# Capstone Project

Devarshi Das

# Introduction

▶ **Background**

Safety is a top concern when moving to a new area. If your safety in an area is not ensured, your stay will not be an enjoyable one.

▶ **Problem**

This project aims to select a relatively safe borough in London based on total crimes, explore the neighborhood for common venues and finally cluster them using k-means.

▶ **Interest**

Expats who are considering to relocate to London will be interested to identify the safer boroughs in London and explore its neighborhoods.

# Data Acquisition and Cleaning

► Data Acquisition

The data acquired for this project is a combination od data from 3 sources:

- The first source uses a London crime data that show crime per borough in London.

- The second source is scraped from a Wikipedia page containing the list of London boroughs which includes additional information for each borough.

- The third data source is the list of neighborhood in the London Borough of Richmond upon Thames as found on Wikipedia.

# ▶ Data Cleaning

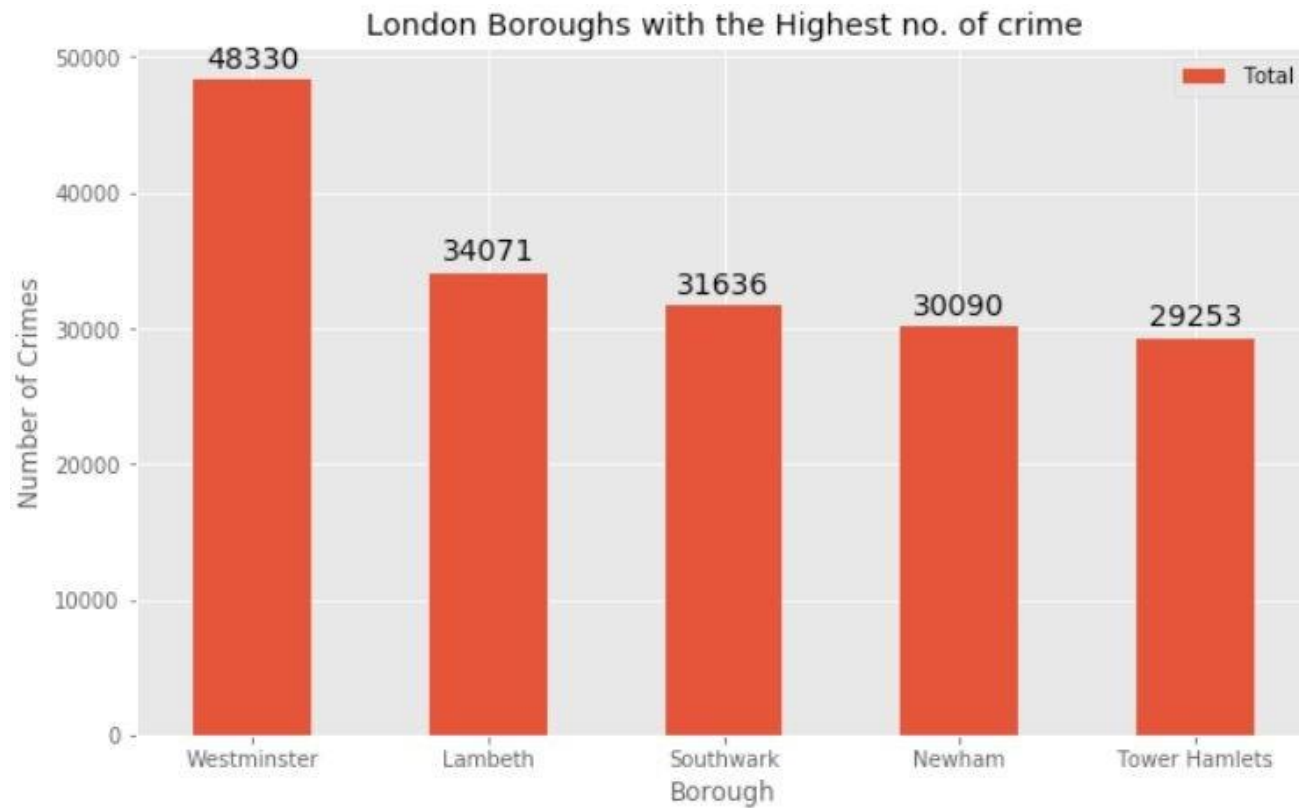The data cleaning process for each source is done separately.

- From London crime data, crimes of most recent year (2016) is only used. Major categories of crimes are summed to get total crimes per borough.

- The second data is scraped from a Wikipedia page using Beautiful Soup library.

- The two data sets are merged on Borough names to form a new data set in order to visualize the crime rates in each borough.

- After visualization, we select a borough and the name of the neighborhoods are created with the borough name and obtaining latitude and longitude using Google Maps API geocoding.

- The final data set is used to generate the 10 most common venues for each neighborhood using Foursquare API, and ultimately using k-means algorithm to cluster similar neighborhoods together.
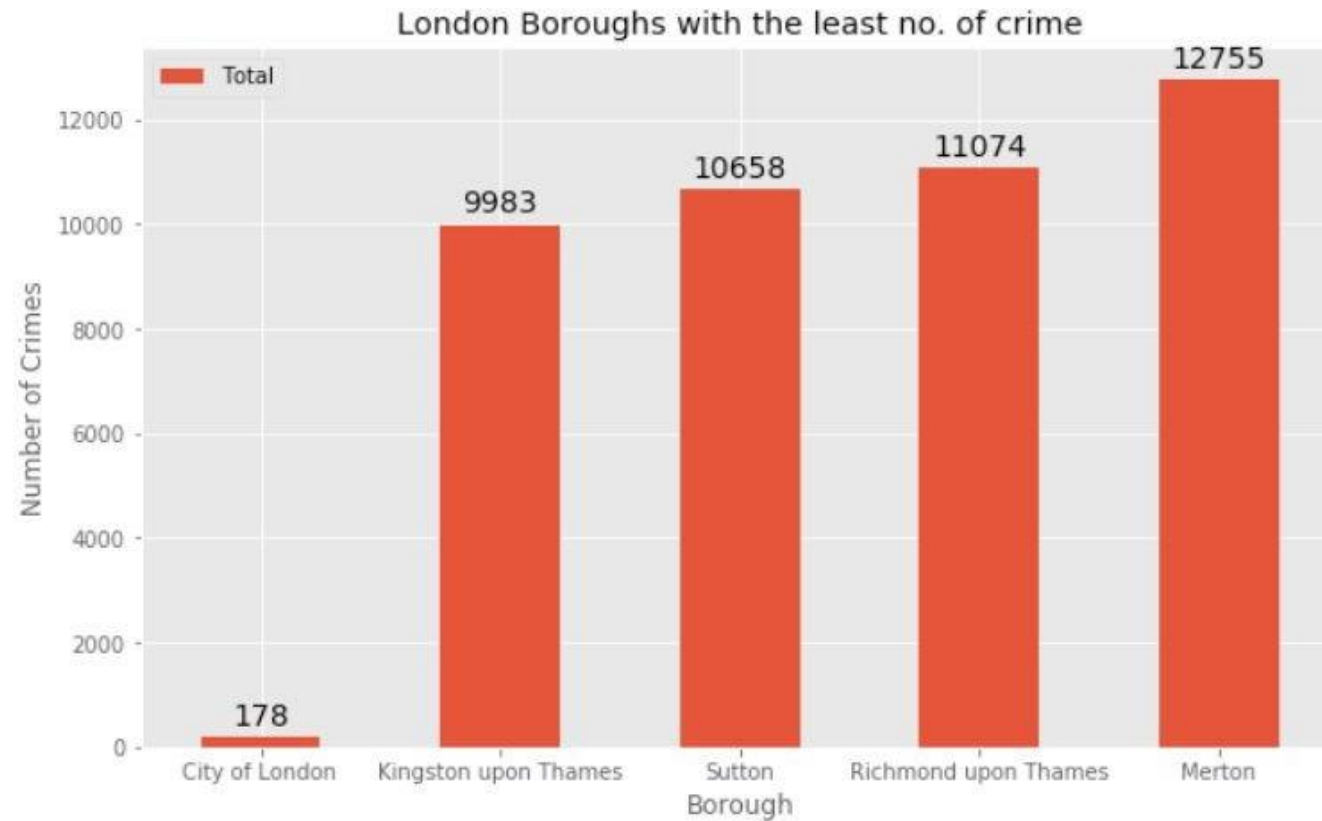
# Methodology

▶ 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 |

The count for each of the major categories return 33, which is the number of London boroughs. Theft and Handling is the highest crime during the year 2016 followed by Violence against person and Criminal damage. The lowest recorded crimes are Drugs, Robbery and Other Notifiable offenses.
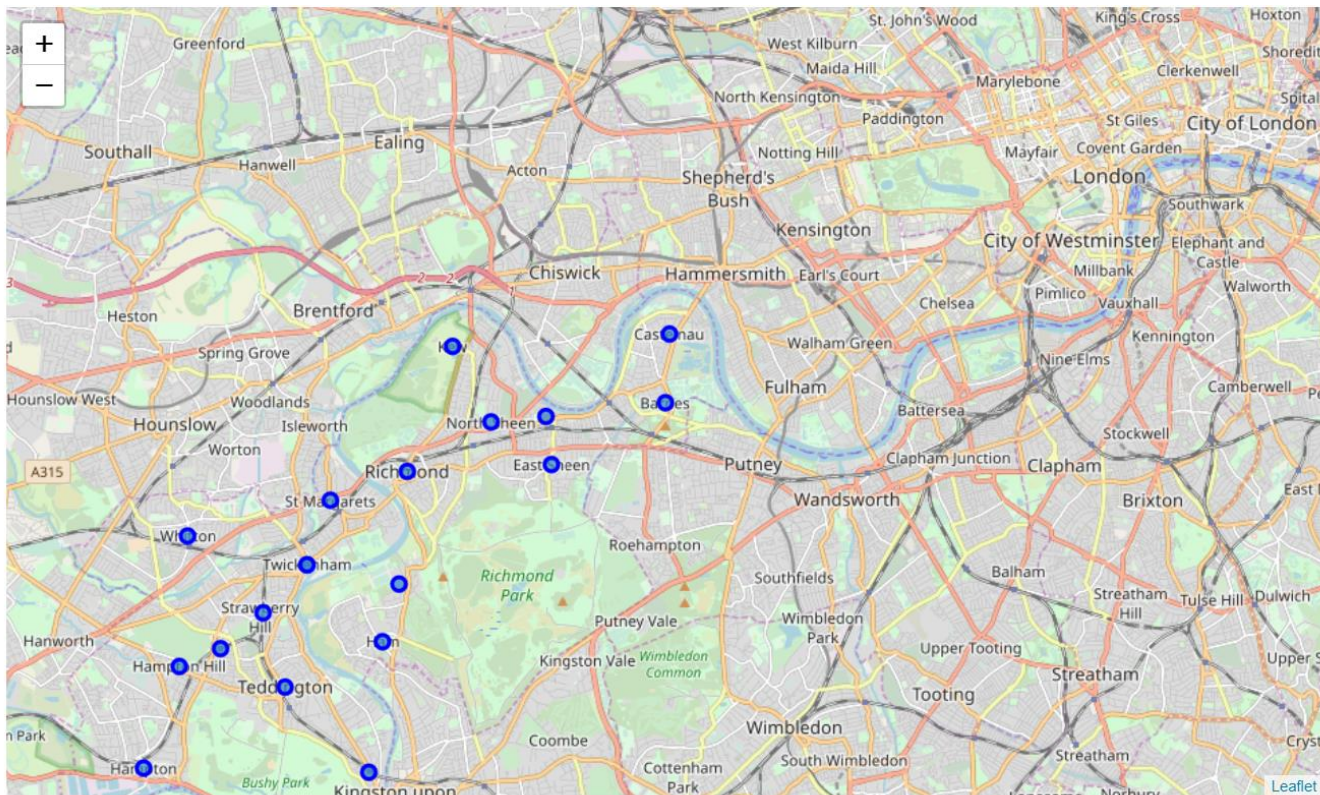
London Boroughs with the Highest no. of crime

Comparing five boroughs with highest crime during the year 2016, it is evident that Westminster has the highest crime recorder, followed by Lambeth, Southwark, Newham and Tower Hamlets.

London Boroughs with the least no. of crime

Comparing borough with lowest crime recorded, City of London has the lowest numbers, followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.
City of London is not considered as it of a very small area.
Hence, we select Richmond upon Thames for this study.

There are 19 neighborhoods in the borough of Richmond upon Thames. They are visualized using folium on python.
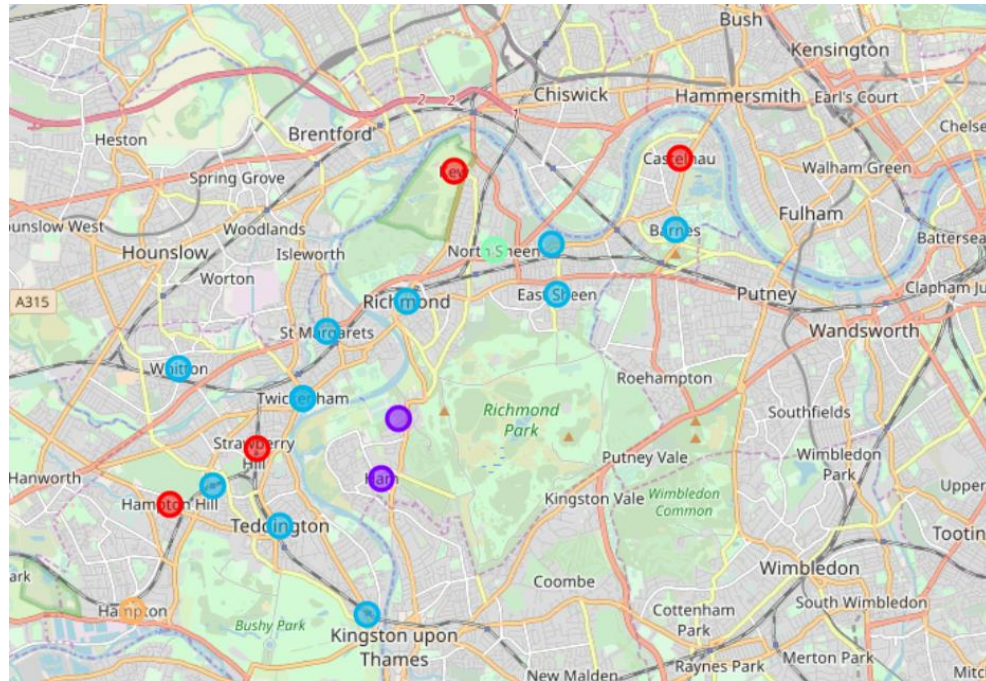
# Modelling

- Using the final data set, 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 | Barnes | 51.471896 | -0.238744 | Olympic Studios Cafe + Dining Room | 51.475158 | -0.240333 | Indie Movie Theater |
| 1 | Barnes | 51.471896 | -0.238744 | ArteChef | 51.474705 | -0.241282 | Pizza Place |
| 2 | Barnes | 51.471896 | -0.238744 | Awesome Thai Cuisine | 51.474905 | -0.240909 | Thai Restaurant |
| 3 | Barnes | 51.471896 | -0.238744 | Alma Café | 51.474880 | -0.239207 | Breakfast Spot |
| 4 | Barnes | 51.471896 | -0.238744 | Olympic Cinema | 51.475171 | -0.240435 | Movie Theater |

- One hot encoding is done on the venues column. The venues column is then grouped by neighborhood and mean of the venues I calculated, finally the 10 common venues are calculated for each neighborhoods.

- To help people find similar neighborhoods, we will be clustering similar neighborhoods using k-means clustering which a form of unsupervised machine learning algorithm that clusters data based on a predefined cluster size.

- We use a cluster size of 5 for this project. This is done so people can choose the area based on the available venues/amenities of each neighborhood.

# Results

After running the k-means clustering, we can access each cluster to see which neighborhoods were assigned to each cluster. Visualizing is carried on a map using folium library.



Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the blue cluster which is the third cluster. Three neighborhoods have their own cluster (Red, Purple and Yellow), these are clusters one, two and five. The green cluster consists of one neighborhood, which is the fourth cluster

# 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 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Castelnau | Richmond upon Thames | 51.482695 | -0.237688 | 0 | Café | French Restaurant | Indian Restaurant | Grocery Store | Italian Restaurant | Recreation Center | Lake | Cosmetics Shop | R |
| 7 | Hampton Hill | Richmond upon Thames | 51.430996 | -0.360211 | 0 | Café | Burger Joint | Pub | Grocery Store | Park | Food & Drink Shop | Cycle Studio | Deli / Bodega | |
| 9 | Kew | Richmond upon Thames | 51.480663 | -0.291929 | 0 | Botanical Garden | Café | Garden | Playground | Wine Shop | Restaurant | Coffee Shop | Historic Site | |
| 15 | Strawberry Hill | Richmond upon Thames | 51.439168 | -0.339301 | 0 | Wine Shop | Train Station | History Museum | Pub | Thai Restaurant | Café | Convenience Store | Fish Market | |

The cluster one consists of 4 neighborhoods out of 19 in the borough Richmond upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Cafes, Restaurants and Gardens.

# 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | Ham | Richmond upon Thames | 51.434764 | -0.309299 | 1 | Pub | Convenience Store | Bakery | Park | Café | Fish Market | Creperie | Cycle Studio | De Bode |
| 12 | Petersham | Richmond upon Thames | 51.443691 | -0.305293 | 1 | Pub | Bus Station | Garden Center | Café | Sports Club | Playground | Fish Market | Creperie | Cy Stu |

The second cluster has one neighborhood which consists of Venues such as Pubs, Stores, Bakeries, and Cafes.

# 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Barnes | Richmond upon Thames | 51.471896 | -0.238744 | 2 | Park | Bookstore | Farmers Market | Pub | Café | Food & Drink Shop | Italian Restaurant | Breakfast Spot |
| 2 | East Sheen | Richmond upon Thames | 51.462371 | -0.267094 | 2 | Coffee Shop | Pub | Pizza Place | Plaza | Stationery Store | Italian Restaurant | Middle Eastern Restaurant | Creperie |
| 3 | East Twickenham | Richmond upon Thames | 51.446744 | -0.328189 | 2 | Pub | Coffee Shop | Italian Restaurant | Indian Restaurant | Café | Garden | Japanese Restaurant | Fish Market |
| 4 | Fulwell | Richmond upon Thames | 51.433748 | -0.349685 | 2 | Pizza Place | Garden Center | Pub | Diner | Seafood Restaurant | Chinese Restaurant | Café | Liquor Store |
| 8 | Hampton Wick | Richmond upon Thames | 51.414452 | -0.312674 | 2 | Hotel | Pub | Plaza | Train Station | Coffee Shop | Park | Indian Restaurant | Italian Restaurant |
| 10 | Mortlake | Richmond upon Thames | 51.469887 | -0.268523 | 2 | American Restaurant | Platform | Grocery Store | Golf Course | Pub | Coffee Shop | Tapas Restaurant | Park |
| 13 | Richmond | Richmond upon Thames | 51.461353 | -0.303277 | 2 | Coffee Shop | Italian Restaurant | Pub | Restaurant | Gift Shop | Gastropub | Waterfront | Juice Bar |
| 14 | St Margarets | Richmond upon Thames | 51.456709 | -0.322412 | 2 | Pub | French Restaurant | Comedy Club | Grocery Store | Deli / Bodega | Italian Restaurant | Coffee Shop | Park |
| 16 | Teddington | Richmond upon Thames | 51.427784 | -0.333653 | 2 | Café | Coffee Shop | Pub | Hotel | Mediterranean Restaurant | Grocery Store | Italian Restaurant | Pharmacy |
| 17 | Twickenham | Richmond upon Thames | 51.446744 | -0.328189 | 2 | Pub | Coffee Shop | Italian Restaurant | Indian Restaurant | Café | Garden | Japanese Restaurant | Fish Market |
| 18 | Whitton | Richmond upon Thames | 51.451169 | -0.357976 | 2 | Coffee Shop | Grocery Store | Pizza Place | Bakery | Indian Restaurant | Pub | Café | Steakhouse |

The third cluster is the biggest one, consisting of 11 out of 19 neighborhoods. The most common places are Parks, Cafes, Pubs and Coffee Shops

# 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | North Sheen | Richmond upon Thames | 51.468886 | -0.282283 | 3 | Park | Supermarket | Coffee Shop | Bus Stop | Food & Drink Shop | Cycle Studio | Deli / Bodega | Diner | Farmers Market |

The fourth cluster has one neighborhoods, having common venues such as Parks, Supermarket, Coffee Shops 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 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | Hampton | Richmond upon Thames | 51.415027 | -0.369141 | 4 | Grocery Store | Train Station | Pizza Place | Park | Soccer Stadium | Wine Shop | Fish & Chips Shop | Creperie | Cycle Studio | |

The fifth cluster has one neighborhood which consists of Venues such as Grocery shops, Train Stations, Restaurants, Parks etc.

# Discussion

▶ The aim of this project is to help people who want to relocate to the safest borough in London, expats can choose 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 Cluster 2 and 5 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 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Super Markets, Bus Stops and Restaurants which is ideal for a family.

▶ The preference of venues will vary from person to person.

# 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 safest 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, such as filtering areas based on a predefined budget.