



# **THE BATTLE OF NEIGHBOURHOODS**

**By Prabhat Ranjan**

# INTRODUCTION

- **Problem Description** : Due to crime rate in London , how do we determine which borough in London will be the safest for staying?
- **Objective**: Our objective for this assignment is to find 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**.
- **Target Audience**: The target audience for the report will be the people who are looking to relocate to London or looking for safest borough in London. In order to finalize a neighborhood to hunt for an apartment, safety is considered as a top concern when moving to a new place. The crime statistics will provide an insight into this issue..

# DATA ACQUISTION & CLEANING

- **Data Acquisition:** The data acquired for this project is a combination of data from three sources:
  - 1) The first data source of the project uses a London crime data that shows the crime per borough in London.
  - 2) The second source of data is scrapped from a Wikipedia page that contains list of London boroughs. This page contains additional information about the boroughs.
  - 3) The third data source is the list of Neighbourhoods in the royal borough of Kingston upon Thames as found on the Wikipedia page.

# DATA ACQUISTION & CLEANING CONT..

- **Data Cleaning** : The data cleaning for each of the three sources of data sources are done:
- From the 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 the boroughs for each major category.
- The second data is scraped from a Wikipedia page using the BeautifulSoup library in python. Using this library we can extract the data in the tabular format as shown in the website.
- 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 recorded 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 the neighborhoods and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.
- The new data set is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

# 3.METHODOLOGY

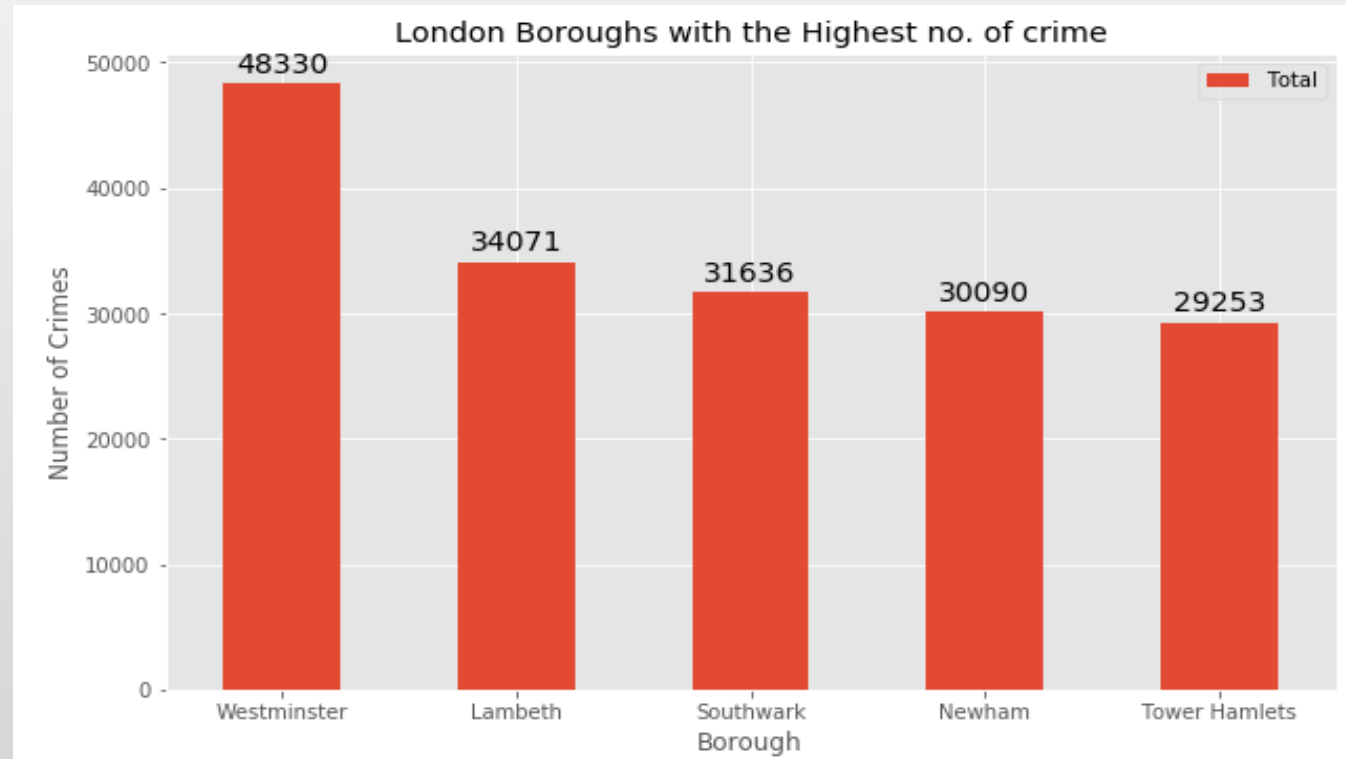
Exploratory data analysis:

Statistical summary of crimes

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
<b>count</b>	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
<b>mean</b>	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
<b>std</b>	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
<b>min</b>	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
<b>25%</b>	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
<b>50%</b>	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
<b>75%</b>	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
<b>max</b>	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

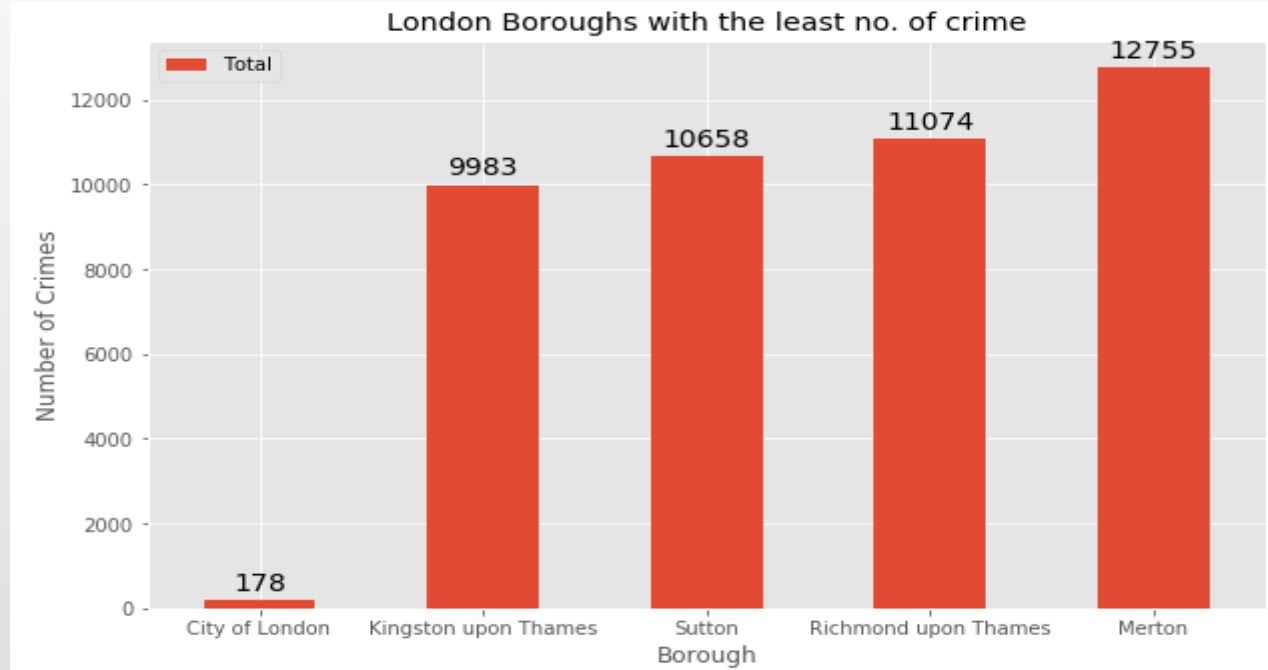
The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and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

## Boroughs with the highest crime rates



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

## Boroughs with the lowest crime rates

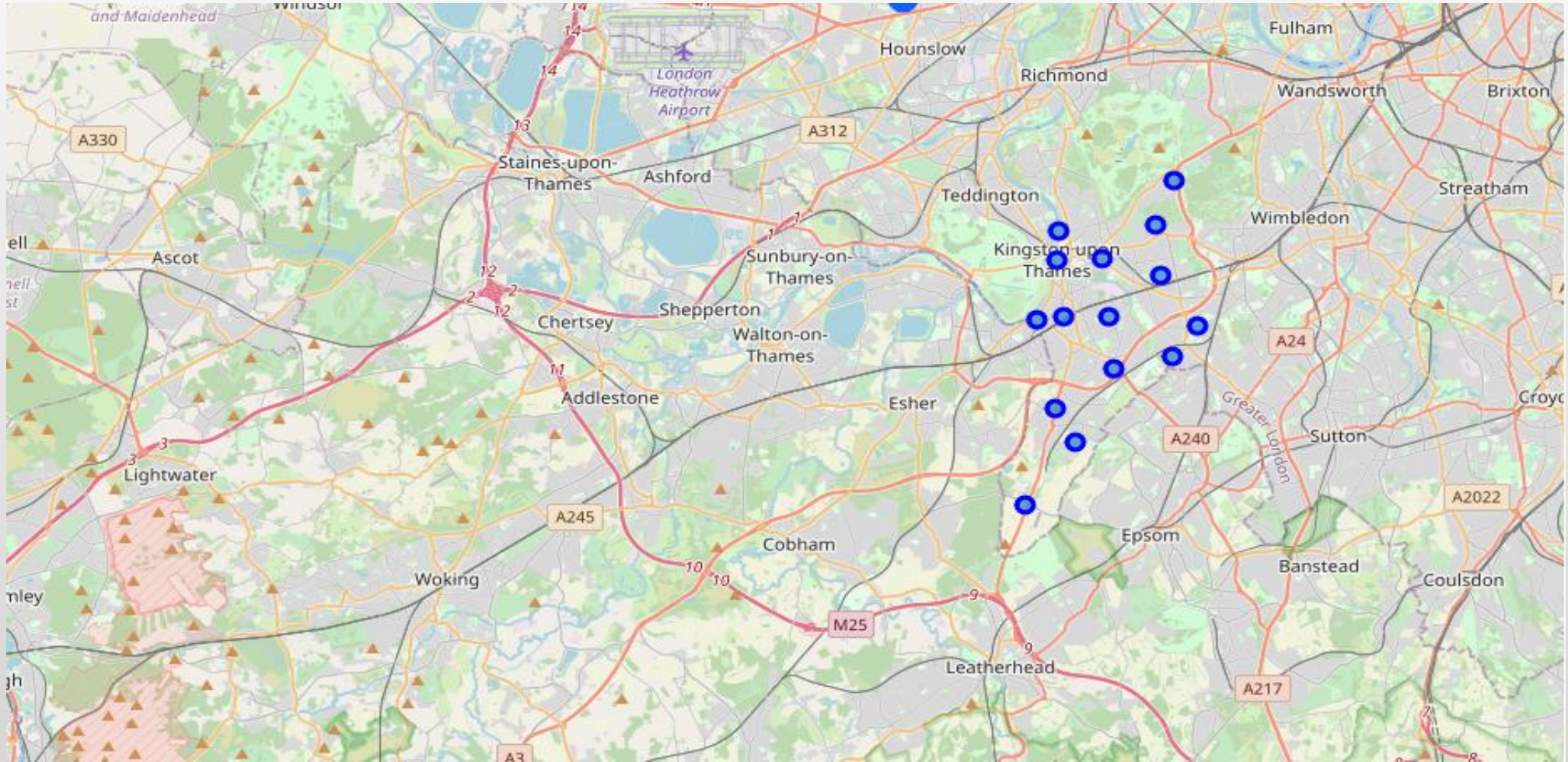


Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.

- City of London has a significantly lower crime rate because it 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.
- We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.



## 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.



Modelling:

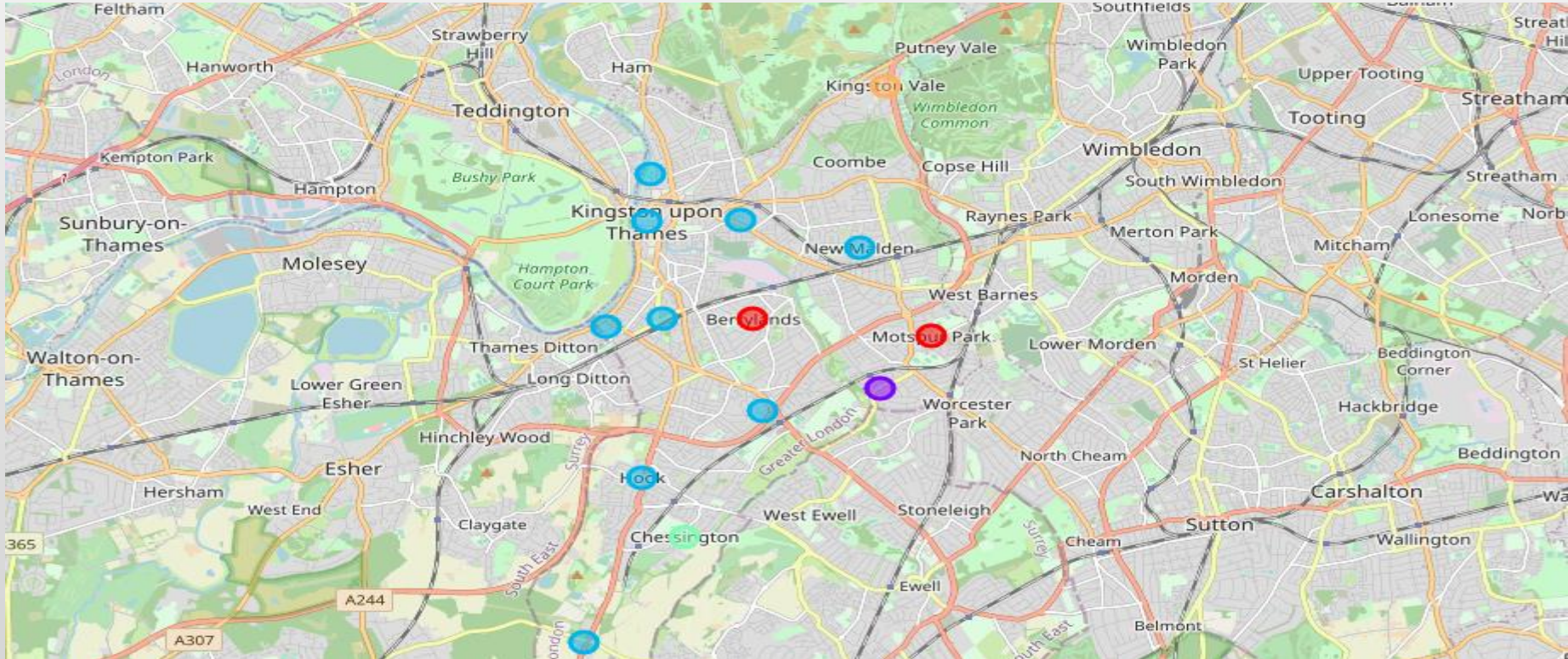
Using the final data set containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API.

	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	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

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

## 4.RESULTS

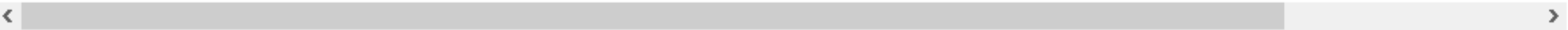
After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.



Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood 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 two neighborhoods which is the 4th cluster.

Cluster 1: Looking into the neighborhoods . 1n 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	0	Wine Shop	Park	Bus Stop	Gym / Fitness Center	Thai Restaurant	Golf Course	Gastropub	Garden Center
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	0	Gym	Bus Stop	Park	Restaurant	Soccer Field	Furniture / Home Store	Fried Chicken Joint	French Restaurant



The cluster one has 2 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are wine shop , gym , park , bus stop etc.

Cluster 2: 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	8th Most Common Venue
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	1	Food	Pub	Construction & Landscaping	Train Station	German Restaurant	Gastropub	Garden Center	Furniture / Home Store

<

>

The second cluster has one neighborhood which consists of Venues such as food, pub , construction & landscaping etc.



Cluster 3: 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	2	Pub	Gym / Fitness Center	Spa	Plaza	Café	Shop & Service	Hotel
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Bakery	Indian Restaurant	Supermarket	Fish & Chips Shop	Wine Shop	Food	Department Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	2	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Asian Restaurant	Furniture / Home Store
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Fast Food Restaurant	Deli / Bodega
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Gastropub	Bar	Sushi Restaurant	Supermarket	Indian Restaurant	Chinese Restaurant	Korean Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	2	Food	Indian Restaurant	Italian Restaurant	Platform	Pub	Wine Shop	Rental Car Location

The third cluster has nine neighborhoods which consists of Venues such as Pub,grocery strores, coffee shop , food, Restaurants, bakery shops etc.



Cluster 4: 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	8th Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	3	Construction & Landscaping	Food	Deli / Bodega	Department Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

<

>

The fourth cluster has one neighborhood in it, this neighborhood have common venues such as construction/landscaping , food . deli bodega , electronic stores , wine shop etc.

Cluster 5: 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	8th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Soccer Field	Grocery Store	Bar	Sandwich Place	Food	Department Store	Electronics Store	Farmers Market

<

>

The fifth cluster has one neighborhood which consists of Venues such as soccer field , grocery store , bar , sandwich place , food , farmers market etc.

## 5.DISCUSSIONS

- The aim of this project is to help people who want to relocate to the safest borough in London, expats 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 1 and 2 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 third cluster is suitable.
- For a family I feel that the neighborhoods in Cluster 1,4 & 5 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.
- The preference of venues may vary from person to person, they can select a neighborhood based on ones priorities.

## 6.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 taking other factors such as cost of living in the areas into consideration to shortlist the borough based on safety and a predefined budget.