Capstone Project
The Battle of Neighborhoods

By
Jagapathi Babu Batchu

Introduction

- **Background:** The first and foremost concern when moving to a new location is safety. So to enjoy any kind of lifestyle our home should be safely located.

- **Problem:** This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of the borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using K-means.

- **Interest:** Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighborhood and common venues around each neighborhood.

Data Acquisition and Cleaning

Data Acquisition: The data used for this project comprises from three sources

- The first data source of the project is from Kaggle that shows the crime per borough in London.

- The second source of data is scraped from Wikipedia page containing the list of boroughs and its additional information respectively.

- The third data source is the list of neighborhoods in the Royal Borough of Kingston upon Thames as found on the Wikipedia page .

Data Acquisition and Cleaning

Data Cleaning: The data cleaning process is done separately

- From the London crime data, the crimes during the latest year 2016 are only analyzed per the boroughs for each major category.

- The second data is scrapped from Wikipedia page using beautiful soup library in python

- These two data sets ate merged to form a new data set of new borough names

- After visualizing the crime in each borough, a new data set is created with the lowest crime rate recorded and the latitude and longitude are obtained using Google Maps API geocoding.
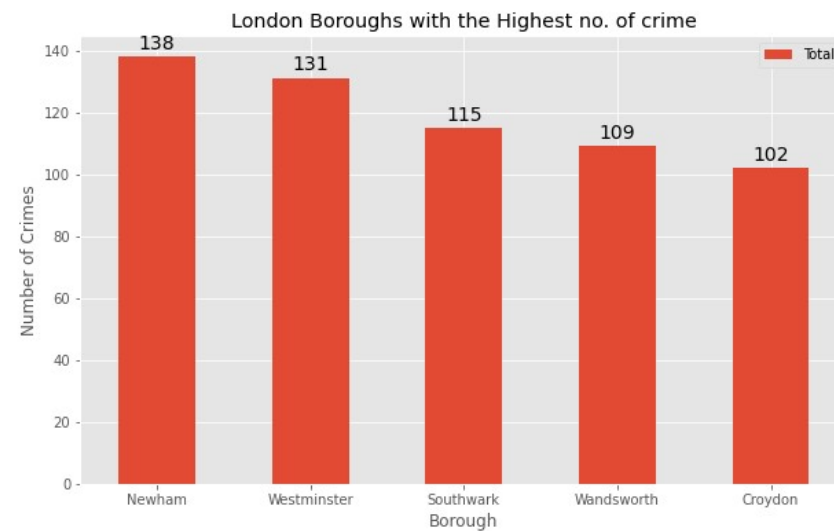
Data Acquisition and Cleaning

- The new data set is used to generate the 10 most common venues for each neighborhood using the Foursquare API.

- K-means algorithm is used to cluster similar neighborhoods together in the lowest crime rate recorded Kingston upon Thames.
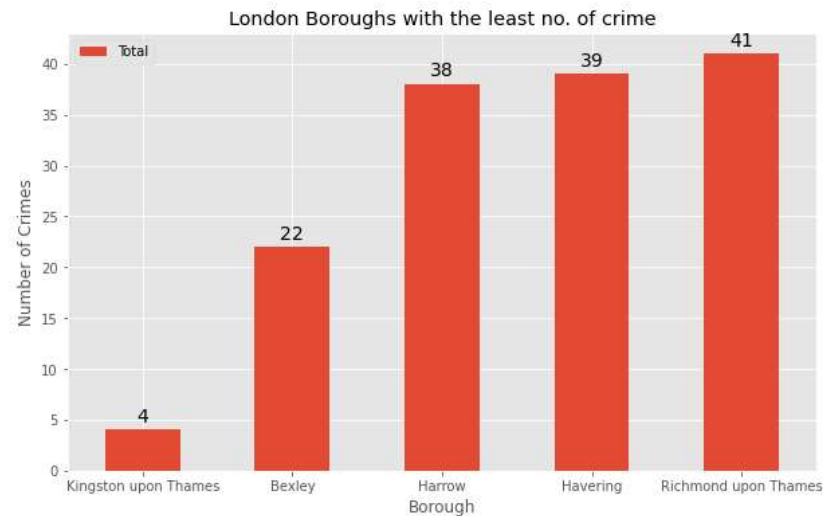
Methodology

- Exploratory Data Analysis (EDA)

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000
mean	6.406250	6.000000	4.593750	1.343750	2.000000	29.625000	20.468750	70.437500
std	4.507051	3.672039	6.100552	1.536846	2.047816	17.916923	11.129688	28.970438
min	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	1.000000	4.000000
25%	3.000000	3.000000	1.000000	0.000000	0.000000	16.000000	12.000000	53.750000
50%	7.000000	5.500000	2.500000	1.000000	1.500000	27.000000	16.500000	70.000000
75%	9.250000	8.000000	5.500000	2.000000	4.000000	36.000000	30.500000	81.250000
max	16.000000	16.000000	31.000000	5.000000	7.000000	78.000000	41.000000	138.000000

- It is observed that the Theft and Handling is the highest recorded crime in the year 2016 followed by the crimes such as Violence against the person.
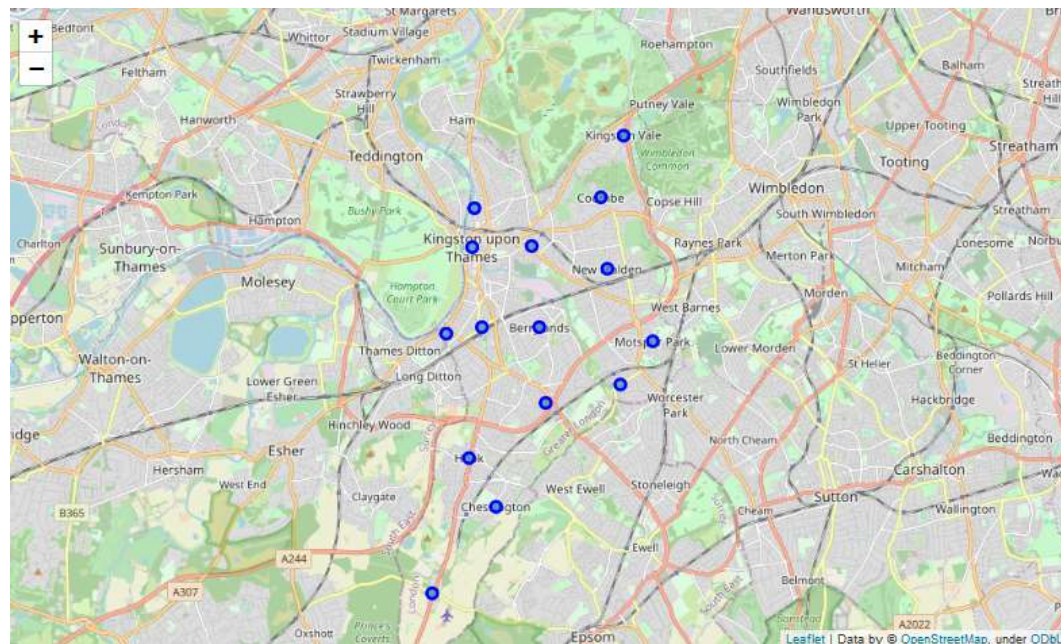
London Boroughs with the Highest no. of crime

- It is observed that the Newnham has the highest crime rate recorded during the year 2016.

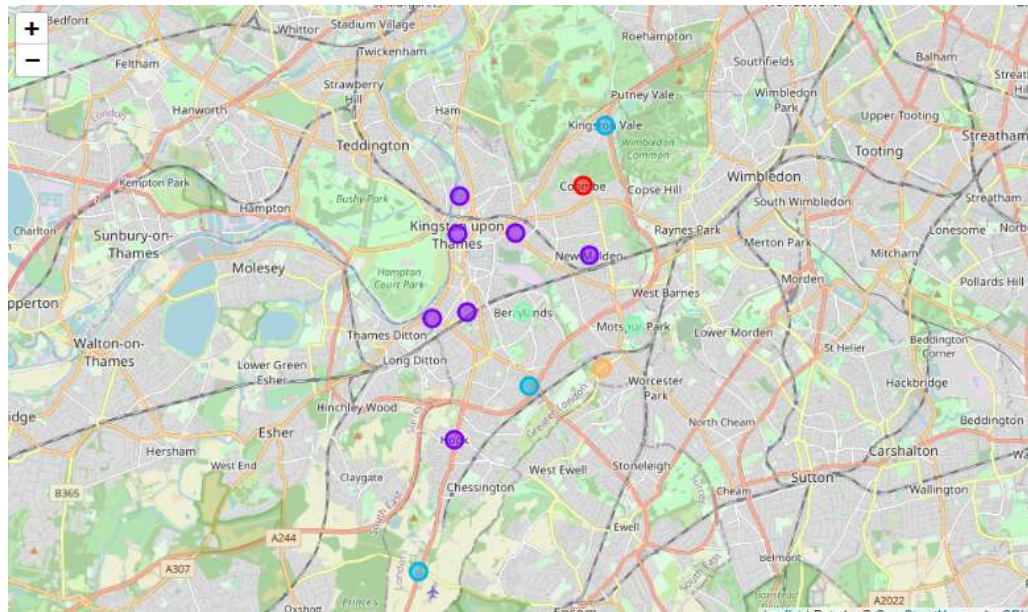London Boroughs with the least no. of crime

- It is observed that the Kingston upon Thames has recorded the lowest crime rate during the year 2016

- There are 15 neighborhoods in the Kingston upon Thames as visualized on a map using Folium on Python and is as shown by the figure below.

Results

- After the K-means clustering we can access each cluster to see which neighborhood were assigned to each of the five clusters listed. Visualizing the clustered neighborhoods on a map using the folium library.

- Cluster 1: The cluster 1 is the lowest cluster with only one neighborhood in the borough Kingston upon Thames and is consisted of venues Health and Beauty services, wine shops, food and departmental stores as shown in figure below.

	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 Mos Commo Venue
3	Coombe	Kingston upon Thames	51.41945	-0.265398	0	Health & Beauty Service	Wine Shop	Food	Deli / Bodega	Department Store	Dessert Shop	Electronics Store	Farmers Market	Fast Foo Restauran

- Cluster 2: This cluster consists of 7 neighborhoods near to Kingston upon Thames with Pub, Gym and parks nearby as shown below.

	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 Mc Comm Vene
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Gym / Fitness Center	Fish & Chips Shop	Park	Café	Plaza	Hotel	Shop Servi
4	Hook	Kingston upon Thames	51.367898	-0.307145	1	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Deli / Bodega	Departme Sto
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Café	Coffee Shop	Department Store	Italian Restaurant	Sushi Restaurant	Burger Joint	Pub	Gyr Fitne Cen
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Indian Restaurant	Korean Restaurant	Gym	Gastropub	Chinese Restaurant	Sushi Restaurant	Supermarket	E
0	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Food	Indian Restaurant	Italian Restaurant	Platform	Pub	Wine Shop	Rental Car Location	Farm Mar
2	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Coffee Shop	Pub	Gym / Fitness Center	Playground	Chinese Restaurant	Café	Restaura
3	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Pharmacy	Platform	Italian Restaurant	Grocery Store	Bistro	Optic Sh

- Cluster 3: This cluster consists of 3 venues and venues such as stores and bars nearby.

	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	
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	2	Sandwich Place	Grocery Store	Bar	Soccer Field	Wine Shop	Deli / Bodega	Department Store	Dessert Shop	El
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	2	Grocery Store	Pub	Garden Center	Restaurant	Wine Shop	Farmers Market	Cosmetics Shop	Deli / Bodega	De
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	2	Grocery Store	Pharmacy	Indian Restaurant	Hotel	Italian Restaurant	Bowling Alley	Restaurant	Coffee Shop	S

- Cluster 4: This cluster consists of 2 venues nearby and consists of wine shop, park and bus stop as shown below.

	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 Com Ve
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Wine Shop	Park	Bus Stop	Gym / Fitness Center	Thai Restaurant	Fast Food Restaurant	Coffee Shop	Cosmetics Shop	D Boo
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Fish & Chips Shop	Deli / Bodega	Department Store	Dessert Shop	Electro S

- Cluster 5: This cluster has just only one venue nearby and has the facilities such as train station, food, pub, shops and departmental store as shown in figure below.

	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 Mo Comm Ven
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	4	Food	Train Station	Pub	Fish & Chips Shop	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Electron Sto

Discussion

- The aim of this project is to help people who want to relocate to the safest borough in London, expacts 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 4 and 5 have Train stations and Bus stops as the most common venues. If a person is looking for a neighbourhood with stores, beauty care and restaurants in a close proximity then the neighbourhoods in the first cluster is suitable. For a family I feel that the neighbourhoods in Clusters 3 and 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.

Conclusion

- 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 borough of London. The future scope 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.