Final Project - The Battle of Neighborhoods

Rizqi Ardhi

Table of Content

Introduction

1

Background:
Safety place is the one of the main concern for people to select place or move to new place. As a person, you have to make sure that your home is safe and comfortable for living.

2

Problem:
The project will be used to know the safest place in the London based on the total crimes, explore the neighborhoods of that borough 10 most common venues. In this case, k-means clustering as a machine learning method will be used to create clusterization.

3

Interest:
People who want to move or relocate to London and/or identify the safest borough in London, explore the neighborhoods, and common venues around the neighborhood.

Data Acquisition & Cleaning

Data Acquisition Process:

The data that has been used to solve the issue is the combination of some sources such as:

1. Data preprocessing using Kaggle Dataset regarding “London Crimes from 2008 to 2016”. Data set URL can be seen here
2. Web scraping using Beautifulsoup library to get additional information regarding differentiation of Boroughs in London from Wikipedia. The Wikipedia link can be seen here
3. Create a new dataset of the neighborhoods of the safest place in London, we will use google maps API to get the coordinate information

Data Cleansing Process:

The process for cleansing the data has been done separately

1. London data crime on 2015 has been selected. The major categories of crime are provided combined with pivoting process to get total crimes per the boroughs for each major category
2. The second data is using List of London Borough from Wikipedia combined with scraping method & beautifulsoup library in python.
3. Then, two data sets above are merged to get new data form. Then, the data will be used to visualized crime rates in each borough and identify the highest 5 and the lowest 5 number of crimes during 2015
4. After the visualization process, the lowest number of crime will be used to create the forth data set. Then after that, Google Maps API geocoding will be used to get latitude and longitude data.
5. The new data set is used to generate the 10 most common venues for each neighborhoods using the Foursquare API and combined with K-Means clustering algorithm.

Methodology - Exploratory Data Analysis

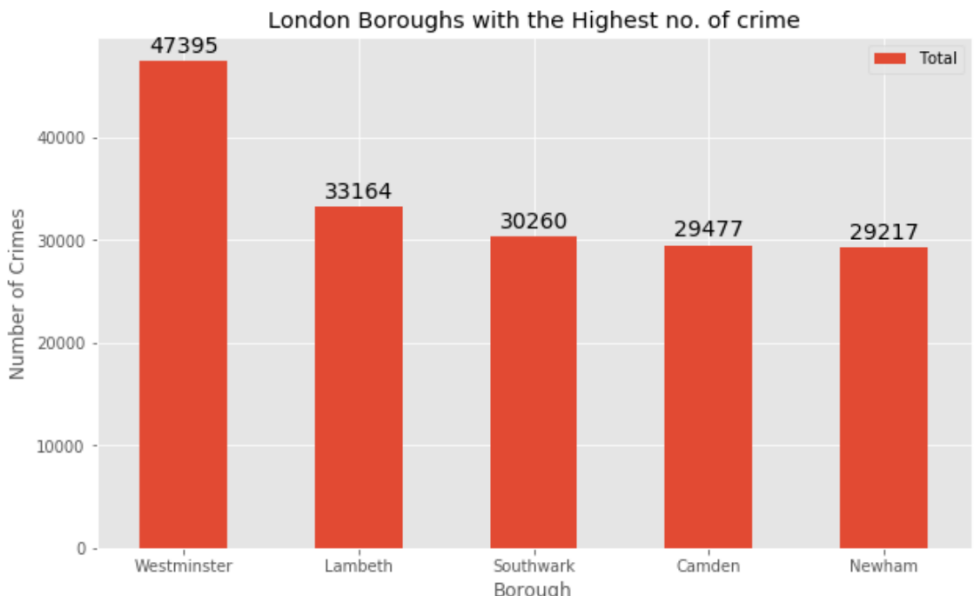
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2136.030303	1908.363636	1205.606061	431.181818	647.969697	8606.727273	6628.484848	21564.363636
std	743.587482	609.458972	572.721078	184.367916	406.333875	4686.355622	2385.072956	8605.496648
min	5.000000	3.000000	6.000000	0.000000	5.000000	104.000000	28.000000	151.000000
25%	1629.000000	1565.000000	809.000000	359.000000	323.000000	5850.000000	5033.000000	16346.000000
50%	2245.000000	2033.000000	1231.000000	421.000000	624.000000	8467.000000	6924.000000	22076.000000
75%	2569.000000	2295.000000	1568.000000	513.000000	979.000000	10143.000000	8441.000000	26941.000000
max	3853.000000	3113.000000	2652.000000	1008.000000	1497.000000	27980.000000	10491.000000	47395.000000

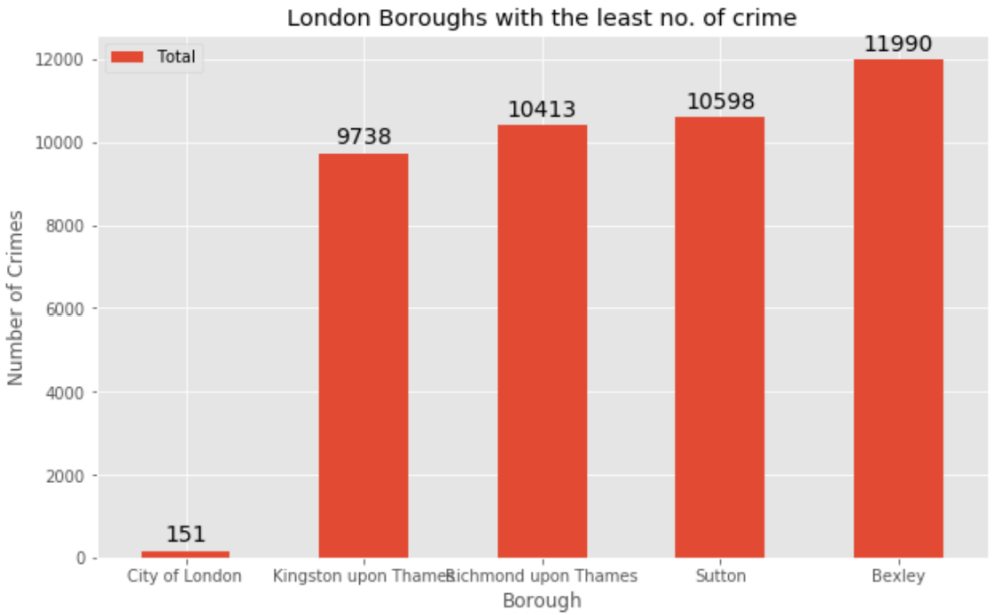
Notes:

The number of boroughs are 33 with the reported crime from high to low as below:

1. Theft and Handling
2. Violence Against the Person
3. Burglary
4. Criminal Damage
5. Drugs
6. Robbery
7. Others

Methodology - Exploratory Data Analysis





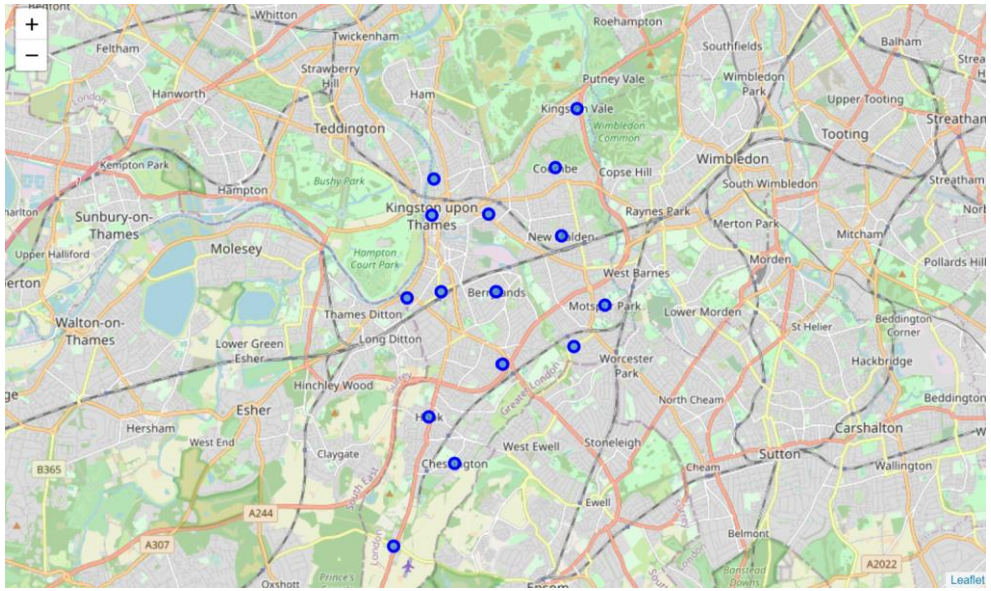
Comparing with the five boroughs with the highest crime rate during 2015, Westminster has the highest crimes rate followed by Lambeth, Southwark, Camden, and Newham.

Comparing with the five boroughs with the lowest crime rate during 2015, City of London has the lowest crimes rate followed by Kingston upon Thames, Richmond upon Thames, Sutton, and Bexley

Methodology - Modelling

The City of London has significantly the lowest crimes rate compare to the other. The City of London is the 33rd principal division of Greater London but it's not a London borough, it only has 1,12 square miles with the population of 7000. In this case, we can categorize the City of London as a small area and exclude it from safest place analysis.

According to the previous reason, the second lowest crimes rate can be considered as the safest borough in London which is Kingston upon Thames. The visualization of Kingston upon Thames can be seen below:



There are 15 neighborhoods around Kingston upon Thames

Methodology - Modelling

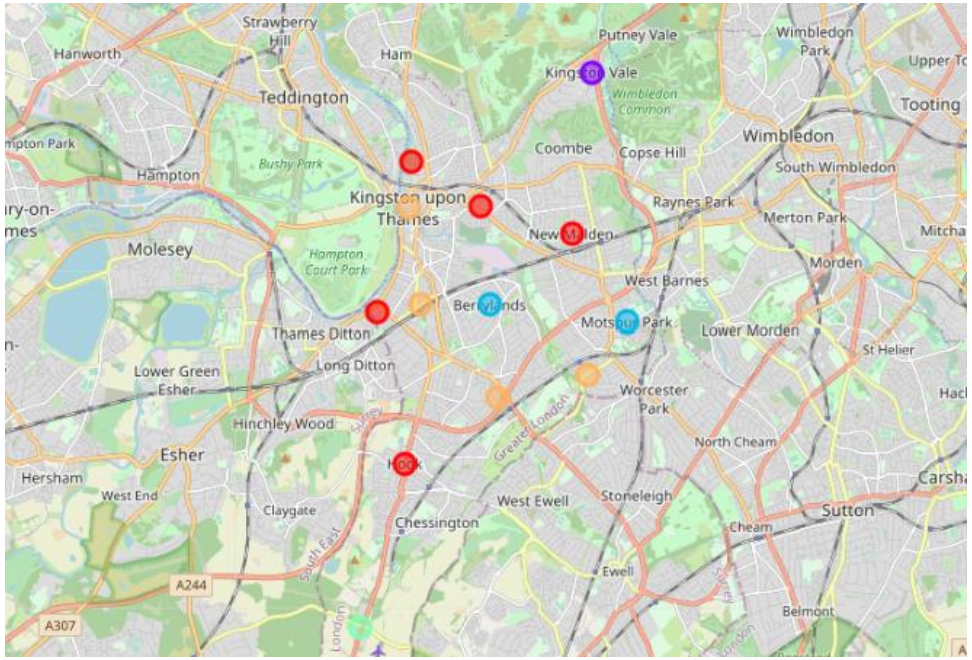
Using the final data set, we can find all venues around 500m of each neighborhood, grouping the venues by the neighborhood to get mean of the venues, get 10 common venues for each neighborhood.

	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	51.417250	-0.305631	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417250	-0.305631	The Boater's Inn	51.418546	-0.305915	Pub

To help people find similar neighborhood, we will create clusterization using K - Means Clustering algorithm. The size of cluster is 5, The consideration for choosing K - Means clustering is to cluster the neighborhoods with the similar venues together so that the people can move to the area based on their interest place/venues/need.

Results

The visualization of clustered neighborhoods can be seen below:



According to the result above, we have:

1. Red cluster as the majority result
2. Blue cluster
3. Purple cluster
4. Orange cluster
5. Green cluster

We will analyze each cluster to identify the characteristics of each cluster and neighborhoods of them

Results

Red 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	Canbury	Kingston upon Thames	51.417250	-0.305631	0	Pub	Park	Fish & Chips Shop	Japanese Restaurant	Supermarket	Spa	Café	Shop & Service	Gym / Fitness Center
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Indian Restaurant	Pub	Bakery	Supermarket	Fish & Chips Shop	Fast Food Restaurant	Deli / Bodega	Department Store	Dry Cleaner
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Chinese Restaurant	Sushi Restaurant	Supermarket	Bar	Korean Restaurant	Indian Restaurant	Farmers Market
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Food	Italian Restaurant	Pub	Indian Restaurant	Wine Shop	Japanese Restaurant	Hotel	Hardware Store	Fried Chicken Joint
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Pub	Hotel	Pet Café	Chinese Restaurant	Café	Italian Restaurant	Restaurant

Red cluster is the biggest cluster with 9 of 13 neighborhoods in the borough Kingston upon Thames. We can see that the most common venues in these neighborhoods are Pub, Indian Restaurants, Gastropub, park, gym, supermarket, and Food

Results

Purple 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
6	Kingston Vale	Kingston upon Thames	51.43185	-0.258138	1	Grocery Store	Sandwich Place	Bar	Soccer Field	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store	Farmers Market

Purple cluster has one neighborhood which consists of venues such as restaurant, soccer field, dept. store

Blue 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	2	Gym / Fitness Center	Park	Bus Stop	Wine Shop	Fast Food Restaurant	Department Store	Dry Cleaner	Electronics Store	Farmers Market
Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Gym	Park	Bus Stop	Soccer Field	Wine Shop	Fast Food Restaurant	Department Store	Dry Cleaner	Electronics Store

Blue cluster has two neighborhoods which consist of venues such as park, gym, bus stop

Results

Green 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
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	3	Grocery Store	Pub	Garden Center	Restaurant	Farmers Market	Deli / Bodega	Department Store	Dry Cleaner	Electronics Store

Green cluster has one neighborhood which consists of venues such as restaurant, garden, food.

Orange 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 upon Thames	Kingston upon Thames	51.409627	-0.306262	4	Coffee Shop	Café	Pub	Sushi Restaurant	Burger Joint	German Restaurant	Gift Shop	Gym / Fitness Center	Electronics Store
Old Malden	Kingston upon Thames	51.382484	-0.259090	4	Coffee Shop	Pub	Food	Construction & Landscaping	Train Station	Bakery	Bar	Gift Shop	German Restaurant
Surbiton	Kingston upon Thames	51.393756	-0.303310	4	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Breakfast Spot	French Restaurant	Train Station	Gym / Fitness Center
Tolworth	Kingston upon Thames	51.378876	-0.282860	4	Grocery Store	Restaurant	Sandwich Place	Train Station	Hotel	Italian Restaurant	Coffee Shop	Pharmacy	Café

Orange cluster is the second biggest cluster with 4 of 13 neighborhoods in the borough Kingston upon Thames. We can see that the most common venues in these neighborhoods are Pub, Restaurants, Bus Stop, Hotel, and Food

Discussion

The project will be used for people who want to move or relocate to London and/or identify the safest borough in London, explore the neighborhoods, and common venues around the neighborhood.

If the person is looking for a neighborhood with good connectivity with train and bus station, they can consider orange cluster.

For family, red cluster can be considered as the main option because it has parks, gym, food, restaurant, groceries, etc.

The preference of venues may vary for each person. People can select a neighborhood based on their need

Conclusion

Safety place is the one of the main concern for people to select place or move to new place. As a person, you have to make sure that your home is safe and comfortable for living. We have just taken safety combined with the venues around the borough of London.

The preference of venues may vary for each person. People can select a neighborhood based on their need. This project can be improved with combining/considering additional aspect such as cost of living in the areas.


THANK YOU