Clustering the Neighbourhoods of London

1. Introduction

London is a quite the popular tourist and vacation destination for people all around the world. It is diverse and multicultural and offer a wide variety of experiences that is widely sought after. We try to group the neighbourhoods of London and draw insights to what they look like now.

2. Business Problem

The aim is to help tourists choose their destinations depending on the experiences that the neighbourhoods have to offer and what they would want to have. This also helps people make decisions if they are thinking about migrating to London or even if they want to relocate neighbourhoods within the city. Our findings will help stakeholders make informed decisions and address any concerns they have including the different kinds of cuisines, provision stores and what the city has to offer.

3. Data

We require geographical location data for both London and Paris. Postal codes in each city serve as a starting point. Using Postal codes we use can find out the neighborhoods, boroughs, venues and their most popular venue categories.

London

We scrape our data from https://en.wikipedia.org/wiki/List_of_areas_of_London on:

- borough : Name of Neighbourhood
- town : Name of borough
- post_code : Postal codes for London.

and we use ArcGIS API:

- ArcGIS API

ArcGIS Online enables you to connect people, locations, and data using interactive maps. Work with smart, data-driven styles and intuitive analysis tools that deliver location intelligence. Share your insights with the world or specific groups.

More specifically, we use ArcGIS to get the geo locations of the neighbourhoods of London. The following columns are added to our initial dataset which prepares our data.

latitude : Latitude for Neighbourhood longitude : Longitude for Neighbourhood

Foursquare API Data

We will need data about different venues in different neighbourhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a

location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

- Neighbourhood : Name of the Neighbourhood
- Neighbourhood Latitude : Latitude of the Neighbourhood
- Neighbourhood Longitude : Longitude of the Neighbourhood
- Venue : Name of the Venue
- Venue Latitude : Latitude of Venue
- Venue Longitude : Longitude of Venue
- Venue Category : Category of Venue

4. Methodology

The approach taken here is to explore each of the cities individually, plot the map to show the neighbourhoods being considered and then build our model by clustering all of the similar neighbourhoods together and finally plot the new map with the clustered neighbourhoods. We draw insights and then compare and discuss our findings.

Neighbourhoods of London

We begin to start collecting and refining the data needed for our business solution to work.

	borough	town	post_code
0	Bexley, Greenwich	LONDON	SE2
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4
2	Croydon	CROYDON	CR0
3	Croydon	CROYDON	CR0
4	Bexley	BEXLEY, SIDCUP	DA5, DA14
...	...	...	...
526	Greenwich	LONDON	SE18
527	Sutton, Kingston upon Thames	WORCESTER PARK	KT4
528	Hammersmith and Fulham	LONDON	W12
529	Hillingdon	HAYES	UB4
530	Hillingdon	WEST DRAYTON	UB7

531 rows × 3 columns

We currently have 531 records and 3 columns of our data. It's time to perform Feature Engineering

Feature Engineering

	borough	town	post_code
0	Bexley, Greenwich	LONDON	SE2
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4
6	City	LONDON	EC3
7	Westminster	LONDON	WC2
9	Bromley	LONDON	SE20
...	...	...	...
521	Redbridge	LONDON	IG8, E18
522	Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8
525	Barnet	LONDON	N12
526	Greenwich	LONDON	SE18
528	Hammersmith and Fulham	LONDON	W12

308 rows × 3 columns

We now have only 308 rows. We can proceed with our further steps. Getting some descriptive statistics.

Geolocations of the London Neighbourhoods

ArcGis API

We need to get the geographical co-ordinates for the neighbourhoods to plot out map. We will use the arcgis package to do so.

Arcgis doesn't have a limitation on the number of API calls made so it fits our use case perfectly.

Defining London arcgis geocode function to return latitude and longitude

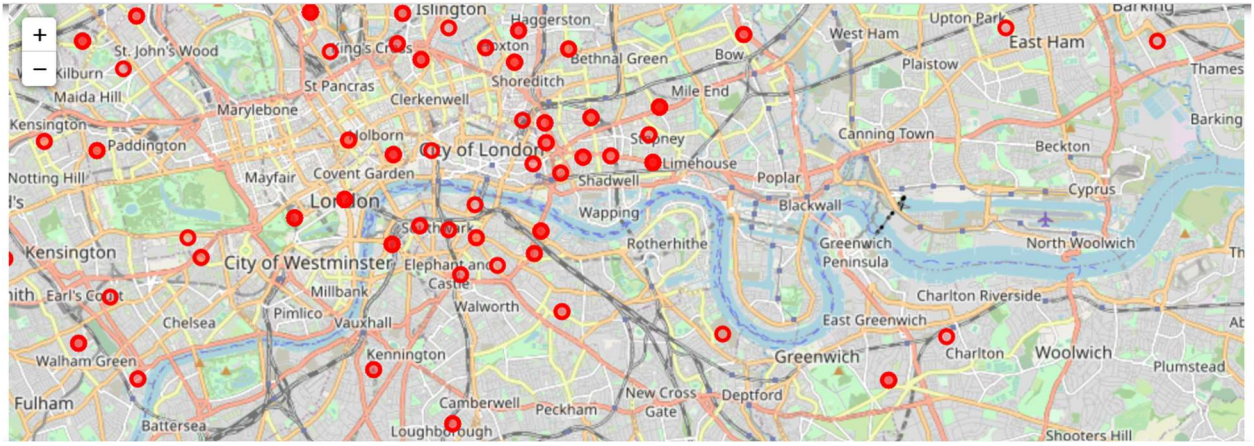
We proceed with Merging our source data with the geographical co-ordinates to make our dataset ready for the next stage

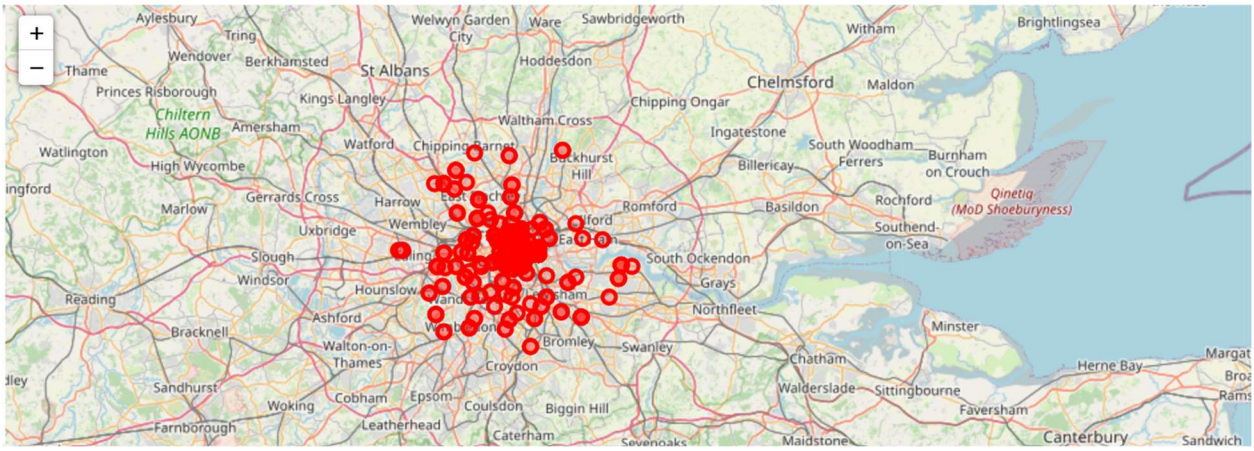
	borough	town	post_code	latitude	longitude
0	Bexley, Greenwich	LONDON	SE2	51.499741	0.124061
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.497765	-0.255852
6	City	LONDON	EC3	51.513145	-0.078733
7	Westminster	LONDON	WC2	51.514625	-0.114860
9	Bromley	LONDON	SE20	51.482490	0.119194
...	...	...	...	...	...
521	Redbridge	LONDON	IG8, E18	51.511800	-0.071290
522	Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8	51.507408	-0.127699
525	Barnet	LONDON	N12	51.542635	-0.098581
526	Greenwich	LONDON	SE18	51.503129	-0.108025
528	Hammersmith and Fulham	LONDON	W12	51.515085	-0.242696

308 rows × 5 columns

Visualize the Map of London

To help visualize the Map of London and the neighbourhoods in London, we make use of the folium package.





Venues in London

To proceed with the next part, we need to define Foursquare API credentials.

Using Foursquare API, we are able to get the venue and venue categories around each neighbourhood in London.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Category
0	Bexley, Greenwich	51.499741	0.124061	Southmere Lake	Lake
1	Ealing, Hammersmith and Fulham	51.497765	-0.255852	Lara Restaurant	Mediterranean Restaurant
2	Ealing, Hammersmith and Fulham	51.497765	-0.255852	Hack & Veldt	Coffee Shop
3	Ealing, Hammersmith and Fulham	51.497765	-0.255852	Good Boy Coffee	Coffee Shop
4	Ealing, Hammersmith and Fulham	51.497765	-0.255852	Chief Coffee	Coffee Shop

```
venues_in_London.shape
```

```
(12339, 5)
```

This will definitely make the clustering interesting.

Model Building

K Means

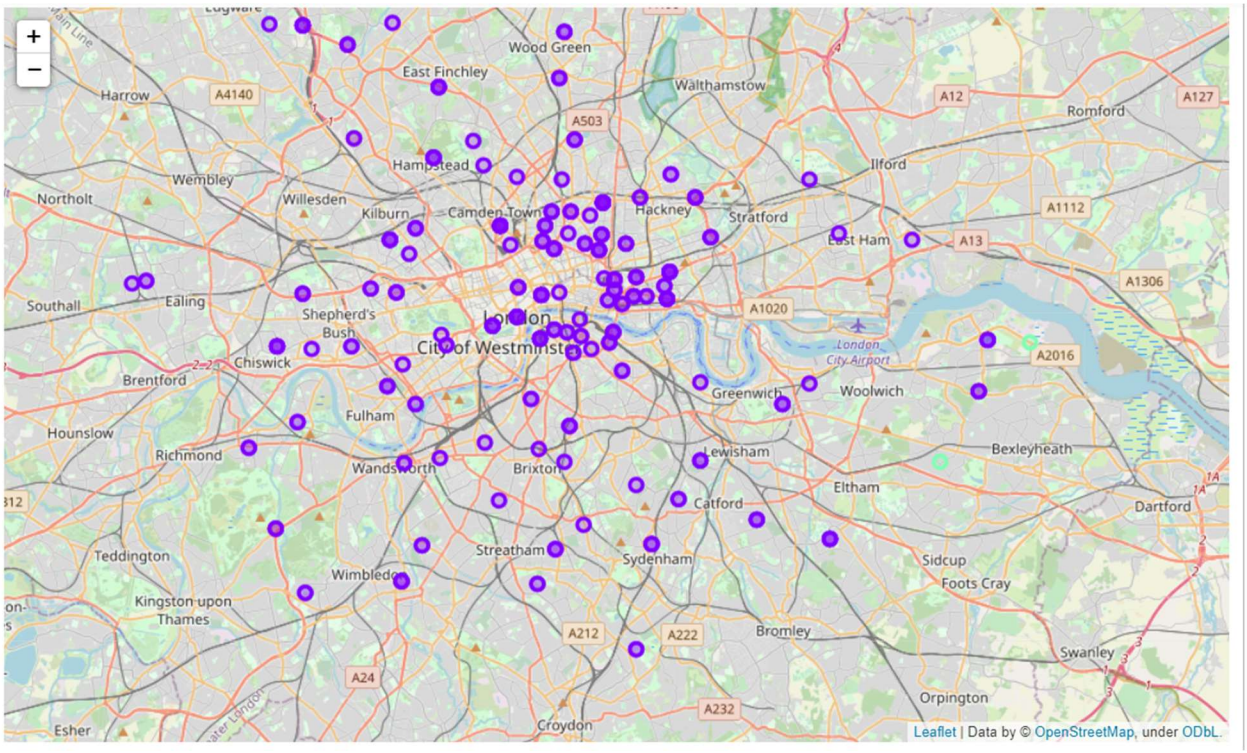
Let's cluster the city of london to roughly 5 to make it easier to analyze.

We use the K Means clustering technique to do so.

	borough	town	post_code	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	Bexley, Greenwich	LONDON	SE2	51.499741	0.124061	2	Lake	Yoga Studio	Filipino Restaurant	Event Space	Exhibit	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.497765	-0.255852	1	Coffee Shop	Park	Playground	Health Food Store	Fish & Chips Shop	French Restaurant	Fruit & Vegetable Store	Mediterranean Restaurant	Garden Center	Comedy Club
6	City	LONDON	EC3	51.513145	-0.078733	1	Coffee Shop	Hotel	Italian Restaurant	Pub	Gym / Fitness Center	Wine Bar	Cocktail Bar	French Restaurant	Restaurant	Sandwich Place
7	Westminster	LONDON	WC2	51.514625	-0.114860	1	Hotel	Pub	Coffee Shop	Café	French Restaurant	Sandwich Place	Theater	Lounge	Art Gallery	Restaurant
9	Bromley	LONDON	SE20	51.482490	0.119194	1	Bus Station	Campground	Forest	Athletics & Sports	Café	Gym / Fitness Center	Coffee Shop	Convenience Store	Park	Japanese Restaurant

Visualizing the clustered neighbourhood

Let's plot the clusters



Results and Discussion

The neighbourhoods of London are very multicultural. There are a lot of different cusines including Indian, Italian, Turkish and Chinese. London seems to take a step further in this direction by having a lot of Restaurants, bars, juice bars, coffee shops, Fish and Chips shop and Breakfast spots. It has a lot of shopping options too with that of the Flea markets, flower shops, fish markets, Fishing stores, clothing stores. The main modes of transport seem to be Buses and trains. For leisure, the neighbourhoods are set up to have lots of parks, golf courses, zoo, gyms and Historic sites.

Overall, the city of London offers a multicultural, diverse and certainly an entertaining experience.

Conclusion

The purpose of this project was to explore the city of London, and see how attractive it is to potential tourists and migrants. We explored the city based on its postal codes and then extrapolated the common venues present in each of the neighbourhoods finally concluding with clustering similar neighbourhoods together.

We could see that each of the neighbourhoods in London have a wide variety of experiences to offer which is unique in it's own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion.

London seems to offer a vacation stay or a romantic gateaway with a lot of places to explore, beautiful landscapes and a wide variety of culture.