**Using Data Science to open a Wine bar in Paris.**

# Introduction & Business problem

In case my career as Data scientist fails (let's hope it doesn't), I want to open a wine bar in Paris, France.

Of course wine, **I'm french !**

The problem is that, from my experience, Paris has multiple areas where people go out for a drink and these areas are not concentrated but rather spread around the city.

Therefore, where is the best location to open a new wine bar to ensure enough clients to be successful ?

To ensure success, I need the bar to be in a location where the concentration of venues such as theaters, cinemas, restaurants demonstrates an active life in the area. Using the Foursquare data, I will geolocate the venues and find the best spot to open my wine bar.

# Data section

To provide an analytical answer to the business problem of where to open my future wine bar in Paris I will do :

* A segmentation of Paris inner-city using a .geojson file : available here [data.gouv](https://www.data.gouv.fr/fr/datasets/arrondissements-1/)
* Venues data related to the neighborhoods using Foursquare API (Category of the venue, customer rating, ...)

# Methodology

In order to find the right spot to open our wine bar, we will proceed as follow:

* We will import Paris neighborhoods coordinates from a geojson file and place the markers on a map
* We will use the Foursquare API to retrieve venues based on the Night life and Culture and leisure categories
* We will also identify the number of wine bars in each neighborhood to analyse the comptetitors based on the density
* From these results, we will use the **K-mean clustering** method to find identify different types of neighborhoods based on the kind of cultural sites and night life venues that they offer
* Finally, we will display a choropleth map highlighting the density of wine bars in Paris and with the identified clusers as markers

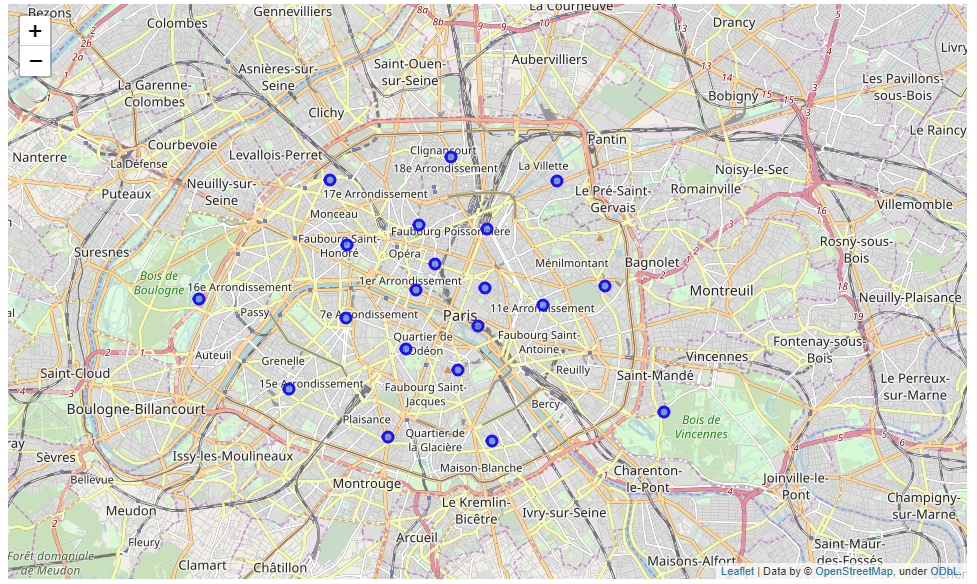
# Analysis

## **Loading the Paris coordinates**

From a .geojson file arrondissements.geojson from available the French national open data web site [data.gouv](https://www.data.gouv.fr/fr/datasets/arrondissements-1/) website, we will first display a map of Paris, France and place markers on the centers of each of Paris' 20 neighborhoods (arrondissments).

We fill the DataFrame dfparis with data from the .geojson file and mark these coordinates on a map.



*The above map shows a detailed map of Paris' neighborhoods.*

Now let's identify the venues around each of these center coordinates of the city using the **Foursquare API**.

## **Foursquare**

We're interested in venues from the **Night life**, the **Culture and leisure** and the **Wine bar** categories. The density will indicate the activity of the area: it will show the areas where tourists make visits (cultural places) and also where locals and tourists go during night time.   
We are interested in areas where there is a good density of bars, nightclubs and pubs but less wine bars (less competitors).

Let's send a query to retrieve the venues using Foursquare API. To do so, we will send a query to Foursquare for each Paris' neighborhood coordinates and look for venues in the three selected categories: *Night life*, *Culture and leisure* and *Wine bars*.

**Night life** category includes venues like :

* Bars
* Cocktail bars
* Whisky bars
* Lounge
* etc...

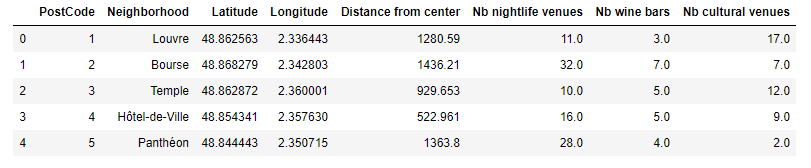
**Culture and leisure** category includes venues like :

* Art gallery
* Casino
* Theater
* Concert venue
* Aquarium
* Cinema
* Museums
* etc...

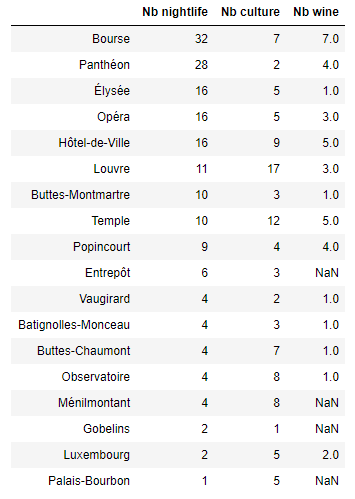
Finally, the **Wine bar** category is a sub category itself.   
We use this last category to isolate the results in each neighborhood to evaluate the density.

#### **Night life, Culture and leisure and wine bars type of venues:**

We create a dictionary to isolate the venues which categories are most likely to be our competitors. We also want to store the number of venues from each category per neighborhood in the initial dataframe dfparis.



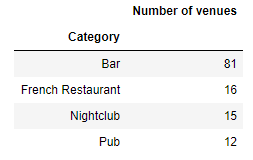
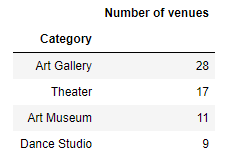
We now have the number of nightlife venues, of wine bars and of cultural venues per neighborhood.



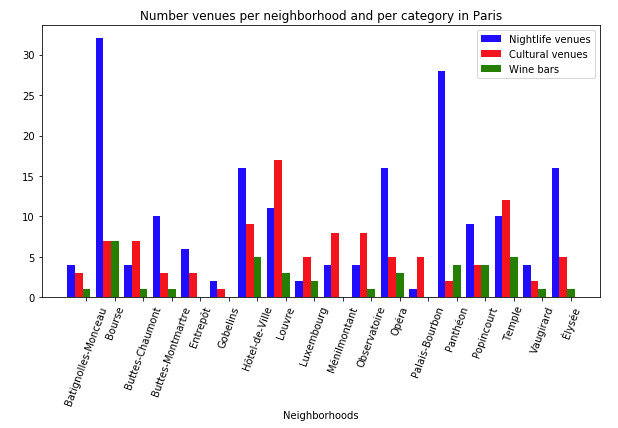
The neighborhoods "Bourse", "Panthéon" and "Elysée" have the largest number of Nighlife venues.   
The neighborhoods "Louvre", "Temple" and "Hotel de ville" have the largest number of Cultural venues.   
The neighborhoods "Bourse", "Hotel de Ville" and "Temple" have the largest number of wine bars venues.

These later neighborhoods will certainly not be interesting as they are competitive areas. There are probably too many bars and wine bars already.

If we count the number of venues per **Category :**

In the **Nightlife** category, on the left-hand side, the most frequent venues are Bar, French restaurant, Nightclub and Pub.   
In the **Cultural** category, on the right-hand side, the most frequent venues are Art gallery, Theater, Art museum and Dance studio.

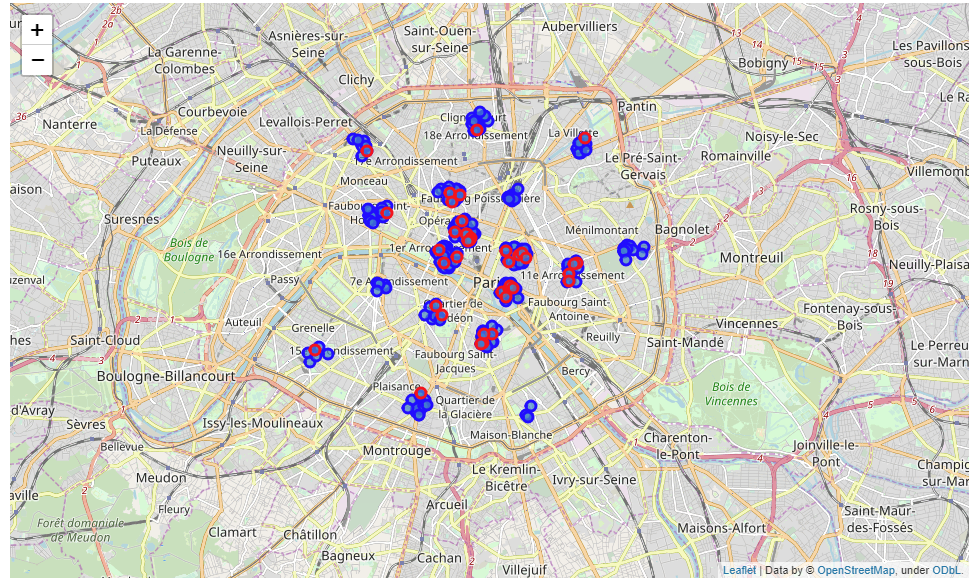


Bar chart showing the number of venues per neighborhoods and per category.   
We see that the Nightlife venues are predominant in the city. We also find cultural venues in every location.   
Nevertheless, we see that the number of wine bars is sometimes very low or missing. 

Let's dig more to find out why...

#### **Let's visualize these results on a map.**

**In red**, you can see the location of wine bars across the city. They are mostly concentrated around the city center.  
**In blue** are the other venues from the Foursquare categories "Night life" and "Culture and leisure" on a Paris Map



This map shows that every location where we have wine bars, we also identify groups of other venues such as cultural and nightlife venues. We can already notice that there are areas where there are cultural places and nightlife venues but no or very few wine bars: 10th arrondissement, 20th arrondissement (east), 8th arrondissment near the champs elysees and 18th arrondissement near Montmartre.

In order to dig further, we will prepare our data for a K-mean clustering method.

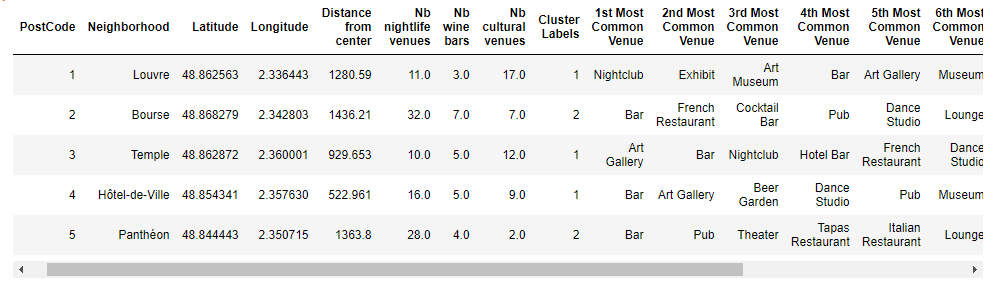
## **K-Mean Clustering**

Let’s define k-means to cluster the neighborhood into 3 clusters.

After computing the k-mean algorithm with *k=3,* each neighborhood is assigned to a cluster based on similarities of the nightlife and cultural venues in Paris.

The following table shows the 10 most common venues in each neighborhood.

The column name “**Cluster Labels**” indicates the cluster to which the area is assigned to.



|  |  |
| --- | --- |
|  |  |

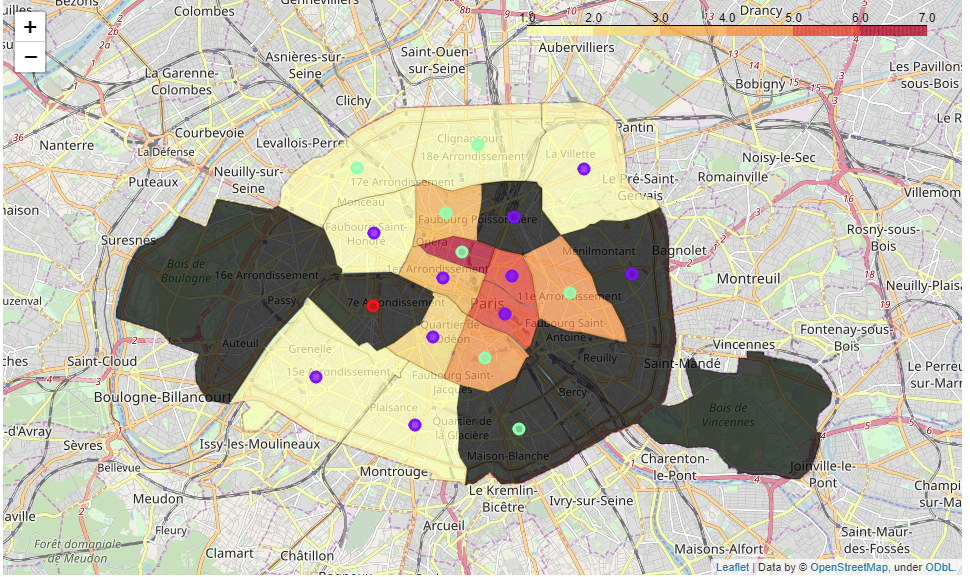
This bar chart indicates the most common venues in each of the three clusters:

* **Cluster 0** is mostly Cultural sites with "History museum"
* **Cluster 1** contains also lots of cultural venues but also some nighclubs and bars. These areas seem to be very active during the day time as well as at night time
* **Cluster 2** is reflects a very active nightlife with music venues and a large number of bars.

Based on the analysis so far and on the requirements, we might want to avoid Cluster 2 areas because the daytime activity does not seem to be very active and the number of competitors would be high. Only one area is identified as Cluster 0 which we might not be very interesting in terms of popular location.

## **Visualization of clusters overlapping wine bars density**

Now that we have understood the meaning of the 3 clusters, let's have a look at Paris using a choropleth map. Each neighborhood is highlighted with de density of wine bars from non-colored (0 results) to red (high density of wine bars).  
We have just seen that Clusters 0 and 2 might not be very interesting areas. Therefore, an ideal location would be a neighborhood affected to **Cluster 1** and with a low density of wine bars.



*Choropleth map of Paris highlighting the density of wine bars and the clusters obtained using k-mean clustering.*

# Results & Discussion

From our exploratory analysis, we first identified that neighborhoods like Bourse, Hotel de Ville and Temple would **not** be interesting areas based on th number of competitors. The choropleth map shows that this assumption was true considering that they are a high wine bar density areas and they belong to clusters 1 and 2 which have an active night life.

Furthermore, the non-supervised K-Mean clustering method enabled to identify the neighborhoods where both cultural and nightlife activities are active. Locations like Hotel de Ville and Temple fit this profile and we can see from the choropleth map that wine bars have successfully opened in these areas.

One very central location has no wine bars identified, it is the Palais-Bourbon neighborhood. It is in Cluster 0, the one having mostly cultural sites and very few nightlife venue. This can also be explained because this area has lots of embassy and government buildings. Therefore there are very few nightlife. A sleepy area is not what we seek.

Now, chances are to open a wine bar in a similar location but with less competitors. The choropleth map highlights two locations that fullfil this requirement: **Entrepôt** which is the (10th arrondissement) and **Menilmontant** (20th arrondissement).

* The first (**Entrepôt**) is close to very central Paris. It is a popular neighborhood getting gentrified which borders the highly popular and competitive areas of Bourse and Temple. This area is very promissing.
* The second (**Menilmontant**) is quite further away from the city-center but offers a very diverse part of Paris. Young and active middle class with affordable real estate prices is also a good target. Yes there are less tourists despite a few cultural sites and music venues.

# Conclusion

Paris being a large mutli-cultural city with an active nightlife, competition might be harsh to be successfull in opening a wine bar.

Despite a large number of bars, clubs, pubs we were also interested in the presence of cultural sites in the city which would give an indicator of a tourist presence.  
Our results show that we identified fairly central locations where the opening of a wine bar has good chances of being successfull.

For a further analysis, it might be interesting to zoom in at the neighborhood level and at the street level of both **Entrepôt** and **Ménilmontant**.

Thank you for reading my work.

**Cheers !**