

Coursera Data Science Capstone Project

Barcelona vs Paris - How (dis)similar are they?

Quynh Vo

September 14, 2020

1. Problem and discussion of the background.

Barcelona and Paris are two of my most favourite cities in Europe. But why is it that I like these cities the most? Are they similar or very dissimilar. Both cities have different backgrounds, but what makes them so popular. Let's find out if we can use data to solve my problem and to find out which city should be chosen for the next visit.

2. How will data be used to solve the problem.

Now we have chosen to use data to find out the (dis)similarities between the two cities, we need to decide which/what data to use. Let's use data from Wikipedia and FourSquare and find out if they can help us to find the answer!

To specify our quest a little bit more we will look for venues and more specifically the number of Vietnamese, Chinese, and Japanese restaurants in both cities as well as list down the 10 most common venues in both cities Neighbourhood wise. I have chosen these types of cuisines as I have an South-East Asian background. Foursquare is a perfect provider to supply us this data.

Wikipedia will be used for getting a start on geospatial information on the districts of Barcelona and arrondissements of Paris.

The outcome of the study will be used to determine to which city the next visit will be organized.

3. Let the battle begin. The used methodology for approaching my problem.

First we need to get data on all boroughs and neighbourhoods of Barcelona and Paris. For Barcelona I created a CSV-file, which I constructed by gathering all needed data from Wikipedia.

For Paris a JSON-file is available on the internet. The city of Paris has made this available through the following website: https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin

For both cities the WGET-package is used to download the data and the Folium package is used to show every neighbourhood on the map. Also the GeoPy Nominatim package is used to determine the city centers latitude and longitude so the cities are displayed centered on the map.

Foursquare is then used to retrieve a Venue Top 100 per neighbourhood to perform two analysis. The first analysis is to compare both cities on their similarities and dissimilarities based on all venue categories available. And the second analysis is to compare both cities based on some selected categories, as these have a personal preference as mentioned before.

To perform both analysis two functions are defined and used. The first function has no condition for the category, whereas the second function does contain the condition for the category.

As for some categories Foursquare offers the possibility to further distinguish the category an extra lists of hierarchical categories is used to retrieve the 'main category'. For example the 'Japanese restaurant' has for example the subcategory 'sushi restaurant', 'ramen restaurant' and 'yakitori restaurant'. As we to use the 'main' category, a hierarchical list is retrieved from Foursquare to make the desired selection and analysis possible.

For clustering the neighbourhoods the K-Means clustering method and to visualize the charts the matplotlib package is used.

4. Results and discussion

Our analysis on the overall venues show that Paris has a lot more venues when compared to Barcelona. In Paris there are more than 5000 venues, 5248 to be exact, whereas Barcelona has almost 3000 venues, 2936 to be exact.

```
print(barcelona_venues.shape)
barcelona_venues.head(5)

(2936, 7)
```

Fig 1. Number of venues in Barcelona.

```
print(paris_venues.shape)
paris_venues.head(5)

(5248, 7)
```

Fig 2. Number of venues in Paris

When comparing the number of neighbourhoods there's little difference between the two cities. Paris has 80 neighbourhoods, whereas Barcelona has 75. Paris though, has twice the number of boroughs when comparing both cities. Paris has twenty (20) boroughs, called arrondissements in French, and Barcelona has ten (10) boroughs, called districts.

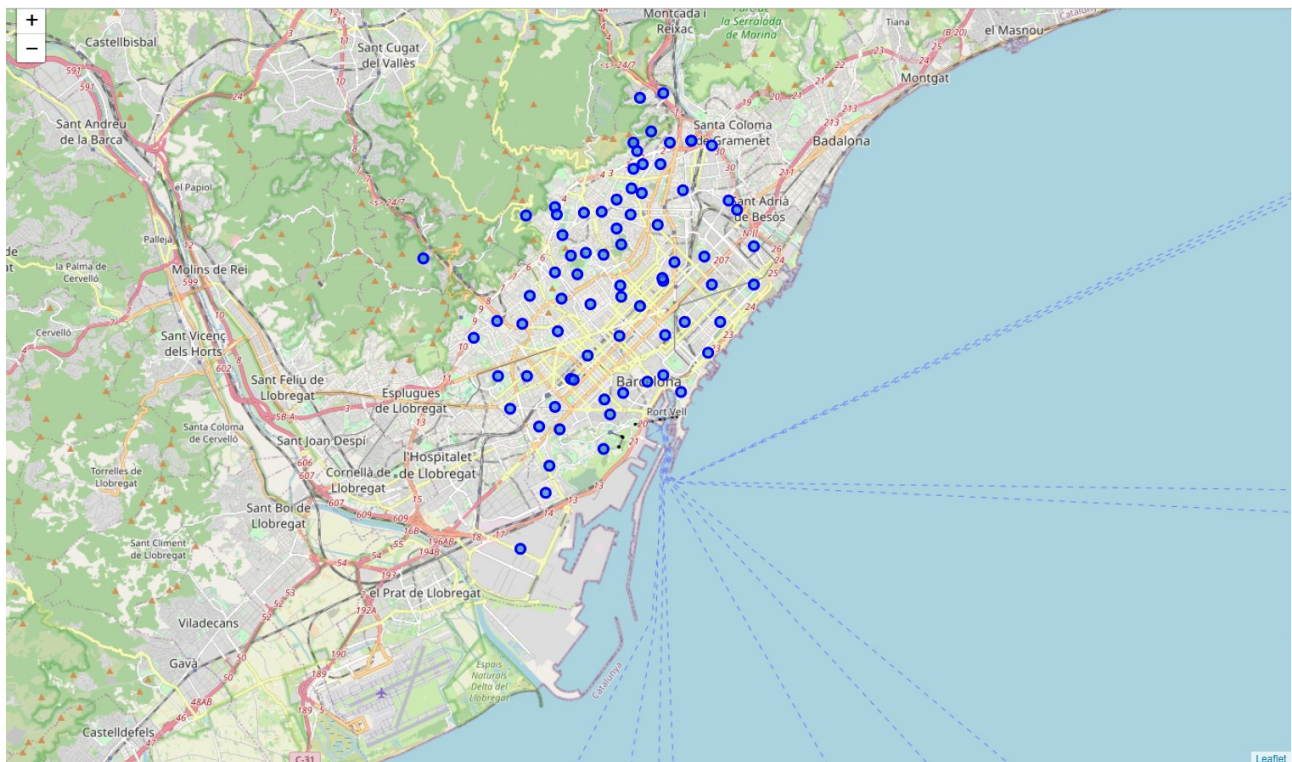


Fig. 3. Barcelona and it's 75 neighbourhoods on the map.

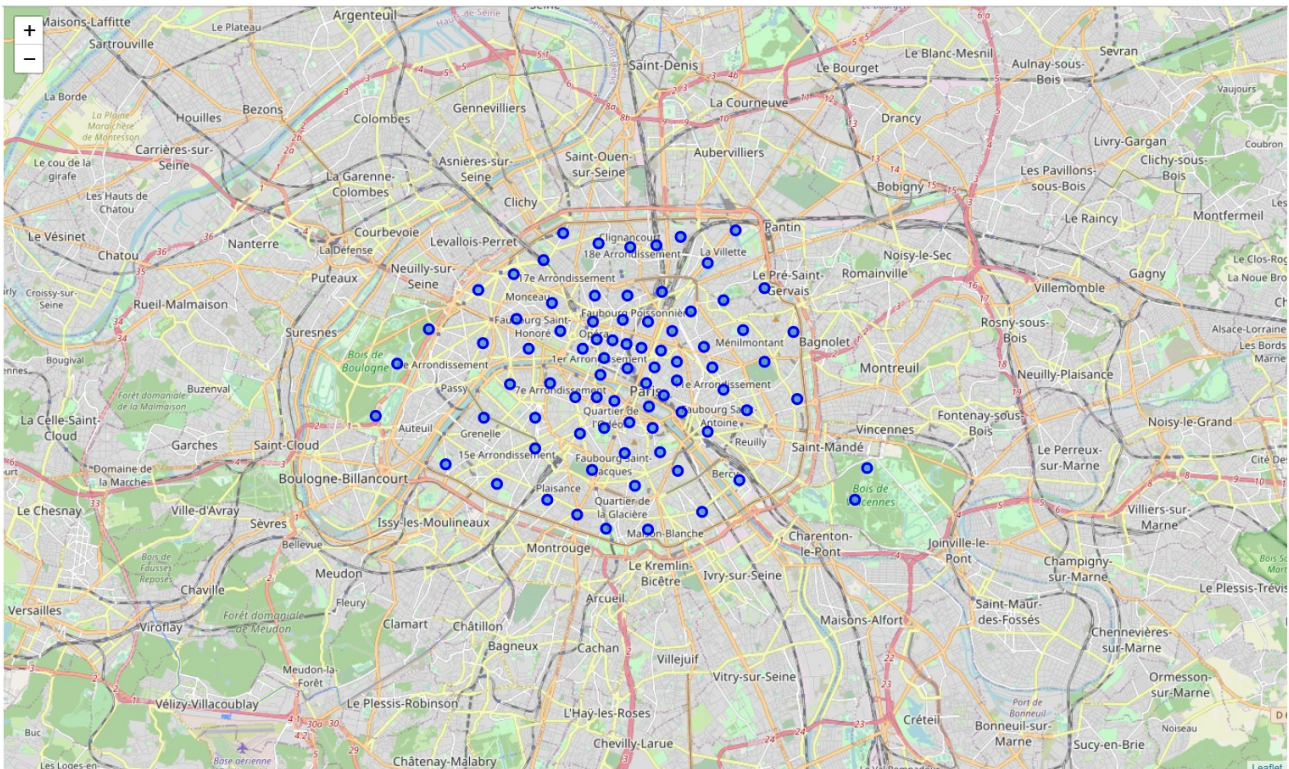


Fig. 4. Paris and its 80 neighbourhoods on the map.

When examining the clusters for Barcelona, we can see that cluster 6 is the biggest cluster and that most of the neighbourhoods can be found in the center of Barcelona. The second biggest cluster seems to be cluster 2 and most of it's neighbourhoods can be found outside of the city center of Barcelona. This can be seen in figure 5.

When examining the clusters for Paris, we can see there are three big clusters. These are cluster 4, 6 and 1 are the biggest clusters. Cluster 4 has 28 neighbourhoods is most of them are located on the western side of Paris. Cluster 6 has 18 neighbourhoods and most of them are located on the eastern-center side of Paris. Cluster 1 has 15 neighbourhoods and these are more spread over Paris. The clusters for Paris can be seen in figure 6.

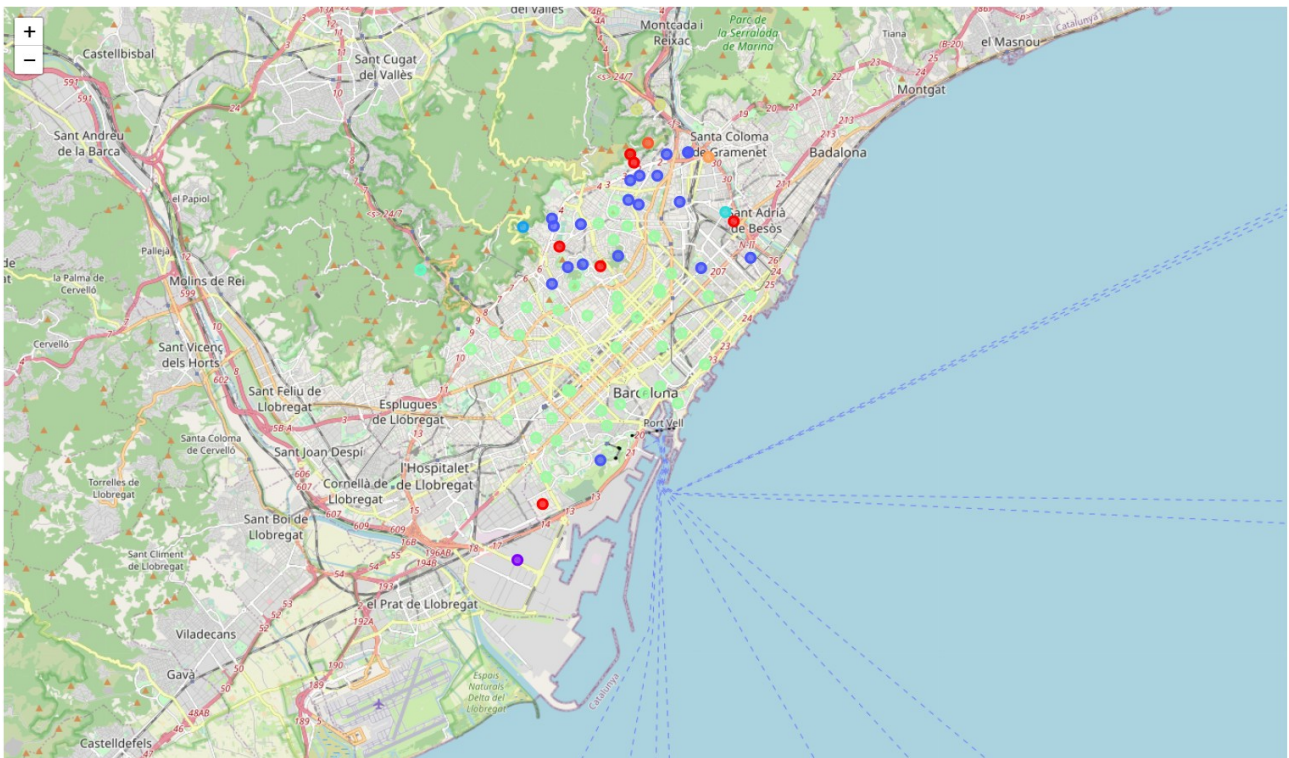


Fig. 5. All Barcelona venues split into 10 clusters with K-Means and shown on the map.

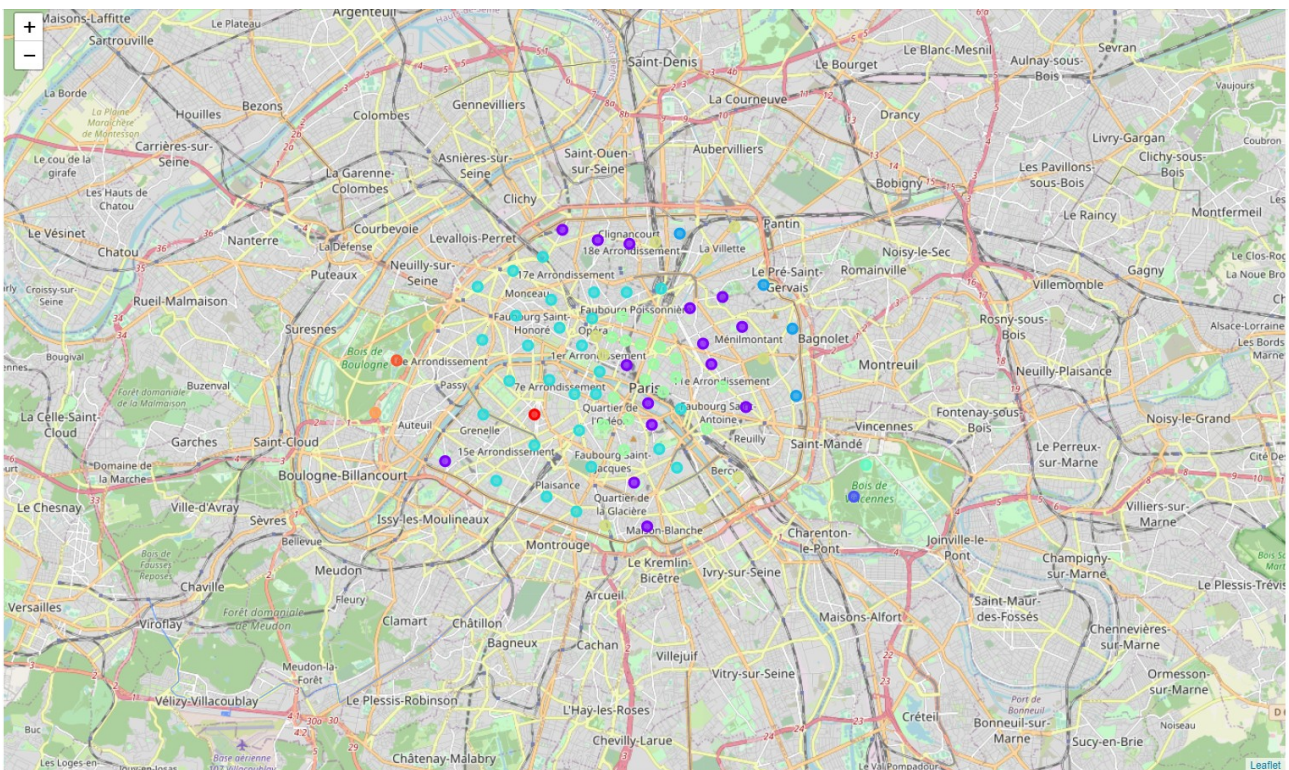


Fig. 6. All Barcelona venues split into 10 clusters with K-Means and shown on the map.

In our second analysis among the venues of Paris and Barcelona, a closer look has been done for the Chinese, Japanese and Vietnamese restaurants. In Paris a number of 193 of these restaurants can be found, whereas in Barcelona there can be found only 108.

When looking at the generated bar charts we can see that for Paris in most neighbourhoods restaurants can be found of each category. For Barcelona this is not the case. From the Barcelona bar chart we can also see that even for a lot of districts no Vietnamese restaurants can be found as well.

For both cities we can see that there are a lot of Japanese restaurants and this category can be found the most in some neighbourhoods as well, as the Japanese restaurants has the most peaks in both bar charts.

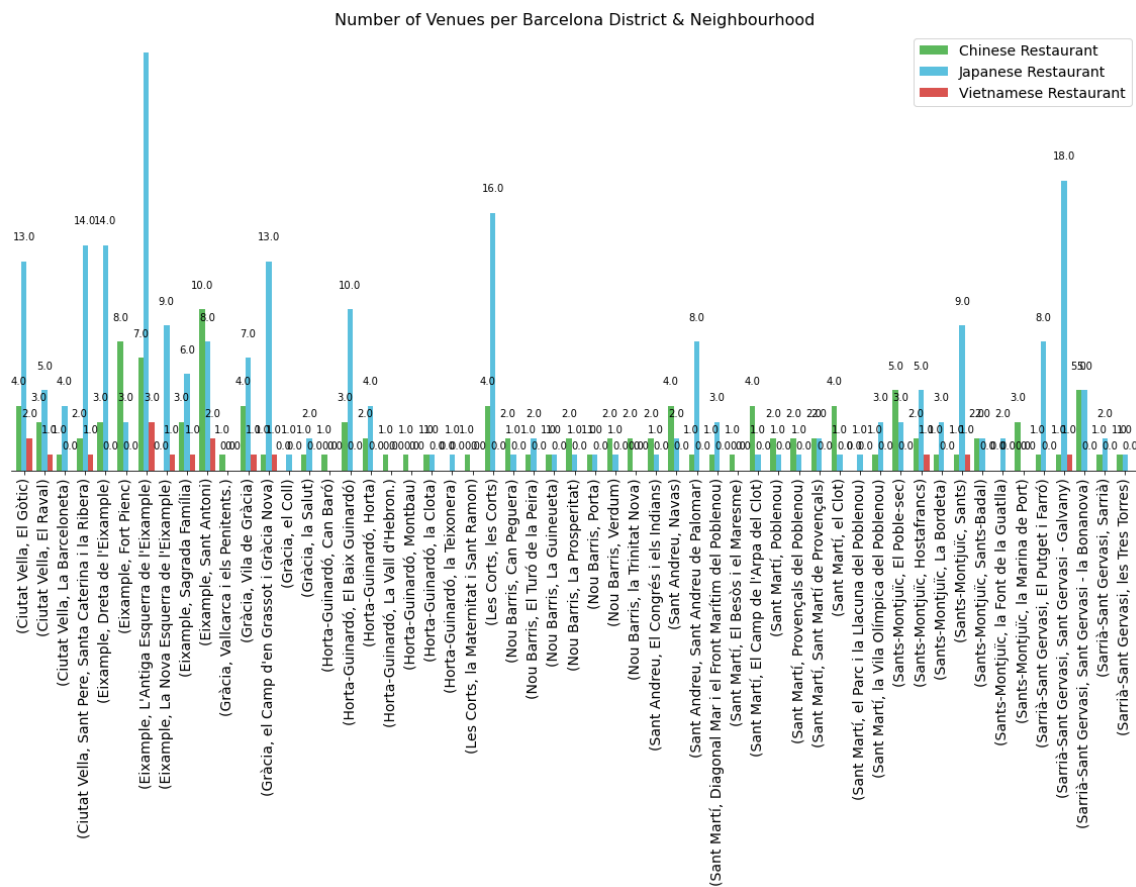


Fig. 7. Number of venues per Barcelona 'District' and neighbourhood.

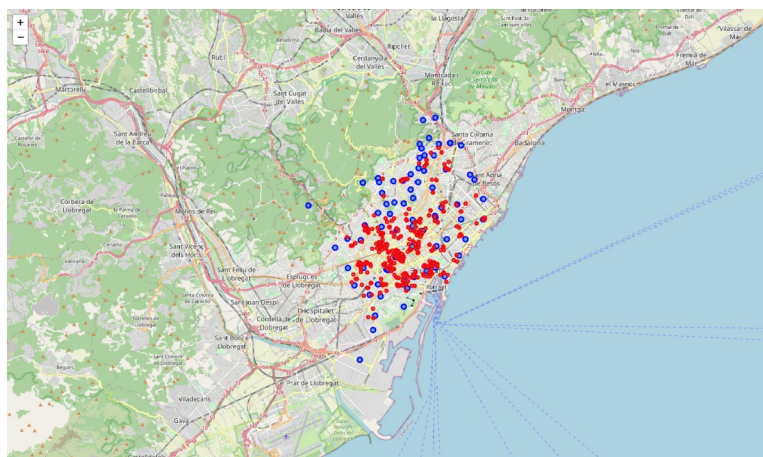


Fig. 8. Selected category venues in Barcelona.

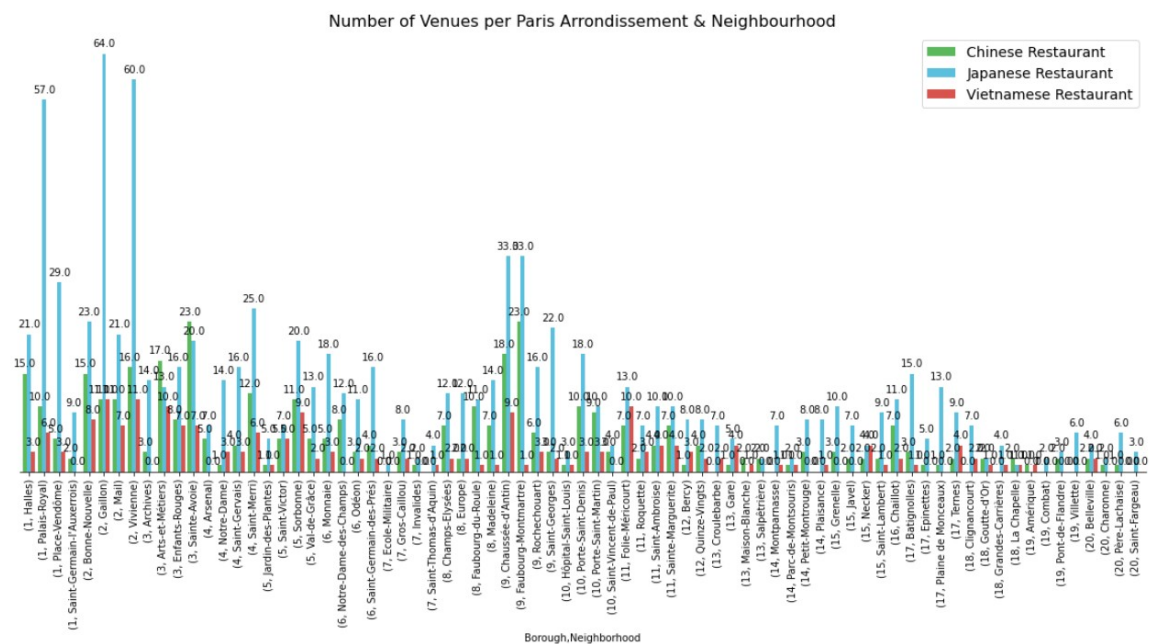


Fig. 9. Number of venues per Paris 'Arrondissement' and neighbourhood.

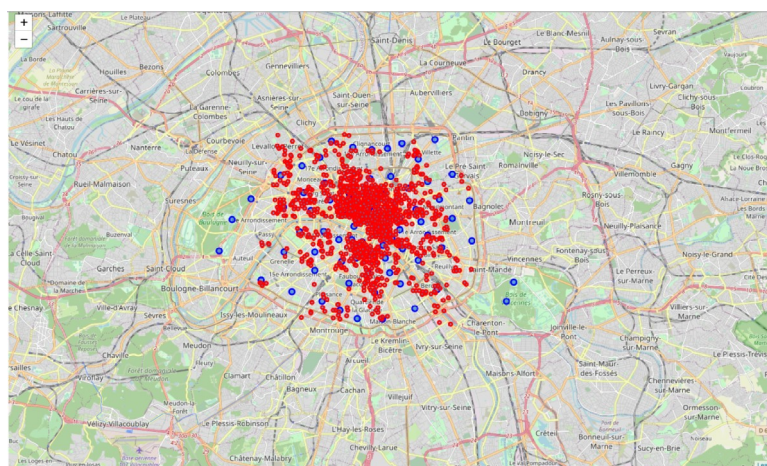


Fig. 10. Selected category venues in Barcelona.

5. Conclusion

Having a look at the clusters for both cities we can see the biggest clusters can be found in the centers of both cities. Barcelona has one big cluster and Paris has two big clusters which together also can be found in the center of the city.

When having a look at the selected categories, we can see that Paris is better represented than Barcelona. Almost every neighbourhood has all three restaurant categories and Japanese restaurants are well represented in both cities.

So are both cities similar or dissimilar? When looking at the cluster examination both cities are quite similar as on the map the biggest clusters are located in the city centers. However when looking at the selected categories, Paris is definitely the better choice as all three restaurant categories are better represented.

So the final conclusion is that **Paris** should be the next destination!