

The background of the slide is decorated with numerous realistic water droplets of various sizes. Some are large and prominent, while others are small and scattered. They have a glossy, reflective surface with highlights and shadows, giving them a three-dimensional appearance. The droplets are distributed across the slide, with a higher concentration in the top-left and bottom-right corners.

CAPSTONE PROJECT – THE BATTLE OF NEIGHBOURHOODS

CONTENT

1. INTRODUCTION
2. DATA ACQUISITION AND CLEANING
3. METHODOLOGY
4. RESULTS
5. DISCUSSION

1.INTRODUCTION

BACKGROUND: UNDERSTANDING OF NEW AREA BEFORE RELOCATING IS VERY IMPORTANT FOR PERSON WHO PLAN TO RELOCATE, THIS INCLUDES MANY ASPECTS FROM THE NEW LOCALITY, SAFETY IS CONSIDER AS A PRIMARY CRITERIA WHILE RELOCATING.

PROBLEM: THIS PROJECT HELPS TO UNDERSTAND DIFFERENT AREAS WITHIN LONDON AND ALSO HELPS YOU TO UNDERSTAND SAFETY STANDARD BASED ON THE TOTAL CRIMES WITHIN DIFFERENT BOROUGH. THIS ALSO HELPS TO FIND 10 MOST COMMON VENUES IN EACH NEIGHBOURHOODS AND FINALLY CLUSTER THE NEIGHBOURHOODS USING K-MEAN CLUSTER

2. DATA ACQUISITION AND CLEANING

DATA EXTRACTED FROM THREE DIFFERENT SOURCES

1. LONDON CRIME DATA THAT SHOWS CRIME PER BOROUGH IN LONDON
2. WIKIPEDIA DATA CONTAINS LIST OF LONDON BOROUGHES
3. LIST OF NEIGHBOURHOODS IN THE ROYAL BOROUGH OF KINGSTON UPON THAMES AS FOUND ON THE WIKIPEDIA PAGE

Data Cleaning:

- London Crime data, the crimes during the most recent year (2016) are only selected, The major categories of crime are pivoted to get Total crimes per the boroughs for each major category
- Wikipedia page using the beautiful soap library in python.
- Two data sets merged on the Borough names to form a new data set. The purpose of this data set is to visualise the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016
- The Third dataset is created, with the names of the neighbourhoods 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 neighborhoods using Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together

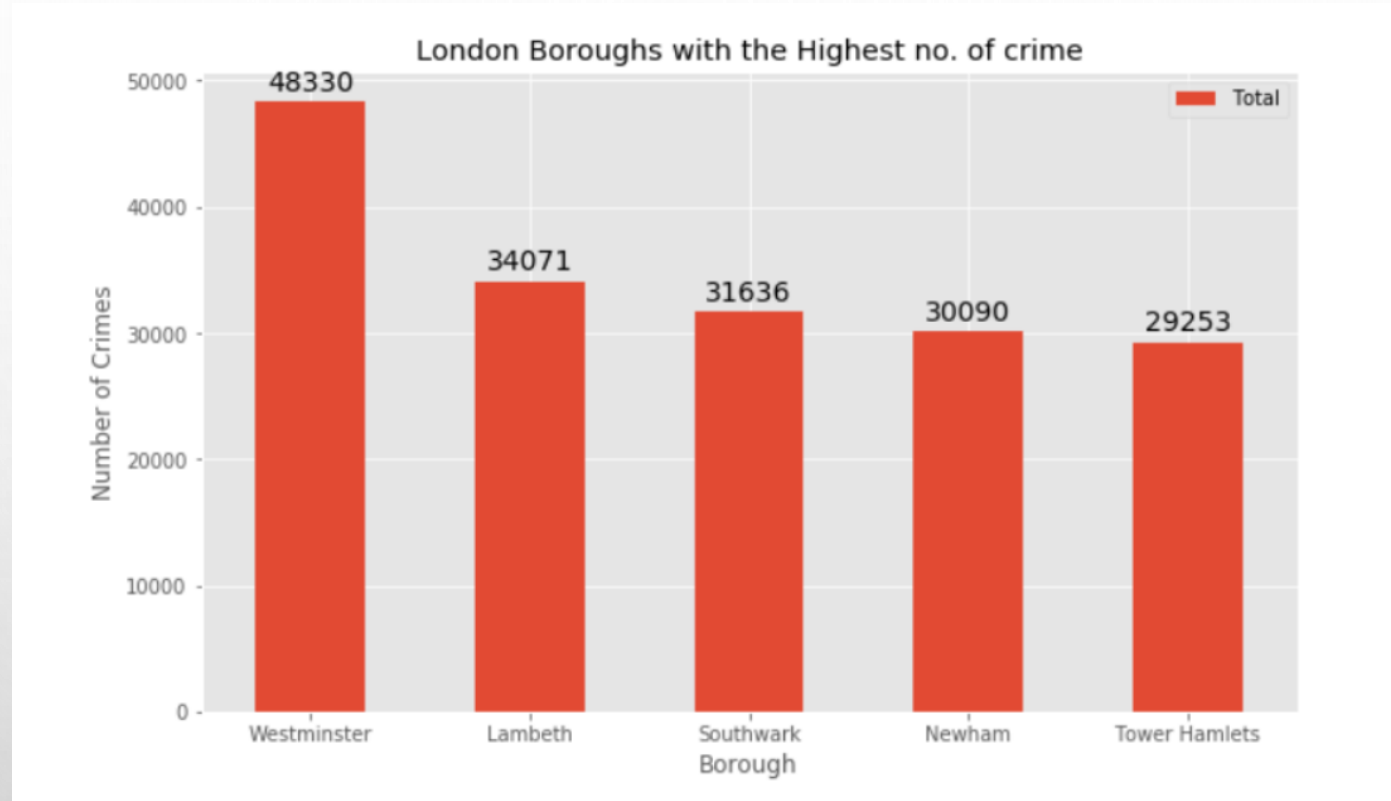
3. METHODOLOGY

Statistical summary of Crimes

	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	<u>2069.242424</u>	<u>1941.545455</u>	<u>1179.212121</u>	479.060606	682.666667	<u>8913.121212</u>	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	<u>4620.565054</u>	<u>2513.601551</u>	<u>8828.228749</u>
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	<u>1531.000000</u>	1650.000000	743.000000	378.000000	377.000000	<u>5919.000000</u>	5936.000000	16903.000000
50%	<u>2071.000000</u>	1989.000000	1063.000000	490.000000	599.000000	<u>8925.000000</u>	7409.000000	22730.000000
75%	<u>2631.000000</u>	<u>2351.000000</u>	<u>1617.000000</u>	551.000000	936.000000	10789.000000	<u>8832.000000</u>	27174.000000
max	3402.000000	3219.000000	<u>2738.000000</u>	<u>1305.000000</u>	<u>1822.000000</u>	27520.000000	10834.000000	48330.000000

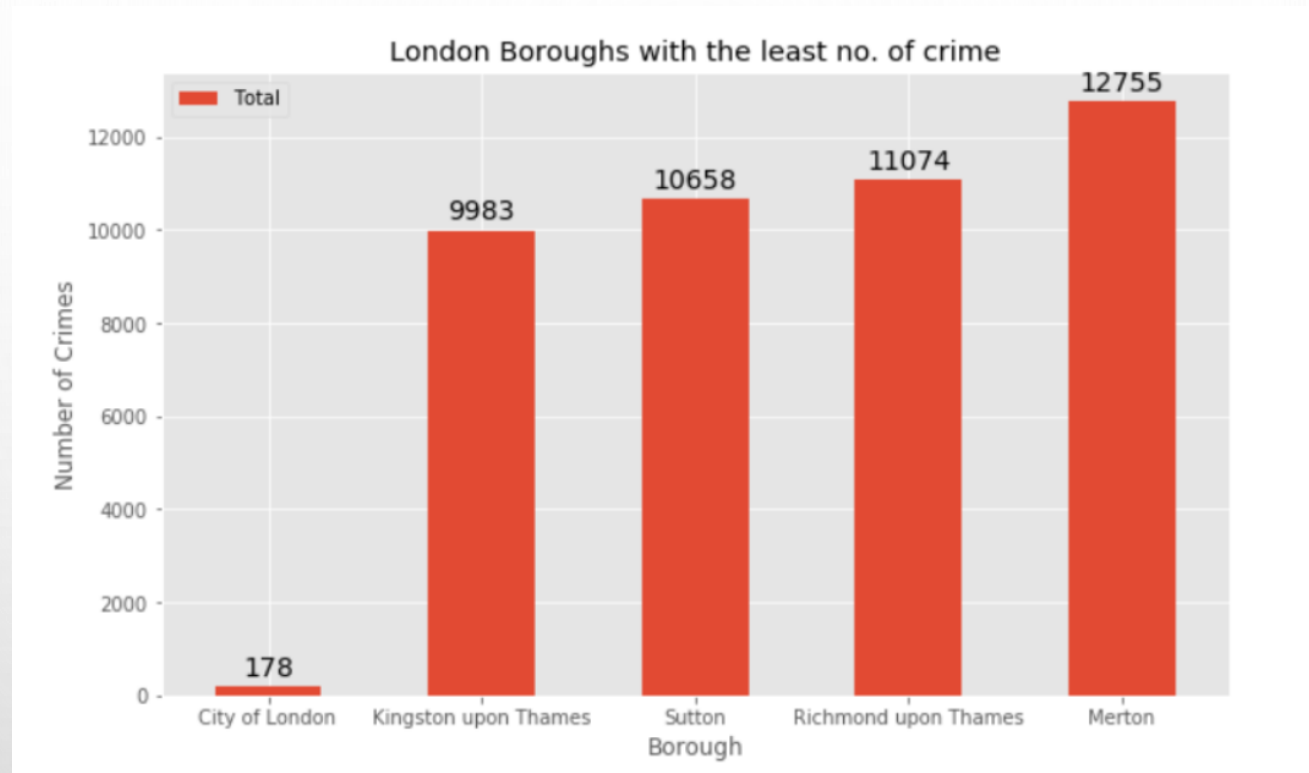
The count for each of the major category of crime returns the value 33 which is the number of London boroughs, Theft and Handling is the highest reported crime during year 2016 followed by ‘violence against person’, ‘Criminal damage’. The lowest recorded crimes are ‘Drugs’, ‘Robbery’ and ‘Other Notifiable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

BOROUGHES WITH 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 other 4 boroughs

BOROUGHS WITH LOWEST CRIME RATES



- City of London has significantly lower crime rate because it is the 33rd principal division of Greater London but it is not a London borough. It has area of 1.12 SQ miles and a population of 7000 as 2013 which suggest 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

Modelling

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

	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 on the venues data, The Venues data is then grouped by the Neighbourhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighbourhoods.
- To help people find similar neighbourhood in the safest borough we will be clustering similar neighbourhoods 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 neighbourhood into 5 clusters. The reason to conduct a k-means clustering is to cluster neighbourhoods with similar venues together so that people can shortlist the area of their interest based on the venues around each neighbourhood

4.RESULTS

The 1st Cluster has one neighbourhood which consists of venues such as train station, restaurants and furniture shop

	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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	0	Train Station	Pub	Food	Wine Shop	French Restaurant	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Fried Chicken Joint

The 2nd Cluster hascluster 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, Supermarkets, and stores.

	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	1	Pub	Hotel	Spa	Plaza	Fish & Chips Shop	Café	Indian Restaurant	Shop & Service	Park	Supermarket
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Bakery	Fish & Chips Shop	Indian Restaurant	Supermarket	Wine Shop	Fried Chicken Joint	Farmers Market	Fast Food Restaurant	Food	French Restaurant
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Pub	Sushi Restaurant	Coffee Shop	Burger Joint	Asian Restaurant	Furniture / Home Store	Turkish Restaurant	French Restaurant	Electronics Store
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Grocery Store	Pub	Restaurant	Garden Center	Food	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gym	Supermarket	Grocery Store	Chinese Restaurant	Korean Restaurant	Gastropub	Indian Restaurant	Sushi Restaurant	Bar	German Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Food	Indian Restaurant	Italian Restaurant	Platform	Pub	Wine Shop	Rental Car Location	Fried Chicken Joint	Hardware Store	Hotel
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Pub	Coffee Shop	Harbor / Marina	Chinese Restaurant	Gym	Gym / Fitness Center	Turkish Restaurant	Hotel	Italian Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	Gym / Fitness Center	French Restaurant	Hotel	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1	Grocery Store	Pharmacy	Coffee Shop	Italian Restaurant	Train Station	Thai Restaurant	Furniture / Home Store	Pizza Place	Soccer Field	Bowling Alley

The 3rd cluster has one neighborhood which consists of Venues such as Grocery Stores, Restaurants, and Store and Markets.

	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	2	Grocery Store	Bar	Sandwich Place	Soccer Field	Gift Shop	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Department Store

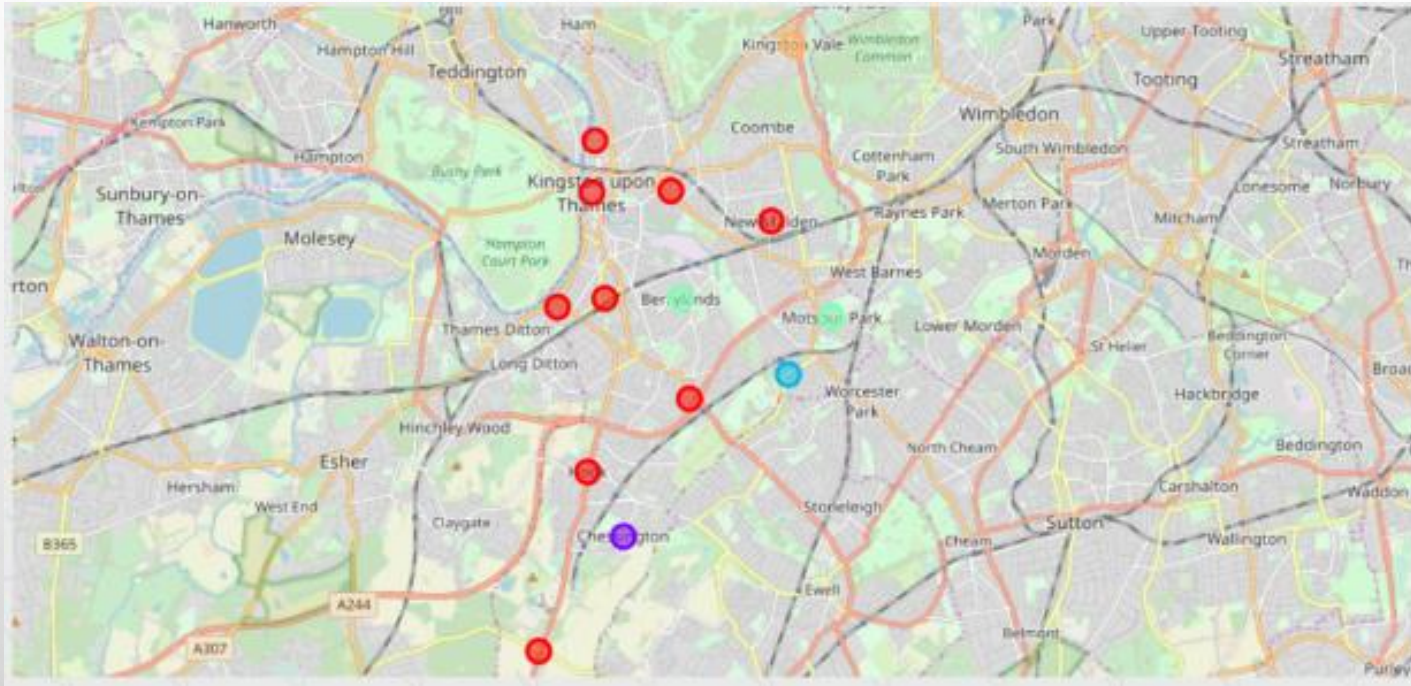
The 4th Cluster cluster has one neighborhoods in it, these neighborhoods have common venues such as Indian Restaurant, Bakery, Supermarket, fish and chip shop food, Electronics Stores and Dry cleaner etc.

	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
4	Hook	Kingston upon Thames	51.367898	-0.307145	3	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Food	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant

The fifth cluster has Two neighborhood , these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

	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	4	Gym / Fitness Center	Park	Bus Stop	Turkish Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fish & Chips Shop
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Park	Gym	Bus Stop	Soccer Field	Restaurant	Turkish Restaurant	Farmers Market	Discount Store	Dry Cleaner	Electronics Store

Visualising the Clustered neighbourhoods on a map using the folium library



Each cluster is colour coded, we can see that majority of the neighbourhood falls in the red cluster which is the second cluster.

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 3 and 4 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 4 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.

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.