Capstone Project
The Battle of Neighborhoods
Sachin Krishan Khanna

Introduction

This is capstone project for Applied Data Science in which we will consider the data set of London and its neighbours, as many people move to new cities but doesn't know where to stay as they don't know which neighboring city is safest for them.

Background:
As I've studies an average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down or do our needs change over time, we eventually leaving the town. What brings us satisfaction in the new area we are going to? Or do we often move to new area without knowing anything about that; leads to discomfort and unhappiness.

Problem:
The crime stats in this dataset of London found on Kaggle has crimes in each Boroughs of London from 2008-2016. The 2016 being the latest one, we will be considering the data of the year which is actually old information as of now. The crime rates in each borough may have changed over time. This project is aimed to select the safest borough in London based on the total crimes, explore the neighborhoods of the borough to find the most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.
People who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighbourhood and common venues around each other.

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 dataset has 7 columns :
Isoa_code, Borough, Major_category, Minor_catergory, Value, Year, Month

The second source of data is carped from a Wikipedia page that contains the list of London boroughs. This page contains additional information about the boroughs, the dataset contains 10 columns: Borough, Inner, Status, Local Authority, Political Control, Headquarters

Area Population:

Then we have third data source which is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as well was their on Wikipedia page. The dataset contains 4 columns: Neighborhood, Borough, Latitude, Longitude

Data Cleaning:

The data is prepared for each of the three sources of data is also done separately as to maintain no confusion. From London crime data, the crimes done in year 2016 were taken in consideration. The major categories of crime are here pivoted to get the total as per all the boroughs for each category.

	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

The second data was extracted from the Wikipedia using beautiful soup library in python. Using this library we can extract the data in the tabular format. After web scraping, string manipulation is required to get the names of boroughs in the correct. Then merge the two datasets according to Boroughs names.

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Populati (2013 e
0	Barking and Dagenham	1287	1949	919	378	534	5607	6067	16741	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	1943
1	Barnet	3402	2183	906	499	464	9731	7499	24684	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	3690
2	Bexley	1123	1673	646	294	209	4392	4503	12840	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	2366
3	Brent	2631	2280	2096	536	919	9026	9205	26693	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	3172
4	Bromley	2214	2202	728	417	369	7584	6650	20164	NaN	NaN	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	3178

Now the two datasets are merged on the borough names as new dataset is formed that combines the necessary information in one dataset. The purpose of this merger 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 lowest crime rate and hence the borough is the safest borough. The third source of data is acquired from the list of neighborhoods is the safest. The dataset is created from scratch with pandas data frame and the latitude and longitude of their corresponding neighborhoods. These co-ordinates are generated through Google API geocoding.

	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

Exploratory Data Analysis

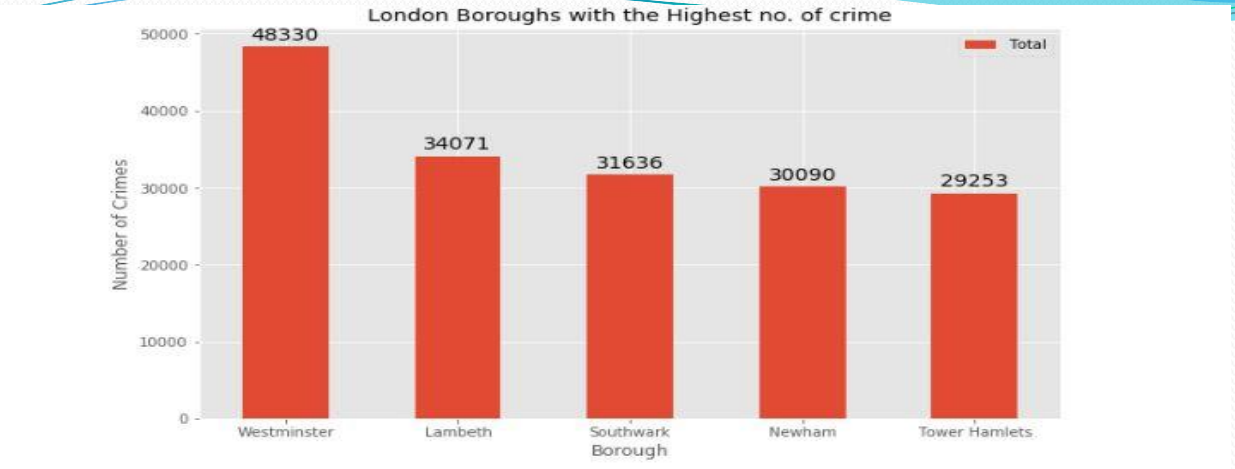
Statistical summary of crimes

The describe function in python is used to get statistics of the London crime data which returns the mean, std. deviation, minimum, maximum for each of the major categories of crime.

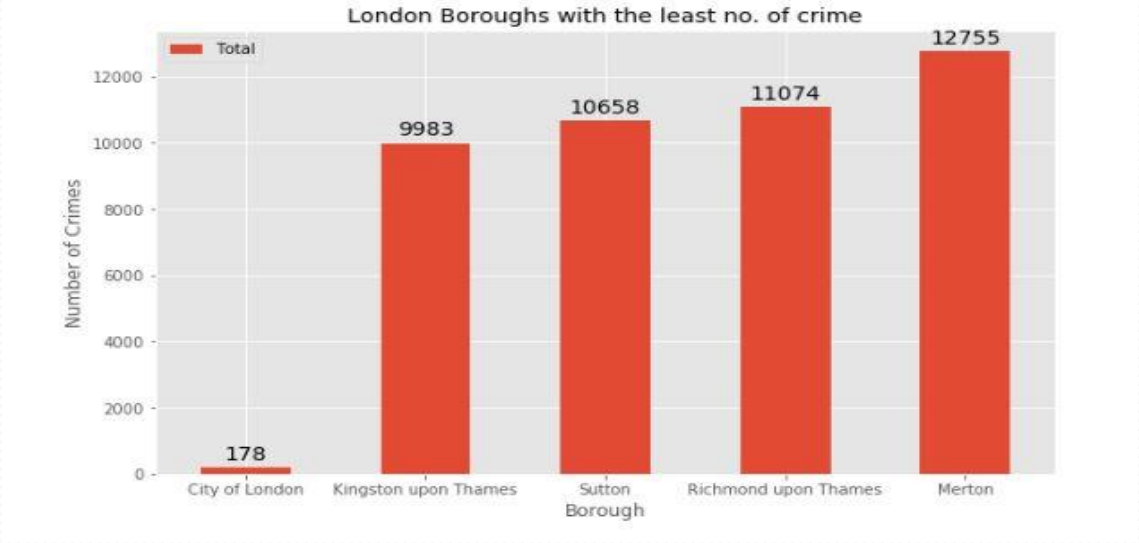
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	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	1063.000000	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

The count for each of the major categories of crime returns the value which is the number of London boroughs. All these crimes were done in year 2016. The lowest recorded crimes are Drugs, Robbery and Other Notifiable Offenses.

By comparing boroughs with the highest crime rate during 2016, it has shown us the Westminster has the highest crimes recorded as followed by Lambeth, Southwark, Newham and Tower Hamlets.

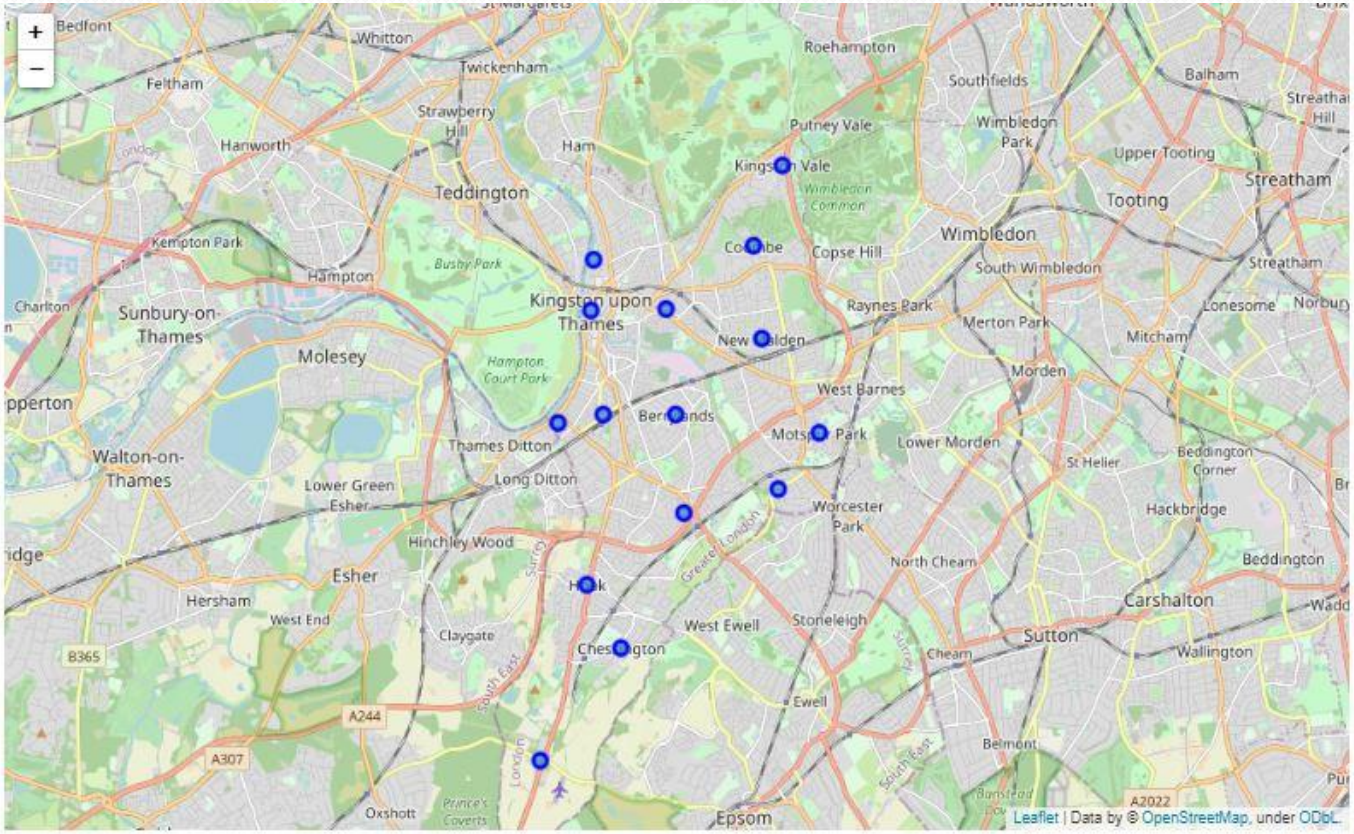


Now after comparing with all boroughs we can record lowest crimes in the City of London followed by Kingston upon Thames, Sutton, Richmond and Merton



As we can see the City of London has lowest crime rate but it is on 33rd division of Greater London not on London. Hence we will consider the next up to it to Kingston Upon Thames.

There are some neighborhoods in the royal borough of Kingston Upon Thames, they are visualized on a map using folium.



Modeling

Now we have a final dataset containing the neighbourhoods in Kingston Upon Thames along with latitude and longitude, now we can find venues in the 500 m radii using Foursqaure API. From this we can have a data frame with all the names corresponding to their co-ordinates.

	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

I’ve used one hot coding so that variables are converted in the form so that ML algorithm can use it easily can do more accurate prediction. Then these venues are grouped by the neighborhoods. This will help people to locate the safest borough and find similar neighborhoods using K-Means Clustering Algorithm. The main reason of using this algorithm is that people can locate the venues they want to visit or they need to visit or they want near their home.

Results

After doing Clustering we can access each one cluster to see their respective neighborhood venues.
Now I'll share 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	1(C
3	Coombe	Kingston upon Thames	51.41945	-0.265398	0	Health & Beauty Service	Wine Shop	Food	Deli / Bodega	Department Store	Electronics Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Re

This cluster has one neighborhood which has venues like Wine Shop, Departemental Store & Restaurants

Now we will see 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Café
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Food	Deli / Bodega	Department Store	Electronics Store
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	Asian Restaurant	Clothing Store	Market
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	1	Gym	Soccer Field	Park	Construction & Landscaping	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Food
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gym	Supermarket	Indian Restaurant	Korean Restaurant	Chinese Restaurant	Bar	Gastropub	Sushi Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Pub	Food	Indian Restaurant	Italian Restaurant	Platform	Wine Shop	Farmers Market	Hardware Store
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Café	Golf Course	Gym / Fitness Center	Harbor / Marina	Fish & Chips Shop
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Pharmacy	Italian Restaurant	Grocery Store	French Restaurant	Train Station	Gastropub

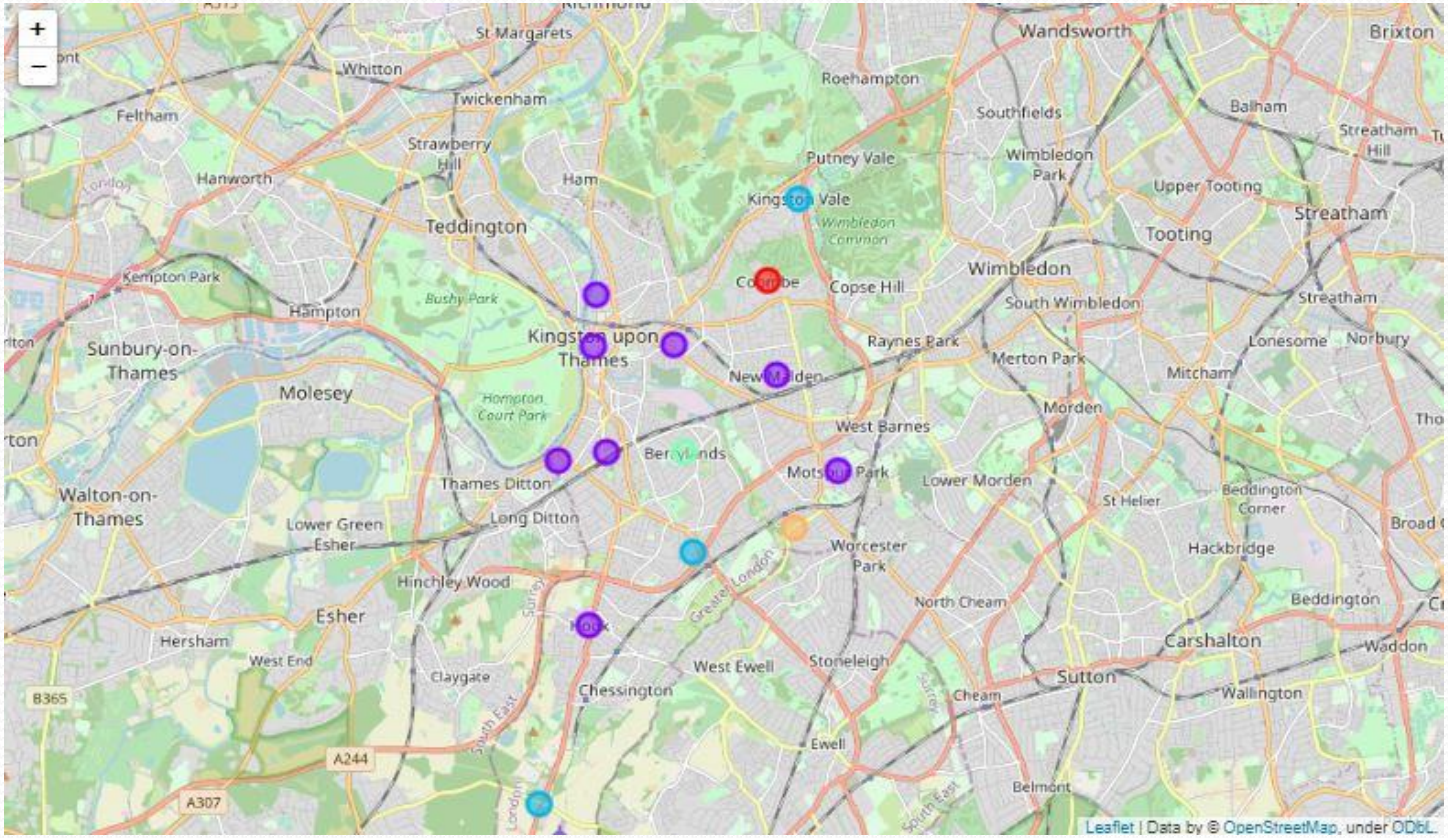
This cluster being the biggest cluster our of all neighbourhoods has some common places to visit as: Pub, Cafes, Departmental Store and Restaurants.

This cluster being the biggest cluster our of all neighbourhoods has some common places to visit as: Pub, Cafes, Departmental Store and Restaurants.
Now I will share the third cluster here and then the visualization of the neighborhoods in a map.

	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 M Comm Ver
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Grocery Store	Sandwich Place	Bar	Soccer Field	Deli / Bodega	Department Store	Electronics Store	Farmers Market	Fast Fc Restaur
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Garden Center	Pub	Restaurant	Fast Food Restaurant	Cosmetics Shop	Deli / Bodega	Department Store	Electror St
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Restaurant	Pizza Place	Sandwich Place	Train Station	Hotel	Coffee Shop	Pharmacy	Bus S

This cluster has 3 neighbourhoods which has some common venues to visit as : Grocery Store, Restaurants & Sandwich Place

Visualization on map



Discussion and Conclusion

Discussion

The aim of this project was to see the safest place to relocate so if a person is going to relocate they can see the safest place and can also get some venues they like to go to near their house, People who are moving out with their family they not on;t need the safest play to live but also there are some needs to be fulfilled, so they can see the clusters and according to the places or venues they like to go or they need to go they can set their place to relocate.

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. I 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.

Thank you