Capstone Project

The battle of Neighborhoods

Final Report

Carlos Molina

1. Introduction

1.1 Problem Relevance

Our client is an Investor Group, who owns several Companies in United States and they want to open two new Businesses Branches in Paris and London. They want to be able to share resources and knowledge across both Business. They want to identify what kind of businesses are successful on both cities and to identify neighborhoods with similar profiles and characteristics.

As a way to leverage what was learned in previous modules from this course, I had chosen to compare how similar or dissimilar are two important European Cities, Financial Capitals of two Countries. As these two cities are geographically near, I chose to compare Paris and London, by obtaining the number and type of Foursquare Venues close to the center coordinates (Latitude and Longitude) of each Neighborhood.

As cities are geographically near, chosen cities, would allow using the K-Means Clustering and Geographical Information viewable in the same map. Clusters will allow grouping Neighborhoods, based on similar characteristics (frequency and types of Foursquare Venues).

1.2 Problem Audience

Our stakeholders are Finance executives who will coordinate a market and business viability studies to determine the expected market and return of investment for the new Business, once our team provides a suggestion and justification on which business should they open and on which neighborhoods.

2. Data

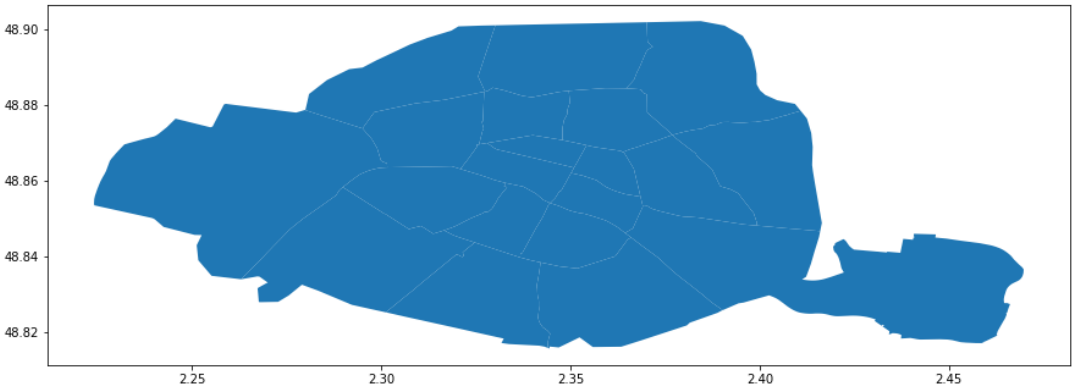
This exercise was a great opportunity to leverage what we learned in previous courses. First step was to obtain, clean and consolidate the data. The following sections summarize the process followed.

2.1 Geographic Data Gathering: Paris

In order to obtain Paris Data, I found out that Paris consists in twelve Districts (arrondissements). I had subscribed free to French Post Office (Le Groupe La Poste). Their address is https://datanova.laposte.fr/page/accueil/

From the site, I had obtained two files: arrondissements.json (geographic JSON File) and arrondissements.zip, including arrondissements.dbf arrondissements.prj, arrondissements.shp and arrondissements.shx Shape Files. These files contain the coordinates of the polygons forming the different districts, as well as the names of the districts and latitude and longitude of their centers.

The following figure shows the Paris districts and their coordinates.



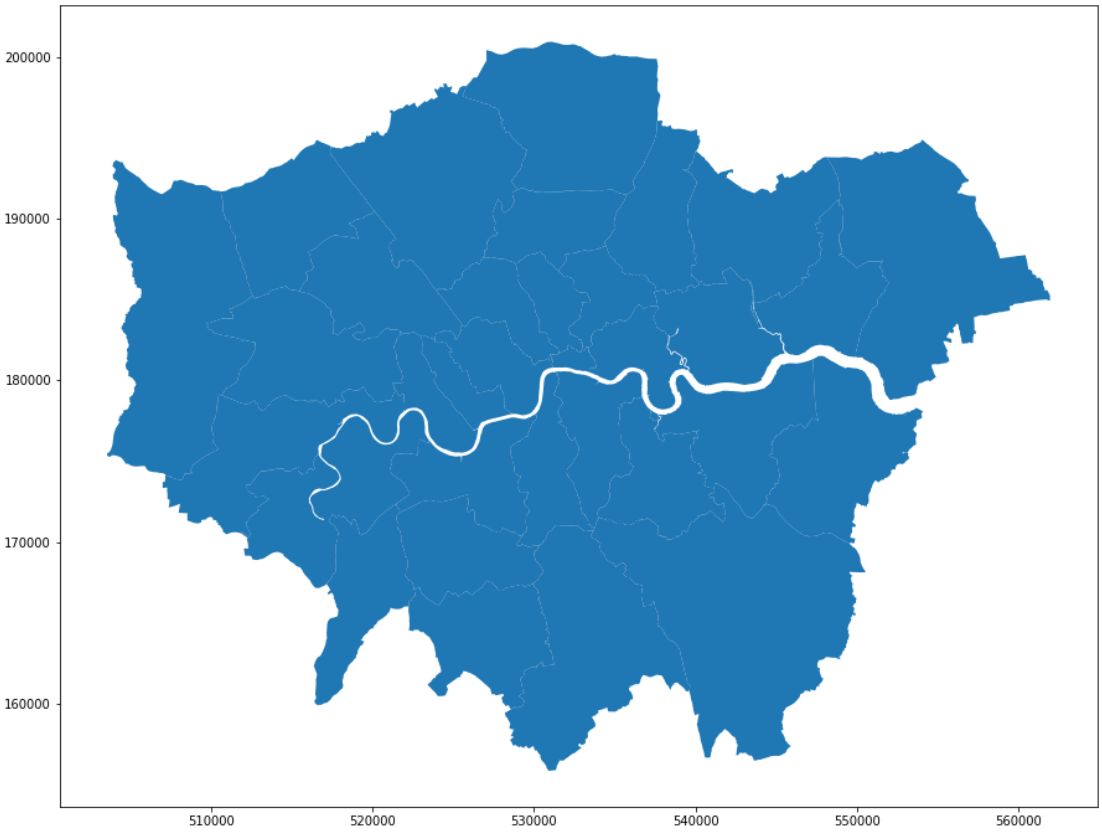
2.2 Geographic Data Gathering: London

London Geographic Data was obtained in Greater London Authority (GLA) Site is located in https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london I had downloaded statistical-gis-boundaries-london.zip and used the following relevant files:

- London_Borough_Excluding_MHW.dbf
- London_Borough_Excluding_MHW.GSS_CODE.atx
- London_Borough_Excluding_MHW.NAME.atx
- London_Borough_Excluding_MHW.prj
- London_Borough_Excluding_MHW.sbn
- London_Borough_Excluding_MHW.sbx
- London_Borough_Excluding_MHW.shp
- London_Borough_Excluding_MHW.shp.xml
- London_Borough_Excluding_MHW.shx

To verity regions of London, I had found a very interesting Article by Benjamin Coley, with a walkthrough on how to leverage Geopandas, Pandas and Matplotlib. Article is called "Let's make a map! Using Geopandas, Pandas and Matplotlib to make a Choropleth map" and it can be found on https://towardsdatascience.com/lets-make-a-map-using-geopandas-pandas-and-matplotlib-to-make-a-chloropleth-map-dddc31c1983d

By using GeoPandas, London Geographic Information was used to create a GeoPandas Data Frame, plotted below.



The challenge was that coordinates system was not Latitude or Longitude based but in Universal Transverse Mercator (UTM) coordinate system.

More details about this coordinates system are included in Wikipedia under the following reference:

https://en.wikipedia.org/wiki/Universal_Transverse_Mercator_coordinate_system

As mentioned before, London Geographic Information was in Universal Transverse Mercator (UTM) coordinate system and transformation to Latitude and Longitude Coordinates was not straight forward, even though there was a library called UTM. As Foursquare requires Latitude and Longitude and we had the geographic shapes of London Boroughs, I had decided to identify the "official" coordinated of each of the Boroughs, instead of using the Centroid of each Polygon.

I had used the Latitude and longitude obtained from DISTANCESTO.COM. For Example, these are the coordinates of Westmister.

https://www.distancesto.com/coordinates/gb/city-of-westminster-latitude-longitude/history/44297.html

I had created and imported an Excel Spreadsheet with the London Borough Coordinates (London_Boroughs_Coordinates.xlsx).

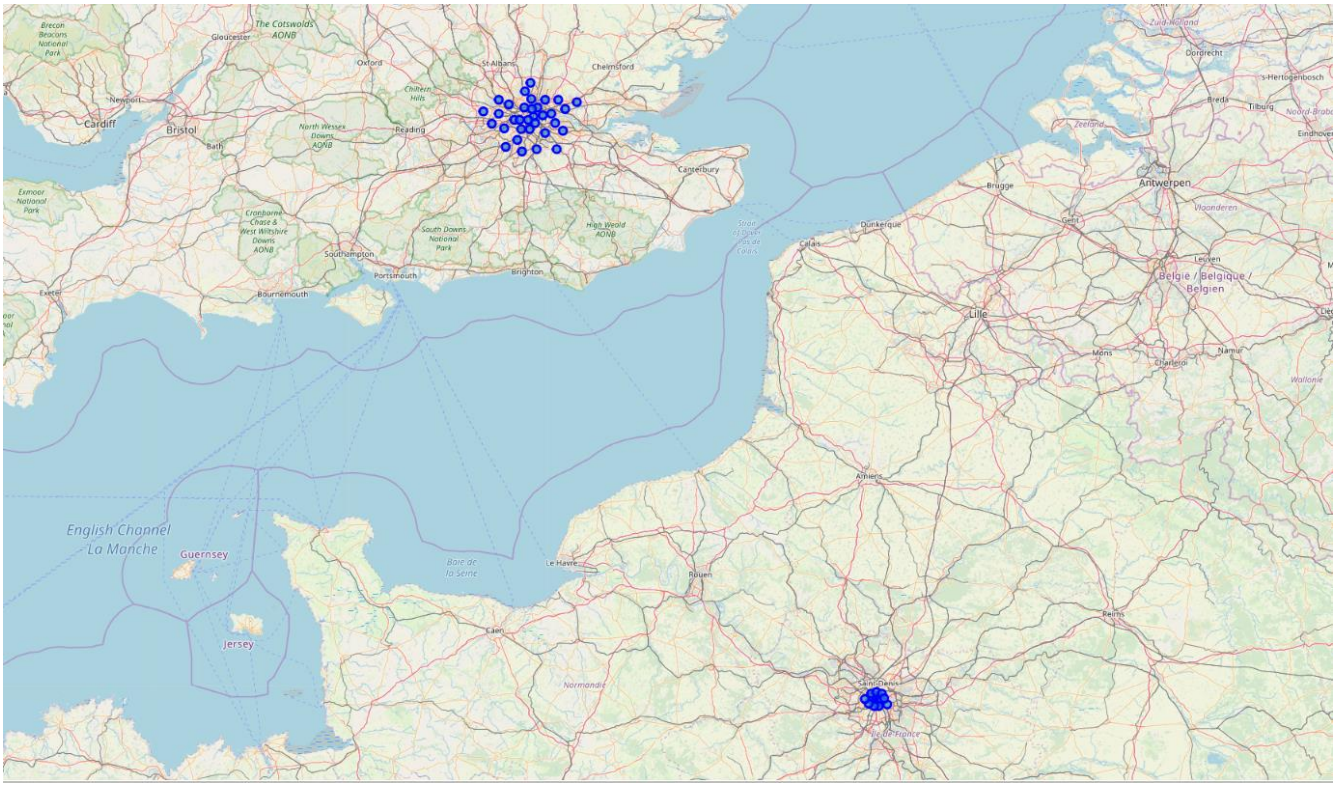
3. Methodology

3.1 Exploratory Data Analysis

In order to be able to compare the London Boroughs and Paris Districts, I had merged them in a single Data Set first, so I had added the "City" column to be able to distinguish Paris and London rows on the Merged Data Set.

First, I had obtained the Latitude and Longitude of Paris Districts, based on the JSON File with Geographic Data.

Figure below shows the obtained map



Having the coordinates of the Paris Districts and London Boroughs, we obtained the nearest Venues from each point with Foursquare API. For London and Paris, there were 245 different Venue Categories. Then we determined the Frequency of each Venue Category for each location, and used the ten most frequent Venue Categories to identify Similarities across Geographic Points.

	NeighborhoodGroup	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	Barking and Dagenham	Construction & Landscaping	Park	Yoga Studio	English Restaurant	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market	Farm
1	Barnet	Café	Bus Stop	Yoga Studio	Forest	Food & Drink Shop	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market
2	Batignolles-Monceau	French Restaurant	Hotel	Italian Restaurant	Café	Plaza	Bakery	Restaurant	Japanese Restaurant	Bistro	Bus Stop
3	Bexley	Indian Restaurant	Lake	Restaurant	Park	Café	Pub	Hot Dog Joint	Hostel	Fast Food Restaurant	Farmers Market
4	Bourse	French Restaurant	Wine Bar	Bistro	Hotel	Japanese Restaurant	Salad Place	Cocktail Bar	Italian Restaurant	Bar	Bakery

3.2 Inferential Statistical Testing, Machine Learnings used and Justification

As we wanted to group similar data points together, based on categories, we used K-means unsupervised Machine Learning Algorithm, which looks for a fixed number (k) of clusters in a dataset.

Results

By applying the K-means clustering, Algorithm, I grouped Paris Districts and London Boroughs in 5 Clusters but only Clusters 1 and 2 existed in both cities.

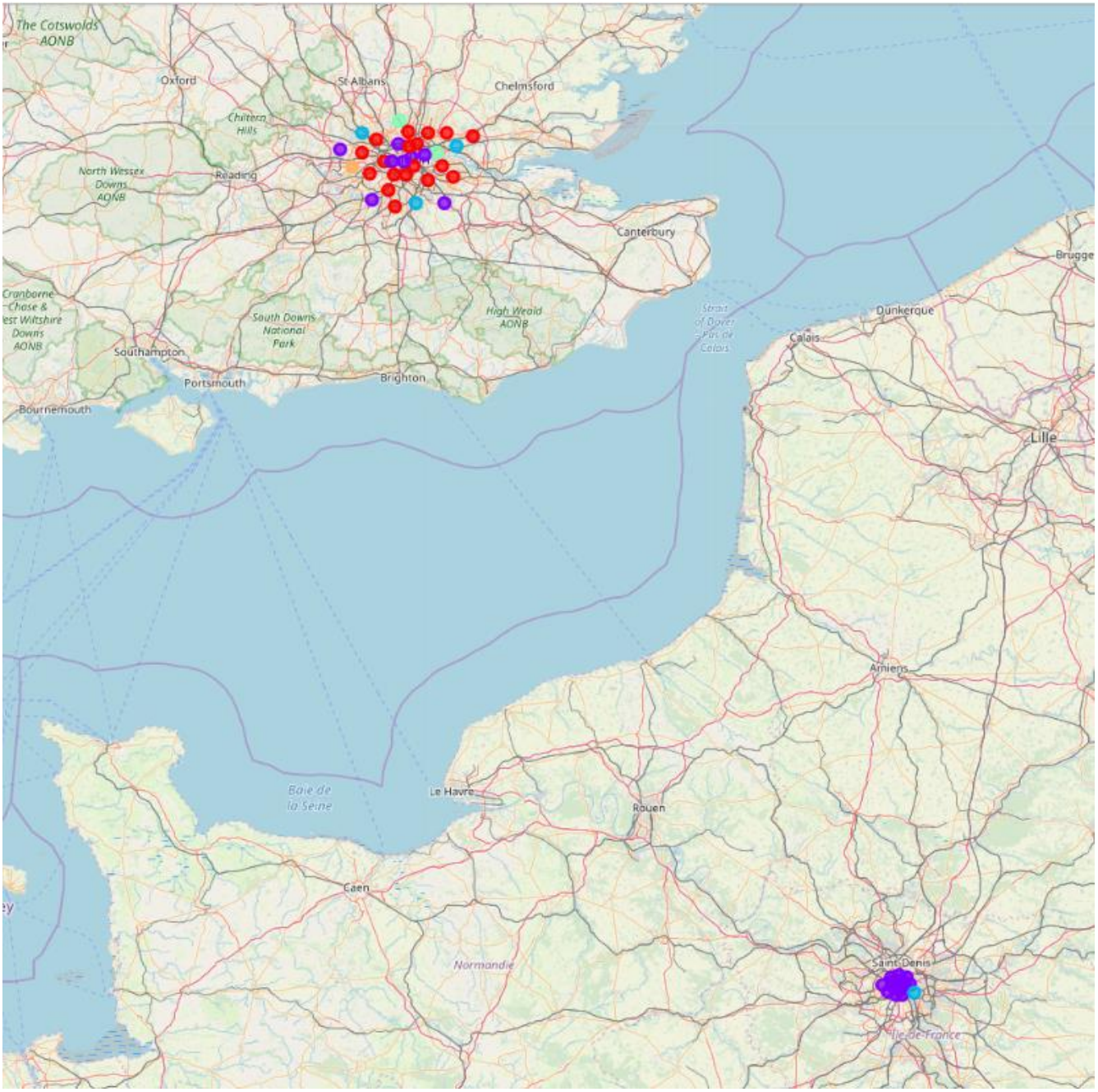
These are the elements of Cluster 1

	NeighborhoodGroup	City	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	Cluster Labels
18	Batignolles-Monceau	Paris	French Restaurant	Hotel	Italian Restaurant	Café	Plaza	Bakery	Restaurant	Japanese Restaurant	Bistro	Bus Stop	1
19	Bourse	Paris	French Restaurant	Wine Bar	Bistro	Hotel	Japanese Restaurant	Salad Place	Cocktail Bar	Italian Restaurant	Bar	Bakery	1
20	Bromley	London	Bed & Breakfast	Construction & Landscaping	Tennis Court	Forest	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	English Restaurant	Farm	1
21	Buttes-Chaumont	Paris	Bar	French Restaurant	Hotel	Restaurant	Seafood Restaurant	Beer Bar	Brewery	Bistro	Farmers Market	Pizza Place	1
22	Buttes-Montmartre	Paris	Bar	French Restaurant	Pizza Place	Convenience Store	Italian Restaurant	Hotel	Bistro	Vietnamese Restaurant	Restaurant	Gastropub	1
23	Camden	London	Pharmacy	Bakery	Gastropub	Café	Coffee Shop	Pizza Place	American Restaurant	Bookstore	Burger Joint	Museum	1
24	City of London	London	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Pub	Seafood Restaurant	Wine Bar	Sandwich Place	Asian Restaurant	Cocktail Bar	French Restaurant	1
25	Entrepôt	Paris	French Restaurant	Bistro	Coffee Shop	Indian Restaurant	Café	Bar	Pizza Place	Japanese Restaurant	Italian Restaurant	Furniture / Home Store	1
26	Gobelins	Paris	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	French Restaurant	Thai Restaurant	Japanese Restaurant	Juice Bar	Pizza Place	Bakery	Butcher	1
27	Hillingdon	London	Chinese Restaurant	Breakfast Spot	Bakery	Gym	Yoga Studio	Falafel Restaurant	Event Space	Exhibit	Fabric Shop	Farmers Market	1
28	Hôtel-de-Ville	Paris	French Restaurant	Ice Cream Shop	Hotel	Pub	Wine Bar	Pastry Shop	Plaza	Bakery	Tea Room	Garden	1
29	Kensington and Chelsea	London	Café	Italian Restaurant	Pub	Clothing Store	Bakery	Garden	Burger Joint	French Restaurant	English Restaurant	Gym / Fitness Center	1
30	Kingston upon Thames	London	Tourist Information Center	Gym	Breakfast Spot	Gym / Fitness Center	Theater	Fabric Shop	Ethiopian Restaurant	Event Space	Exhibit	Farm	1
31	Louvre	Paris	French Restaurant	Hotel	Café	Plaza	Exhibit	Japanese Restaurant	Historic Site	Restaurant	Coffee Shop	Art Museum	1
32	Luxembourg	Paris	French Restaurant	Hotel	Italian Restaurant	Bakery	Wine Bar	Tea Room	Bistro	Ice Cream Shop	Pastry Shop	Japanese Restaurant	1
33	Ménilmontant	Paris	Italian Restaurant	French Restaurant	Bakery	Bistro	Café	Plaza	Bar	Sushi Restaurant	Indian Restaurant	Japanese Restaurant	1
34	Observatoire	Paris	French Restaurant	Hotel	Pizza Place	Bike Rental / Bike Share	Cosmetics Shop	Café	Fast Food Restaurant	Food & Drink Shop	Miscellaneous Shop	Brasserie	1
35	Opéra	Paris	French Restaurant	Hotel	Cocktail Bar	Bar	Bistro	Bakery	Japanese Restaurant	Wine Bar	Lounge	Italian Restaurant	1
36	Palais-Bourbon	Paris	French Restaurant	Hotel	Italian Restaurant	Plaza	Café	Cocktail Bar	History Museum	Coffee Shop	Bakery	Bistro	1
37	Panthéon	Paris	French Restaurant	Bar	Italian Restaurant	Café	Plaza	Hotel	Greek Restaurant	Wine Bar	Museum	Pub	1
38	Passy	Paris	Plaza	French Restaurant	Bus Stop	Boat or Ferry	Lake	Bike Rental / Bike Share	Art Museum	Bus Station	Pool	Flower Shop	1
39	Popincourt	Paris	French Restaurant	Bar	Restaurant	Cocktail Bar	Japanese Restaurant	Theater	Pizza Place	Italian Restaurant	Café	Wine Bar	1
40	Temple	Paris	French Restaurant	Bistro	Coffee Shop	Café	Hotel	Wine Bar	Japanese Restaurant	Boutique	Italian Restaurant	Sandwich Place	1
41	Tower Hamlets	London	Turkish Restaurant	Track Stadium	Museum	Go Kart Track	Canal Lock	Recreation Center	Middle Eastern Restaurant	Athletics & Sports	Thrift / Vintage Store	Coffee Shop	1
42	Vaugirard	Paris	French Restaurant	Hotel	Italian Restaurant	Indian Restaurant	Bistro	Thai Restaurant	Coffee Shop	Park	Lebanese Restaurant	Bakery	1
43	Westminster	London	Hotel	Sandwich Place	Coffee Shop	Outdoor Sculpture	Café	Sushi Restaurant	Plaza	Hotel Bar	Garden	Juice Bar	1
44	Élysée	Paris	French Restaurant	Hotel	Bakery	Japanese Restaurant	Italian Restaurant	Art Gallery	Cocktail Bar	Theater	Bar	Salad Place	1

These are the Elements of Cluster 2

	NeighborhoodGroup	City	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	Cluster Labels
45	Barking and Dagenham	London	Construction & Landscaping	Park	Yoga Studio	English Restaurant	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market	Farm	2
46	Croydon	London	Disc Golf	Park	Soccer Field	Yoga Studio	Ethiopian Restaurant	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market	2
47	Harrow	London	Convenience Store	Park	Yoga Studio	English Restaurant	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market	Farm	2
48	Reuilly	Paris	Monument / Landmark	Supermarket	Park	Bike Rental / Bike Share	Yoga Studio	Food Truck	Flower Shop	Fish & Chips Shop	Field	Fast Food Restaurant	2

Figure Below shows the location of the 5 Clusters identified (Cluster 0 to Cluster 4)



5 Discussion

5.1 Observations

As per Data Visualization, Clusters 1 and 2 happen in both Cities, but there are less elements belonging to Cluster 2.

By Analyzing the Districts in the common Clusters, One option could be to evaluate establishing a French Restaurant in both London and Paris. However, French Restaurant is one of the most common Venue Categories in most of the Paris Districts in Cluster 1, which means more competitors. For this reason, Cluster to can be considered and Option with less competitors.

Below I include additional findings about French District.

What I could find on https://worldinparis.com/paris-arrondissement-guide regarding the Paris Reuilly District is below:

Paris 12 is one of the main Paris residential areas. Also, it has the honor to be the greenest district of Paris, home of Parc Floral, Bois de Vincennes, and Parc de Bercy. This is a very sleepy district but still, sights like La Coulée Verte or a glass of wine in good company at Bercy Village are worth the visit.

MAIN SIGHTS: Le Marché d'Aligre

BEST MUSEUMS: Cinema Museum at Cinémathèque Française, Cité nationale de l'histoire de l'immigration, Musée des Arts Forains

PARKS AND GARDENS: Parc Floral, Promenade Plantée (Coulée Verte), Parc de Bercy

5.2 Recommendations

From the most common Venue Categories in Cluster 2, there are several options that suggest healthy and environment friendly lifestyle, such as Farmers Markets, Fields, Gold Yoga, Bike Rental / Bike Sharing and Parks.

As per the previous Analysis, I would recommend to coordinate a Market and Feasibility Study for a Fast Food Restaurant with Healthy Options as a brand differentiator in Croydon, London and Reuilly Paris.

6 Conclusion

Paris and London are Financial Capitals of their Countries and relatively near of each other (456 km and 4 hours driving).

As per our Geographic and Venue Categories Analysis, obtained with Foursquare API, and by using the K-means Clustering Machine Learning, unsupervised algorithm, to compare how similar are France Districts to Paris Boroughs, in terms in Venue Categories, our recommendation was to coordinate a Market and Feasibility Study for a Fast Food Restaurant with Healthy Options as a brand differentiator, in Croydon, London and Reuilly Paris.