THE BEST LOCATION FOR A CRAFT BEER BAR IN PARIS

Using data science to find the best location

Introduction:

Background

The craft beer movement, a trend imported from the United States, is spreading all around Europe and is growing steadily in France, originally considered as a wine country. The insatiable demand for quality craft beer led to the creation of more and more microbreweries and beer bars specialized in craft beer over the past 15 years.

Goals

Through data science, the following analysis tries to define the most favorable districts to establish a craft beer bar in Paris, France. Using data analysis, data visualization and machine learning, we will narrow down neighborhoods in Paris based on different criteria:

First, we want our bar to be located in an attractive and vibrant area of Paris, with an active nightlife. The presence of other nearby venues will attract people and create traffic. On the other hand, as our concept represents a niche (craft beer lovers), we want to avoid the presence of direct competitors in the same neighborhood. At last, Paris is a very expensive city and the exorbitant real estate prices can be a barrier to entry for opening a business. Thus, we will focus on neighborhoods whose prices match our budget.

This analysis represents a draft market study for a business plan or an executive summary. It is intended for potential entrepreneurs who want to better understand the market of craft beer pubs in Paris by providing a data-driven insight into the best locations, based on specific criteria.

Data:

Paris is divided in 20 boroughs called "Arrondissements" and 80 neighborhoods. Data about boroughs and neighborhoods is available online. Names and geographical coordinates for each neighborhood can be found here: https://www.data.gouv.fr/fr/datasets/quartiers-administratifs/.

The average price per Square Meter is provided by DataFrance: https://datafrance.info/paris-75000.

To compare neighborhoods among each other's, we will use the Foursquare API that list all venues in specific areas. We will then be able to cluster the neighborhoods based on their similarities using unsupervised machine learning technique to select the best fit for our craft beer pub.

I. Collecting the data

The data used is collected from two sources : https://datafrance.info/paris-75000 and https://www.data.gouv.fr/fr/datasets/quartiers-administratifs/ as JSON files.

After cleaning, the master data looks like this:

|	Neighborhood	Arrondissement	Latitude	Longitude	Price
0	Place-Vendome	1	48.867019	2.328582	9816
1	Halles	1	48.862289	2.344899	9816
2	Palais-Royal	1	48.864660	2.336309	9816
3	Saint-Germain-l'Auxerrois	1	48.860650	2.334910	9816
4	Gaillon	2	48.869307	2.333432	11262
...	...	...	...	...	...
75	Amerique	19	48.881638	2.395440	7909
76	Charonne	20	48.854760	2.407430	7853
77	Saint-Fargeau	20	48.871035	2.406172	7853
78	Belleville	20	48.871531	2.387549	7853
79	Pere-Lachaise	20	48.863719	2.395273	7853

80 rows × 5 columns

It contains the necessary of information: Neighborhood, Borough, Latitude, Longitude and average square meter price. Note that the table shows 80 rows representing the 80 neighborhoods.

II. Segmenting the data

A- Creating a map of Paris and its neighborhoods

I used the geographical data of each neighborhood to visualize them on the map with the Folium library in Python.
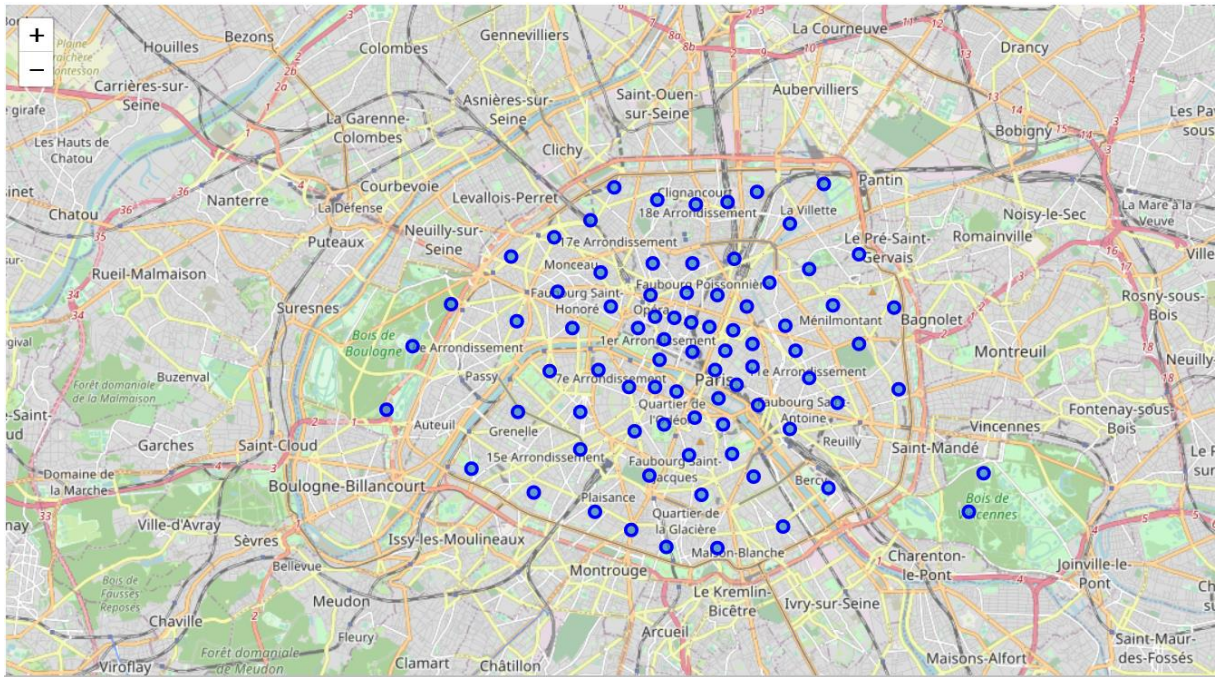
## B- Using Foursquare API to get the venues

We want our bar to be located in a good area with an active night life and similar venues around to attract customers. To explore and define the type of neighborhood, I use the Foursquare API to get all nearby venues around. I set the limit to 500 in a radius of 1km.

Here is a head of the list with the venues name, category, latitude and longitude information from Forsquare API:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Place-Vendome	48.867019	2.328582	Hôtel Ritz	48.868106	2.328892	Hotel
1	Place-Vendome	48.867019	2.328582	Place Vendôme	48.867798	2.329741	Plaza
2	Place-Vendome	48.867019	2.328582	Hôtel Mandarin Oriental	48.866987	2.327178	Hotel
3	Place-Vendome	48.867019	2.328582	Charvet	48.868312	2.330282	Men's Store
4	Place-Vendome	48.867019	2.328582	Ladurée	48.866121	2.328449	Dessert Shop

I put all the venues into a data frame. We notice the table has 5036 rows which means that the Foursquare request found 5036 nearby venues with 297 unique venues categories.

We can check how many venues are returned for each neighborhood to see the most active ones. We can visualize them using a bar chart:
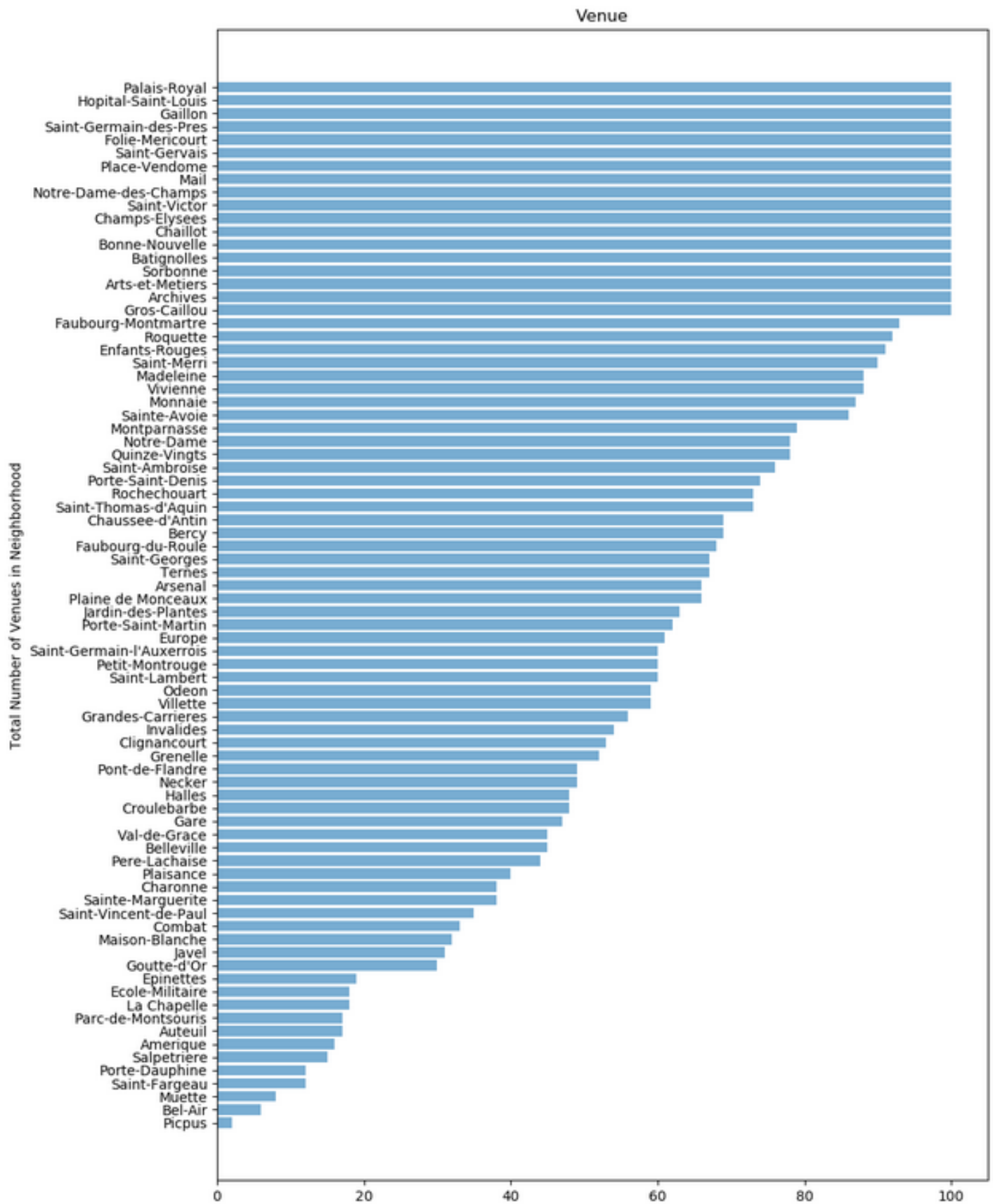
Venue

C- Creating a new data frame with the top 10 venues for each neighborhood

|	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	Amerique	Supermarket	Pool	French Restaurant	Health Food Store	Bistro	Park	Bus Stop	Plaza	Bed & Breakfast	Café
1	Archives	French Restaurant	Hotel	Coffee Shop	Italian Restaurant	Bar	Art Gallery	Clothing Store	Japanese Restaurant	Plaza	Cocktail Bar
2	Arsenal	French Restaurant	Hotel	Plaza	Park	Vegetarian / Vegan Restaurant	Gastropub	Boat or Ferry	Thai Restaurant	Pedestrian Plaza	Tapas Restaurant
3	Arts-et-Metiers	French Restaurant	Hotel	Cocktail Bar	Wine Bar	Chinese Restaurant	Italian Restaurant	Restaurant	Coffee Shop	Japanese Restaurant	Vietnamese Restaurant
4	Auteuil	Tennis Court	Stadium	Racecourse	Outdoors & Recreation	Sporting Goods Shop	Museum	French Restaurant	Botanical Garden	Office	Garden

In this table, we discover the 10 most common venues for each neighborhood. This already gives us some insight about the type of neighborhood and if it can be a good fit for a craft beer bar.

III. Clustering similar neighborhoods using K-means

Now that we know some characteristics of Parisian neighborhoods, we can create clusters based on their similarities using an unsupervised machine learning tool called K-means.

A- Defining the best number of clusters

To define the most appropriate number of clusters, I use the silhouette score. The best number of clusters is 6.
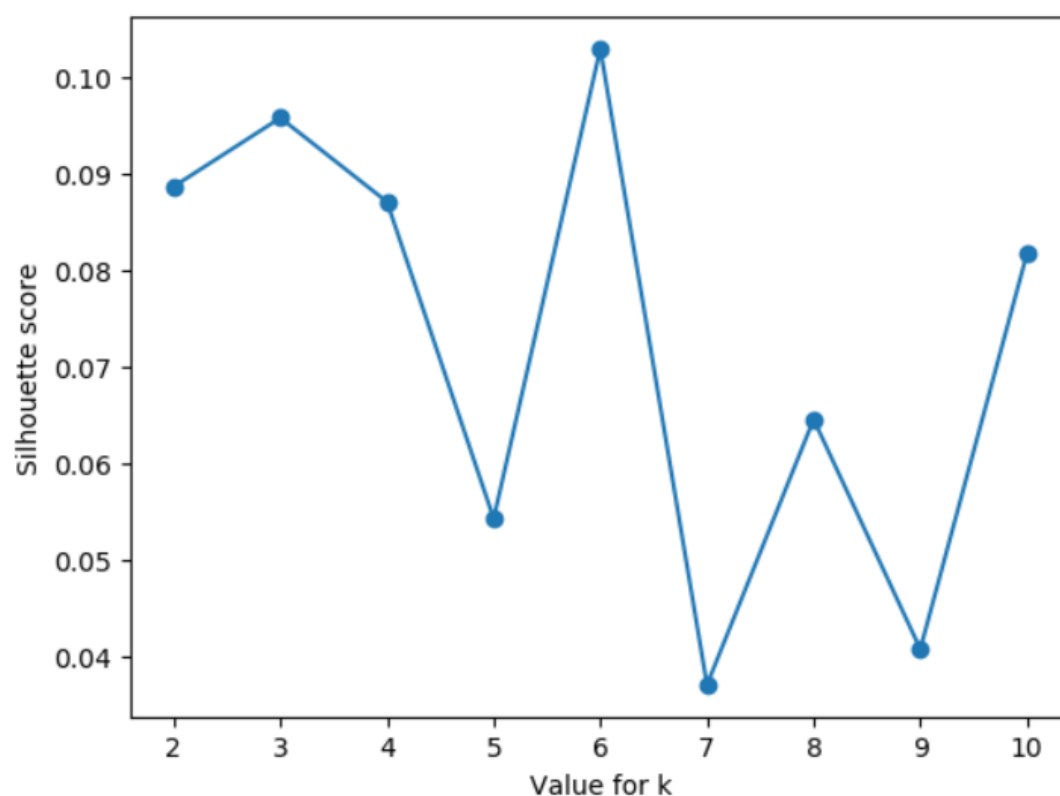
B. Clustering the neighborhoods in Paris

   K-Means will assign each neighborhood to a cluster. I added the cluster label in a new column of the data frame. Now it is time to visualize the 6 different clusters on the map, using Folium.
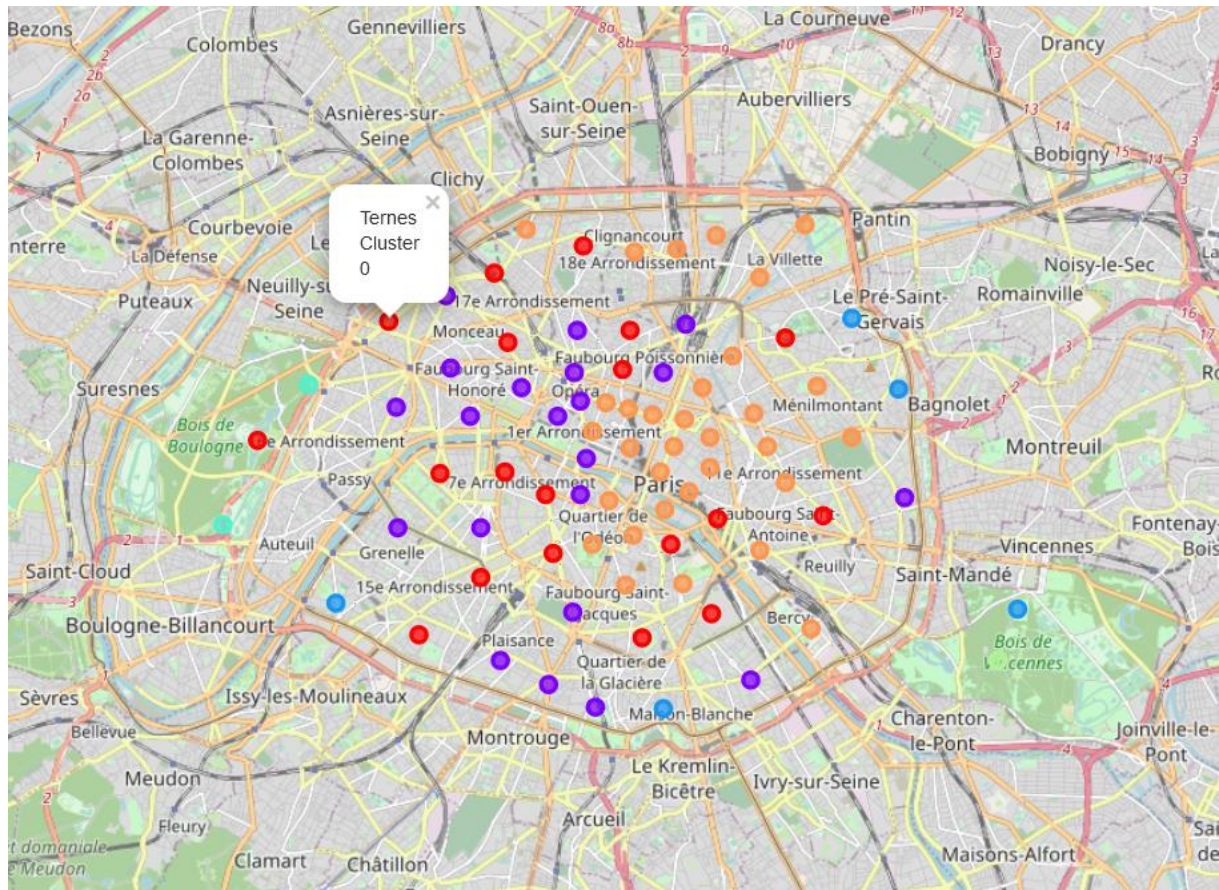
Neighborhoods are grouped into 6 clusters based on their characteristics (top 10 venues). To open our beer bar, we are looking for an active location, where people go out. So let's analyze each clusters:

Cluster 0 (red) : neighborhoods with a majority of restaurants as nearby venues.

Cluster 1 (purple) : neighborhoods with a majority of hotels and some restaurants and bars.

Cluster 2 (blue) : not interesting neighborhoods to open a bar

Cluster 3 (turquoise) : not interesting neighborhoods to open a bar

Cluster 4 (green) : not interesting neighborhoods to open a bar

Cluster 5 (orange) : neighborhoods with a majority of bars, café and restaurants

After analyzing each cluster, we want our bar to be in a neighborhood belonging to the cluster 0, 1 or 5. Those are the neighborhoods where we find numerous restaurants, bars, hotels, and cafés.

IV. Selecting Neighborhoods based on their average square meter price

Now that we narrowed down neighborhoods based on the nearby venues, it is important to look at the average square meter price for each neighborhood. Some neighborhoods in Paris are extremely expensive. Let us get rid of locations where the average price is above 9,8k€ per m² so it can match our budget. We end up with 30 rows.

Neighborhood	Arrondissement	Latitude	Longitude	Price	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	Com V
Necker	15	48.842711	2.310777	9583	5	French Restaurant	Hotel	Café	Dessert Shop	Gym / Fitness Center	Bakery	Restaurant	Pet Café	Supermarket	Ch Resta
Javel	15	48.839060	2.278076	9583	5	French Restaurant	Pizza Place	Café	Sushi Restaurant	Bakery	Thrift / Vintage Store	Spanish Restaurant	Supermarket	Moroccan Restaurant	
Saint-Lambert	15	48.834294	2.296920	9583	5	French Restaurant	Hotel	Bakery	Plaza	Thai Restaurant	Italian Restaurant	Café	Supermarket	Burger Joint	Book
Grandes-Carrieres	18	48.892578	2.334363	8874	5	French Restaurant	Hotel	Bar	Restaurant	Pizza Place	Park	Thai Restaurant	Bistro	Italian Restaurant	Sta
Combat	19	48.878639	2.380127	7909	5	French Restaurant	Italian Restaurant	Park	Restaurant	Pool	Coffee Shop	Scenic Lookout	Bar	Beer Garden	

V. Dropping neighborhoods with competitors

To avoid competing with another already established craft beer pub, we need to check which selected neighborhoods already have a Beer bar.

'0' means there is no competitor for our craft beer project in this neighborhood. '1' means there is at least a beer bar already in place. We get rid of those neighborhoods.

[33]:

	Neighborhood	Beer Bar
31	Villette	1
6	Folie-Mericourt	1
1	Bercy	1
25	Roquette	1
11	Hopital-Saint-Louis	1
10	Grenelle	1
9	Grandes-Carrieres	1
29	Sainte-Marguerite	1
3	Clignancourt	1
27	Saint-Lambert	0
26	Saint-Ambroise	0
19	Petit-Montrouge	0
28	Saint-Vincent-de-Paul	0
24	Quinze-Vingts	0
23	Porte-Saint-Martin	0
22	Porte-Saint-Denis	0
30	Salpetriere	0
21	Pont-de-Flandre	0
20	Plaisance	0
0	Belleville	0
18	Pere-Lachaise	0
17	Parc-de-Montsouris	0
15	Montparnasse	0
14	Maison-Blanche	0
13	La Chapelle	0
12	Javel	0
8	Goutte-d'Or	0
7	Gare	0
5	Croulebarbe	0
4	Combat	0
2	Charonne	0

VI. Results and discussion

We then narrow down neighborhoods with no beer bar already present. At this stage, we can show a list of areas in Paris that have an active night life, within our budget range, with no direct competitor for our craft beer pub. Let's merge those variables with our master data table.

	Neighborhood	Arrondissement	Latitude	Longitude
37	Porte-Saint-Denis	10	48.873618	2.352283
38	Saint-Vincent-de-Paul	10	48.880735	2.357471
39	Porte-Saint-Martin	10	48.871245	2.361504
42	Saint-Ambroise	11	48.862345	2.376118
47	Quinze-Vingts	12	48.846916	2.374402
48	Salpetriere	13	48.837406	2.363319
49	Croulebarbe	13	48.833734	2.347673
51	Gare	13	48.827527	2.372398
52	Parc-de-Montsouris	14	48.823453	2.337070
53	Plaisance	14	48.830317	2.315305
54	Petit-Montrouge	14	48.826653	2.326437
55	Montparnasse	14	48.837623	2.331784
56	Necker	15	48.842711	2.310777
59	Saint-Lambert	15	48.834294	2.296920
69	La Chapelle	18	48.894012	2.364387
70	Goutte-d'Or	18	48.892138	2.355536
72	Combat	19	48.878639	2.380127
74	Pont-de-Flandre	19	48.895556	2.384777
76	Charonne	20	48.854760	2.407430
78	Belleville	20	48.871531	2.387549
79	Pere-Lachaise	20	48.863719	2.395273

And visualize it on the map:
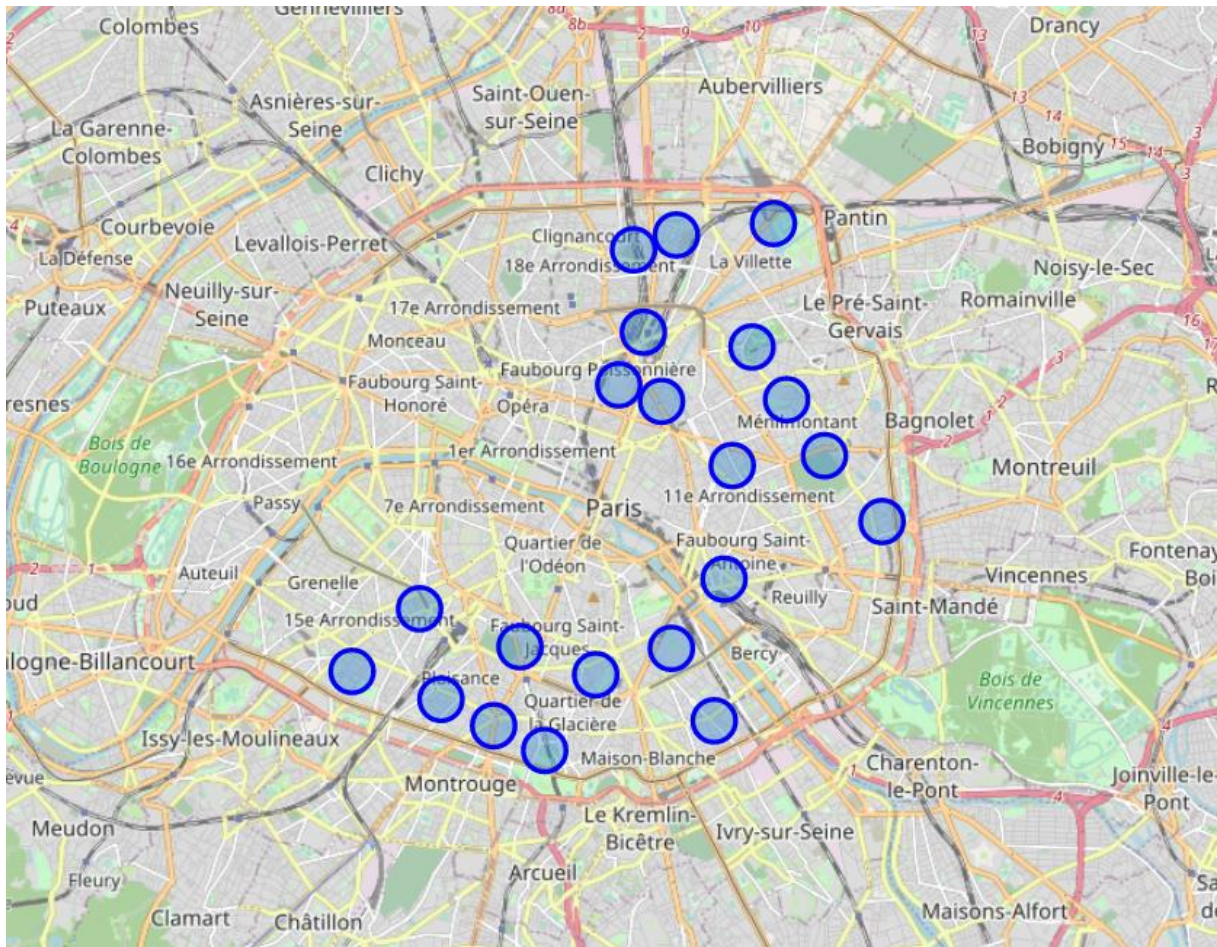
This last map shows the best location for opening a craft beer pub in Paris based on our criteria.

After sorting out neighborhoods based on their average price and the competition already established, we notice that the best places are mostly located outside of the city's hyper center. The result is still quite large. 21 neighborhoods are shown on the map in 9 different boroughs. We could be more precise by adding other criteria depending on the project, such as the population in each district and the average standard of living of the inhabitants. Nonetheless, this analysis provides an overall view of the trends in each neighborhood. This study represents a first step to understand the best location for a beer pub in Paris but does not answer all the questions: how far is the location from my supplier? are there any premises available in the area? And what about accessibility?

VII. Conclusion

Paris is a large and heterogeneous city. To determine the best place to open a beer bar, we need to know the characteristics of each district. To do so we can visualize them on a map using their geographical coordinates. The Foursquare API can obtain the venues nearby and draw up a portrait of each neighborhood based on the assigned venues. The use of K-Means classifies the neighborhoods into different clusters based on their similarities and thus allows us to visualize the neighborhoods by categories. In this case study, the neighborhoods selected are characterized by the presence of bars and restaurants nearby - attesting of the attractiveness of the area - its average price and preexistence of competitors. Several other criteria can be used to define the best location for a business. None of that would have been possible without data. The use of data enables entrepreneurs to find out useful hidden information that help them make more informed decision. A lot of data and tools are accessible and can grant an edge for smarter decision making. No surprise data analysis is becoming increasingly important to companies.

Please check my GitHub repository for more details and code -> Here