

IBM APPLIED DATA SCIENCE

THE BATTLE FOR NEIGHBOURHOOD

CAPSTONE PROJECT

JAGADEESAN NEELAKANDAN

ABOUT ME



I'm a Business Analytics Leader with over 10+ years of experience. I lead the Data and Analytics Team for one of fortune 500 company. I have held numerous marketing analytic positions. Within my experience, I have built Data Science teams and I have helped to grow tomorrow's analytics leaders.

INTRODUCTION

THE BATTLE OF NEIGHBOURHOOD

Background: 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.

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

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

DATA PREPARATION

THE BATTLE OF NEIGHBOURHOOD

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

METHODOLOGY

THE BATTLE OF NEIGHBOURHOOD

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	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

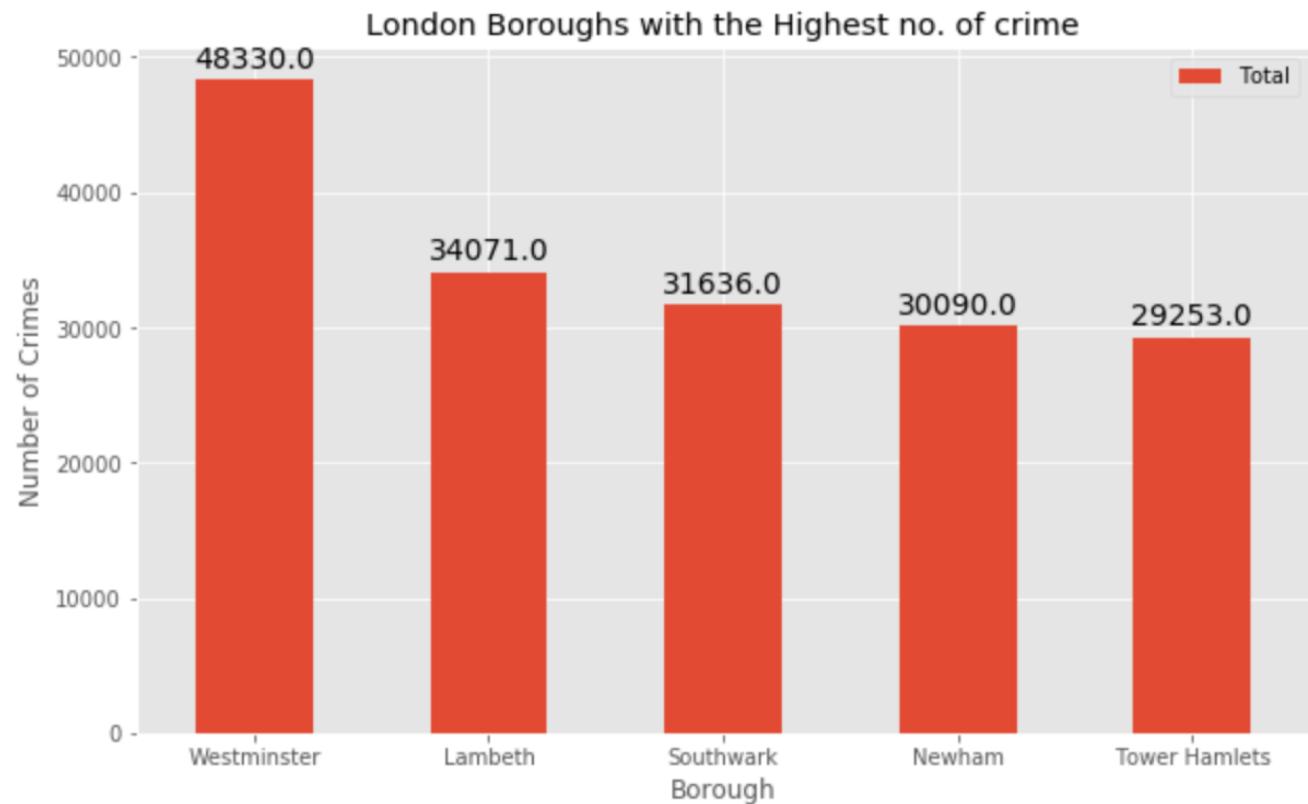
The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'

METHODOLOGY

THE BATTLE OF NEIGHBOURHOOD

Highest Crime Rate:

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.



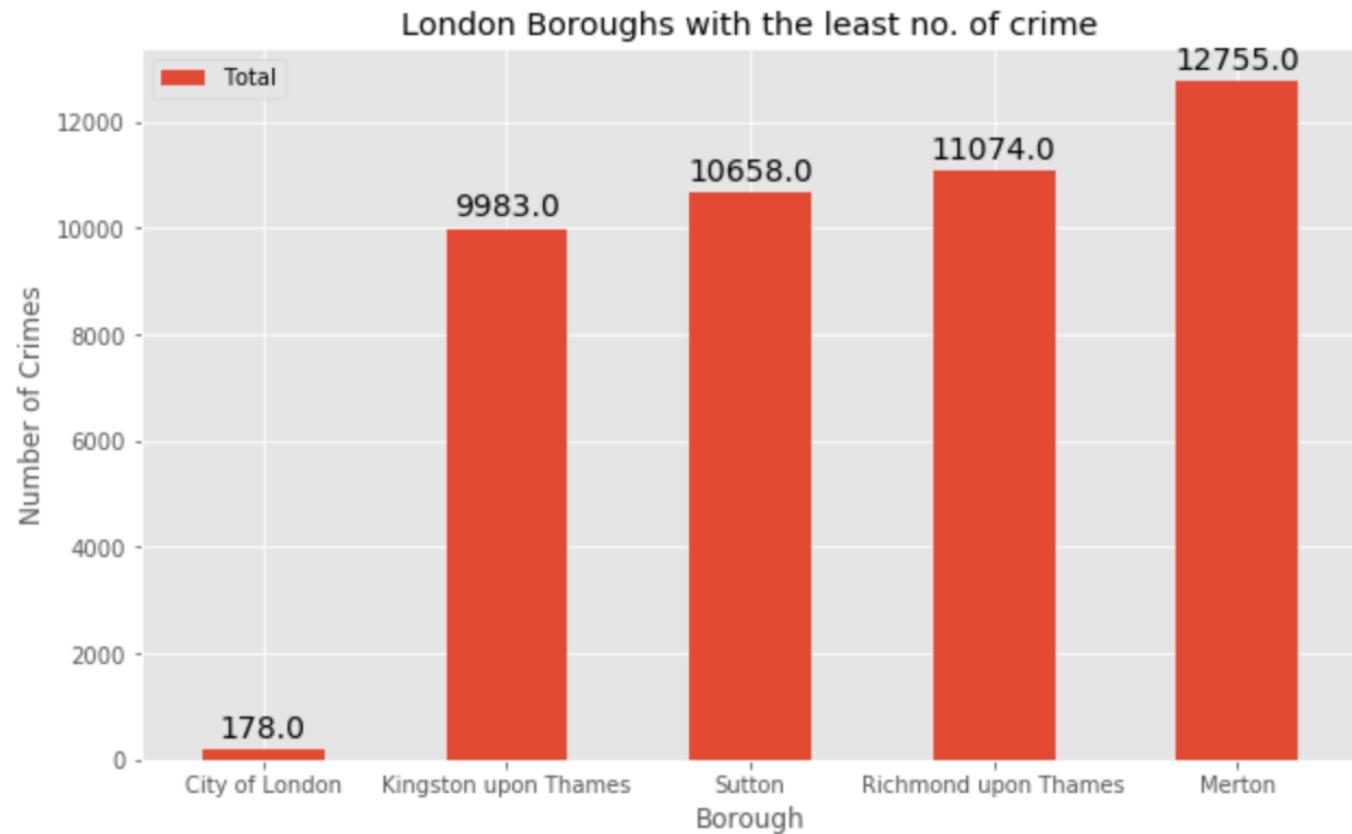
METHODOLOGY

THE BATTLE OF NEIGHBOURHOOD

Lowest Crime Rate:

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.

- City of London has a significantly lower crime rate because it 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.
- We will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

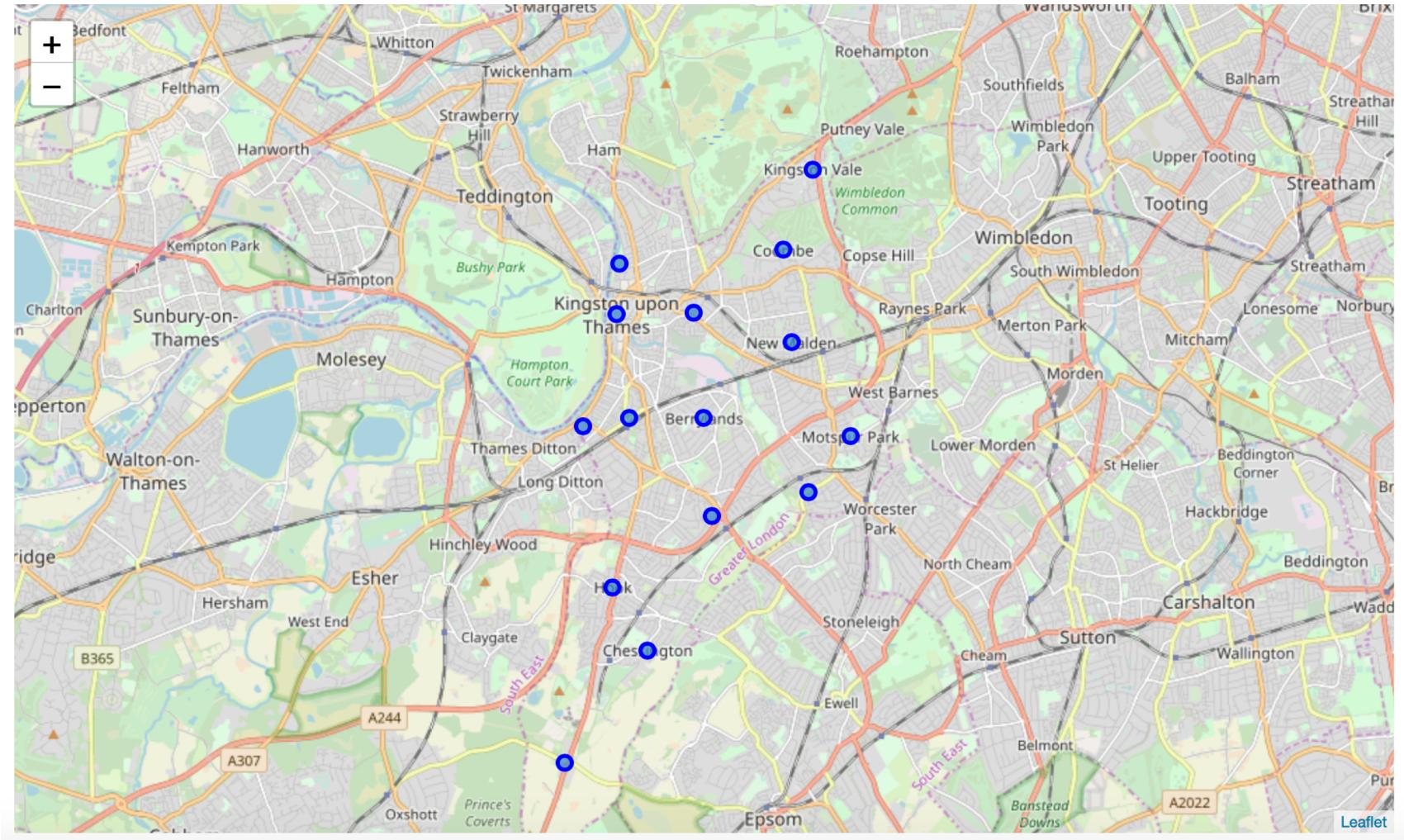


METHODOLOGY

THE BATTLE OF NEIGHBOURHOOD

Neighborhoods in Kingston upon Thames

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



METHODOLOGY

THE BATTLE OF NEIGHBOURHOOD

Modelling

- Using the final data set containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API
- One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods
- To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods 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 neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

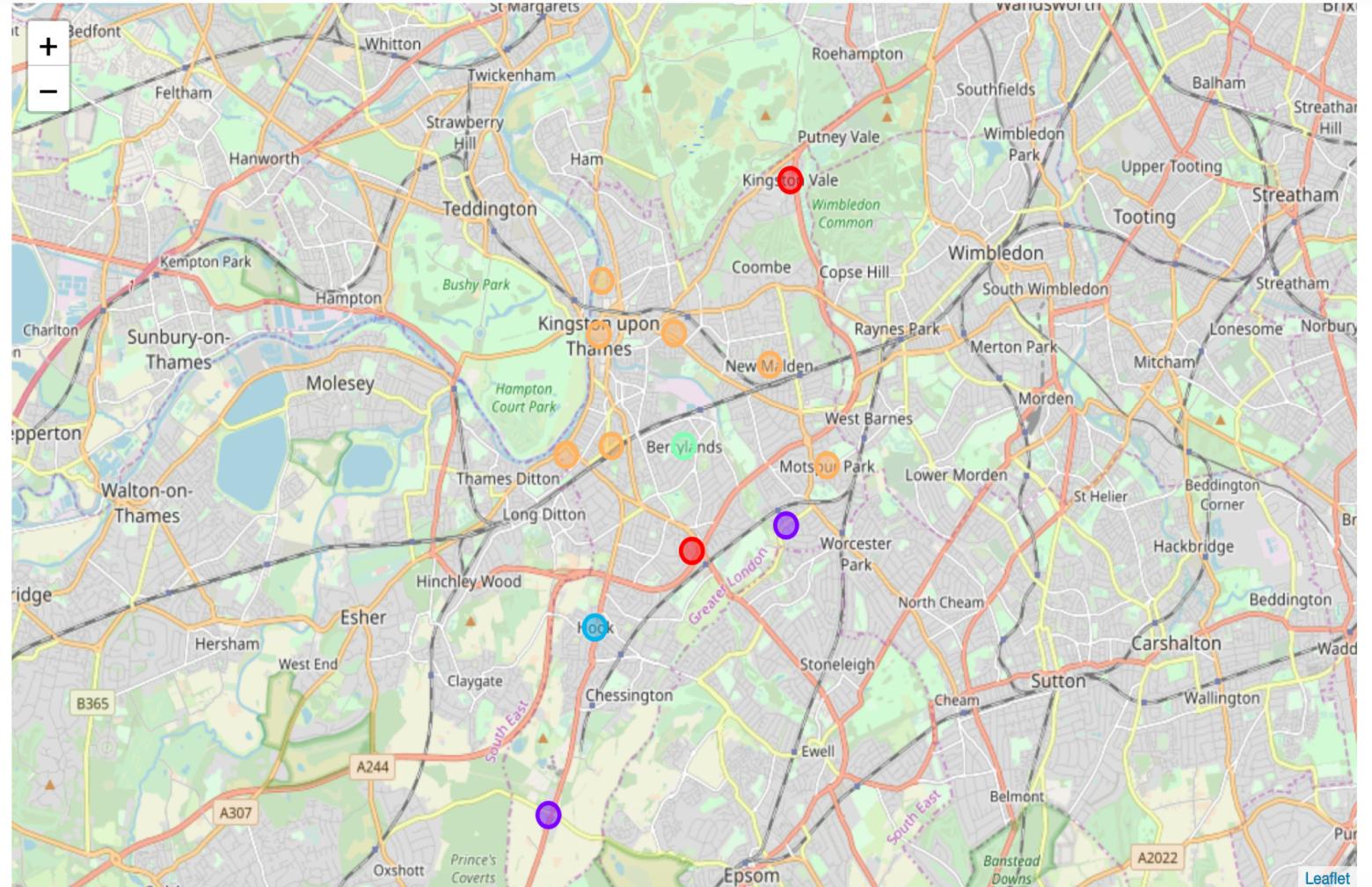
RESULT

THE BATTLE OF NEIGHBOURHOOD

K-Means Clustering

The aim of this project is to help people who want to relocate to the safest borough in London, expats can chose the neighborhoods to which they want to relocate based on the most common venues in it.

K-Means Clustering grouped the similar neighbourhood in 5 Clusters based on their similarity. The clusters are visualized on a map using the folium library.



DISCUSSION

THE BATTLE OF NEIGHBOURHOOD

Recommendation:

Cluster 5 (Yellow) :
 More suitable for families due to the common venues such as Parks, Gym/Fitness Centers, Coffee Shops 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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4	Pub	Park	Indian Restaurant	Fish & Chips Shop	Supermarket
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	4	Coffee Shop	Café	Sushi Restaurant	Burger Joint	Pub
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	4	Gym	Park	Soccer Field	Construction & Landscaping	Wine Shop
9	New Malden	Kingston upon Thames	51.405335	-0.263407	4	Indian Restaurant	Gastropub	Gym	Sushi Restaurant	Supermarket
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	4	Pub	Food	Indian Restaurant	Italian Restaurant	Platform
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	4	Indian Restaurant	Coffee Shop	Pub	Café	Golf Course
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	4	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy

DISCUSSION

THE BATTLE OF NEIGHBOURHOOD

Recommendation

Cluster 2 and 4 are more suitable for people who is looking for good connectivity and public transportation as the Train stations and Bus stops as the most common venues in this neighborhoods

	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
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	1	Convenience Store	Garden Center	Pub	Restaurant	Farmers Market
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	1	Train Station	Food	Pub	Convenience Store	Garden Center
	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
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Park	Gym / Fitness Center	Bus Stop	Wine Shop	Fast Food Restaurant

DISCUSSION

THE BATTLE OF NEIGHBOURHOOD

Recommendation

Cluster 1 and 3 (Red and Blue): These clusters have the own attributes and people tend to visit convenient stores, grocery stores and restaurants 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
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	0	Italian Restaurant	Grocery Store	Bar	Soccer Field	Wine Shop	Convenience Store
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Pizza Place	Sandwich Place	Italian Restaurant	Convenience Store
	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
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Wine Shop	Convenience Store

CONCLUSION

THE BATTLE OF NEIGHBOURHOOD

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. We 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.