Capstone Project: Battle of Neighbourhoods

1. Introduction

Modernization is taking its toll on the world today. The increase in the number of facilities, infrastructure etc. in a region has led in the increase of population in that area. People keep on moving to areas where these facilities are found in abundance. Also it depends on what the people like. Sometimes they prefer to be near a good school, some prefer to be near train stations, bus stops and so on. So the moving is based on things which make people happy and necessary.

And people do a lot of research on the places so they don't waste time and money. Safety is a major factor in deciding this process.

Problem

The aim is to select the top ten places in each neighbourhood with low crime rates for a particular country. Here we use the kaggle data set for crime rates in London Boroughs from 2008-2016. We use K-means Clustering for this purpose.

2. Data Acquisition and Cleaning

The data acquired for this project is from three sources:

The first is a dataset of London crimes which contains the following columns:

- Isoa_code: Area code in London
- Borough: Name of London borough
- Major category: High level crimes
- Minor category: Low level crimes

- Value: Monthly report of crimes
- Year: Year between 2008-16
- Month: Month between 1-12

The second is data scraped from a Wikipedia page that contains London Boroughs.

- Boroughs-Names of London Boroughs
- Inner-Checking if it is inner or outer London.
- Status-Checking if it is royal, city etc.
- Local Authority-Local Authority assigned to the Borough
- Political Control-Political party that controls the Borough
- Headquarters-Headquarters of the Borough
- Area-Area of the Borough in miles.
- Co-ordinates-Latitude and Longitude of the Boroughs

The third is also from a Wikipedia page which is created from scratch. It is the list of neighbourhoods in Royal Borough of Kingston.

- Neighbourhood-Name of the Neighbourhood in the Borough
- Borough-Name of the Borough
- Latitude-Latitude of the Borough
- Longitude-Longitude of the Borough

Data Cleaning:

For the London Crimes Dataset the crimes during the most recent year (2016) are only selected.

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741
1	Barnet	3402	2183	906	499	464	9731	7499	24684
2	Bexley	1123	1673	646	294	209	4392	4503	12840
3	Brent	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	2214	2202	728	417	369	7584	6650	20164

For the second dataset the web scraping is done using Beautiful Soup library. The table is obtained and after string manipulation the following table is created.

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]	Co-ordinates	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E	1287	1949	919	378	534	5607	6067	16741
1	Barnet	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W	3402	2183	906	499	464	9731	7499	24684
2	Bexley	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E	1123	1673	646	294	209	4392	4503	12840
3	Brent	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E	2214	2202	728	417	369	7584	6650	20164

The third source of data is acquired from the list of neighbourhoods in the safest borough on Wikipedia. This dataset is created from scratch; the panda's data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude.

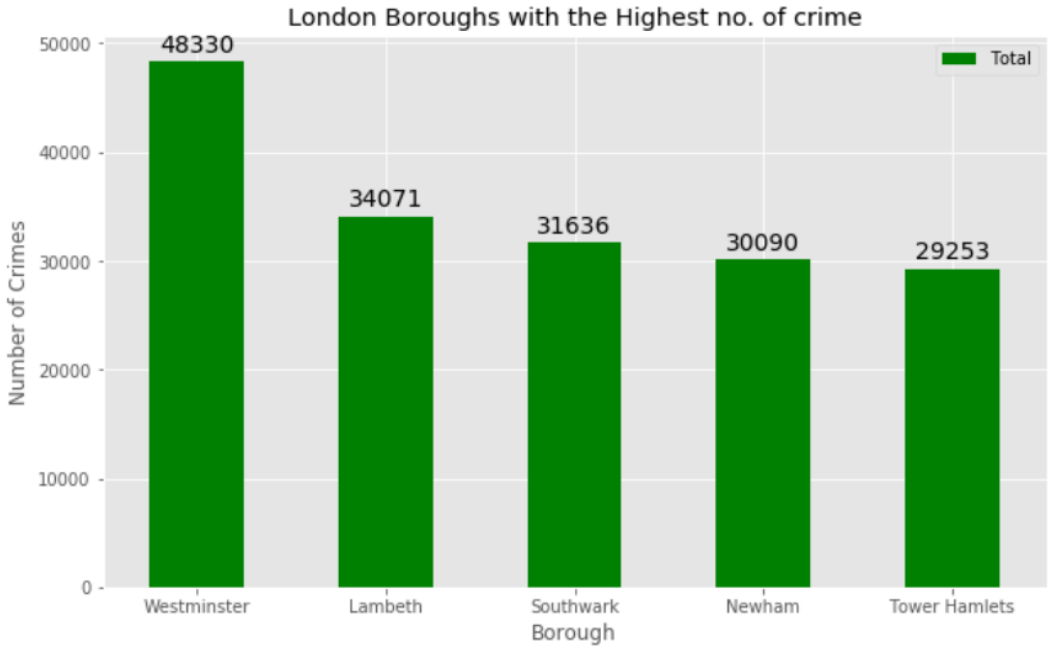
	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames	51.393781	-0.284802
1	Canbury	Kingston upon Thames	51.417499	-0.305553
2	Chessington	Kingston upon Thames	51.358336	-0.298622
3	Coombe	Kingston upon Thames	51.419450	-0.265398
4	Hook	Kingston upon Thames	51.367898	-0.307145
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898
9	New Malden	Kingston upon Thames	51.405335	-0.263407
10	Norbiton	Kingston upon Thames	51.409999	-0.287396
11	Old Malden	Kingston upon Thames	51.382484	-0.259090
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366
13	Surbiton	Kingston upon Thames	51.393756	-0.303310
14	Tolworth	Kingston upon Thames	51.378876	-0.282860

Methodology

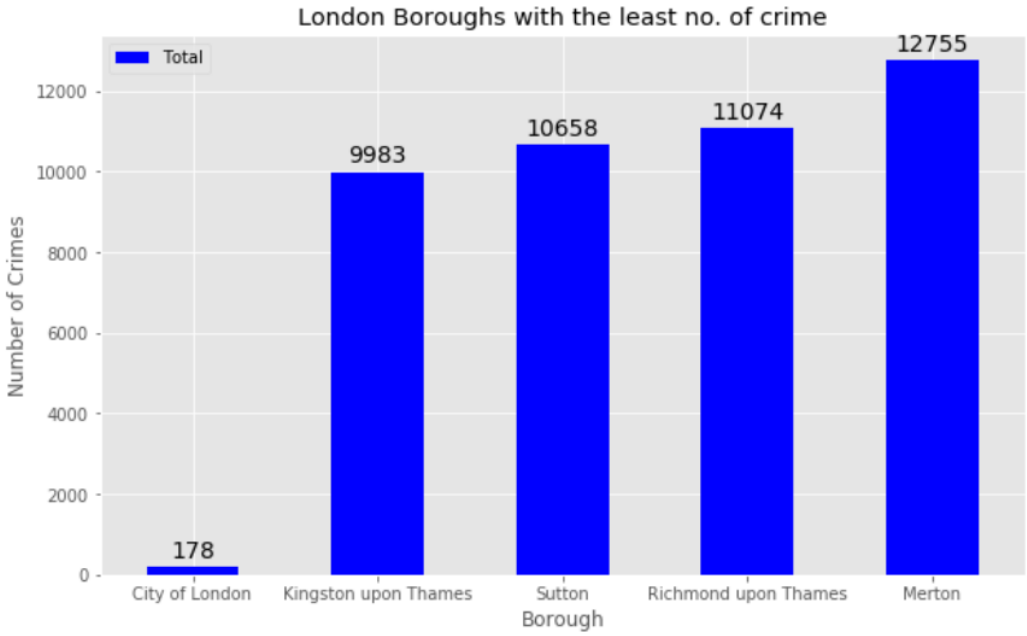
EDA

The describe function in python is used to get statistics of the London crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime.

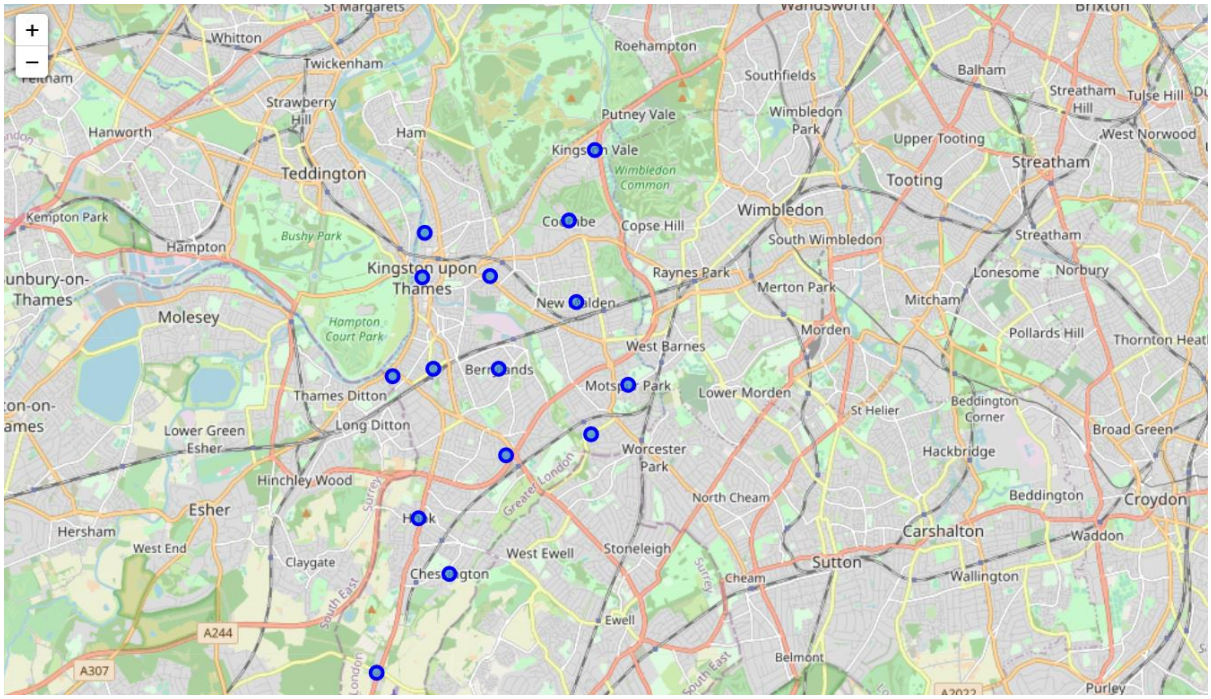
Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.



Also, 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.



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



Results

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five clusters. Looking into the neighbourhoods 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	
2	Chessington	Kingston upon Thames	51.358336	-0.298622	0	Park	Wine Shop	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Fast Food Restaurant	C L

The first cluster consists of only one neighbourhood which consists of fast food restaurant, department stores 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	
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Café	Shop & Service	Plaza	Indian Restaurant	Hotel	Park	Gym / Fitness Center	
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Indian Restaurant	Pub	Bakery	Supermarket	Fish & Chips Shop	Wine Shop	Cosmetics Shop	Deli / Bodega	De
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Pub	Sushi Restaurant	Burger Joint	Coffee Shop	Gift Shop	Department Store	Furniture / Home Store	Re
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gastropub	Korean Restaurant	Sushi Restaurant	Supermarket	Bar	Indian Restaurant	Wine Shop	Electronics Store	
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Indian Restaurant	Food	Italian Restaurant	Pub	Breakfast Spot	Fried Chicken Joint	Dry Cleaner	Grocery Store	Con
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Park	Café	Pet Café	Fish & Chips Shop	Fast Food Restaurant	Gol
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Bistro	Gym / Fitness Center	Platform	

The second cluster consists of seven neighborhoods.These neighbourhoods consist of many places like cafes, pubs, electronics store 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 Com V
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Cosmetics Shop	Deli / Bodega	Department Store	Cl
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Garden Center	Grocery Store	Pub	Restaurant	Electronics Store	Convenience Store	Cosmetics Shop	Deli / Bodega	Depar
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Park	Gym	Bus Stop	Soccer Field	Wine Shop	Electronics Store	Cosmetics Shop	Deli / Bodega	Depar
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Pharmacy	Restaurant	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Coffee Shop	Pizza

The third cluster consists of four neighbourhoods consisting of gyms, hotels 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	10(Cc
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Wine Shop	Gym / Fitness Center	Café	Bus Stop	Farmers Market	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Fa Res

The fourth cluster consists of one neighbourhood.

	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	Old Malden	Kingston upon Thames	51.382484	-0.25909	4	Train Station	Pub	Food	Construction & Landscaping	Farmers Market	Cosmetics Shop	Deli / Bodega	Department Store	Dry Cleaner

The fifth cluster also consists of one neighbourhood.

Conclusion

So we can find out the most suitable place to live in based on these clusters. Each cluster consists of neighbourhoods, which consist of different venues. So people can choose the locations they prefer. With the advancement of technology it would be useful for people to use it in situations which are important to them in terms of money and time.