

# Less popular activities in Paris, London and Madrid

## I. Introduction:

An Entrepreneur wants to start a business in one of the most visited cities in the world. He has an idea for 3 cities, Paris, Madrid and London.

For his orientation, he would like to look among similar cities, the activity which is already growing rapidly in several cities, but is still lagging in some.

For this mission, we will collect the data (cities, districts, latitude and longitude) of these three cities and save them in a single dataset.

After observing this data on a geographic map, we will use an unsupervised machine learning algorithm (the Kmeans) to group the different districts of these cities according to the activities present. From these clusters, we will easily identify the neighborhoods that are similar and determine the activities with high potential.

## II. Data collection

### A. extraction of districts without coordinates (Latitude and logitude)

We have extracted the districts without coordinates of the different cities individually in the websites below:

Given the reduced size of the different data sources, we opted for the manual copy / paste method in order to save time, unlike the other methods (BeautifulSoup ...) which are advantageous in the case of a very large number of data.

The extracted data will be gathered in a single CVS file.

- Extraction of Madrid neighborhoods from the site: <https://fr.db-city.com/Espagne--Madrid--Madrid--1>
- Extraction of paris districts: [https://fr.wikipedia.org/wiki/Liste\\_des\\_quartiers\\_administratifs\\_de\\_Paris](https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris)
- Extraction of the districts of London: [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)

### B. Update of our district file with GPS coordinates.

Two methods seemed more affordable to us for collecting location data.

- The method with the Geocoding API: <https://developers.google.com/maps/documentation/geocoding/start>
- The method with the tools <https://geocode.localfocus.nl/>

To save time (although we practice with the geocoding method), we have used the <https://geocode.localfocus.nl/> tool where we have uploaded our previously extradited data.

After some restatements, we have obtained, the cities, the districts and the GPS coordinates of the different cities (Paris, Madrid and London) attached

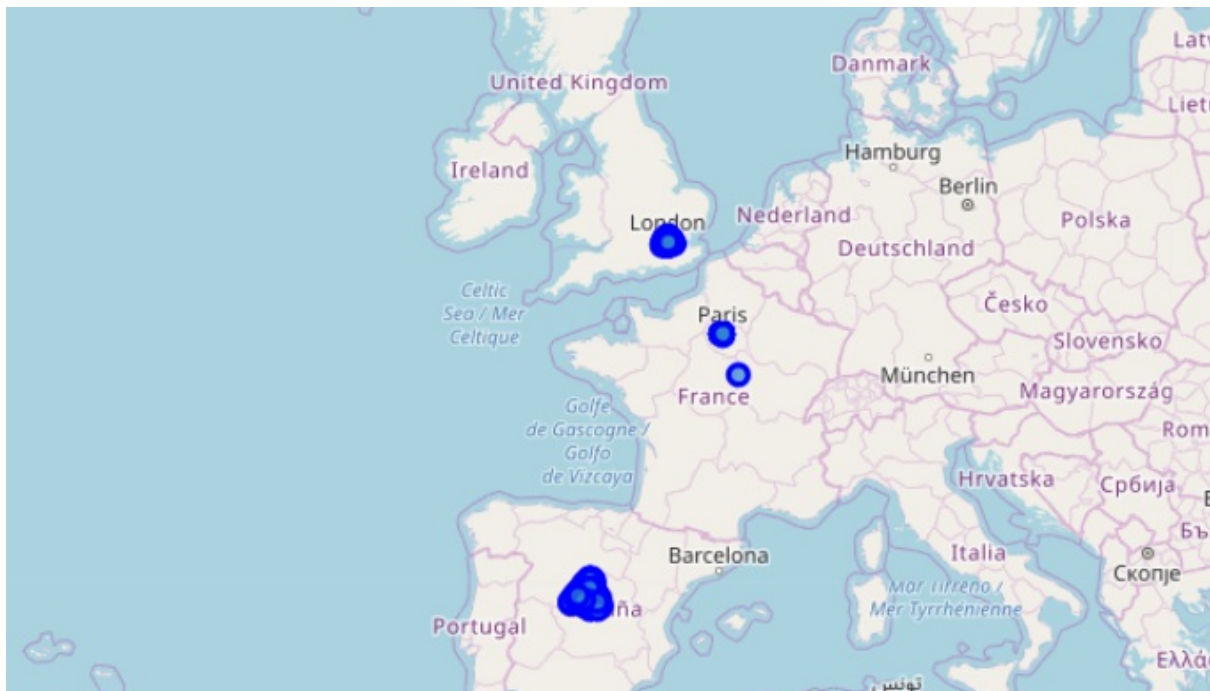
[https://s3.amazonaws.com/csvpastebin/uploads/b219de95b1e1a73911b6443d5fd85cc1/Data\\_final.csv](https://s3.amazonaws.com/csvpastebin/uploads/b219de95b1e1a73911b6443d5fd85cc1/Data_final.csv)

### III. Methodology

We observe the data contained in our database with as columns the districts, the cities and their coordinates (latitude and longitude). This data is shared in the link above.

	Quartiers	villes	Latitude	Longitude
0	Quartier Saint-Germain-l'Auxerrois	Paris	48.860562	2.336489
1	Les Halles	Paris	48.862567	2.345809
2	Quartier du Palais-Royal	Paris	48.863357	2.336062
3		Paris	48.859116	2.331839
4		Paris	48.859116	2.331839

Overview of the geographic distribution of neighborhoods on a map based on our data.



With the Foursquare API, we have collected all the places present in these different districts. The tables below show the total number of places and their grouping by district.

(17761, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Quartier Saint-Germain-l'Auxerrois	48.860562	2.336489	Cour Carrée du Louvre	48.860360	2.338543	Pedestrian Plaza
1	Quartier Saint-Germain-l'Auxerrois	48.860562	2.336489	Musée du Louvre	48.860847	2.336440	Art Museum
2	Quartier Saint-Germain-l'Auxerrois	48.860562	2.336489	La Vénus de Milo (Vénus de Milo)	48.859943	2.337234	Exhibit
3	Quartier Saint-Germain-l'Auxerrois	48.860562	2.336489	Cour Napoléon	48.861172	2.335088	Plaza
4	Quartier Saint-Germain-l'Auxerrois	48.860562	2.336489	Place du Palais Royal	48.862523	2.336688	Plaza

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Ajalvir	5	5	5	5	5	5
Alameda del Valle	5	5	5	5	5	5
Alcala de Henares	31	31	31	31	31	31
Alcobendas	10	10	10	10	10	10
Alcorcón	10	10	10	10	10	10
Aldea del Fresno	8	8	8	8	8	8
Algete	8	8	8	8	8	8
Alpedrete	5	5	5	5	5	5
Anchuelo	1	1	1	1	1	1
Aranjuez	33	33	33	33	33	33
Arganda del Rey	4	4	4	4	4	4

## Analysing Each Neighborhood

For this analysis we will group each district and determine the Average Occurrence of each category in it.

The table below presents some details.

	Neighborhood	Advertising Agency	Afghan Restaurant	African Restaurant	Airport	Alsation Restaurant	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Ente
0	Ajalvir	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
1	Alameda del Valle	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.200000	0.000000	0.000000	0.000000	
2	Alcala de Henares	0.000000	0.000000	0.000000	0.0	0.00	0.032258	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
3	Alcobendas	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
4	Alcorcón	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
5	Aldea del Fresno	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
6	Algete	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
7	Alpedrete	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	
8	Anchuelo	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	

## Now, let display the top 10 venues for each Neighbourhood

	Neighborhood	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	Ajalvir	Dessert Shop	Mediterranean Restaurant	Spanish Restaurant	Hotel	Furniture / Home Store	Zoo Exhibit	Exhibit	Factory	Falafel Restaurant	Farm
1	Alameda del Valle	Restaurant	Food	Spanish Restaurant	Argentinian Restaurant	Zoo Exhibit	Fast Food Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
2	Alcala de Henares	Tapas Restaurant	Bar	Mediterranean Restaurant	Spanish Restaurant	Pub	Restaurant	Hotel	Coffee Shop	Café	Frozen Yogurt Shop
3	Alcobendas	Stadium	Restaurant	Spanish Restaurant	Mediterranean Restaurant	Food	Grocery Store	Gym	Pharmacy	Park	Pizza Place
4	Alcorcón	Concert Hall	Restaurant	Tennis Court	Grocery Store	Gym	Other Great Outdoors	Bus Station	Department Store	Bakery	Plaza

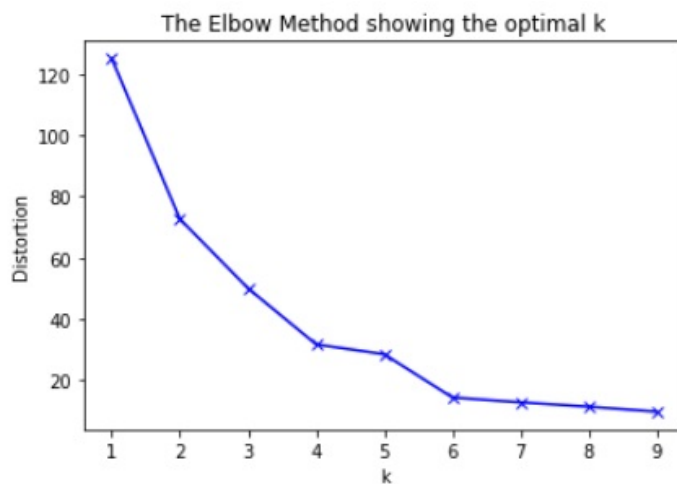
## Determination of similar clusters:

### Identification of K

After reprocessing of data, the description of the new values is presented to us

Venue	
count	205.000000
mean	86.639024
std	321.157185
min	1.000000
25%	4.000000
50%	6.000000
75%	34.000000
max	2812.000000

We use the Elbow methods to determine the optimal K.



### The best k=5

## Run K-means algorithm to cluster the neighborhood into 5 clusters

	Neighborhood	cities	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 Common Venue
0	Quartier Saint-Germain-l'Auxerrois	Paris	48.860562	2.336489	1.0	French Restaurant	Café	Plaza	Hotel	Coffee Shop	Bar	Museum	Bakery	Art Museum	Historic Site
1	Les Halles	Paris	48.862567	2.345809	1.0	French Restaurant	Bakery	Pizza Place	Cocktail Bar	Wine Bar	Clothing Store	Tea Room	Hotel	Chinese Restaurant	Art Museum
2	Quartier du Palais-Royal	Paris	48.863357	2.336062	1.0	French Restaurant	Japanese Restaurant	Café	Hotel	Coffee Shop	Plaza	Bakery	Historic Site	Art Museum	Rest
3		Paris	48.859116	2.331839	1.0	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	
4		Paris	48.859116	2.331839	1.0	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	

After examining the different clusters, we find that cluster n ° 1 is the one in which our client can position themselves because in the latter, there is much resemblance between the different districts of the three cities.

Similarly, in this cluster, we have 302 registrations for London, 64 for Paris and 74 for Madrid.



## IV. Result

After analyzing the different cases, the most appropriate cluster is cluster N ° 1 with many similarities between the three cities. He is the only one to have similarities between the districts of the three cities.

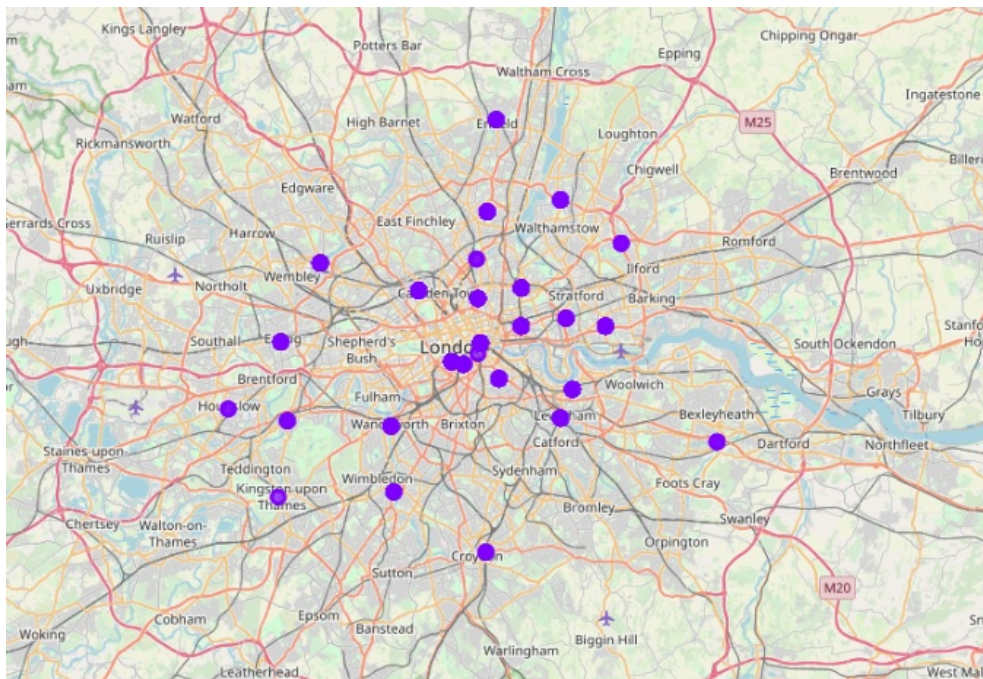
4	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
5	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
6	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
7	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
8	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
9	Paris	French Restaurant	Coffee Shop	Restaurant	Art Gallery	Burger Joint	Café	Cocktail Bar	Pastry Shop	Deli / Bodega	Bakery
10	Paris	French Restaurant	Art Gallery	Plaza	Hotel	Clothing Store	Ice Cream Shop	Bakery	Coffee Shop	Pub	Bar
11	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop
12	Paris	French Restaurant	Café	Hotel	Art Museum	Exhibit	Cosmetics Shop	Italian Restaurant	Historic Site	Bakery	Coffee Shop

### Group venue by city in cluster 1

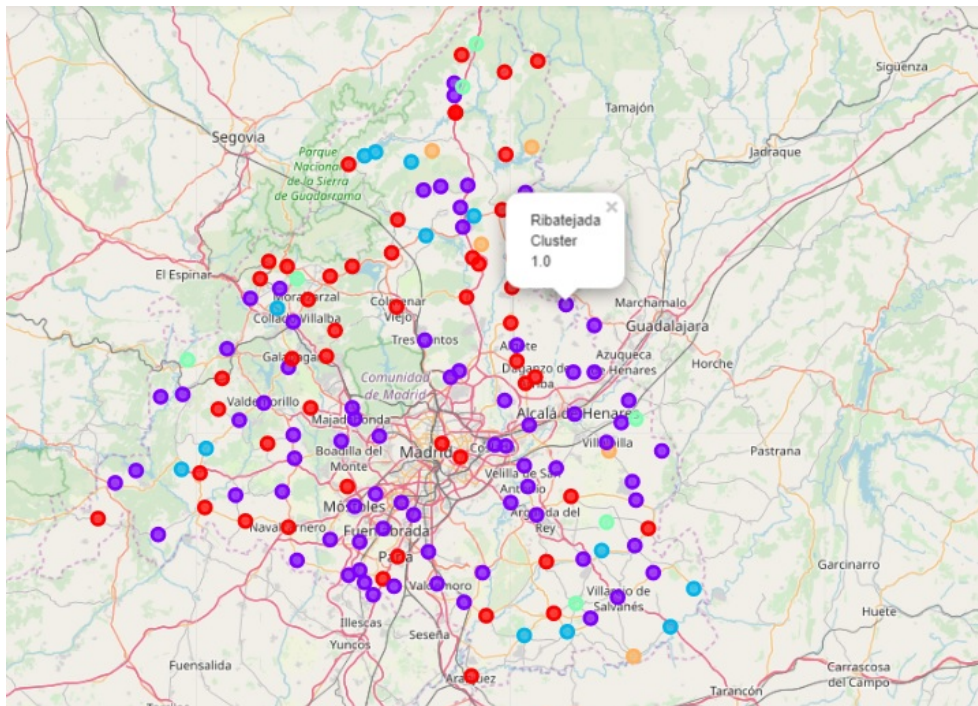
cities	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
London	302	302	302	302	302	302	302	302	302	302
Paris	64	64	64	64	64	64	64	64	64	64
Spain	74	74	74	74	74	74	74	74	74	74

### Geographic representation of venue in cluster 1

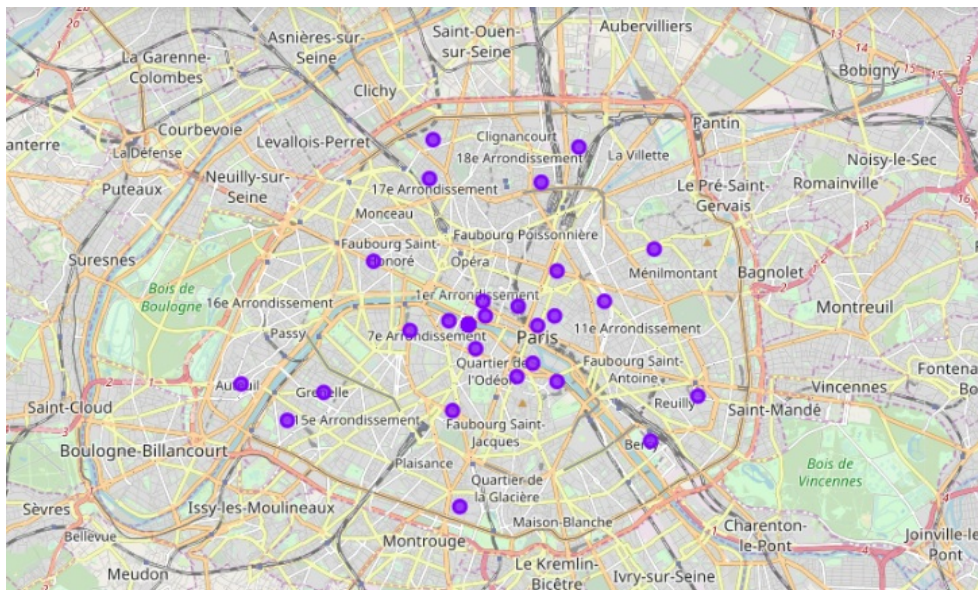
- London (Cluster 1 concerned in purple)



- Madrid (Cluster 1 concerned in purple)



- Paris (Cluster 1 concerned in purple)





## Counting in the 10th Most Common Venue

<i>Counting in the 1st Most Common Venue</i>	<i>Counting in the 2nd Most Common Venue</i>
[('Café', 67), ( 'Coffee Shop', 59), ( 'Pub', 59), ( 'French Restaurant', 56), ( 'Hotel', 50), ( 'Clothing Store', 18), ( 'Gym / Fitness Center', 16), ( 'Garden', 16), ( 'Bar', 14), ( 'Gym', 12)]	[('Café', 75), ( 'Pub', 53), ( 'Coffee Shop', 46), ( 'Brewery', 25), ( 'Clothing Store', 20), ( 'Convenience Store', 20), ( 'Restaurant', 19), ( 'Sandwich Place', 19), ( 'Grocery Store', 18), ( 'Burger Joint', 15)]
<i>Counting in the 3rd Most Common Venue</i>	<i>Counting in the 4th Most Common Venue</i>
[('Hotel', 85), ( 'Scenic Lookout', 46), ( 'Coffee Shop', 46), ( 'Pub', 27), ( 'Chinese Restaurant', 21), ( 'Fast Food Restaurant', 20), ( 'Bar', 18), ( 'Bakery', 16), ( 'Historic Site', 16), ( 'Eastern European Restaurant', 16)]	[('Gym / Fitness Center', 57), ( 'Exhibit', 41), ( 'Supermarket', 26), ( 'Bus Stop', 21), ( 'Café', 20), ( 'Historic Site', 20), ( 'Hotel Bar', 20), ( 'Cocktail Bar', 19), ( 'Grocery Store', 19), ( 'Pub', 18)]
<i>Counting in the 5th Most Common Venue</i>	<i>Counting in the 5th Most Common Venue</i>
[('Pub', 46), ( 'Art Museum', 39), ( 'Coffee Shop', 33), ( 'Bakery', 25), ( 'Café', 21), ( 'Deli / Bodega', 20), ( 'Event Space', 18), ( 'Bus Stop', 18), ( 'Platform', 18), ( 'Supermarket', 17)]	[('Art Museum', 46), ( 'Italian Restaurant', 40), ( 'Restaurant', 39), ( 'Grocery Store', 36), ( 'Exhibit', 21), ( 'Gym / Fitness Center', 21), ( 'Cocktail Bar', 20), ( 'Outdoor Sculpture', 19), ( 'Sushi Restaurant', 16), ( 'Warehouse Store', 15)]
<i>Counting in the 7th Most Common Venue</i>	<i>Counting in the 8th Most Common Venue</i>
[('Italian Restaurant', 45), ( 'Historic Site', 39), ( 'Market', 36), ( 'Park', 26), ( 'Film Studio', 26), ( 'Burger Joint', 26), ( 'Plaza', 23), ( 'Falafel Restaurant', 22), ( 'Bakery', 20), ( 'Gym / Fitness Center', 18)]	[('Café', 49), ( 'Cosmetics Shop', 38), ( 'Exhibit', 33), ( 'Garden', 31), ( 'Asian Restaurant', 22), ( 'Farm', 22), ( 'Italian Restaurant', 21), ( 'Supermarket', 19), ( 'Brewery', 18), ( 'Falafel Restaurant', 16)]
<i>Counting in the 9th Most Common Venue</i>	<i>Counting in the 10th Most Common Venue</i>
[('Pedestrian Plaza', 45), ( 'Gift Shop', 38), ( 'Falafel Restaurant', 34), ( 'Fast Food Restaurant', 32), ( 'Gastropub', 25), ( 'Farmers Market', 23), ( 'Beer Bar', 22), ( 'Soccer Field', 21), ( 'Bar', 20), ( 'Hotel Bar', 19)]	[('Coffee Shop', 59), ( 'Burger Joint', 45), ( 'Farm', 30), ( 'Fast Food Restaurant', 24), ( 'Park', 22), ( 'Market', 21), ( 'Pharmacy', 18), ( 'Clothing Store', 18), ( 'Museum', 17), ( 'Farmers Market', 16)]

## **V. Discussion**

We believe that all of the neighborhoods have characteristics of tourist spots, dominated by cafes and restaurants. Exhibitions such as gift shops and pharmacies are very few and certainly represent areas for close study.

## **VI. Conclusion:**

Given the lowest number of occurrences in Paris, we believe this is the ideal place to invest in one of the areas listed above. Again, we will have to choose where. If Gift shops are rare enough, it would still be important to check how many are present in Paris before getting started.

Code on github: [https://github.com/Chrischri/Less-popular-activities/blob/master/Paris\\_Madrid\\_Londre\\_final.ipynb](https://github.com/Chrischri/Less-popular-activities/blob/master/Paris_Madrid_Londre_final.ipynb)

## **VII. Référence:**

- [www.wikipedia.com](http://www.wikipedia.com)
- <https://fr.db-city.com/Espagne--Madrid--Madrid--1>
- [https://fr.wikipedia.org/wiki/Liste\\_des\\_quartiers\\_administratifs\\_de\\_Paris](https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris)
- [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)
- <https://geocode.localfocus.nl/>

COURSERA: Capstone Project - The Battle of Neighborhoods  
Less-popular-activities  
By NOUBISSI Christian Martial