CAPSTONE PROJECT – THE BATTLE OF THE NEIGHBORHOODS - PRESENTATION

Jonadab Rajkumar

APRIL 18, 2020

COURSERA

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 values 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 crimes during the recent years(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 Beautiful Soup 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 the k means clustering algorithm to cluster similar neighborhoods together.

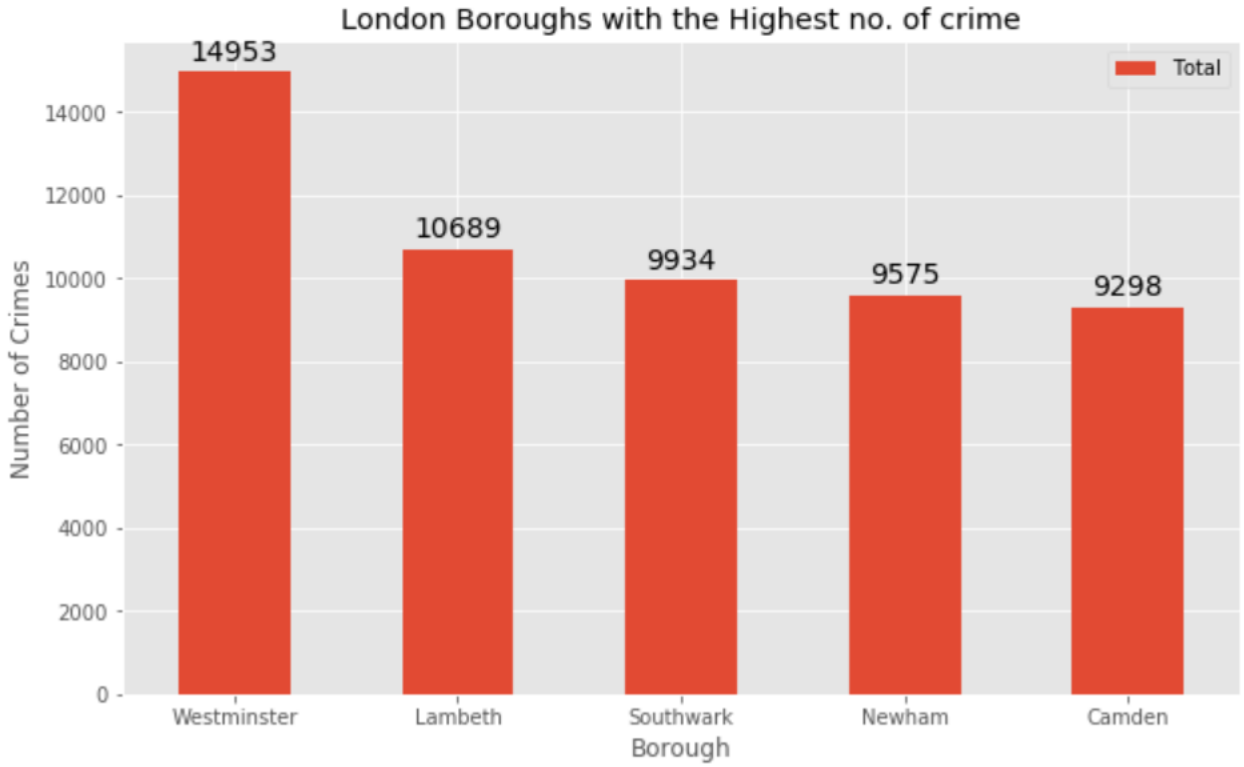
3. Methodology

Exploratory Data Analysis

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.00000	33.00000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	652.212121	616.30303	375.30303	143.484848	210.515152	2786.939394	2200.333333	6985.090909
std	235.347317	195.90136	198.32273	58.680342	133.856995	1432.377262	777.511040	2758.177074
min	0.000000	1.00000	3.00000	0.000000	3.000000	40.000000	10.000000	57.000000
25%	454.000000	510.00000	249.00000	102.000000	117.000000	1877.000000	1830.000000	5428.000000
50%	692.000000	628.00000	363.00000	144.000000	188.000000	2811.000000	2349.000000	7072.000000
75%	824.000000	727.00000	509.00000	171.000000	289.000000	3480.000000	2792.000000	8638.000000
max	1099.000000	1024.00000	968.00000	296.000000	517.000000	8397.000000	3439.000000	14953.000000

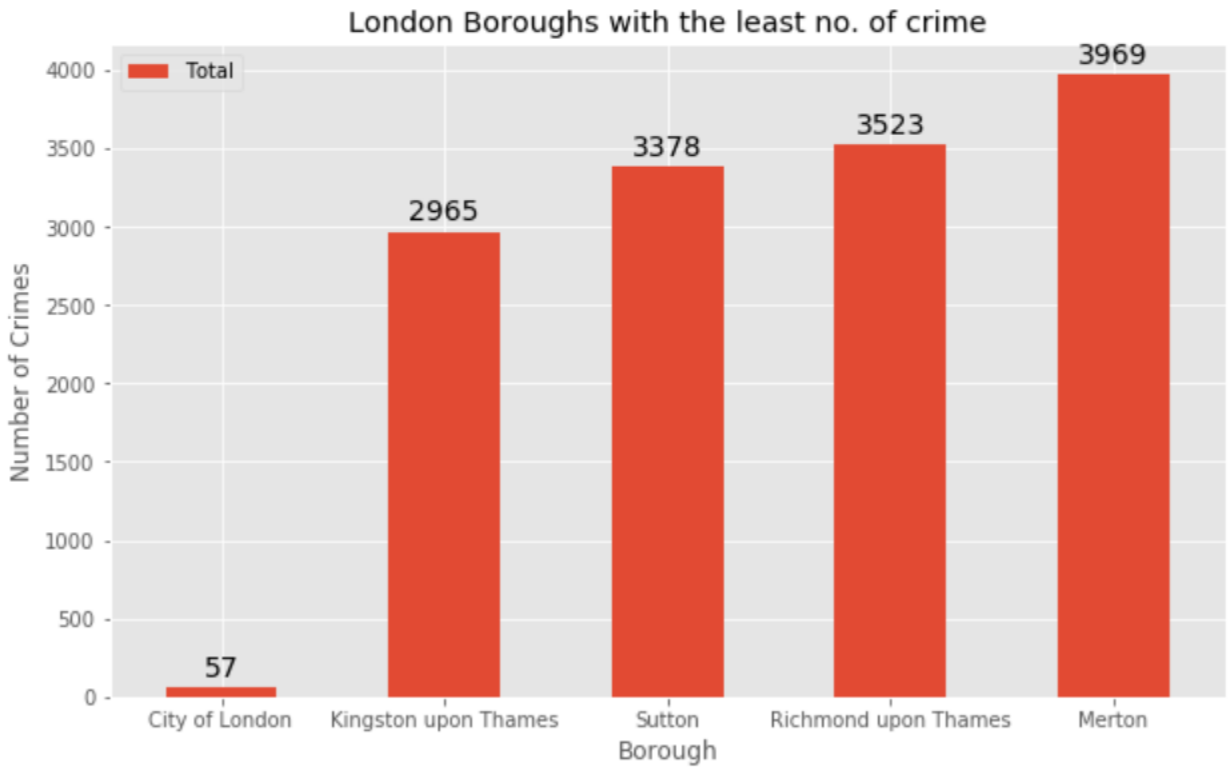
Statistical summary of crimes

The count for each of the major categories of crime returns the 33 which is the number of London boroughs. The ‘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 Notable offenses’



Boroughs with the highest crime rates

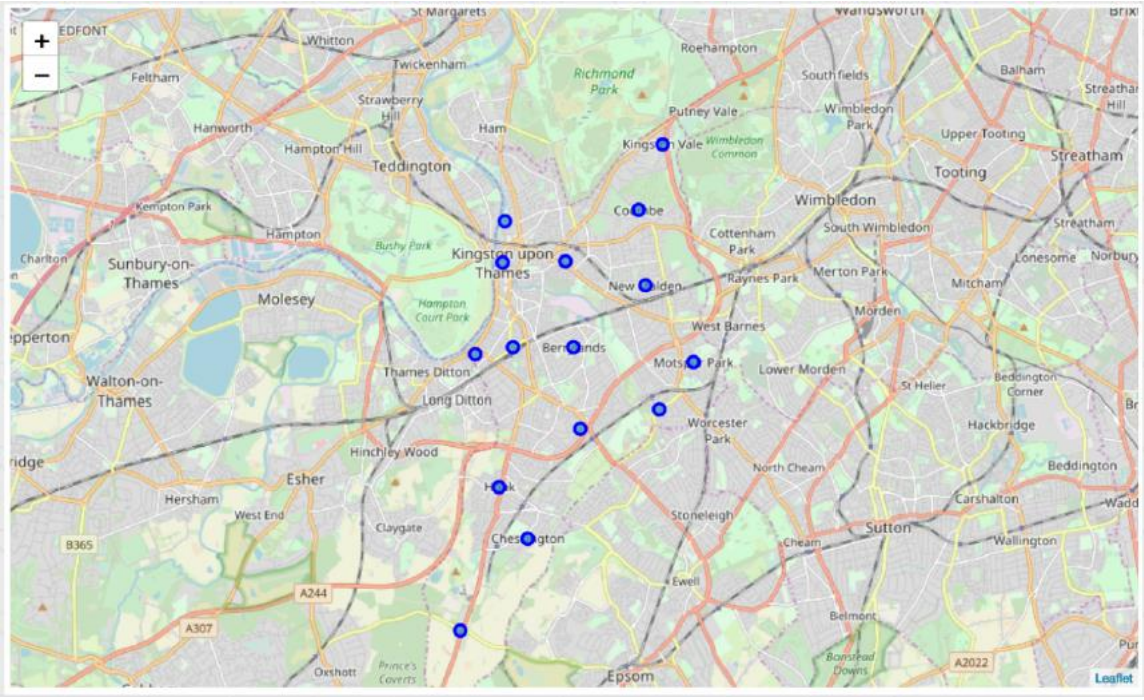
Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.



Boroughs with least 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 sq 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 Kngston 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	Rob Taylor's Ultimate Gay Sauna	51.390022	-0.284894	Sauna / Steam Room
4	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park

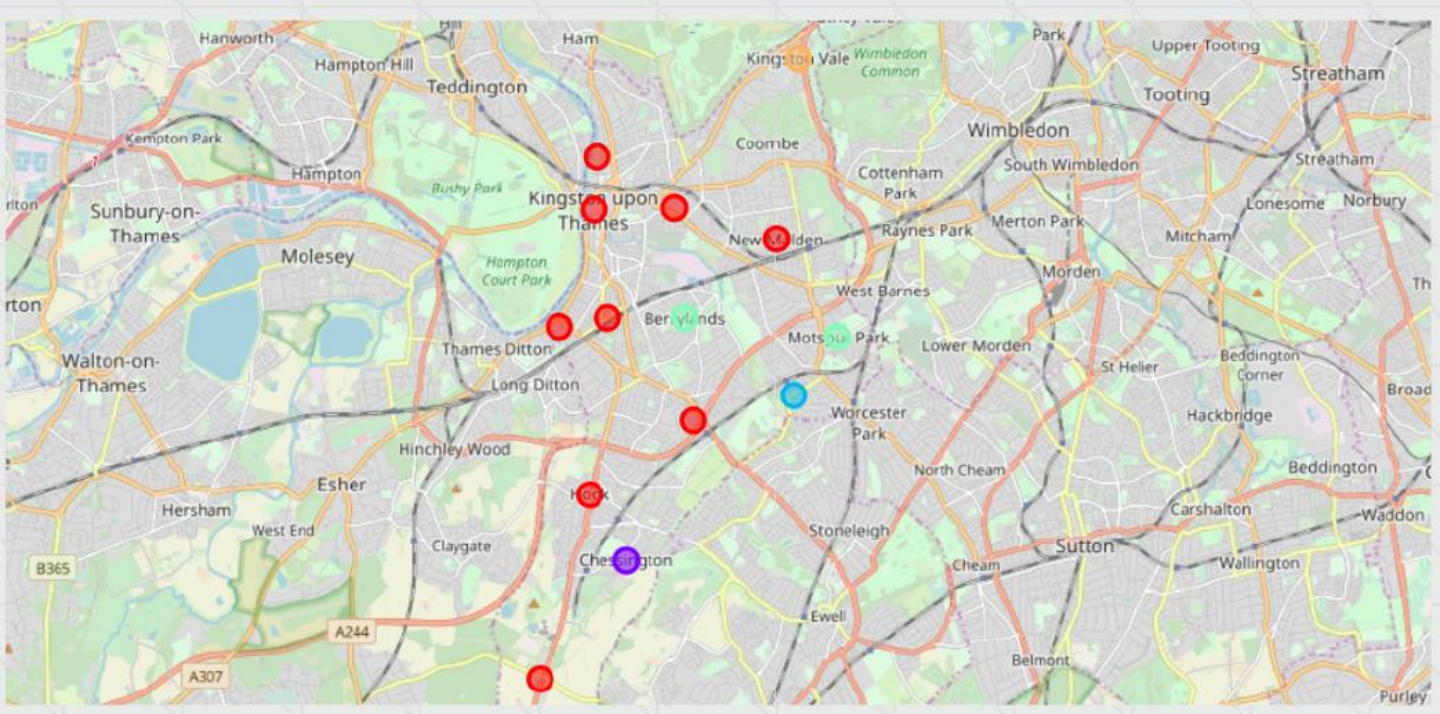
- One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the Neighborhoods.

- 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 neigborhoods 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 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
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	51.409627	-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	-0.287396	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	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

The cluster 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, Café, Supermarkets and stores.

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	9th Most Common Venue
Berrylands	Kingston upon Thames	51.393781	-0.284802	1	Park	Gym / Fitness Center	Bus Stop	Sauna / Steam Room	Wine Shop	Farmers Market	Department Store	Discount Store	Dry Cleaner
Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Gym	Park	Soccer Field	Bus Stop	Wine Shop	Farmers Market	Department Store	Discount Store	Dry Cleaner

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.

ghborhood	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
Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Garden Center	Pub	Restaurant	Convenience Store	Wine Shop	Electronics Store	Deli / Bodega	Department Store	Discount Store

The third cluster has one neighborhood which consists of Venues such as Train Stations, Restaurants and Furniture shops.

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	9th Most Common Venue
Coombe	Kingston upon Thames	51.41945	-0.265398	3	Tea Room	Wine Shop	Fast Food Restaurant	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market

The fourth cluster has two neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields 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	9th Most Common Venue
Kingston Vale	Kingston upon Thames	51.43185	-0.258138	4	Grocery Store	Bar	Soccer Field	Sandwich Place	Wine Shop	Fast Food Restaurant	Department Store	Discount Store	Dry Cleaner

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

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 neighborhoods in the 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.
- 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 neighborhoods.
- 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.