BATTLE OF NEIGHBORHOODS FINAL REPORT

For the final project of the Applied Data Science Capstone – Applied Data Science Specialization.

INTRODUCTION

decided to start this little project basically for myself, however it would be helpful for other people – travelers mainly – as well. I have always wondered which is the best city where to party, but I couldn't think of the right answer without any data. Discovering such information is important for whoever decides to plan a holiday and is interested in the nightlife of a certain city. Therefore, I will try to solve my doubt by using Foursquare API and find the number of bars in five different cities (Los Angeles, New York, London, Milan and Barcelona) in order to determine which of them would probably be the most chaotic and interesting one. In addition, I will k-cluster the bars in 5 groups for each city. Each cluster represents an area of the city (North, South, East, West, Center) and I will try to understand what is the most partying area among all the 25 clusters and what is the best area where to book a hotel in each city – according to the groups.

DATA

I will not be using any external database or dataset as all the data I need can be obtained by using Foursquare API. I will call the nightspots existing in the cities concerned and I will use the data to create some dataframe to cluster the bars.

Precisely, I will import this url:

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&near={}&limit={}&categoryId={}

and I will use this code: 4d4b7105d754a06376d81259

which represents nightspots venues in the Foursquare API.

```
Total number of nightlife spots Los Angeles =  135
Total number of nightlife spots New York =  193
Total number of nightlife spots London =  118
Total number of nightlife spots Milan =  101
Total number of nightlife spots Barcelona =  111
```

As a result, we obtain the number of nightspots in each of the city concerned.

METHODOLOGY

I firstly use the data to display the folium map of each city simple-marking the top 100 nightspots.

Second, to have a better understanding of the data obtained, I create a dataframe for each city. I display them using the *.head()* function. Each dataframe shows the venues.

Third, I cluster each dataframe in five different groups, this in order to find the best area for each metropolis where to grab a drink.

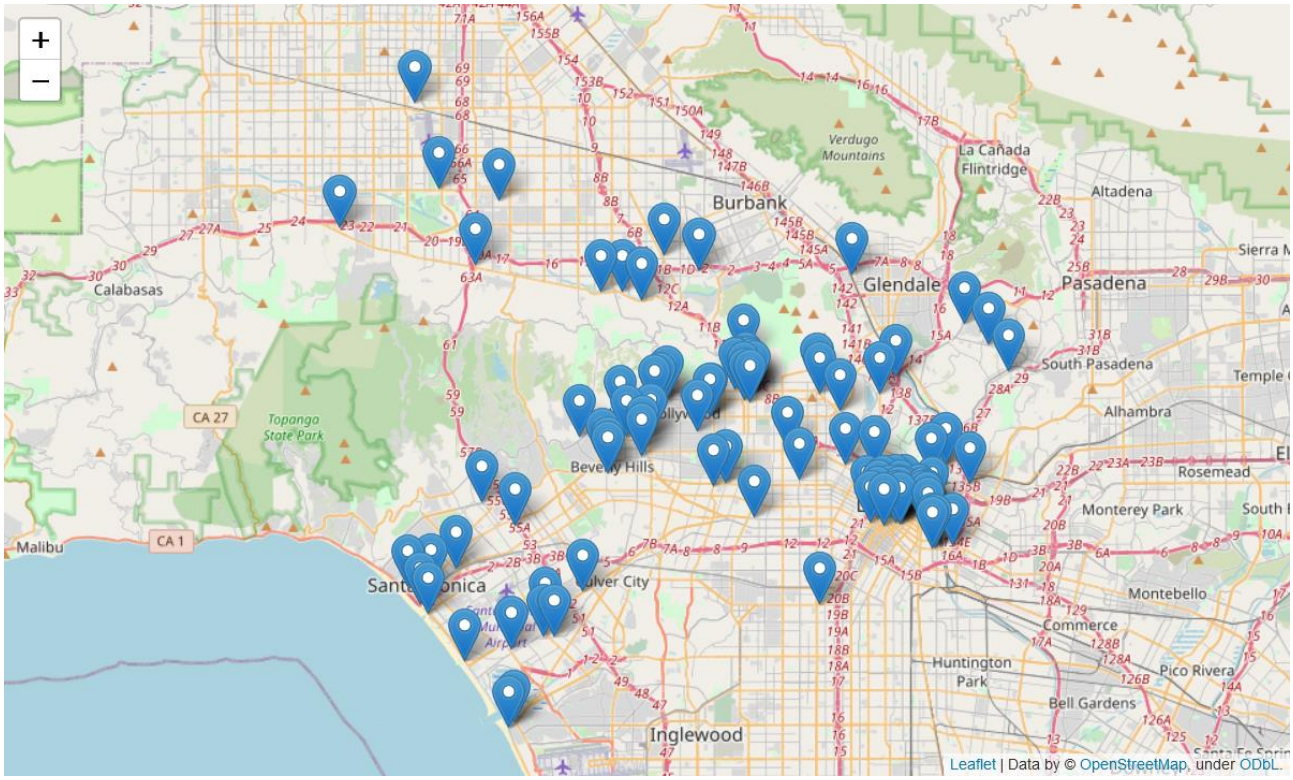
Then, I create a folium map with the venues clustered in the groups.

Finally, I complete the fourth step by creating a dataframe sorting the cluster labels, from the maximum amount of bar per cluster to the minimum amount.

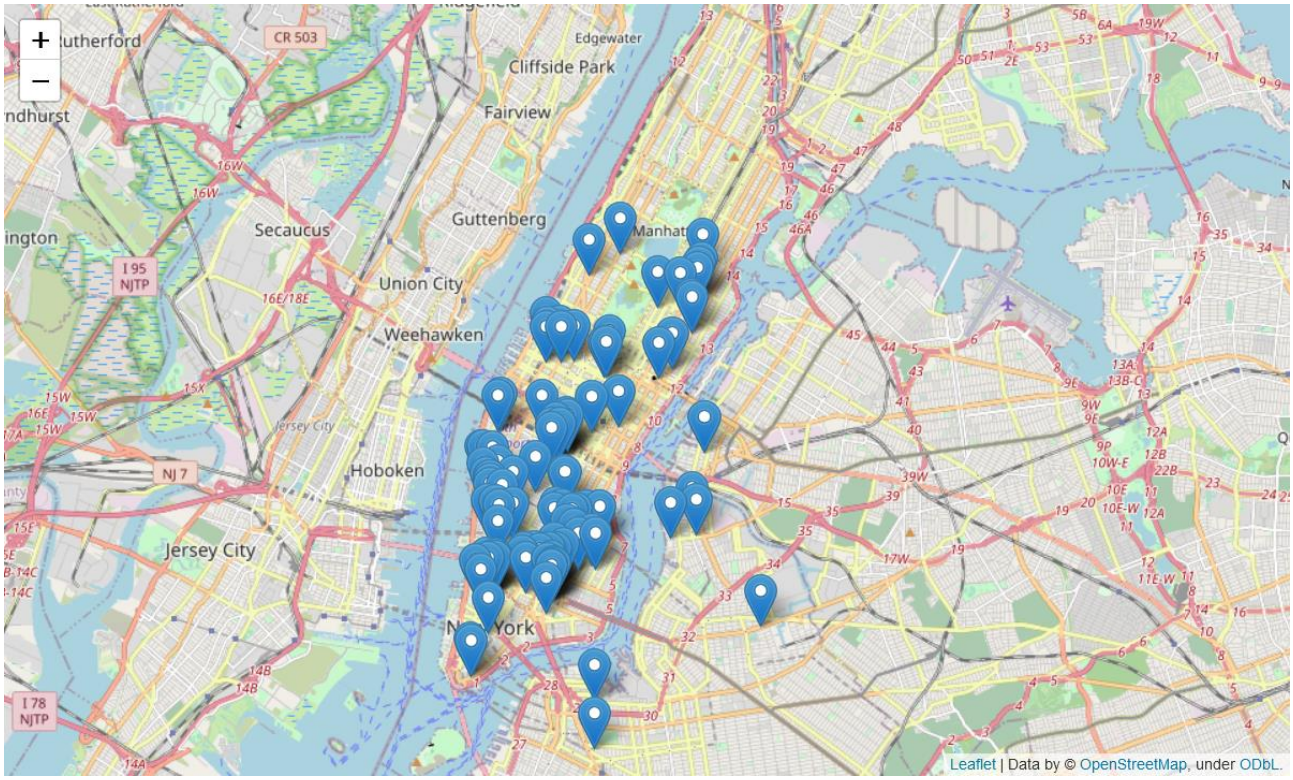
P.S. each point will be better explained in the result section.

RESULTS

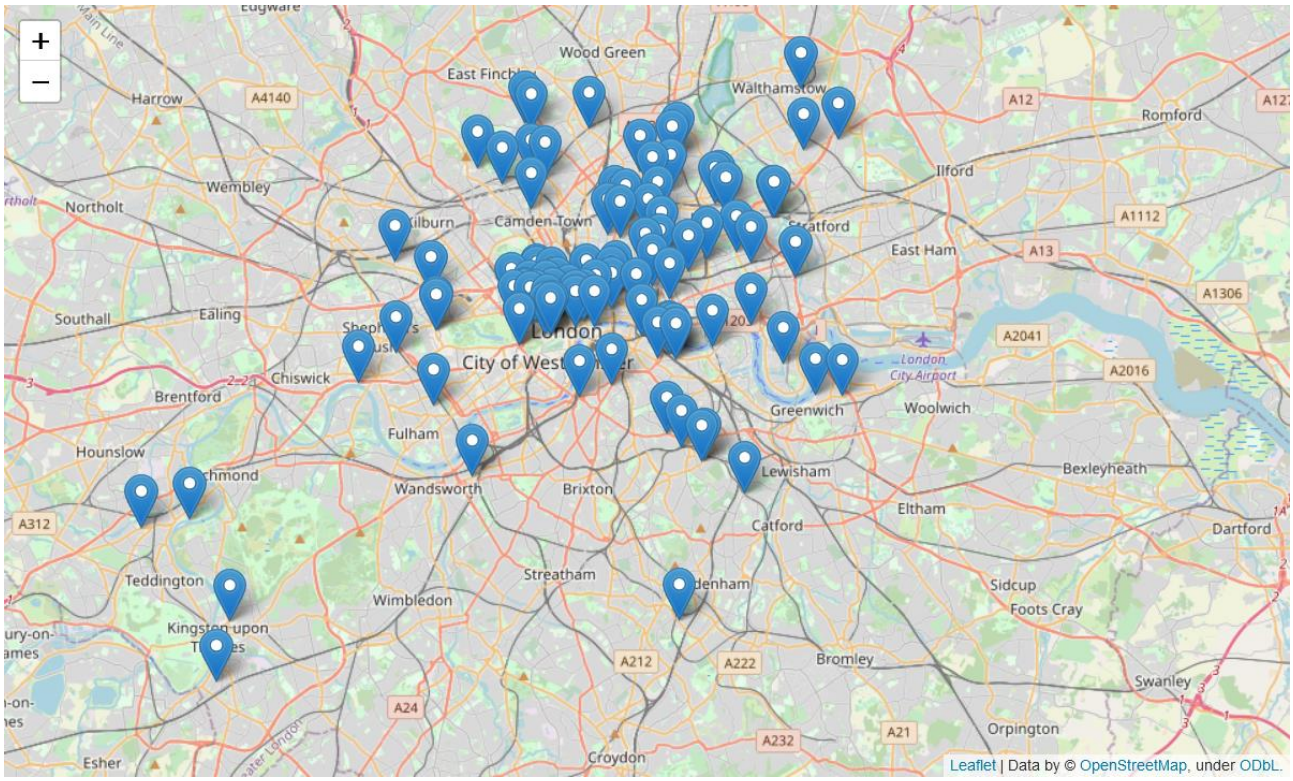
In Los Angeles:



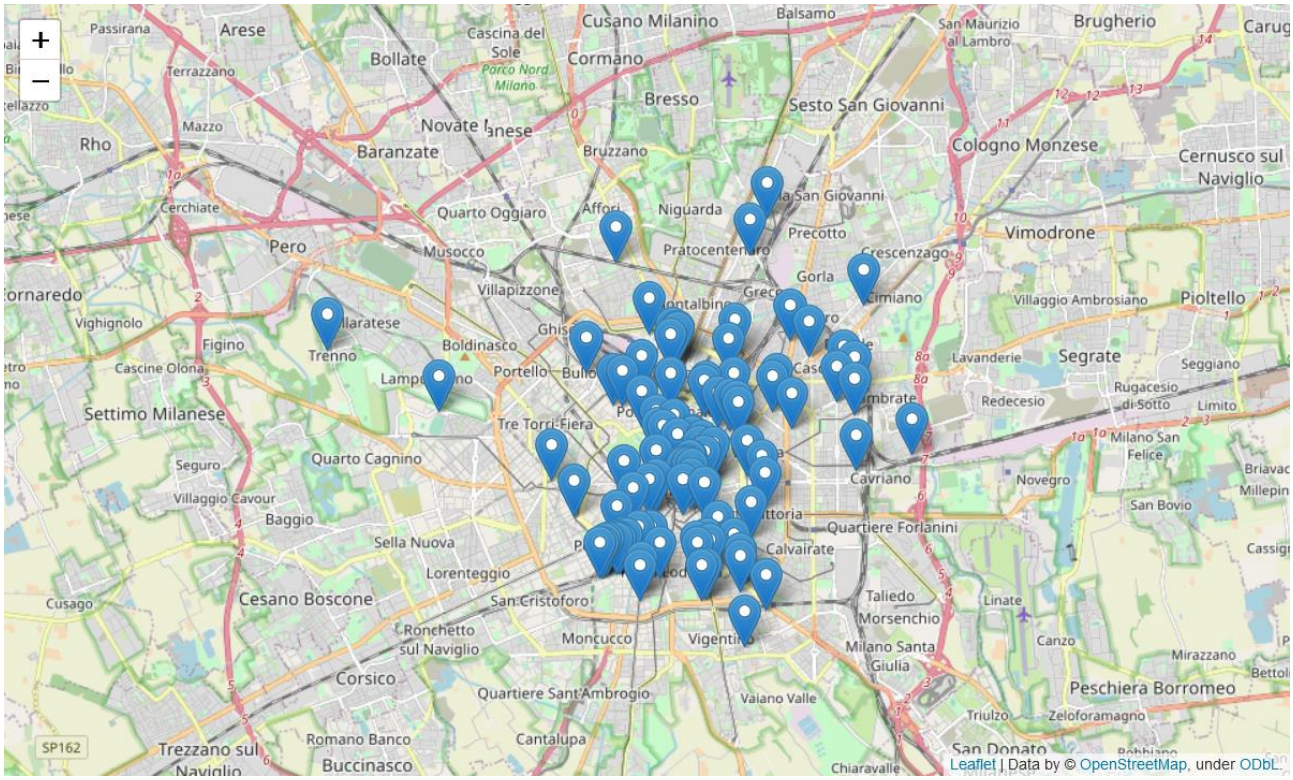
New York:



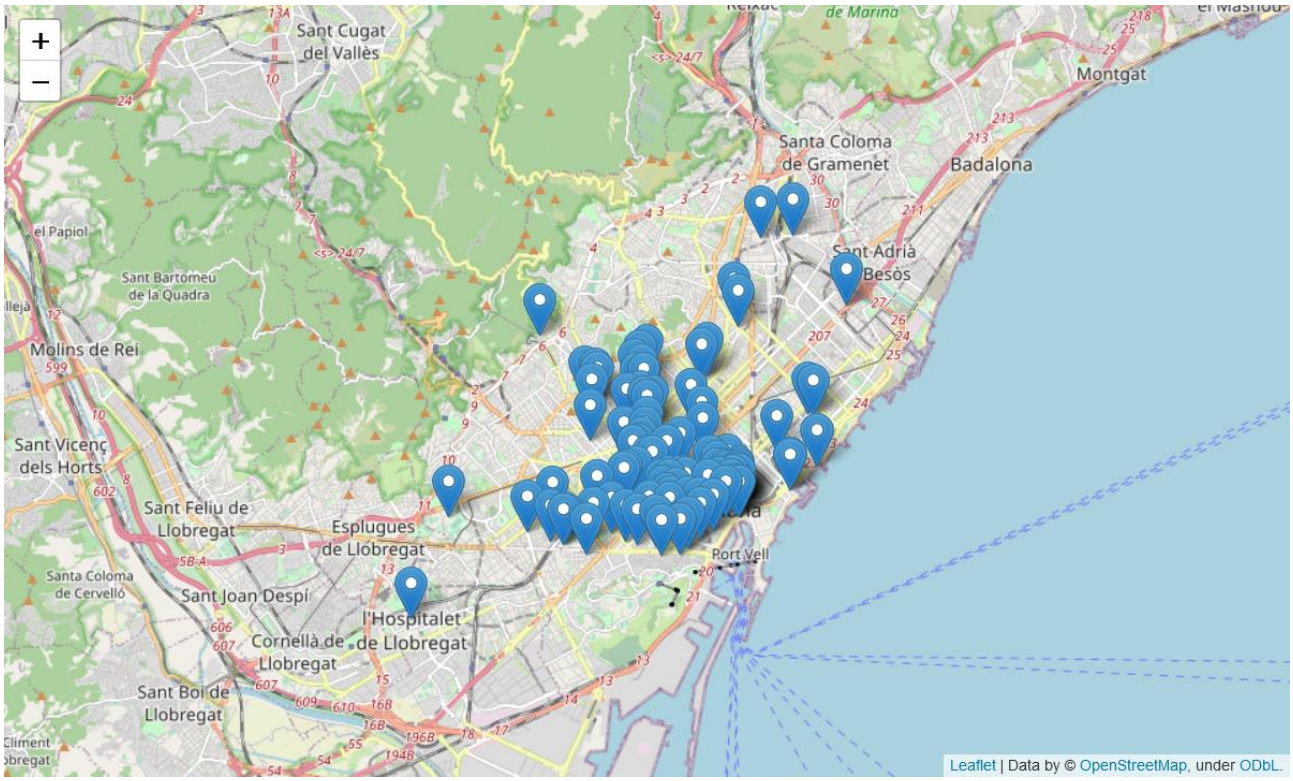
London:



Milan:



Barcelona:



Then, let's create a dataframe for each city's venues – using the *.head()* function.

Los Angeles:

	Name	Address	Lat	Lng
0	Wally's	447 N. Canon Drive	34.071475	-118.401782
1	Silver Lake Wine	2395 Glendale Blvd	34.100003	-118.259315
2	K&L Wine Merchants	1400 Vine St	34.096524	-118.326393
3	Met Her At A Bar	759 S La Brea Ave	34.060557	-118.344744
4	MacLeod Ale Brewing Co.	14741 Calvert St	34.181907	-118.454576

New York:

	Name	Address	Lat	Lng
0	NR	339 E 75th St	40.770082	-73.954978
1	The Bar Room at Temple Court	123 Nassau St	40.711448	-74.006802
2	Dear Irving	55 Irving Pl	40.736089	-73.987324
3	Interboro Spirits and Ales	942 Grand St	40.712771	-73.937030
4	Attaboy	134 Eldridge St	40.718981	-73.991540

London:

	Name	Address	Lat	Lng
0	Scarfes Bar	252 High Holborn	51.517813	-0.118184
1	The Foyer & Reading Room	49 Brook St	51.512577	-0.147663
2	The Chesham Arms	15 Mehetabel Rd	51.547727	-0.050131
3	The Punch Room	10 Berners St	51.516905	-0.136151
4	The Beer Shop London	40 Nunhead Green	51.465729	-0.057836

Milan:

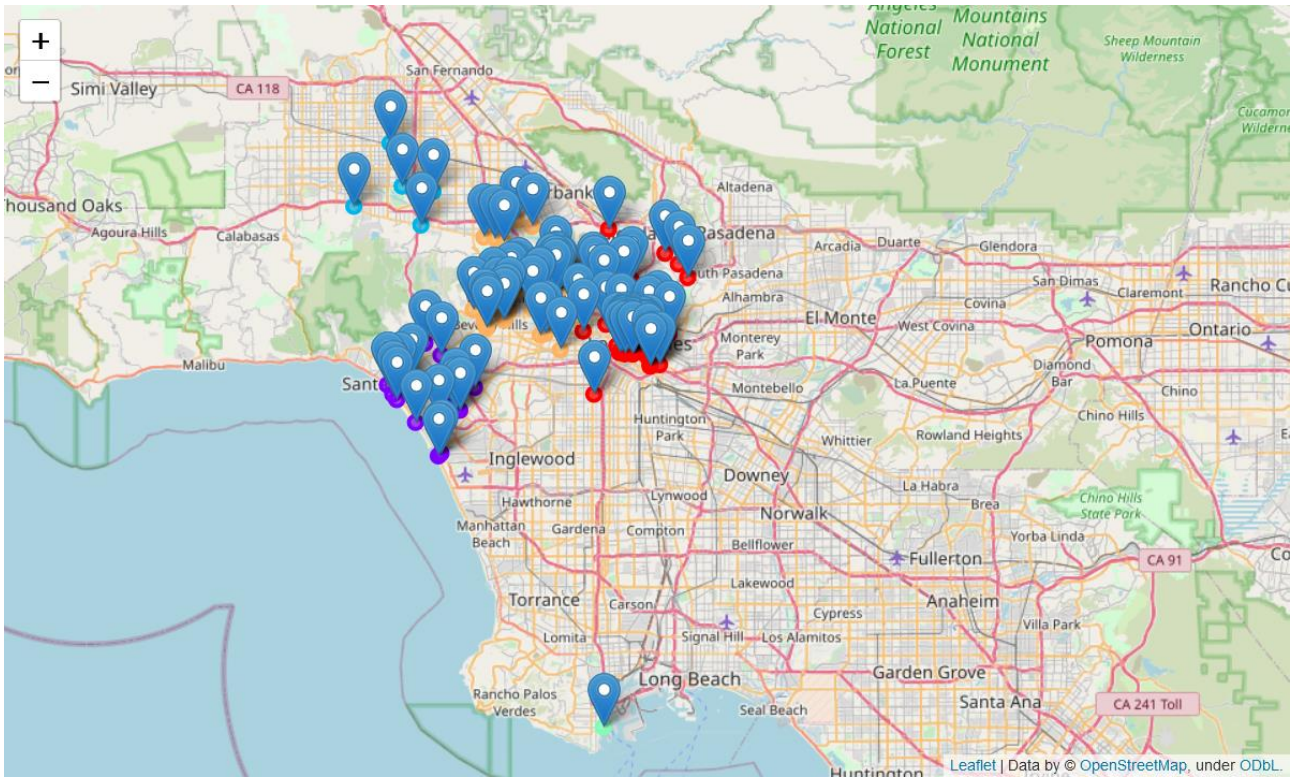
	Name	Address	Lat	Lng
0	Bulgari Lounge Bar	Via Privata Fratelli Gabba	45.470014	9.188943
1	Ceresio 7 Pools & Restaurant	Via Ceresio, 7	45.484025	9.179849
2	La Prosciutteria	Corso Giuseppe Garibaldi, 53	45.474152	9.183449
3	B Café	Via San Maurilio 20	45.462640	9.183381
4	Bicerin	Via Panfilo Castaldi 24	45.476988	9.202432

Barcelona:

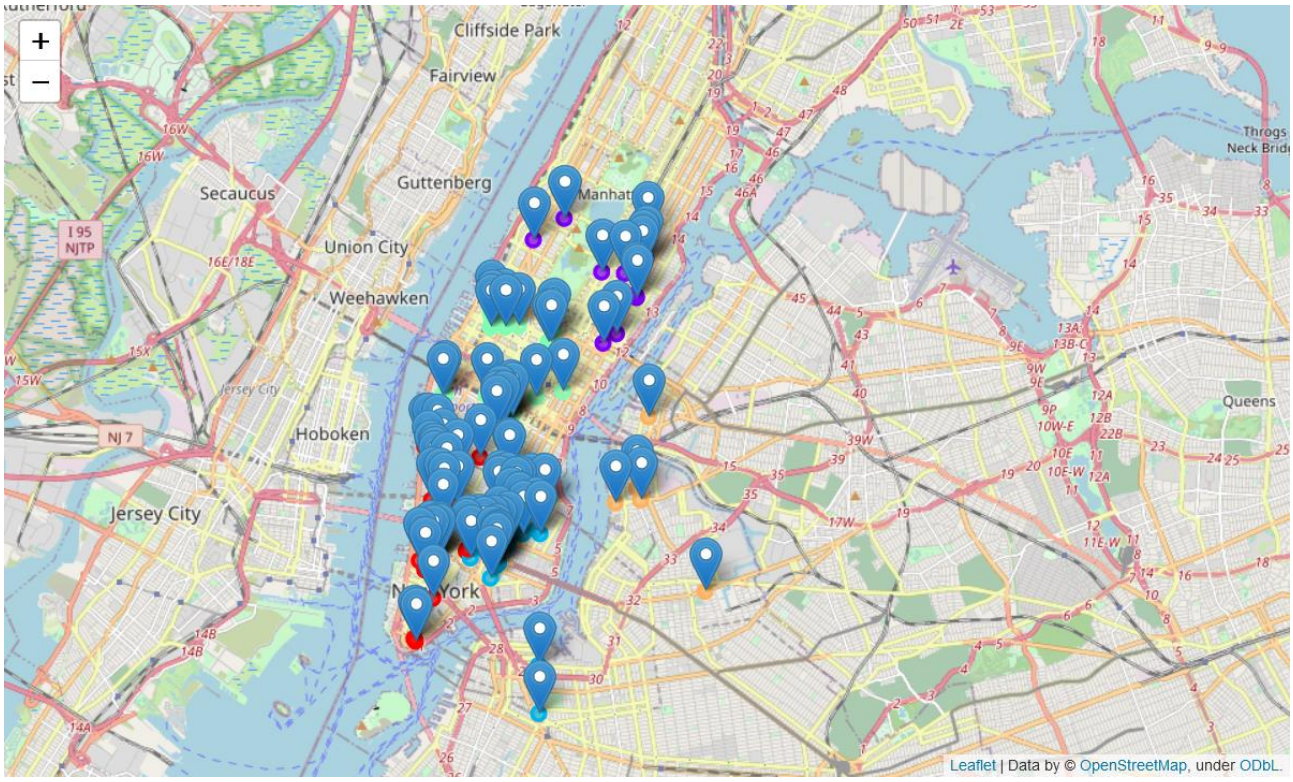
	Name	Address	Lat	Lng
0	Lo Pinyol	C. Torrent de L'Olla, 7	41.398598	2.161148
1	El Quinto Moño	Carrer Villaroel, 96	41.384935	2.157213
2	Hotel Arts Ritz Club Lounge	NaN	41.386596	2.196598
3	Chivuo's	C. Pintor Fortuny, 15	41.382961	2.169948
4	Blavis	Saragossa 85	41.403303	2.147298

Then we finally k-cluster each dataframe in 5 different groups.

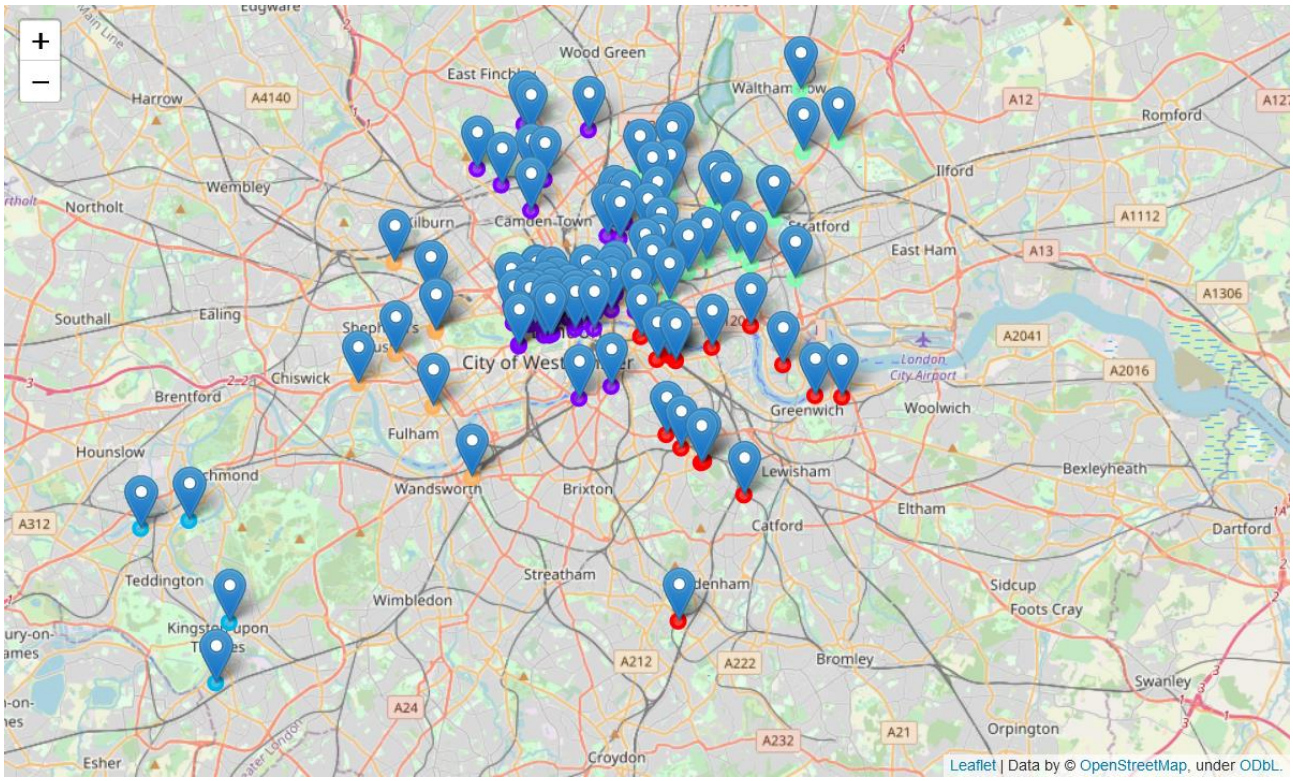
Los Angeles:



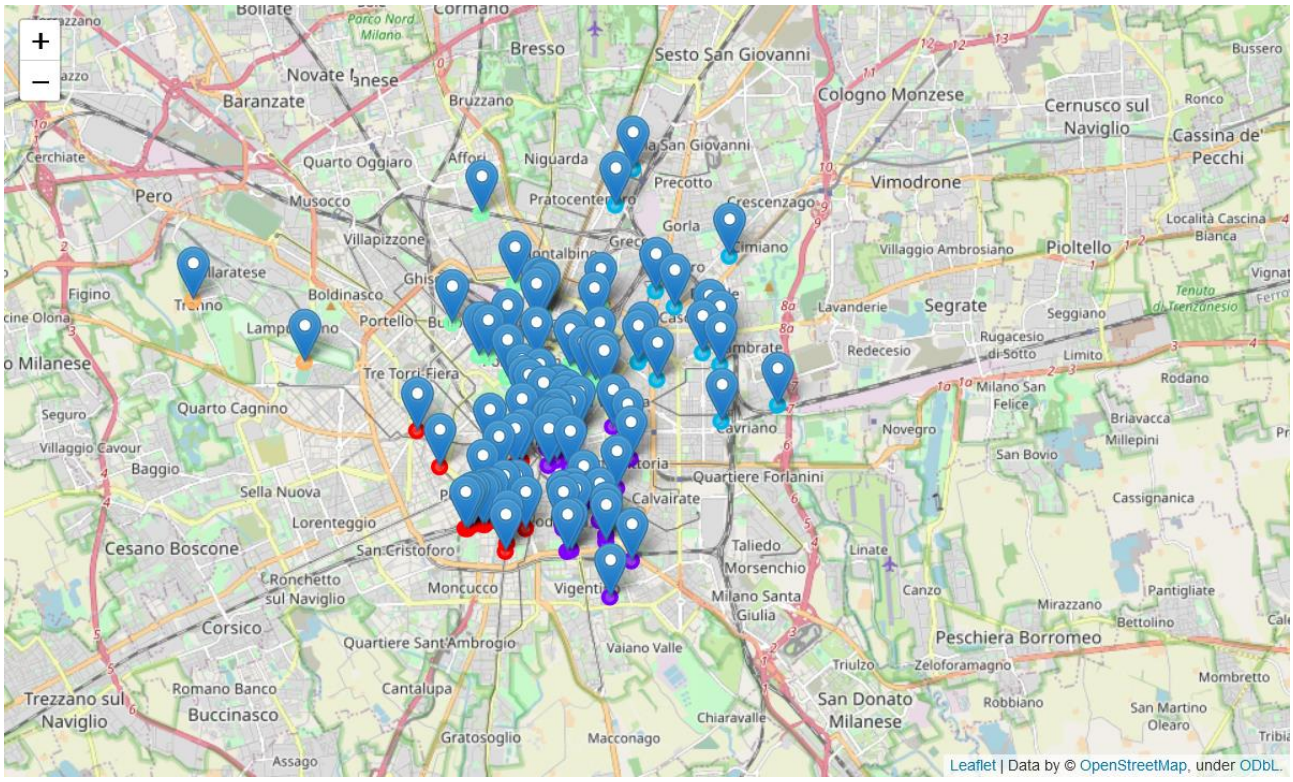
New York:



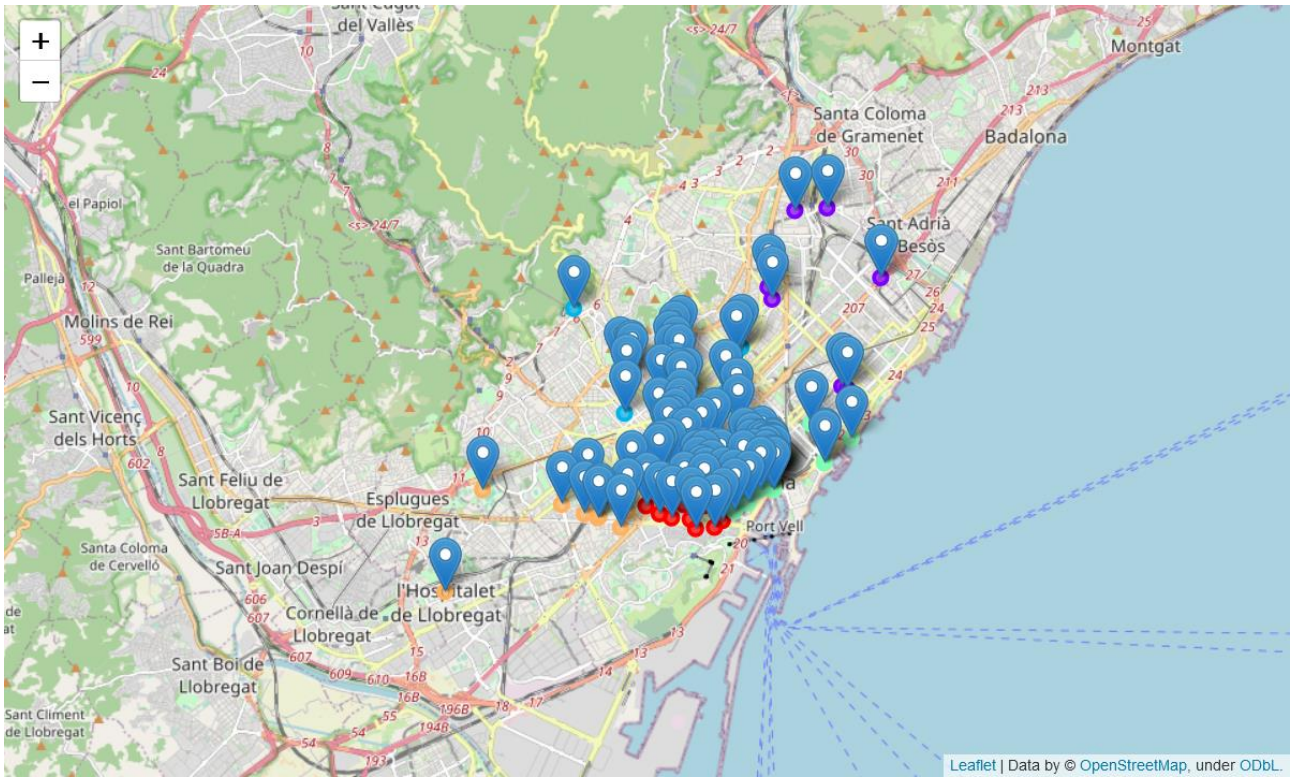
London:



Milan



Barcelona:



Then I finally create a dataframe to display the number of nightspots in each cluster.

Los Angeles:

	Bar Number
0	45
4	34
1	15
2	5
3	1

New York:

	Bar Number
2	32
0	29
3	22
1	12
4	5

London:

	Bar Number
1	45
3	26
0	17
4	8
2	4

Milan:

	Bar Number
1	31
3	30
0	22
2	14
4	2

Barcelona:

	Bar Number
0	45
4	34
1	15
2	5
3	1

DISCUSSION OF THE RESULTS

The city with more nightspots is New York: 193. Follow LA (135), London (118), Barcelona (111) and Milan (101).

Therefore, in a quantitative way the most chaotic city is New York, in case someone would choose to take a party-trip.

According to the clusters:

In LA venues are more concentrated on the Downtown Los Angeles (red cluster, 0) and Western Los Angeles (orange cluster, 4). To avoid Long Beach area and Southern LA (green cluster, 3).

In all New York City there is a high density of nightspots, particularly Lower Manhattan, represented by the blue (2) and red (0). Luckily all the other cluster are close.

The Center of London is the city's area with the greatest number of nightspots and it is represented by the purple cluster (1), which also includes the North part of the metropolis. Cluster 1 is close to the second ranked group: cluster 3 (green). Other areas are sort of dispersive.

About Milan, cluster 1 (purple) and 3 (green) have almost the same quantity of venues and both are in the East side, in the center of the city. Cluster 0 (red), which has an interesting number of bars, is in the center too, however in the West side.

In Barcelona, the green cluster, representing the Barceloneta area, is the group with the highest density of bar, however the red cluster has the greatest number of nightspots.

CONCLUSION

Personally, a traveler should choose New York. A wide amount of places where to party, all close to each other.

However, a person going to LA should book a hotel between Downtown Los Angeles and Western Los Angeles.

A person going to London should book an accommodation in the Center of London, which basically coincides with the cluster number 1.

A person going to Milan could stay basically everywhere in the center of the city, as it is close to the top 3 areas.

Finally, a person going to Barcelona should book a place in the Barceloneta (also because it is really amazing over there, I have been once).