Applied Data Science Capstone
Presentation

Battle of the Neighbourhoods

Paris and London

1. **Introduction and Business Problem**
   The purpose of this project is to help tourists choose their destinations depending on what experiences the neighbourhoods have to offer and what the neighbourhoods would want to have. This also helps people make decisions if they are thinking about migrating to Paris or London or even if they want to relocate between neighbourhoods within the city.

   It also will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life. 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.
   - Work Flow
     Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.
   - Clustering Approach
     To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

2. **Data Description**
   We require geolocation data for both Paris and London. Postal codes in each city serve as a starting point. Using Postal codes we can find out the neighbourhoods, boroughs, venues and their most popular venue categories.
   - Paris
     To derive our solution, We scrape our data from
     https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e

     The JSON file has data about all the neighbourhoods in France, we limit it only Paris.

     1. *postal_code* : Postal codes for France
     2. *nom_comm* : Name of Neighbourhoods in France
     3. *nom_dept* : Name of the boroughs, equivalent to towns in France
     4. *geo_point_2d* : Tuple containing the latitude and longitude of the Neighbourhoods.

   - 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 only 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 us 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. 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.

  4. *latitude* : Latitude for Neighbourhood
  5. *longitude* : Longitude for Neighbourhood

  Based on all the information collected for both Paris and 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 obtain information we 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

- Libraries Which are Used to Develope the Project:

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- Matplotlib: Python Plotting Module.

3. Methodology

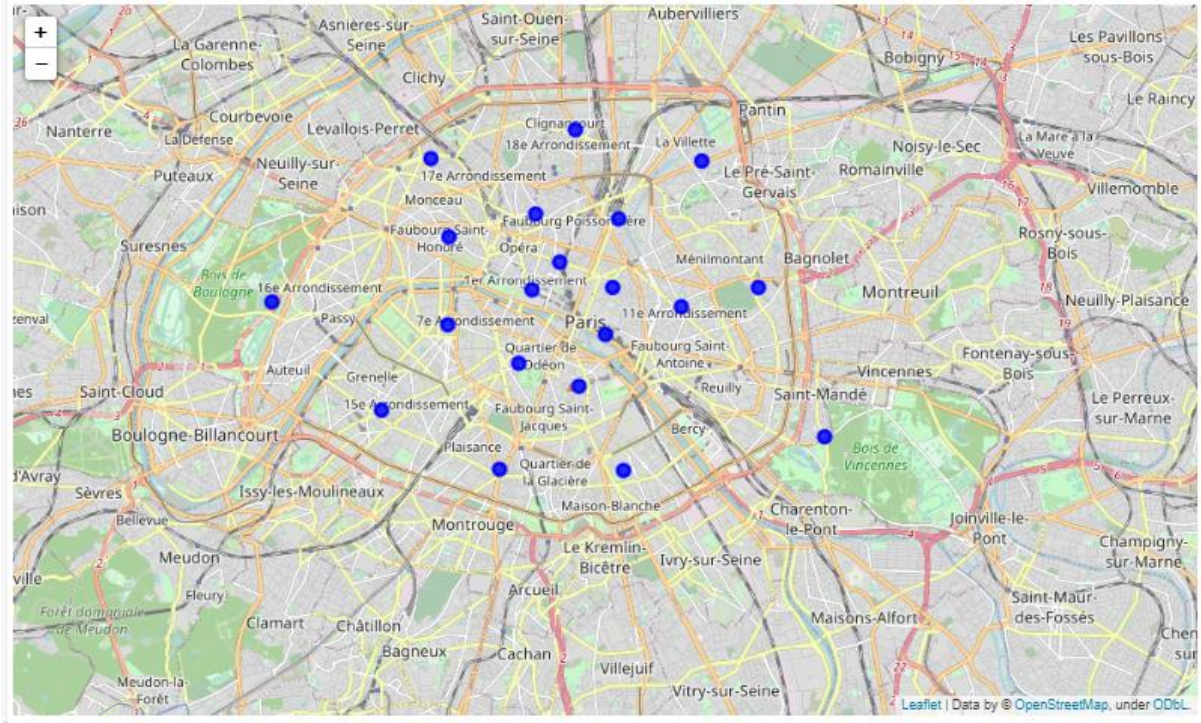

Image 1 Neighbourhood of Paris

	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	PARIS-10E-ARRONDISSEMENT	French Restaurant	Hotel	Bistro	Café	Coffee Shop	Pizza Place	Indian Restaurant	Italian Restaurant	Asian Restaurant	Japanese Restaurant
1	PARIS-11E-ARRONDISSEMENT	Restaurant	Vietnamese Restaurant	French Restaurant	Bakery	Café	Italian Restaurant	Pastry Shop	Bar	Gastropub	Mexican Restaurant
2	PARIS-12E-ARRONDISSEMENT	Zoo Exhibit	Bistro	Monument / Landmark	Supermarket	Zoo	Argentinian Restaurant	Garden	Furniture / Home Store	Frozen Yogurt Shop	French Restaurant
3	PARIS-13E-ARRONDISSEMENT	Asian Restaurant	Vietnamese Restaurant	French Restaurant	Chinese Restaurant	Thai Restaurant	Hotel	Juice Bar	Dessert Shop	Sandwich Place	Coffee Shop
4	PARIS-14E-ARRONDISSEMENT	French Restaurant	Hotel	Bistro	Japanese Restaurant	Tea Room	Pizza Place	Pet Store	Café	Fast Food Restaurant	Food & Drink Shop

Image 2 Most Common Venue of Paris

K Means

Let's cluster the city of Paris to roughly 5 to make it easier to analyze. We use the K Means clustering technique to do so.

```
#set number of clusters
k_num_clusters = 5

Paris_grouped_clustering = Paris_grouped.drop('Neighbourhood', 1)

#run k-means clustering
kmeans_Paris = KMeans(n_clusters=k_num_clusters, random_state=0).fit(Paris_grouped_clustering)
kmeans_Paris
```

Image 3 Model Building

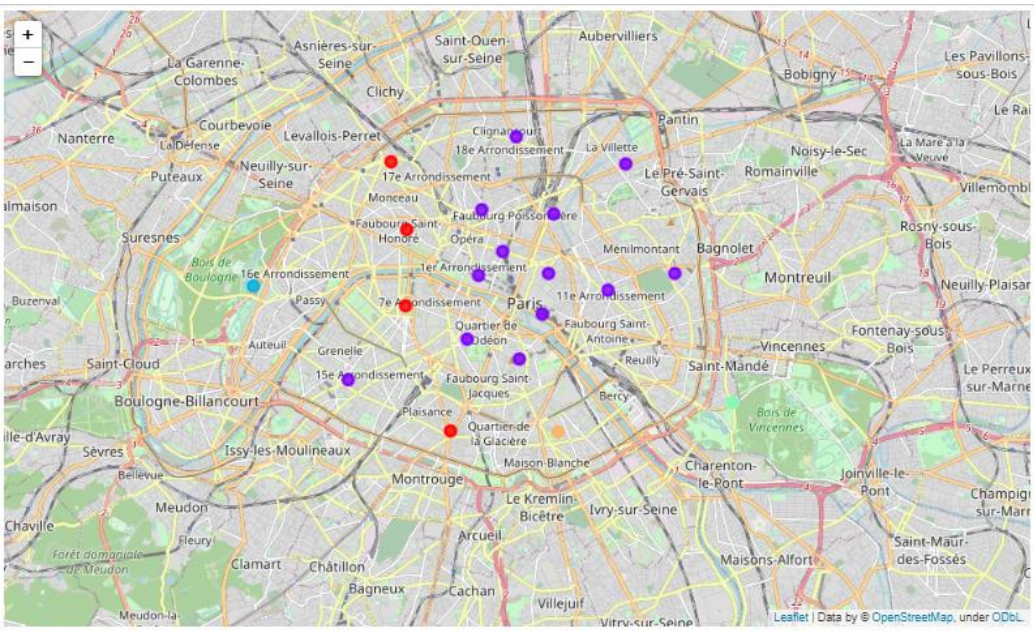

Image 4 Clustered Neighbourhood of Paris

	nom_comm	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	PARIS-9E-ARRONDISSEMENT	1	French Restaurant	Hotel	Japanese Restaurant	Bistro	Bakery	Lounge	Wine Bar	Cocktail Bar	Restaurant	Café
1	PARIS-2E-ARRONDISSEMENT	1	French Restaurant	Cocktail Bar	Coffee Shop	Wine Bar	Bakery	Hotel	Creperie	Italian Restaurant	Thai Restaurant	Sandwich Place
2	PARIS-11E-ARRONDISSEMENT	1	Restaurant	Vietnamese Restaurant	French Restaurant	Bakery	Café	Italian Restaurant	Pastry Shop	Bar	Gastropub	Mexican Restaurant
3	PARIS-3E-ARRONDISSEMENT	1	French Restaurant	Japanese Restaurant	Bakery	Italian Restaurant	Gourmet Shop	Coffee Shop	Cocktail Bar	Art Gallery	Sandwich Place	Wine Bar
4	PARIS-6E-ARRONDISSEMENT	1	Pastry Shop	Chocolate Shop	French Restaurant	Bakery	Wine Bar	Plaza	Fountain	Tea Room	Sandwich Place	Lebanese Restaurant
5	PARIS-4E-ARRONDISSEMENT	1	French Restaurant	Ice Cream Shop	Park	Clothing Store	Hotel	Wine Bar	Pastry Shop	Pedestrian Plaza	Plaza	Bistro
6	PARIS-1ER-ARRONDISSEMENT	1	Japanese Restaurant	French Restaurant	Plaza	Hotel	Italian Restaurant	Café	Art Museum	Bakery	Garden	Korean Restaurant
11	PARIS-5E-ARRONDISSEMENT	1	French Restaurant	Hotel	Italian Restaurant	Plaza	Bakery	Coffee Shop	Pub	Café	Science Museum	Bar
12	PARIS-19E-ARRONDISSEMENT	1	French Restaurant	Bar	Hotel	Supermarket	Bistro	Seafood Restaurant	Beer Bar	Brewery	Burger Joint	Café
13	PARIS-20E-ARRONDISSEMENT	1	Bakery	Plaza	Bistro	Japanese Restaurant	French Restaurant	Supermarket	Bar	Italian Restaurant	Food & Drink Shop	Café
14	PARIS-10E-ARRONDISSEMENT	1	French Restaurant	Hotel	Bistro	Café	Coffee Shop	Pizza Place	Indian Restaurant	Italian Restaurant	Asian Restaurant	Japanese Restaurant
16	PARIS-18E-ARRONDISSEMENT	1	French Restaurant	Bar	Italian Restaurant	Hotel	Café	Restaurant	Plaza	Pizza Place	Bistro	Vietnamese Restaurant
18	PARIS-15E-ARRONDISSEMENT	1	Hotel	Italian Restaurant	French Restaurant	Indian Restaurant	Japanese Restaurant	Thai Restaurant	Restaurant	Lebanese Restaurant	Brasserie	Park

Image 5 Cluster 1 of Paris

	nom_comm	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
15	PARIS-16E-ARRONDISSEMENT	2	Plaza	Lake	Park	French Restaurant	Bus Station	Art Museum	Boat or Ferry	Bike Rental / Bike Share	Zoo Exhibit	Fountain

Image 6 Cluster 2 of Paris

	nom_comm	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
10	PARIS-12E-ARRONDISSEMENT	3	Zoo Exhibit	Bistro	Monument / Landmark	Supermarket	Zoo	Argentinian Restaurant	Garden	Furniture / Home Store	Frozen Yogurt Shop	French Restaurant

Image 7 Cluster 3 of Paris

	nom_comm	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
9	PARIS-13E-ARRONDISSEMENT	4	Asian Restaurant	Vietnamese Restaurant	French Restaurant	Chinese Restaurant	Thai Restaurant	Hotel	Juice Bar	Dessert Shop	Sandwich Place	Coffee Shop

Image 8 Cluster 4 of Paris

	nom_comm	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
7	PARIS-17E-ARRONDISSEMENT	5	French Restaurant	Hotel	Italian Restaurant	Bakery	Café	Bistro	Restaurant	Japanese Restaurant	Bus Stop	Plaza
8	PARIS-8E-ARRONDISSEMENT	5	French Restaurant	Hotel	Spa	Cocktail Bar	Grocery Store	Department Store	Corsican Restaurant	Resort	Park	Fast Food Restaurant
17	PARIS-7E-ARRONDISSEMENT	5	French Restaurant	Hotel	Café	Italian Restaurant	Plaza	History Museum	Cocktail Bar	Japanese Restaurant	Bistro	Irish Pub
19	PARIS-14E-ARRONDISSEMENT	5	French Restaurant	Hotel	Bistro	Japanese Restaurant	Tea Room	Pizza Place	Pet Store	Café	Fast Food Restaurant	Food & Drink Shop

Image 9 Cluster 5 of Paris

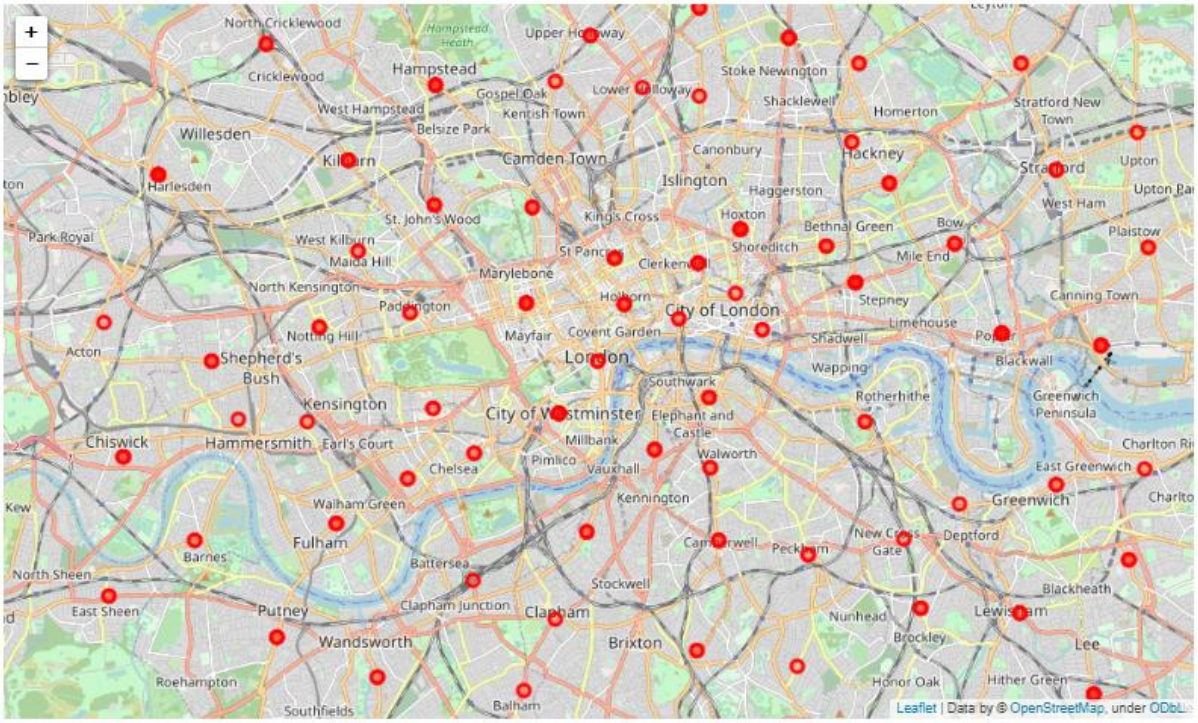

Image 10 Neighbourhood of London

	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	Coffee Shop	Café	Grocery Store	Bus Stop	Pub	Italian Restaurant	Supermarket	Pharmacy	Turkish Restaurant	Pizza Place
1	Barnet, Brent, Camden	Bakery	Bus Stop	Gym / Fitness Center	Clothing Store	Hardware Store	Supermarket	Film Studio	Exhibit	Falafel Restaurant	Farmers Market
2	Bexley	Supermarket	Historic Site	Coffee Shop	Platform	Convenience Store	Train Station	Bus Stop	Golf Course	Construction & Landscaping	Park
3	Bexley, Greenwich	Construction & Landscaping	Golf Course	Daycare	Bus Stop	Sports Club	Historic Site	Park	Massage Studio	Film Studio	Exhibit
4	Bexley, Greenwich	Supermarket	Platform	Convenience Store	Historic Site	Train Station	Coffee Shop	Zoo Exhibit	Event Space	Exhibit	Falafel Restaurant

Image 11 Most Common Venue of London

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.

```
#set number of clusters
k_num_clusters = 5

London_grouped_clustering = London_grouped.drop('Neighbourhood', 1)

#run k-means clustering
kmeans_london = KMeans(n_clusters=k_num_clusters, random_state=0).fit(London_grouped_clustering)
kmeans_london
```

Image 12 Model Building

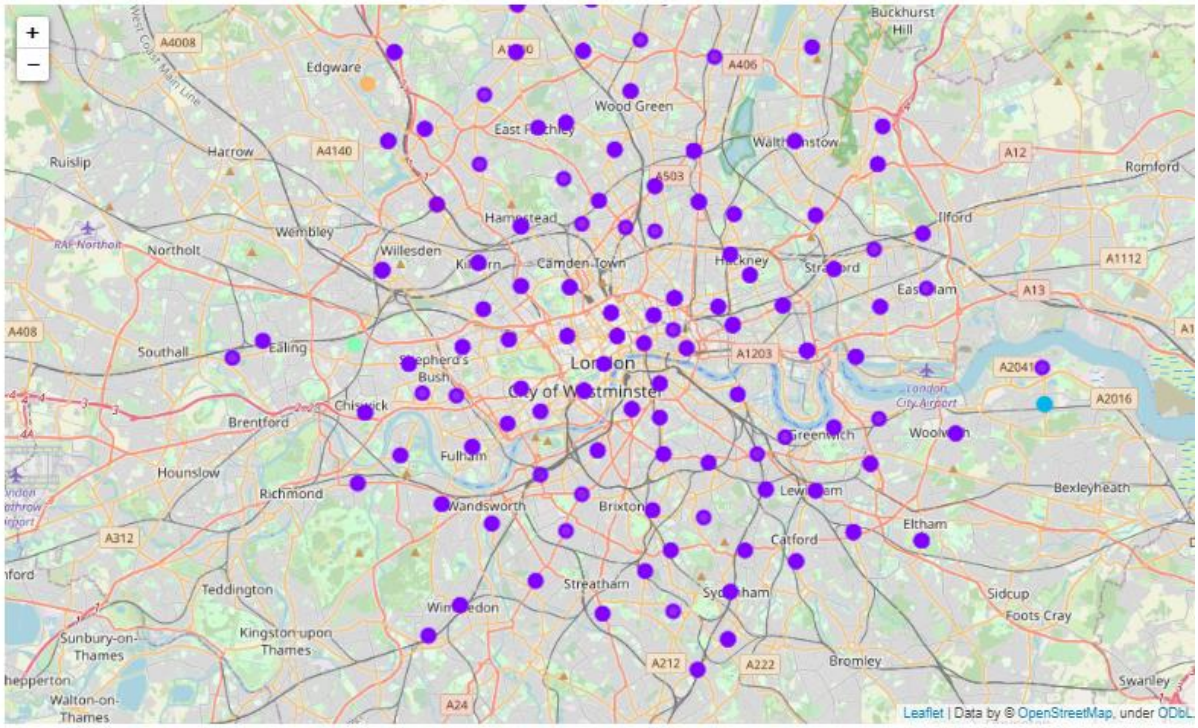

Image 13 Clustered Neighbourhood of London

	town	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	LONDON	1	Hotel	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pub	Restaurant	Cocktail Bar	Wine Bar	Falafel Restaurant	Sandwich Place
7	LONDON	1	Hotel	Coffee Shop	Sandwich Place	Café	Pub	Italian Restaurant	Hotel Bar	Theater	Restaurant	Sushi Restaurant
9	LONDON	1	Supermarket	Hotel	Fast Food Restaurant	Grocery Store	Convenience Store	Park	Gym / Fitness Center	Bus Stop	Fried Chicken Joint	Chinese Restaurant
10	LONDON	1	Coffee Shop	Pub	Café	Food Truck	Vietnamese Restaurant	Italian Restaurant	Park	Gym / Fitness Center	Hotel	Cocktail Bar
12	LONDON	1	Coffee Shop	Pub	Café	Food Truck	Vietnamese Restaurant	Italian Restaurant	Park	Gym / Fitness Center	Hotel	Cocktail Bar
...	...	...	...	...	...	...	...	...	...	...	...	...
521	LONDON	1	Café	Pub	Coffee Shop	Grocery Store	Metro Station	Park	Seafood Restaurant	Pizza Place	Bar	Bakery
522	LONDON, WOODFORD GREEN	1	Hotel	Café	Pub	Monument / Landmark	Plaza	Theater	Garden	Pharmacy	Bakery	English Restaurant
525	LONDON	1	Coffee Shop	Café	Grocery Store	Bus Stop	Pub	Italian Restaurant	Supermarket	Pharmacy	Turkish Restaurant	Pizza Place
526	LONDON	1	Pub	Grocery Store	Bus Stop	Indian Restaurant	Coffee Shop	Fish & Chips Shop	Turkish Restaurant	Golf Course	Supermarket	Park
528	LONDON	1	Pub	Café	Coffee Shop	Grocery Store	Hotel	Gastropub	Park	Pizza Place	Thai Restaurant	Italian Restaurant

300 rows × 12 columns

Image 14 Cluster 1 of London

	town	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	LONDON	2	Supermarket	Platform	Convenience Store	Historic Site	Train Station	Coffee Shop	Zoo Exhibit	Event Space	Exhibit	Falafel Restaurant
45	BEXLEYHEATH, LONDON	2	Supermarket	Historic Site	Coffee Shop	Platform	Convenience Store	Train Station	Bus Stop	Golf Course	Construction & Landscaping	Park
124	LONDON	2	Supermarket	Historic Site	Coffee Shop	Platform	Convenience Store	Train Station	Bus Stop	Golf Course	Construction & Landscaping	Park
291	LONDON, SIDCUP	2	Supermarket	Historic Site	Coffee Shop	Platform	Convenience Store	Train Station	Bus Stop	Golf Course	Construction & Landscaping	Park
505	LONDON	2	Supermarket	Historic Site	Coffee Shop	Platform	Convenience Store	Train Station	Bus Stop	Golf Course	Construction & Landscaping	Park

Image 15 Cluster 2 of London

	town	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	LONDON	3	Grocery Store	Train Station	Breakfast Spot	Park	Indian Restaurant	Department Store	Fish Market	Falafel Restaurant	Farmers Market	Fast Food Restaurant

Image 16 CLuster 3 of London

	town	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
377	HARROW, STANMOREEDGWARE, LONDON	4	Bakery	Gym	Metro Station	Fish & Chips Shop	Exhibit	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Film Studio

Image 17 Cluster 4 of London

	town	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
121	LONDON	5	Bakery	Bus Stop	Gym / Fitness Center	Clothing Store	Hardware Store	Supermarket	Film Studio	Exhibit	Falafel Restaurant	Farmers Market

Image 18 Cluster 5 of London

4. Results and Discussion

Paris is relatively smaller in size geographically. There are a lot of hangout spots including many restaurants, bars, and bistro. There are public transport in Paris such as buses, bikes, boats or ferries. For leisure and sight seeing, there are a lot of plazas, trails, parks, historic sites, clothing shops, art galleries and museums. It has a wide variety of cusines and eateries including French, Thai, Cambodian, Asian, Chinese etc. Overall, Paris seems like the relaxing vacation spot with a mix of lakes, historic spots and a wide variety of cusines to try out.

The neighbourhoods of London are very multicultural. There are a lot of different cusines including Indian, Italian, Turkish and Chinese. London seems takes a step further 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 such as lea markets, flower shops, fish markets, fishing stores, clothing stores. The public transport such as buses and trains. For leisure, the neighbourhoods are set up to have lots of parks, golf courses, zoo, gyms and historic sites. Overall, London offers a multicultural, diverse and certainly an entertaining experience.

5. Conclusion

The purpose of this project was to explore the cities of Paris and London and see how attractive it is to potential tourists and migrants. We explored both the cities 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 in both the cities have a wide variety of experiences to offer which is unique in it's own way.

Both Paris and London seem to offer a vacation stay or a romantic gateaway 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.