Capstone Project
The Battle of Neigbourhoods
Presentation

Anindya Banerjee

1. Introduction

- Background: Investors before making any investment venture would make an assessment about the venue that they would like to set up their business. In case of restaurant owner, this is no different. The choice of venue would be critical to the success of the business.

- Problem: In this instance, a business owner would like to set up a restaurant in the heart of the City of London. He would like to own a little bit about the venue before setting up the business. This is in terms of number of crimes in the neighbourhood, the safety of the venue, the presence of other business in the venue and the general outlook for setting up a business in the area.

- Target audience: This would be of interest to any potential investor in the hospitality sector who is seeking to set up a chain of restaurants in the City of the London and the venue details would be critical to the success of the business.

2. Data Acquisition and Cleaning

- Data Acquisition: The data for the project is gathered from three primary sources:

- Source Data 1: London Crime data showing the crime per borough in London

- Source Data 2: This is scraped from a Wikipedia page that contains the list of London boroughs. It also contains other details of the boroughs.

- Source Data 3: This is a list of neighbourhood in the Royal Borough of Kingston upon Thames.

Data Cleaning.

- Data downloaded or scraped from multiple sources were combined into one table.

- The London crime data is accessed from Kaggle and this is then uploaded to Amazon S3 bucket.

- Data is pre processed for recent crimes and also scraped using the beautiful soup library.

- The two data sets are merged to form a new data set. The reason for doing so is to visualize the crime rates in each of the boroughs and to identify the borough with the least recorded crime.

- After visualizing the crime in each borough, the one with the lowest crime is found out. A third data set is created with the names of the neighbourhood and the name of the borough with the latitude and longitude obtained using Google Maps API geocoding.

- The new data set is used to generate 10 most common venues for each neighbourhood using the Foursquare API and then using k-means clustering algorithm to cluster similar neighbourhoods together.

3. Methodology

- Exploratory Data Analysis

Statistical summary of crimes

	Borough	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	No_of_Crimes	Total
Major_Category										
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	
5	Camden	2652	1935	1493	490	899	14088	7626	29183	
6	City of London	2	2	10	6	4	129	25	178	
7	Croydon	2738	3219	1367	718	1139	9229	10302	28712	
8	Ealing	2492	2562	1355	613	669	10040	9396	27127	
9	Enfield	2541	2136	1063	492	807	8037	7409	22485	
10	Greenwich	1780	2476	867	521	486	8010	8590	22730	
11	Hackney	2719	1981	1353	499	1030	11851	8832	28265	
12	Hammersmith and Fulham	1531	1408	1321	474	397	8925	6118	20174	

The count for each of the major categories gives the value of 33 which is the number of London boroughs. 'Violence against the person' is the highest reported, followed by 'Theft and Handling' followed by 'Robbery. Lowest recorded crimes included 'Burglary', 'Criminal Damage' and ' Drugs'

Boroughs with the highest crime rates

London Boroughs with the Highest no. of crime
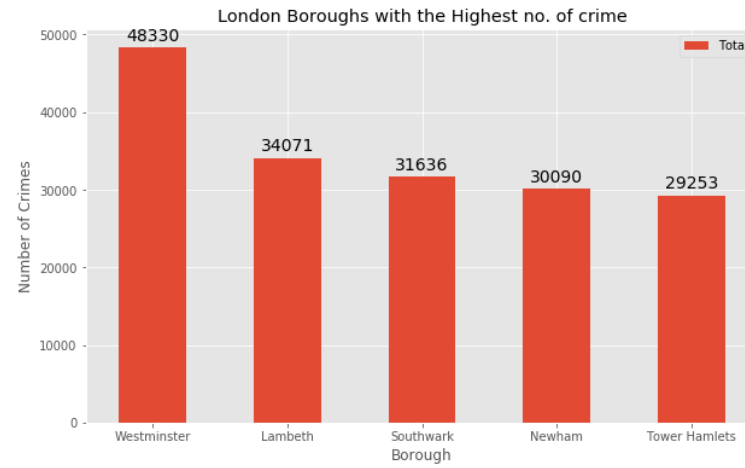
 Comparing the boroughs with the highest crime rate, it is evident that Westminster has the highest recorded crimes followed by the boroughs of Lambeth Southwork, Newham and Tower Hamlets. It is also to be noted that the crime rates at Westminster is higher than the other four boroughs. Based on this advice would be not open any chain of restaurants in this area.

By sorting out the data from the least could we can see the borough with the lowest crime is Kingston upon Thames

Boroughs with the least crime rates

- We can see the Kingston upon Thames has the least crime rates. (City of London is not a borough)

- We will subsequently explore the Borough of Kingston upon Thames and its neighborhoods and venues in the subsequent sections.
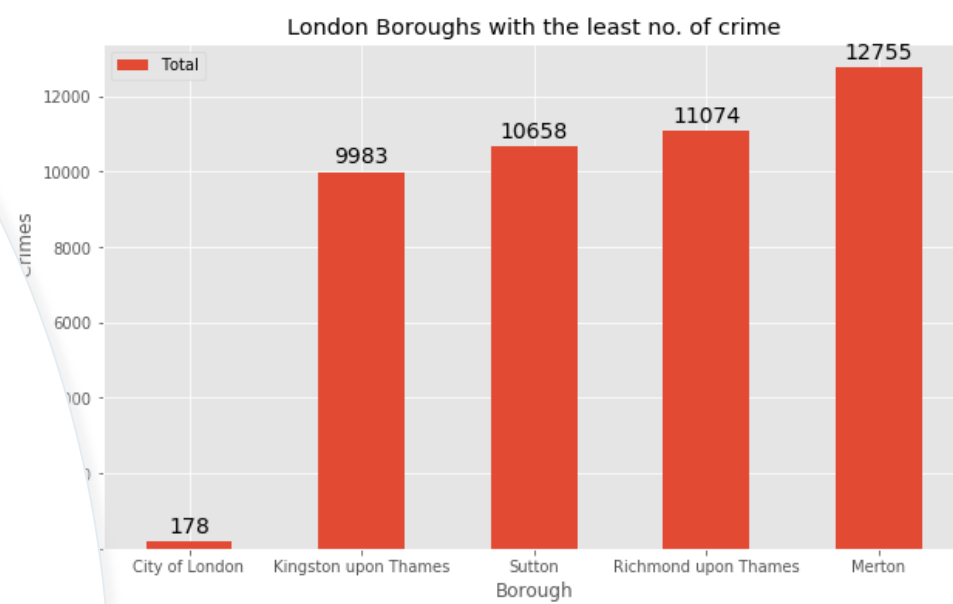
London Boroughs with the least no. of crime
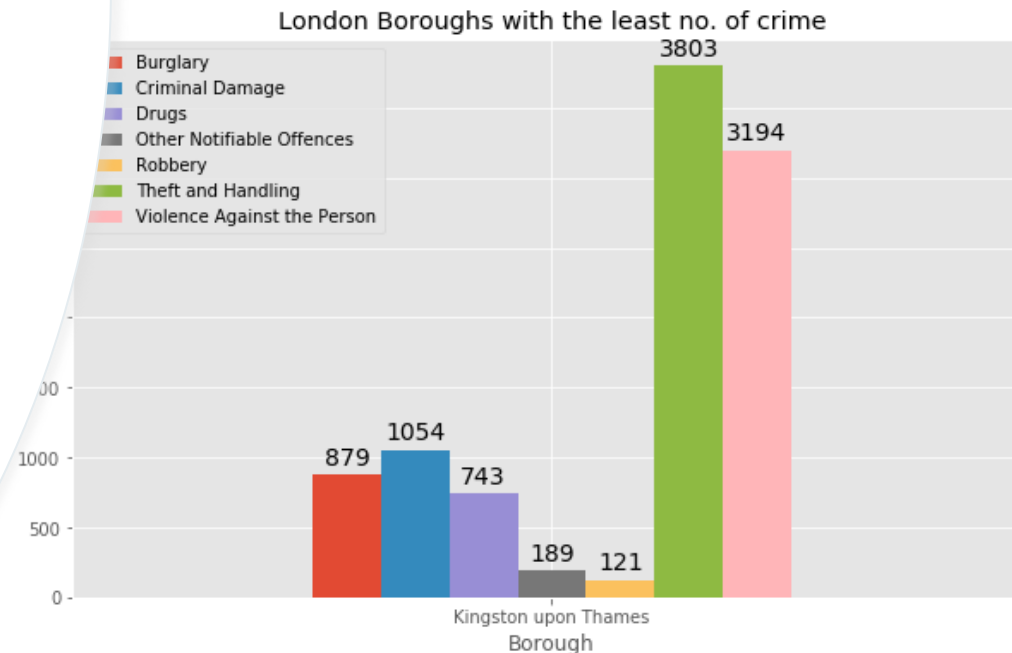
Neighbourhood in Kingston upon Thames
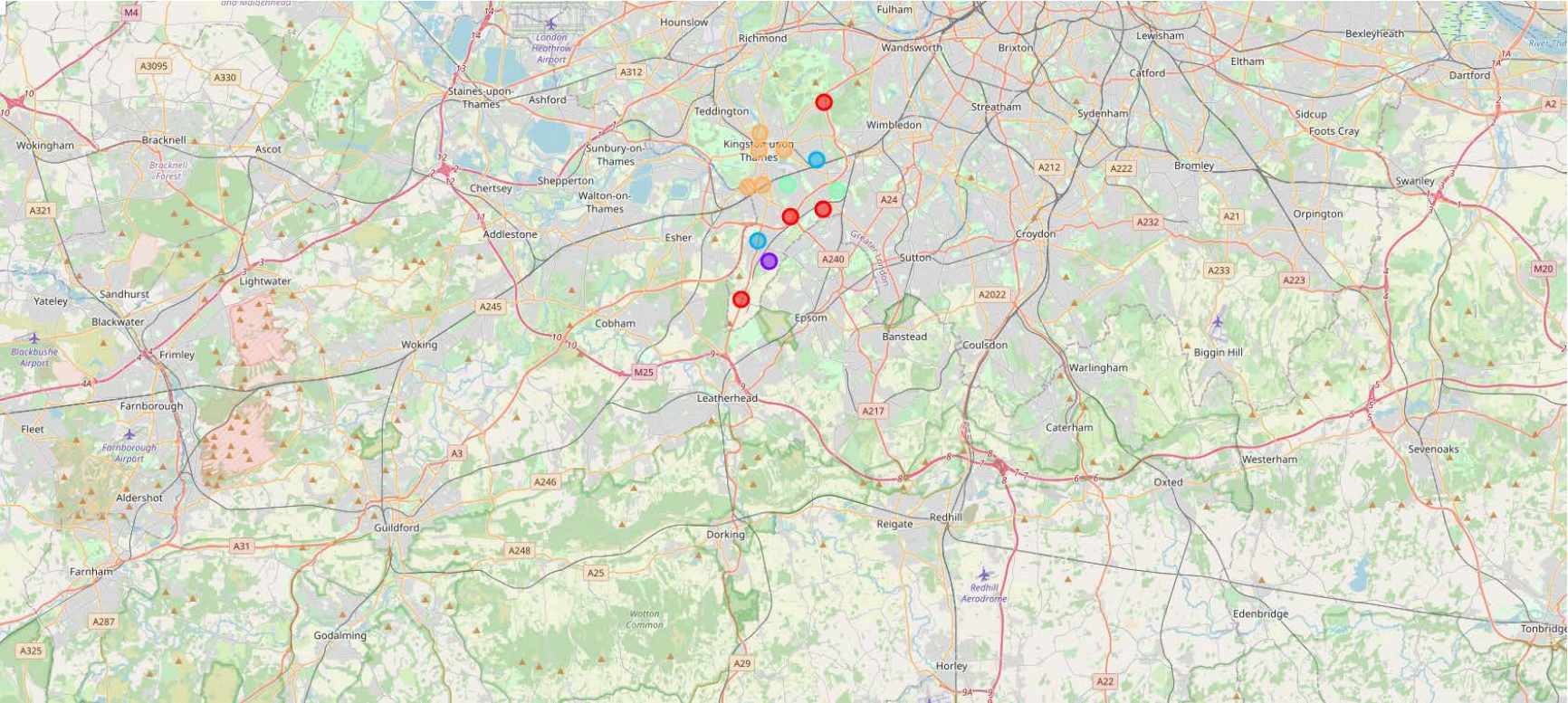
There are 15 neighbourhoods in the royal borough of Kingston upon Thames, they are
visualised on a map using the folium library.

Modelling

- Using the final dataset containing the neighbourhoods in Kingston upon Thames and finding the latitude and longitude, we can then use the Foursquare developer API to find venues within 500 m radius of each of the neighbourhoods.

Out[95]:						
	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berrylands	4	4	4	4	4	4
Canbury	14	14	14	14	14	14
Coombe	1	1	1	1	1	1
Hook	4	4	4	4	4	4
Kingston Vale	4	4	4	4	4	4
Kingston upon Thames	30	30	30	30	30	30
Malden Rushett	4	4	4	4	4	4
Motspur Park	5	5	5	5	5	5
New Malden	6	6	6	6	6	6
Norbiton	27	27	27	27	27	27
Old Malden	3	3	3	3	3	3
Seething Wells	20	20	20	20	20	20
Surbiton	30	30	30	30	30	30
Tolworth	18	18	18	18	18	18

Modelling (contd)

- One hot coding is done on the venues data. This is a process by which categorical values are converted into a form which can be interpreted by machine learning algorithms to do a better job in prediction. The venues data is then grouped by the neighbourhood and the mean of the venues is calculated. As a final task, the 10 common venues are calculated for each of the neighbourhoods.
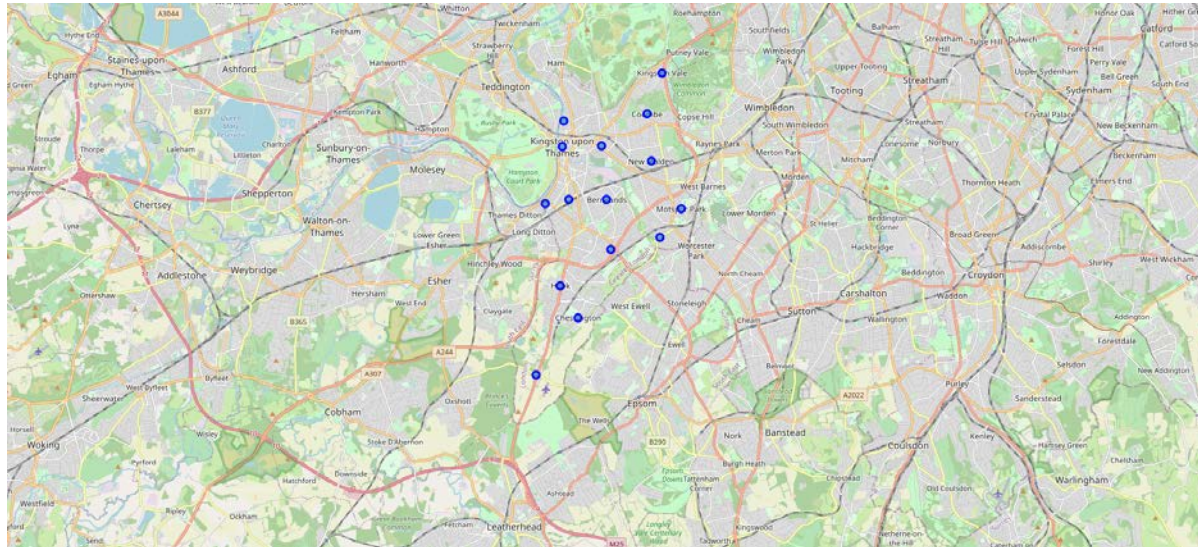
Modelling (contd)

To help restaurant chain owners find similar neighbourhoods in the safest borough, will use 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 five for this project that will cluster 15 neighbourhoods into five clusters. The reason to conduct a k- means clustering is to cluster venues with similar venues so that restaurant owners can shortlist area of interest based on the venues/transport links/amenities around each neighbourhood.

4.Results

- Implementing the k- means clustering we can access each cluster to see which neighbourhoods were assigned to each of the five clusters. We will examine the neighbourhoods in sequence.

Cluster 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	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	0	Grocery Store	Sandwich Place	Bar	Soccer Field	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	Food
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Grocery Store	Pub	Garden Center	Restaurant	Fish & Chips Shop	Deli / Bodega	Department Store	Discount Store	Farmers Market	Fast Food Restaurant
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	0	Construction & Landscaping	Grocery Store	Food	Train Station	Bagel Shop	Department Store	Golf Course	Gift Shop	German Restaurant	Gastropub
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Restaurant	Pharmacy	Pizza Place	Train Station	Hotel	Indian Restaurant	Italian Restaurant	Discount Store	Coffee Shop

Cluster one is the biggest cluster with 9 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Stores, Pubs, Restaurants and Train Station

Cluster 2

	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	10th Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Construction & Landscaping	Food	Deli / Bodega	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Fish & Chips Shop	French Restaurant	Grocery Store

The second cluster has one neighborhood which consists of Venues such as Construction, Food Shops and Departmental Stores.

Cluster 3

	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	10th Most Common Venue
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Turkish Restaurant	Food	Department Store	Discount Store	Farmers Market	Fast Food Restaurant
9	New Malden	Kingston upon Thames	51.405335	-0.263407	2	Gym	Gastropub	Office	Sushi Restaurant	Supermarket	Bar	Korean Restaurant	Indian Restaurant	Fast Food Restaurant	Department Store

The third cluster neighborhoods which consists of Venues such as Gyms, Pubs and Restaurants

Cluster 4

	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	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Gym / Fitness Center	Park	Bus Stop	Turkish Restaurant	Fish & Chips Shop	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Food
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Gym	Park	Bus Stop	Soccer Field	Fish & Chips Shop	Department Store	Discount Store	Farmers Market	Fast Food Restaurant	Food

The fourth cluster has two neighborhood in it, this has common venues such as Train Stations, Pubs, Restaurants, Electronics Stores etc.
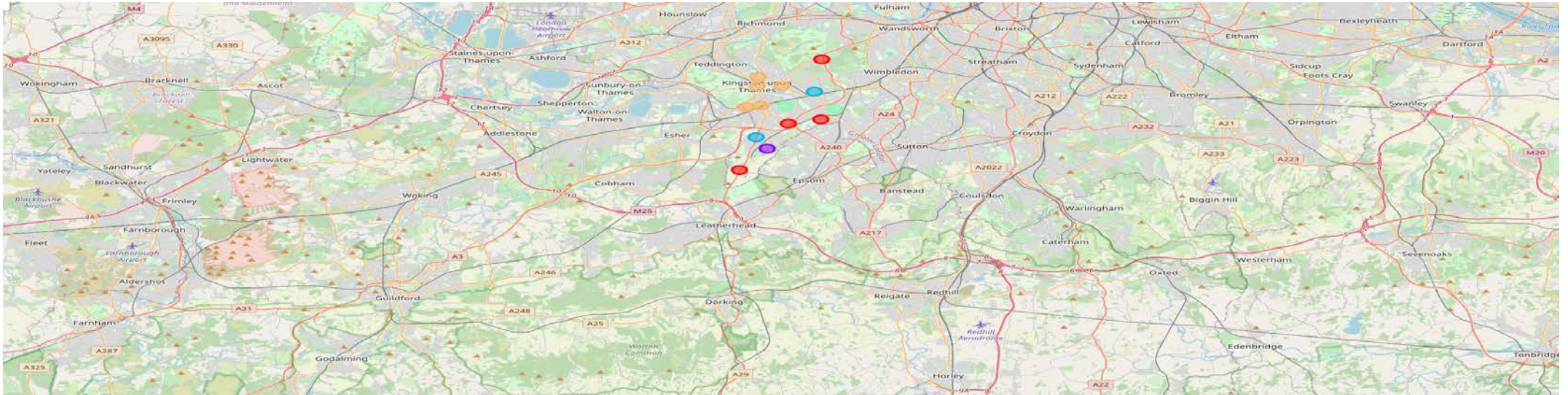
Cluster 5

	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	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	4	Pub	Park	Café	Supermarket	Spa	Gym / Fitness Center	Shop & Service	Fish & Chips Shop	Plaza	Hotel
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	4	Café	Pub	Burger Joint	Sushi Restaurant	Coffee Shop	Turkish Restaurant	French Restaurant	Gym / Fitness Center	Gift Shop	German Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	4	Indian Restaurant	Italian Restaurant	Food	Platform	Pub	Japanese Restaurant	Hotel	Coffee Shop	Hardware Store	Café
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	4	Indian Restaurant	Coffee Shop	Pub	Café	Gym	Chinese Restaurant	Fast Food Restaurant	Fish & Chips Shop	Restaurant	Italian Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	4	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Breakfast Spot	French Restaurant	Train Station	Gym / Fitness Center	Hotel

The fifth cluster has two neighborhood which consists of Restaurants, Coffee Shops, Pubs.

Cluster Neighbourhoods in a folium library.

Each cluster is colour-coded for the ease of presentation. We can see the majority of the neighbourhood falls in the red cluster which is the first cluster. Three neighbourhoods have their own cluster, i.e., blue purple and yellow. These are clusters 2, 3 and 5 the green cluster consists of two neighbourhoods which is 4th cluster.
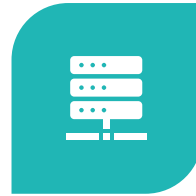
5. Discussion

THE AIM OF THE PROJECT WAS TO HELP BUSINESS INVESTORS WHO WOULD LIKE TO INVEST IN CHAIN OF RESTAURANTS IN THE SAFEST BOROUGH OF LONDON.

THEY WOULD ALSO LIKE TO EXAMINE THE NEIGHBOURHOODS GIVING THEM THE IDEA WHETHER IT IS A SUSTAINABLE FOR RESTAURANT BUSINESS.

THIS COULD BE LOOKING AT TRANSPORT LINKS AND THE NATURE OF OTHER BUSINESSES IN THE AREA.

IF A BUSINESS OWNER IS LOOKING FOR GOOD CONNECTIONS THEN **CLUSTER 1** AND **CLUSTER 4** WOULD BE THE CHOICE.

IF A BUSINESS OWNER IS LOOKING FOR THE PRESENCE OF OTHER BUSINESS IN THE AREA, INCLUDING PUBS RESTAURANTS AND COFFEE SHOPS THEN **CLUSTERS 2, 3, 5** WOULD BE THE IDEAL CHOICE.

BASED ON THE RECOMMENDATIONS, A BUSINESS OWNER WOULD BE ABLE TO DECIDE WHERE TO INVEST THE LOCATION WHERE TO INVEST IN RESTAURANT.

6.Conclusion

This project helps to get a holistic view into the investment decisions that might be critical before investing in a hospitality business.

The venue has got to be safe with regards to crime, it also has to be a place where the potential customers would be able to visit easily. It should also mean the staff work in the restaurant should also be able to commute and get to work in the easiest possible manner.

This project has given the options restaurant owner in terms of the safest boroughs in London and also the location details where are prospective restaurant might thrive the City of London.

Thank you.