



Capstone-Project-The Battle Of Neighborhoods-Presentation

Pavan Srinivas Kambhampati



1.Introduction

Background:Safety is a top concern when moving to anew 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 London based on the total crimes,explore the neighborhoods of that borough to find 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-means 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 communication of data from three sources:

The first data source of the project users a London crime data that shows the crime per borough in London.

The second source of the data is scrapped 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 Neighborhood in the Royal Borough of Kingston upon Thames as found on the wikipedia page.



Data Cleaning: This data cleaning process for each of three sources of data are done separately.

From the London crime data, crimes during most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per 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 data in tabular format as shown website.

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.

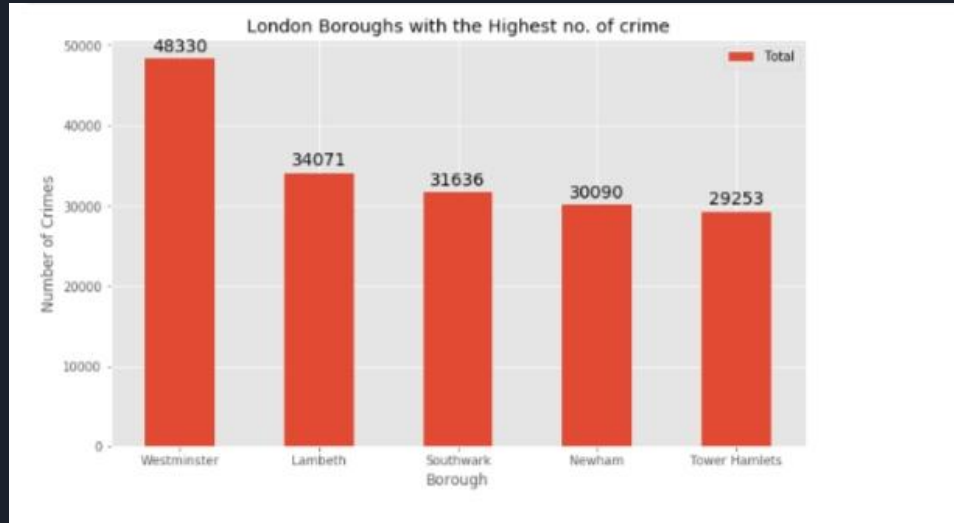
3.Methodology

Exploratory Data Analysis:

Statistical summary of crimes.

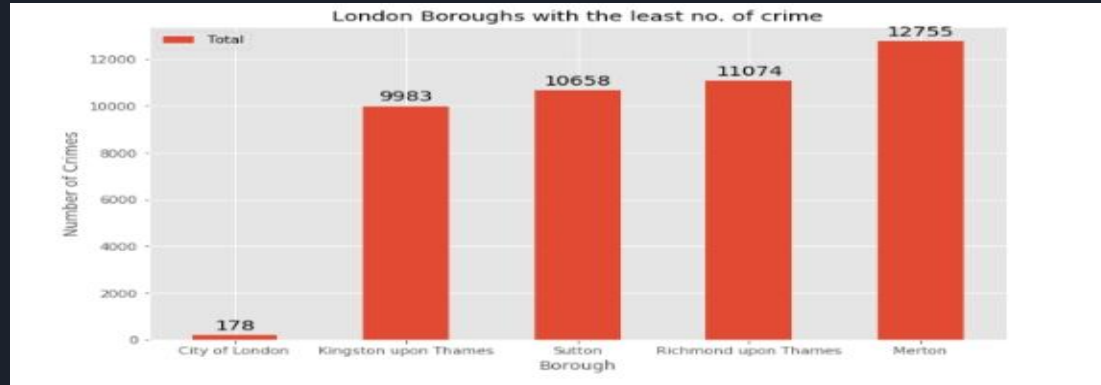
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2089.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448844	625.207070	586.408416	223.298698	441.425368	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	1083.000000	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

Boroughs with highest crime rates.



Comparing five boroughs with 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.

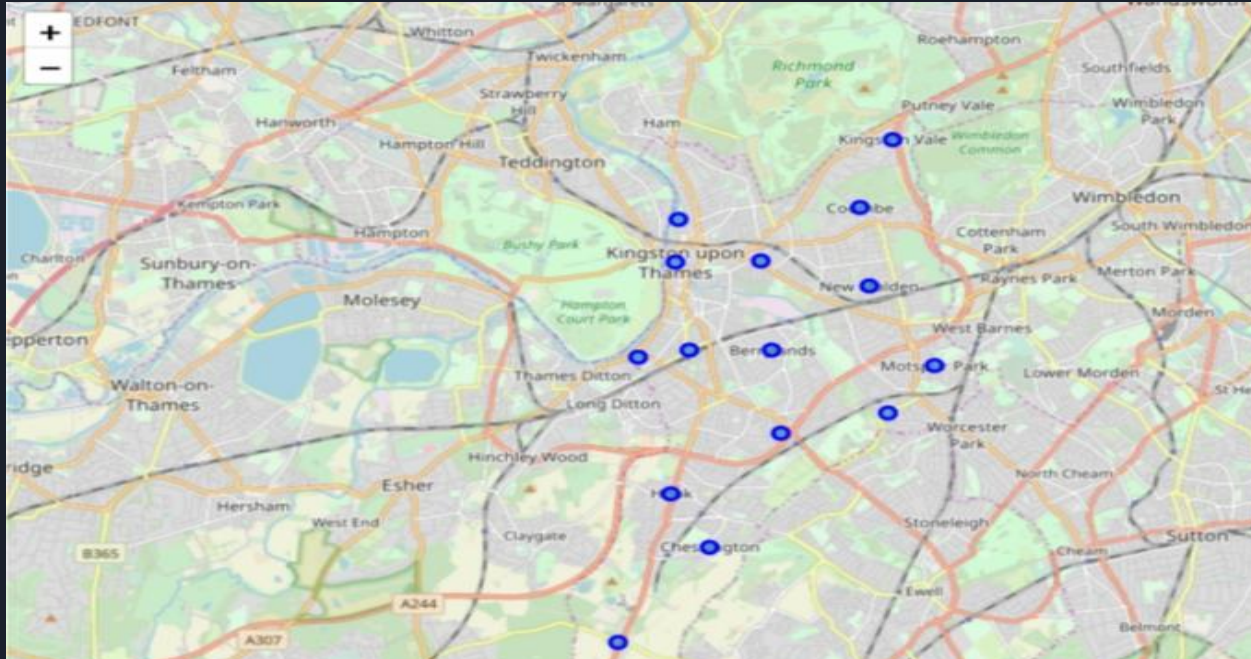
Boroughs with lowest crime rates



Comparing five boroughs with lowest crime rate during the year 2016, City of London which has 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 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 small area.

Neighborhoods in Kingston upon Thames



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

Modelling:

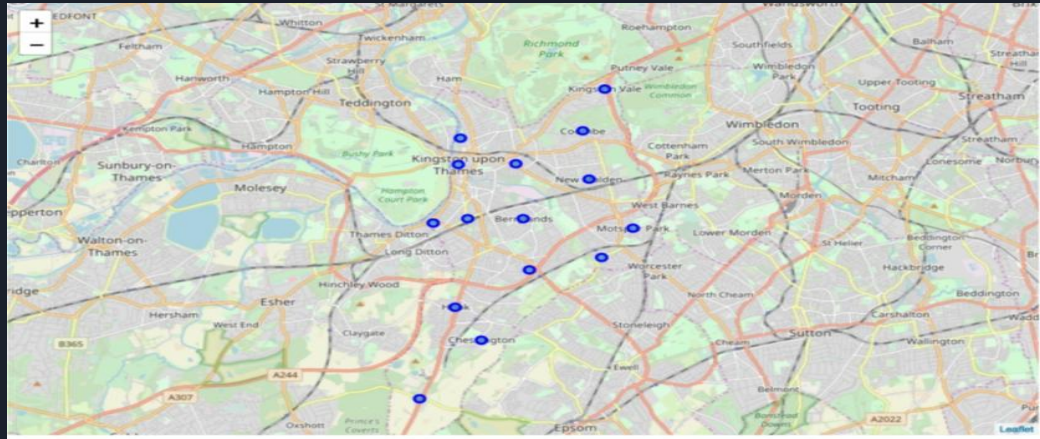
Using final data set containing neighborhoods in Kingston upon Thames along with latitude and longitude, we can find all venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API.

To help people find similar neighborhoods in safest borough we will be clustering similar neighborhoods using K-means clustering which is form of unsupervised learning algorithm. Here we will use size of 5 for this project that will cluster 15 neighborhoods into 5 clusters.

	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	Cafe Rosa	51.390175	-0.282490	Café
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

4.Results

After running K-means clustering we can access each cluster created to see neighborhoods were assigned to each of five clusters. Here each of cluster is color coded for ease of presentation we can see that majority of neighborhood falls in red cluster which is first cluster. 3 neighborhoods have their own cluster (Blue, Purple, Yellow), these are clusters two three and five. Green color has two neighborhoods which is 4th cluster.



Cluster-I: Looking into neighborhoods into 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409527	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.267396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.376676	-0.282660	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

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

Cluster-2: Looking into the neighborhoods in 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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	F

The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses and wine shops.

Cluster-3: Looking into the neighborhoods in the third cluster. As below third cluster has one neighborhood consists of Venues such as Train stations, Restaurants and Furniture shops.

```
In [73]: kut_merged[kut_merged['Cluster Labels'] == 2]
```

```
Out[73]:
```

	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	Cc
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Deli / Bodega	

Clustering-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	9th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Café	Bus Stop	Wine Shop	Fish & Chips Shop	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Park	Gym	Restaurant	Soccer Field	Bus Stop	Wine Shop	Fast Food Restaurant	Dry Cleaner	Electronics Store

The fourth cluster has two neighborhoods in it. these neighborhoods have common venues such Parks, Gym/Fitness centers, Bus stops, Restaurants, Electronics Stores etc


Clustering-5: Looking into the neighborhoods in the fifth cluster

```
kut_merged[kut_merged['Cluster Labels'] == 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Italian Restaurant	Soccer Field	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Department Store

The fifth cluster has one neighborhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops and Department stores.


5. Discussions



The aim of this project is to help people who want to relocate to safest borough in London, expats can choose the neighborhoods to which want to relocated most common venues on it.

For example if a person is looking for a neighborhood with good connectivity and public transportation we can see cluster 3 and 4 have Train stations and Bus stops as most common venues.

If a person is looking for neighborhood with stores and restaurants in a close proximity then the neighborhoods in the first cluster of suitable.



For a family I feel that the neighborhoods in cluster 4 are more suitable for 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 field which is ideal for a family.

The preference of venues may vary from person to person, they can select a neighborhood based on one's 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 ahead i.e. finding out more about places before moving into a neighborhood.

We have just taken safety as a primary concern to shortest 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 shortest the borough,such as filtering areas based on a predefined budget.