

A decorative graphic on the left side of the slide, consisting of white lines and circles on a blue gradient background, resembling a circuit board or a stylized tree structure.

# CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOOD

PRESENTED BY – PRANAV KUMAR

# 1. 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 living there.
- **Problem:** This 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 interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.


## 2. 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 the 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 Wikipedia page.



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

- From the London crime data, the crime during the most recent year (2016) are only selected. The major categories of the 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 Beautiful Soup library in python. Using this library we can extract 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 from Google Map API.
  - 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

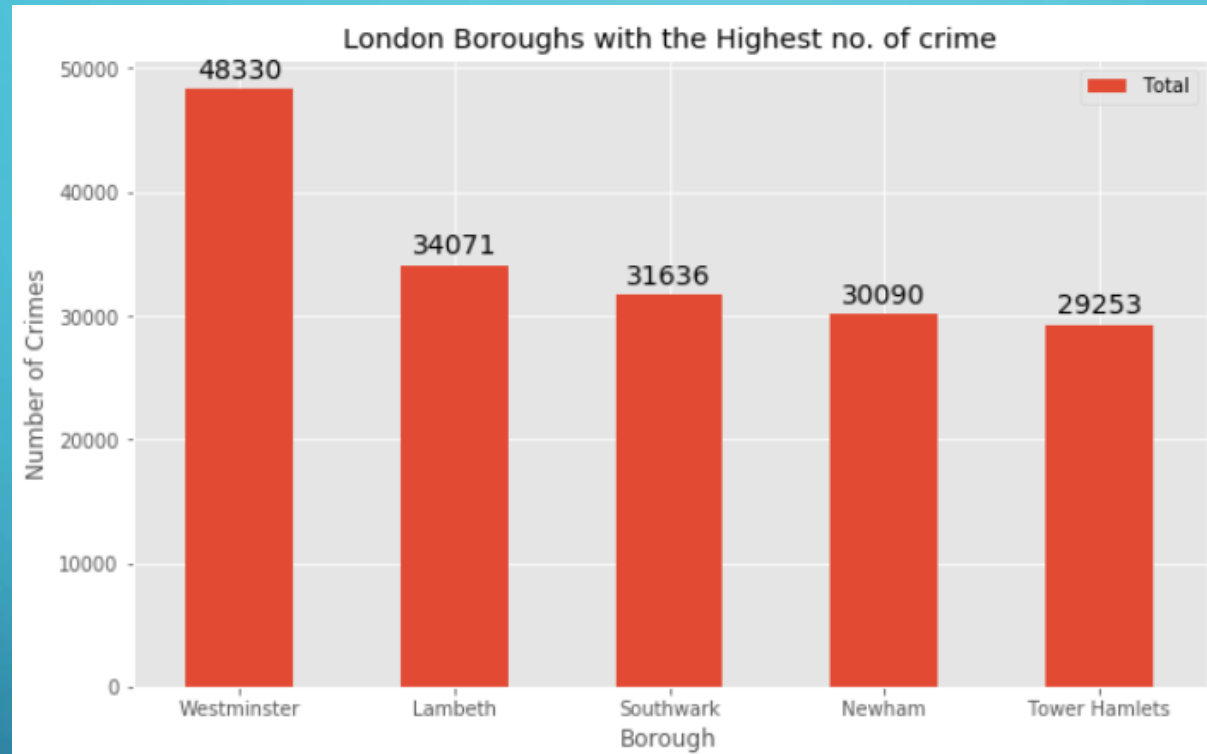
## Exploratory Data Analysis

### Statistical summary of crimes

	Burglary	Criminal Damage	Drugs	Fraud or Forgery	Other Notifiable Offences	Robbery	Sexual Offence	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	33.000000	33.000000
<b>mean</b>	22857.363636	19119.333333	14265.606061	161.363636	3222.696970	7844.636364	38.575758	80662.454545	47214.575758	195386.606061
<b>std</b>	7452.366846	5942.903618	7544.259564	81.603775	1362.107294	4677.643075	15.139002	45155.624776	17226.165191	79148.057551
<b>min</b>	15.000000	16.000000	33.000000	0.000000	17.000000	24.000000	0.000000	561.000000	114.000000	780.000000
<b>25%</b>	18103.000000	17244.000000	8942.000000	106.000000	2358.000000	4744.000000	27.000000	52609.000000	33968.000000	149447.000000
<b>50%</b>	24871.000000	20405.000000	14101.000000	157.000000	3293.000000	7688.000000	40.000000	77940.000000	50943.000000	203879.000000
<b>75%</b>	27980.000000	22755.000000	18389.000000	207.000000	3963.000000	10084.000000	47.000000	92523.000000	59993.000000	228613.000000
<b>max</b>	36981.000000	31218.000000	34031.000000	323.000000	6504.000000	18408.000000	71.000000	277617.000000	72726.000000	455028.000000

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'.

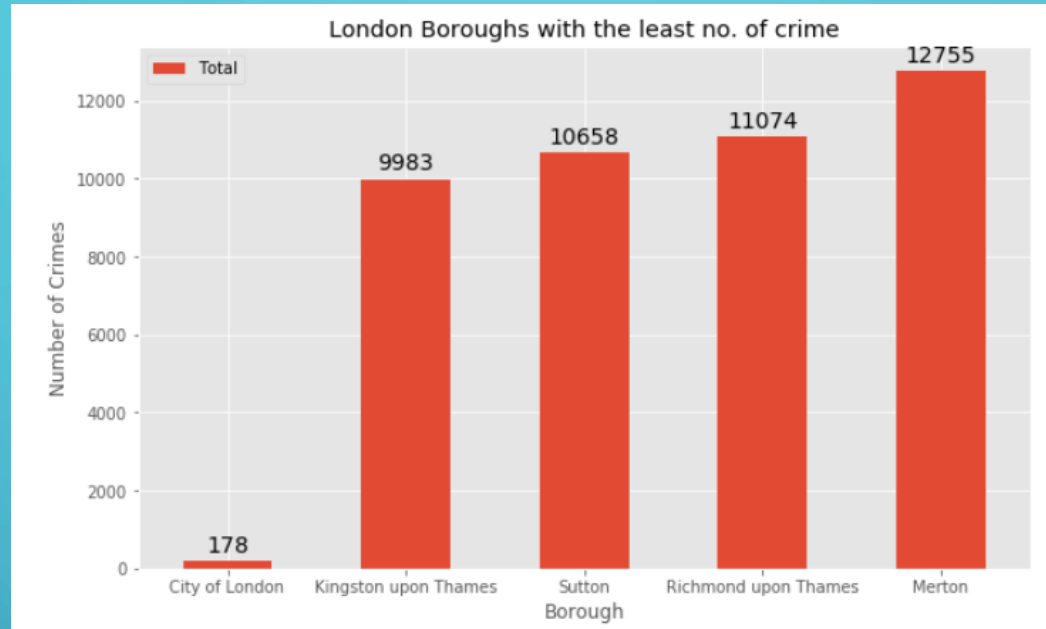
## Boroughs with the highest crime rates



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.



## Boroughs with the lowest crime rates

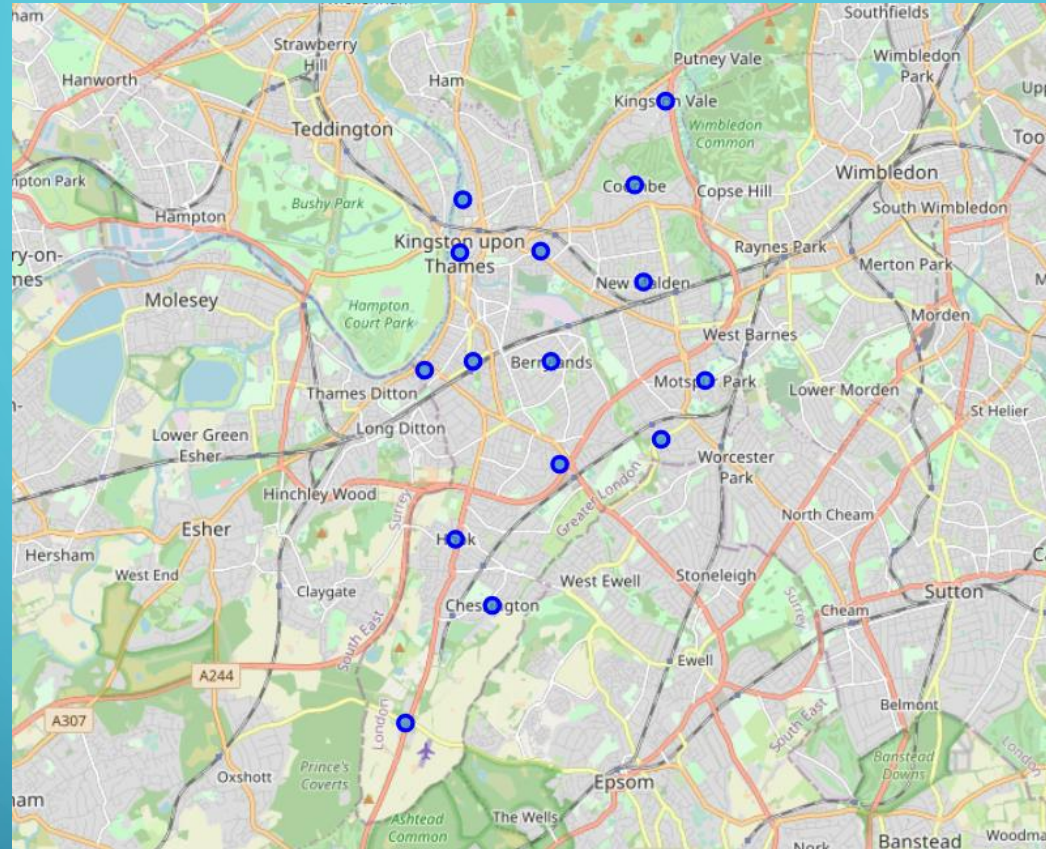


Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.

City of London has a significantly lower crime rate because it is not a London borough. It has an area of 1.12 sq. miles and a population of 7000 as of 2013 which suggests that 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.

## Neighborhood in Kingston upon Thame



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



## Modelling

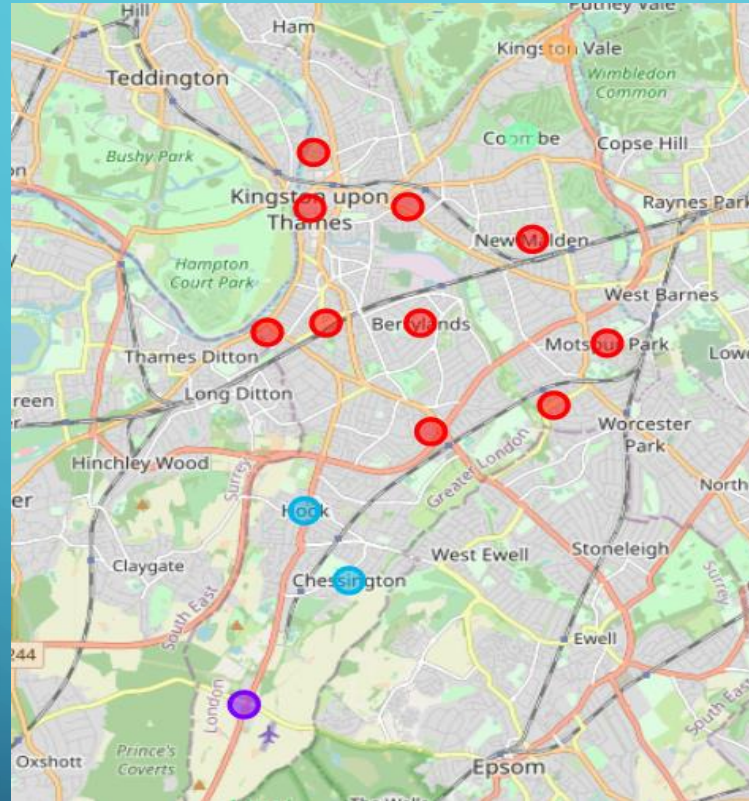
- Using the final data set containing the neighborhoods in Kingston upon Thames along with latitude and longitude, we can find all the venues within a 500 meters radius of each neighborhood by connecting to the Foursquare API.

	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

- 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 of size 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 interest.

## 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 presentation. We can see that majority of neighborhoods falls in the red cluster which is the first cluster.

# CLUSTER 1: NEIGHBORHOOD 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	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	0	Pub	Park	Coffee Shop	Platform	Convenience Store	College Soccer Field	Bus Stop	Gym / Fitness Center	Train Station	Gift Shop
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Coffee Shop	Café	Clothing Store	Hotel	Park	Thai Restaurant	Italian Restaurant	Department Store	Sushi Restaurant
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Pub	Coffee Shop	Café	Clothing Store	Italian Restaurant	Thai Restaurant	Burger Joint	Sushi Restaurant	Park	Department Store
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	0	Park	Bus Stop	Furniture / Home Store	Pub	Grocery Store	Rugby Pitch	Restaurant	Japanese Restaurant	Soccer Field	Mediterranean Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Korean Restaurant	Supermarket	Fast Food Restaurant	Grocery Store	Coffee Shop	Indian Restaurant	Café	Bus Stop	Pub	Department Store
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Pub	Gastropub	Bar	Indian Restaurant	Thai Restaurant	Gym / Fitness Center	Food	Italian Restaurant	Japanese Restaurant	Pizza Place
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Train Station	Park	Steakhouse	Gym / Fitness Center	Grocery Store	Bakery	English Restaurant	Japanese Restaurant	Falafel Restaurant	Fast Food Restaurant
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Pub	Coffee Shop	Indian Restaurant	Platform	Fish & Chips Shop	Italian Restaurant	Gastropub	Bakery	Pharmacy	Grocery Store
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Pub	Coffee Shop	Indian Restaurant	Grocery Store	Park	Italian Restaurant	Hotel	Gym / Fitness Center	Pharmacy	Platform
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Discount Store	Restaurant	Soccer Field	Bowling Alley	Coffee Shop	Climbing Gym	Sandwich Place	Pizza Place

The cluster one is the biggest cluster with 10 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, Supermarkets, Train Station, and stores.

## CLUSTER 2: NEIGHBORHOOD 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	9th Most Common Venue	10th Most Common Venue	
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Theme Park Ride / Attraction	Zoo Exhibit	Pub	Restaurant	Exhibit	Fast Food Restaurant	Fried Chicken Joint	Garden Center	Gift Shop	Grocery Store

The second cluster has one neighborhood which consists of venue like Pub, Restaurant, Theme Parks and Grocery Store.

## CLUSTER 3: NEIGHBORHOOD 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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	2	Train Station	Fast Food Restaurant	Supermarket	Convenience Store	Platform	Golf Course	Breakfast Spot	Zoo Exhibit	Falafel Restaurant	Farmers Market
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Convenience Store	Supermarket	Indian Restaurant	Café	Park	Breakfast Spot	Fish & Chips Shop	Fast Food Restaurant	Bakery	Platform

The third cluster has two neighborhoods which consists of venue like Train Station, Supermarket, Convenience Store and Breakfast Spot.



## CLUSTER 4: NEIGHBORHOOD 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	9th Most Common Venue	10th Most Common Venue	
3	Coombe	Kingston upon Thames	51.41945	-0.265398	3	Hotel	Bus Stop	Rest Area	Sports Club	Stables	Food	Diner	Discount Store	Donut Shop	Electronics Store

The fourth cluster has one neighborhood which consists of venue like Bus Stop, Sports Club, Hotel and Electronic Store.

## CLUSTER 5: NEIGHBORHOOD 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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Stables	Grocery Store	Bar	Coffee Shop	Sandwich Place	Soccer Field	Outdoors & Recreation	Zoo Exhibit	Discount Store	Donut Shop

The fifth cluster has one neighborhood which consists of venue like Grocery Store, Soccer Field and Outdoor & Recreation



## 5. DISCUSSION

- 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 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 3 and 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Bus Stops, Restaurants, Electronics Stores and Grocery Store which is ideal for a family.

## 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.