Capstone Project - Report Battle of the Neighbourhoods

Sri Harsha MV
09 Jun 2020

1. Introduction

1.1 Background

The 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 where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort?

To minimise the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move you'll end up regretting. 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.

1.2 Problem

The crime statistics dataset of London found on Kaggle has crimes in each Boroughs of London from 2008 to 2016. The year 2016 being the latest we will be considering the data of that year which is actually old information as of now. The crime rates in each borough may have changed over time.

This project aims to select the safest borough in London based on the total crimes, explore the neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering.

1.3 Interest

Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighbourhoods and common venues around each neighbourhood.

2. Data Acquisition and Cleaning

2.1 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 contains the following columns:
- lsoa_code : code for Lower Super Output Area in Greater London.
- borough : Common name for London borough.
- major_category : High level categorisation of crime
- minor_category : Low level categorisation of crime within major category.
- value : monthly reported count of categorical crime in given borough year : Year of reported counts, 2008-2016

- month : Month of reported counts, 1-12 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 following are the columns:
- Borough : The names of the 33 London boroughs.
- Inner : Categorising the borough as an Inner London borough or an Outer London Borough.
- Status : Categorising the borough as Royal, City or other borough.
- Local authority : The local authority assigned to the borough.
- Political control : The political party that control the borough.
- Headquarters: Headquarters of the Boroughs.
- Area (sq mi) : Area of the borough in square miles.
- Population (2013 est.)[1] : The population in the borough recorded during the year 2013.
- Co-ordinates : The latitude and longitude of the boroughs.
- Nr. in map : The number assigned to each borough to represent visually on a map. The third data source is the list of Neighbourhoods in the Royal Borough of Kingston upon Thames as found on a wikipedia page. This dataset is created from scratch using the list of neighbourhood available on the site, the following are columns:
- Neighbourhood: Name of the neighbourhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

2.2 Data Cleaning

The data preparation for each of the three sources of data is done separately. From the London crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per the boroughs for each major category (see fig 2.1).

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

Fig 2.1 London crime data after preprocessing

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. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form (see fig 2.1). This is important because we will be merging the two datasets together using the Borough names.

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
0	Barking and Dagenham [note 1]	NaN	NaN	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	25
1	Barnet	NaN	NaN	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	31
2	Bexley	NaN	NaN	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	23
3	Brent	NaN	NaN	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	12
4	Bromley	NaN	NaN	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	20

Fig 2.2 List of London Boroughs

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset (see fig 2.3). The purpose of this dataset is to visualise the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

	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

Fig 2.3 London Borough Crime

After visualising the crime in each borough we can find the borough with the lowest crime rate and hence tag that borough as the safest borough. 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 pandas data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude left blank.

	Neighborhood	Borough	Latitude	Longitude
0	Berrylands	Kingston upon Thames		
1	Canbury	Kingston upon Thames		
2	Chessington	Kingston upon Thames		
3	Coombe	Kingston upon Thames		
4	Hook	Kingston upon Thames		

Fig 2.4 Neighborhoods of the safest borough

3.Analysis & Approach

3.1 Exploratory Data Analysis

3.1.1 Statistical summary of crimes

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.

3.1.2 Boroughs with the lowest 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 (see fig 3.1).

City of London has a significantly lower crime rate because it i is the 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 a small area (see fig 3.4). Hence we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.
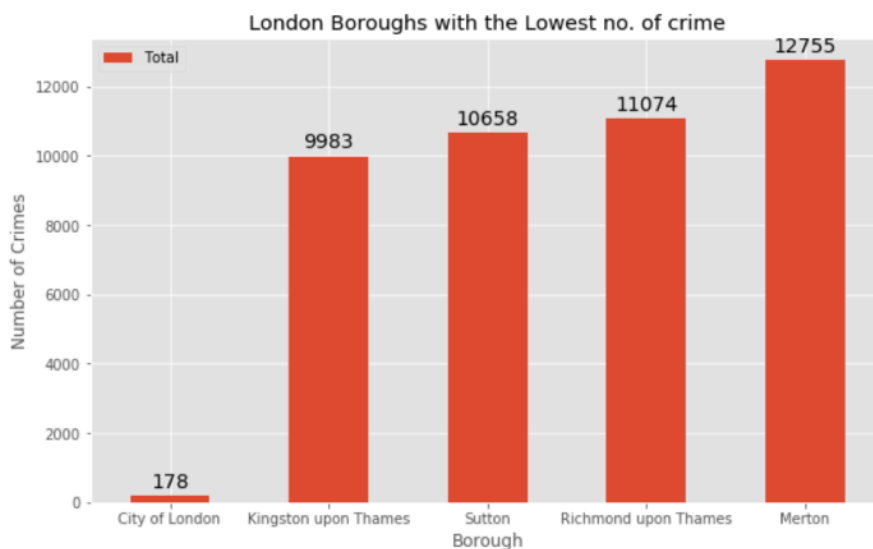

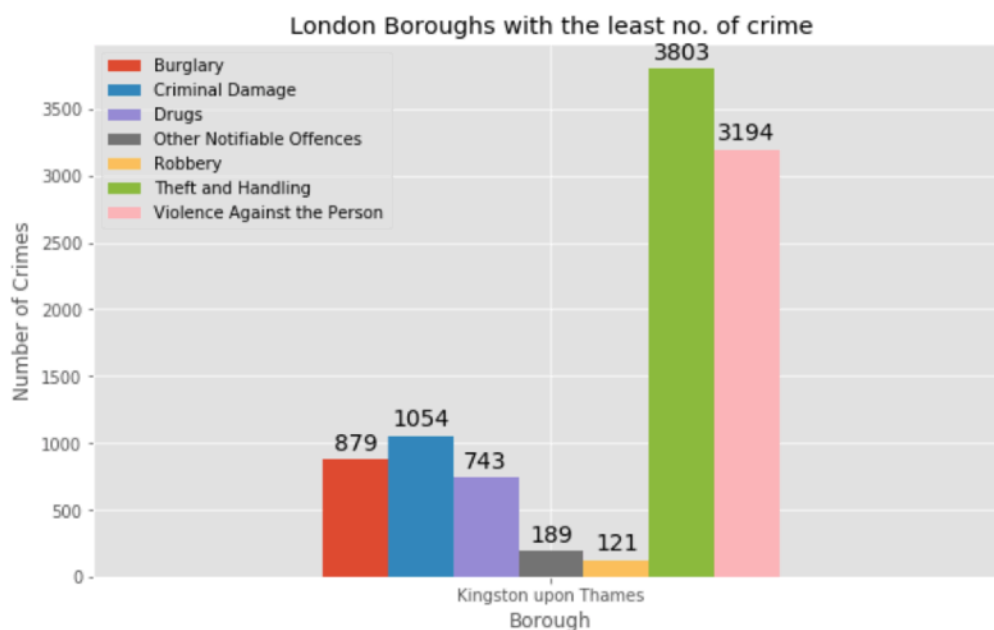Fig 3.1 Boroughs with lowest crime

Fig 3.2 Crimes in Kingston upon Thames

	Borough	Total	Area (sq mi)	Population (2013 est)[1]
6	City of London	178	1.12	7000

Fig 3.4 City of London - Division of greater London

3.1.3 Neighbourhoods in Kingston upon Thames

There are 15 neighbourhoods in the royal borough of Kingston upon Thames, they are visualised on a map using folium on python (see fig 3.5).
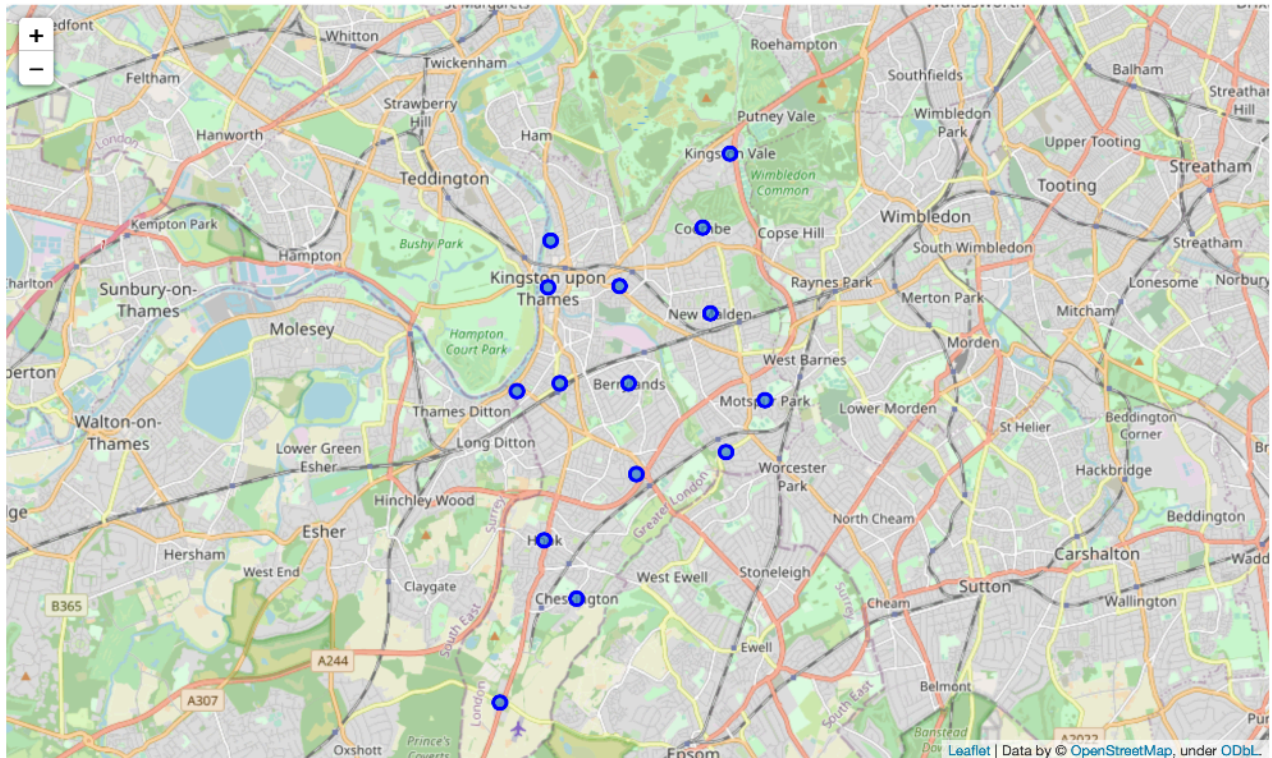

Fig 3.5 Neighbourhoods in Kingston upon Thames

3.2 Modelling

Using the final dataset 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 API. This returns a json file containing all the venues in each neighbourhood which is converted to a pandas dataframe. This data frame contains all the venues along with their coordinates and category (see fig 3.2.1).

	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

Fig 3.2.1 Venue details of each Neighbourhood

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). 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 neighbourhoods 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 neighbourhoods 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 interests based on the venues/amenities around each neighbourhood.

4. Observations

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 (see fig 4.1)

	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
4	Hook	Kingston upon Thames	51.367898	-0.307145	0.0	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Food	Department Store	Discount Store	Electronics Store	Farmers Market

Fig 4.1 Cluster 1 of Kingston Upon Thames Neighbourhoods

	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	51.431850	-0.258138	1.0	Grocery Store	Sandwich Place	Bar	Soccer Field	Deli / Bodega	Department Store	Discount Store	Electronics Store	Farmers Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1.0	Grocery Store	Garden Center	Pub	Restaurant	Farmers Market	Deli / Bodega	Department Store	Discount Store	Electronics Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	1.0	Grocery Store	Pharmacy	Sandwich Place	Train Station	Hotel	Discount Store	Coffee Shop	Café	Bus Stop

Fig 4.2 Cluster 2 of Kingston Upon Thames Neighbourhoods

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	3.0	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Fast Food Restaurant	Department Store	Discount Store	Electronics Store	Farmers Market
Motspur Park	Kingston upon Thames	51.390985	-0.248898	3.0	Gym	Park	Bus Stop	Soccer Field	Wine Shop	Fast Food Restaurant	Department Store	Discount Store	Electronics Store

Fig 4.3 Cluster 3 of Kingston Upon Thames Neighbourhoods

	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	9 C
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4.0	Pub	Fish & Chips Shop	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Plaza	Park	
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	4.0	Café	Coffee Shop	Sushi Restaurant	Burger Joint	Pub	Portuguese Restaurant	French Restaurant	German Restaurant	Ele
9	New Malden	Kingston upon Thames	51.405335	-0.263407	4.0	Gastropub	Gym	Sushi Restaurant	Supermarket	Bar	Chinese Restaurant	Korean Restaurant	Indian Restaurant	Wi
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	4.0	Indian Restaurant	Italian Restaurant	Food	Pub	Wine Shop	Breakfast Spot	Farmers Market	Hardware Store	
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	4.0	Indian Restaurant	Coffee Shop	Pub	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Café	Restaurant	
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	4.0	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Train Station	Hotel	Farmers Market	

Fig 4.4 Cluster 4 of Kingston Upon Thames Neighbourhoods

	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 Venu
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2.0	Train Station	Pub	Food	Construction & Landscaping	Fast Food Restaurant	Deli / Bodega	Department Store	Discount Store	Electronic Stor

Fig 4.5 Cluster 5 of Kingston Upon Thames Neighbourhoods

5. Conclusion

The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighbourhoods to which they want to relocate based on the most common venues in it. For example if a person is looking for a neighbourhood 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 neighbourhood with stores and restaurants in a close proximity then the neighbourhoods in the first cluster is suitable. For a family I feel that the neighbourhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighbourhoods have common venues such as Parks, Gym/Fitness centres, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family. The choices of neighbourhoods may vary from person to person.

This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. 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 neighbourhood. 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.