Neighborhoods of London

Introduction

London is a quite 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.

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.

Data Description

We require geolocation data for London. Postal codes of city serves as a starting point. Using Postal codes we use can find out the neighbourhoods, boroughs, venues and their most popular venue categories.

London

To derive our solution, We scrape our data from https://en.wikipedia.org/wiki/List_of_areas_of_London

This wikipedia page has information about all the neighbourhoods, we limit it London.

1. borough : Name of Neighbourhood
2. town : Name of borough
3. post_code : Postal codes for London

This wikipedia page lacks information about the geographical locations. To solve this problem 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.

1. latitude : Latitude for Neighbourhood
2. longitude : Longitude for Neighbourhood

Based on all the information collected for London, we have sufficient data to build our model. We cluster the neighbourhoods together based on similar venue categories. We then present our observations and findings. Using this data, our stakeholders can take the necessary decision.

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:

1. Neighbourhood : Name of the Neighbourhood
2. Neighbourhood Latitude : Latitude of the Neighbourhood
3. Neighbourhood Longitude : Longitude of the Neighbourhood
4. Venue : Name of the Venue
5. Venue Latitude : Latitude of Venue
6. Venue Longitude : Longitude of Venue
7. Venue Category : Category of Venue

Based on all the information collected, we have sufficient data to build our model. We cluster the neighbourhoods together based on similar venue categories. We then present our observations and findings. Using this data, our stakeholders can take the necessary decision.

Methodology

```
import pandas as pd
import requests
import numpy as np
import matplotlib.cm as cm
import matplotlib.colors as colors
import folium
from sklearn.cluster import KMeans
```

Package breakdown:

1. Pandas : To collect and manipulate data in JSON and HTMl and then data analysis
2. requests : Handle http requests
3. matplotlib : Detailing the generated maps
4. folium : Generating maps of London
5. sklearn : To import Kmeans which is the machine learning model that we are using.

The approach taken here is to explore the city, 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.

Data Collection

```
url_london = "https://en.wikipedia.org/wiki/List_of_areas_of_London"
wiki_london_url = requests.get(url_london)
wiki_london_data = pd.read_html(wiki_london_url.text)
wiki_london_data = wiki_london_data[1]
wiki_london_data
```

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
526	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
527	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
529	Yeading	Hillingdon	HAYES	UB4	020	TQ115825
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

531 rows × 6 columns

Data Preprocessing

For London, We replace the spaces with underscores in the title. The *borough* column has numbers within square brackets that we remove using:

```
wiki_london_data.rename(columns= lambda x: x.strip().replace(" ", "_"), inplace=True)
wiki_london_data
```

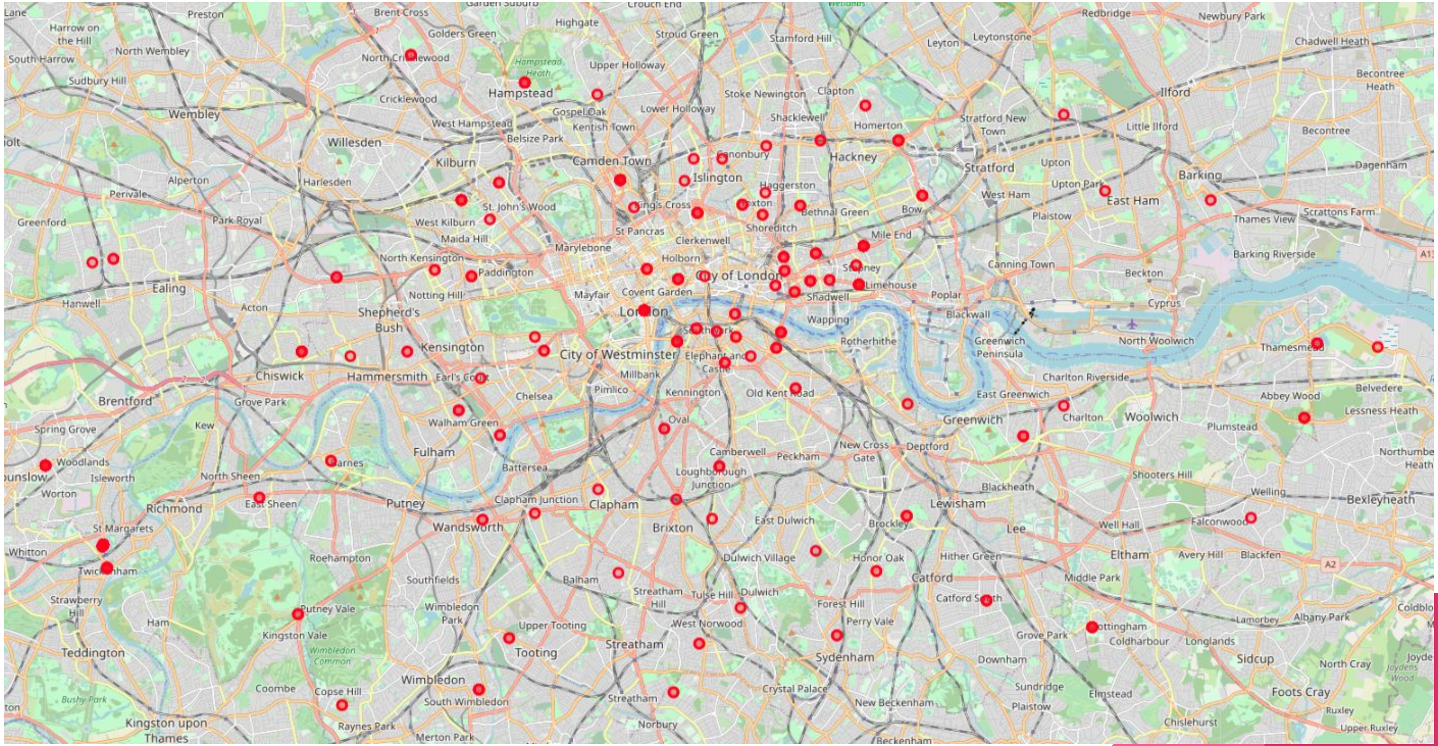
Feature Selection

We need only the boroughs, Postal codes, Post town for further steps. We can drop the locations, dial codes and OS grid.

```
df1 = wiki_london_data.drop( [ wiki_london_data.columns[0], wiki_london_data.columns[4], wiki_london_data.columns[5] ], axis=1)
df1.head()
```

	London borough	Post_town	Postcode district
0	Bexley, Greenwich [7]	LONDON	SE2
1	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4
2	Croydon[8]	CROYDON	CR0
3	Croydon[8]	CROYDON	CR0
4	Bexley	BEXLEY, SIDCUP	DA5, DA14

Visualizing the Neighborhoods in London



One Hot Encoding

Since we are trying to find out what are the different kinds of venue categories present in each neighbourhood and then calculate the top 10 common venues to base our similarity on, we use the One Hot Encoding to work with our categorical datatype of the venue categories. This helps to convert the categorical data into numeric data.

We won't be using label encoding in this situation since label encoding might cause our machine learning model to have a bias or a sort of ranking which we are trying to avoid by using One Hot Encoding.

We perform one hot encoding and then calculate the mean of the grouped venue categories for each of the neighbourhoods.

Top Venues in the Neighborhoods

	Neighbourhood	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	Barnet	Pub	Coffee Shop	Bus Stop	Park	Hotel	Italian Restaurant	BBQ Joint	Breakfast Spot	Café	Grocery Store
1	Barnet, Brent, Camden	Pizza Place	Park	Furniture / Home Store	Yoshoku Restaurant	Fast Food Restaurant	Event Space	Exhibit	Fabric Shop	Falafel Restaurant	Farm
2	Bexley	Lake	Motorcycle Shop	Convenience Store	Gym / Fitness Center	Pet Store	Flea Market	Farmers Market	Exhibit	Food & Drink Shop	Fabric Shop
3	Bexley, Greenwich	Indian Restaurant	Print Shop	Pizza Place	Grocery Store	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Fabric Shop	Falafel Restaurant
4	Bexley, Greenwich	Lake	Yoshoku Restaurant	Escape Room	Event Space	Exhibit	Fabric Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant

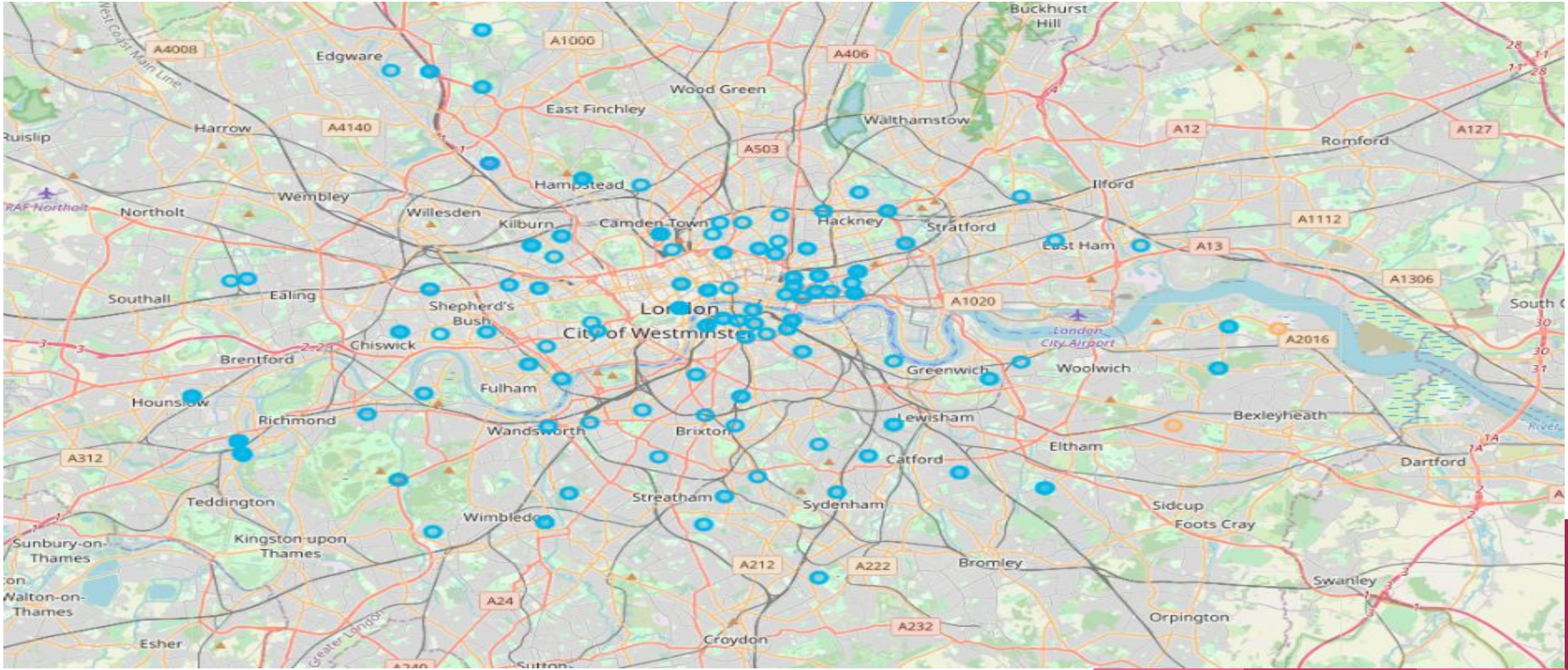
Model Building K-Means

Moving on to the most exciting part - **Model Building!** We will be using K-Means Clustering Machine learning algorithm to cluster similar neighbourhoods together. We will be going with the number of clusters as 5.

	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	3	Lake	Yoshoku Restaurant	Escape Room	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	2	Coffee Shop	Park	Grocery Store	Playground	Comedy Club	Fish & Chips Shop	Spa	Metro Station	French Restaurant	Mediterranean Restaurant
6	City	LONDON	EC3	51.513145	-0.078733	2	Coffee Shop	Pub	Hotel	Sandwich Place	Gym / Fitness Center	Italian Restaurant	Cocktail Bar	French Restaurant	Restaurant	Wine Bar
7	Westminster	LONDON	WC2	51.514625	-0.114860	2	Pub	Hotel	Theater	Coffee Shop	Sandwich Place	Cocktail Bar	Japanese Restaurant	Garden	Plaza	Chinese Restaurant
9	Bromley	LONDON	SE20	51.482490	0.119194	2	Bus Stop	Forest	Campground	Athletics & Sports	Motorcycle Shop	Gym / Fitness Center	Convenience Store	Pet Store	Fast Food Restaurant	Fabric Shop

+ Code + Text

Visualizing the Clustered Neighborhoods



Examining Clusters

Cluster 1

```
london_data_nonan.loc[london_data_nonan['Cluster Labels'] == 1, london_data_nonan.columns[[1] + list(range(5, london_dat
a_nonan.shape[1]))]]
```

Cluster 2

```
london_data_nonan.loc[london_data_nonan['Cluster Labels'] == 2, london_data_nonan.columns[[1] + list(range(5, london_dat
a_nonan.shape[1]))]]
```

Cluster 3

```
london_data_nonan.loc[london_data_nonan['Cluster Labels'] == 3, london_data_nonan.columns[[1] + list(range(5, london_dat
a_nonan.shape[1]))]]
```

Cluster 4

```
london_data_nonan.loc[london_data_nonan['Cluster Labels'] == 4, london_data_nonan.columns[[1] + list(range(5, london_dat
a_nonan.shape[1]))]]
```

Cluster 5

```
london_data_nonan.loc[london_data_nonan['Cluster Labels'] == 5, london_data_nonan.columns[[1] + list(range(5, london_dat
a_nonan.shape[1]))]]
```

Results and Discussion

The neighbourhoods of London are very multicultural. There are a lot of different cuisines including Indian, Italian, Turkish and Chinese. London seems to have 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. For leisure, the neighbourhoods are set up to have lots of parks, gyms and art gallery.

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 their 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 city had 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 getaway with a lot of places to explore, beautiful landscapes and a wide variety of culture. Overall, it's up to the stakeholders to decide which experience they would prefer more and which would more to their liking.