Station	Food	Construction & Landscaping	Women's Store	English Restaurant	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Pub	Café	Gym	Chinese Restaurant	Restaurant	Pet Café	Italian Restaurant	Fish & Chips Shop
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Breakfast Spot	Bistro	Pharmacy	Platform	Post Office	Hotel
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Café	Italian Restaurant	Furniture / Home Store	Coffee Shop	Train Station	Thai Restaurant	Pharmacy	Pizza Place	Restaurant

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



## 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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	1	Gym / Fitness Center	Park	Bus Stop	Women's Store	Farmers Market	Department Store	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Gym	Bus Stop	Park	Soccer Field	Gastropub	Garden Center	Deli / Bodega	Department Store	Dessert Shop	Discount Store

the second cluster has two neighborhoods which consists of venues such as Gyms,Parks,Bus Stops etc.

## 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
7	Malcen Rushett	Kingston upon Thames	51.341052	-0.319076	2	Convenience Store	Pub	Garden Center	Restaurant	English Restaurant	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Discount Store

the third cluster has one neighborhood consisting of venues such as convinience stores, pub,garden center etc.

## 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	
3	Coombe	Kingston upon Thames	51.41945	-0.265398	3	Tea Room	Women's Store	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner

the fourth cluster has one neighborhood in it with venues such as farmers market, cosmetics shops, deli/bodega etc.

## 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Soccer Field	Bar	Italian Restaurant	Grocery Store	Gastropub	Garden Center	Deli / Bodega	Department Store	Dessert Shop	Discount Store

the fifth cluster has one neighborhood consisting of soccer field, bar, restaurant etc.

# RESULTS

- The aim of this project is to help people who want to relocate to the safest borough in London, immigrants can chose 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 2 and 1 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 close proximity then the neighborhoods in the first cluster is suitable.
- For a family I feel that the neighborhoods in Cluster 2 are more suitable dues 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 borough of London. The future of this project includes looking at other factors including cost of houses so that to determine the appropriate budget should someone decide to purchase a house



# THANK YOU !

---

