Finding the Most Promising Neighbourhoods in Paris for a New Bubble Tea Shop Business

Thomas Vanderzande

May 29th, 2021

1. Introduction

1.1. Background

Bubble Tea Shops are relatively new venues in the city of Paris. They sell various kinds of tea beverages (black, green, etc.) with customizable features (toppings, quantity of sugar, flavours, etc.). This allows almost everybody to find their ideal drink. For these reasons, they are getting more and more popular in Eastern Asia (China and Japan especially) and also in Paris.

As these places are still relatively new and more and more popular, it can be advantageous for investors to determine which neighbourhoods are the best candidates to open their new bubble tea shop business, to get a piece of an increasingly large cake.

1.2. Problem

We need to find neighbourhoods with the least amount competition i.e., tea and coffee shops. As these shops are extremely popular in eastern Asia, we will try to find neighbourhoods where people who likes eastern Asian culture and cuisine go.

To go a bit further, we are also interested in places close to park, shopping points and monuments, as these drinks tend to be consumed outside with friends whenever possible.

For this project, we will consider we are in a "non-Covid" environment.

1.3. Interest

The results can be interesting for anyone who would like to open a new business in Paris.

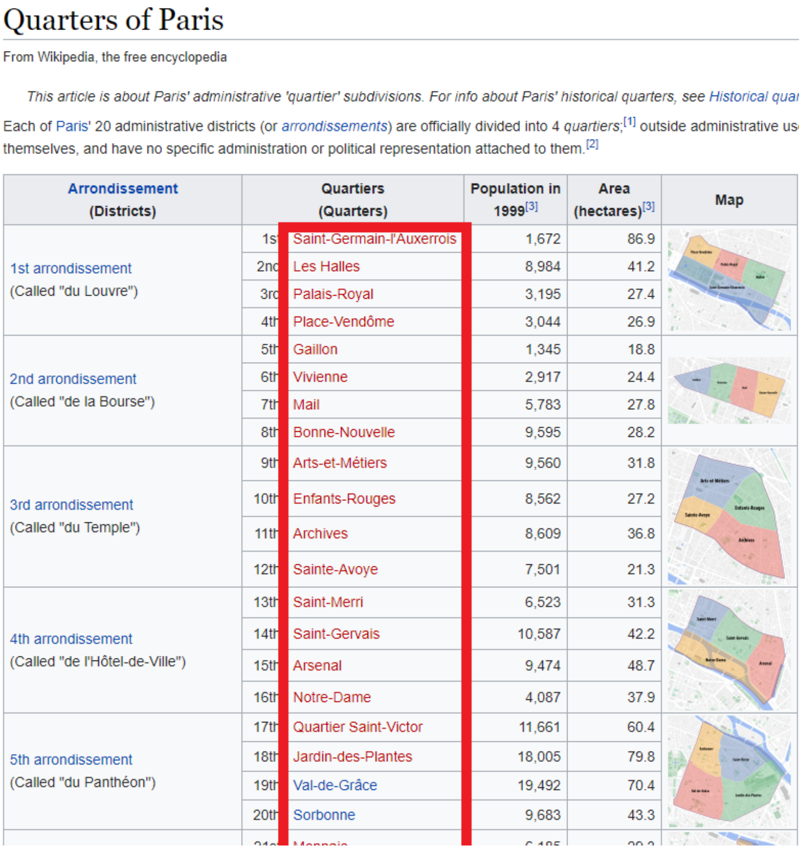
2. Data acquisition

Factors that will influence our decision are:

- Whether tea & coffee shops are common or not in the neighbourhood
- The influence of the Asian culture
- The attractiveness of the neighbourhood (park, garden, shopping points, monuments)
- The GPS coordinates can be extracted in Wikipedia via Web Scraping (French page is available for each neighbourhood), so we will use them for this project.

The complete list of data needed is then the following:

- Name of neighbourhoods (**Web Scraping on [Wikipedia](#)**)



Quarters of Paris

From Wikipedia, the free encyclopedia

This article is about Paris' administrative 'quartier' subdivisions. For info about Paris' historical quarters, see Historical qua

Each of Paris' 20 administrative districts (or *arrondissements*) are officially divided into 4 *quartiers*;[1] outside administrative us themselves, and have no specific administration or political representation attached to them.[2]

Arrondissement (Districts)	Quarters (Quarters)		Population in 1999[3]	Area (hectares)[3]	Map
1st arrondissement (Called "du Louvre")	1st	Saint-Germain-l'Auxerrois	1,672	86.9	
	2nd	Les Halles	8,984	41.2	
	3rd	Palais-Royal	3,195	27.4	
	4th	Place-Vendôme	3,044	26.9	
2nd arrondissement (Called "de la Bourse")	5th	Gaillon	1,345	18.8	
	6th	Vivienne	2,917	24.4	
	7th	Mail	5,783	27.8	
	8th	Bonne-Nouvelle	9,595	28.2	
3rd arrondissement (Called "du Temple")	9th	Arts-et-Métiers	9,560	31.8	
	10th	Enfants-Rouges	8,562	27.2	
	11th	Archives	8,609	36.8	
	12th	Sainte-Avoye	7,501	21.3	
4th arrondissement (Called "de l'Hôtel-de-Ville")	13th	Saint-Merri	6,523	31.3	
	14th	Saint-Gervais	10,587	42.2	
	15th	Arsenal	9,474	48.7	
	16th	Notre-Dame	4,087	37.9	
5th arrondissement (Called "du Panthéon")	17th	Quartier Saint-Victor	11,661	60.4	
	18th	Jardin-des-Plantes	18,005	79.8	
	19th	Val-de-Grâce	19,492	70.4	
	20th	Sorbonne	9,683	43.3	

Figure 1 Extraction of neighbourhoods' name by scraping a Wikipedia page

- GPS coordinates of neighbourhoods (**Web Scraping on Wikipedia, example [here](#)**)

Figure 2 Extraction of the GPS coordinates for a particular neighbourhood using its French Wikipedia page

As coordinates are given in hours, minutes and seconds, we need to make a conversion in order to get decimal values:

$$coord(decimal) = coord(hours) + \frac{coord(minutes)}{60} + \frac{coord(seconds)}{3600}$$

- The venues of interest of each neighbourhood **using Foursquare API**
- The coordinate of Paris centre will be retrieved using **the geocoder library**

Data are readily usable and no data cleaning is necessary for this project.

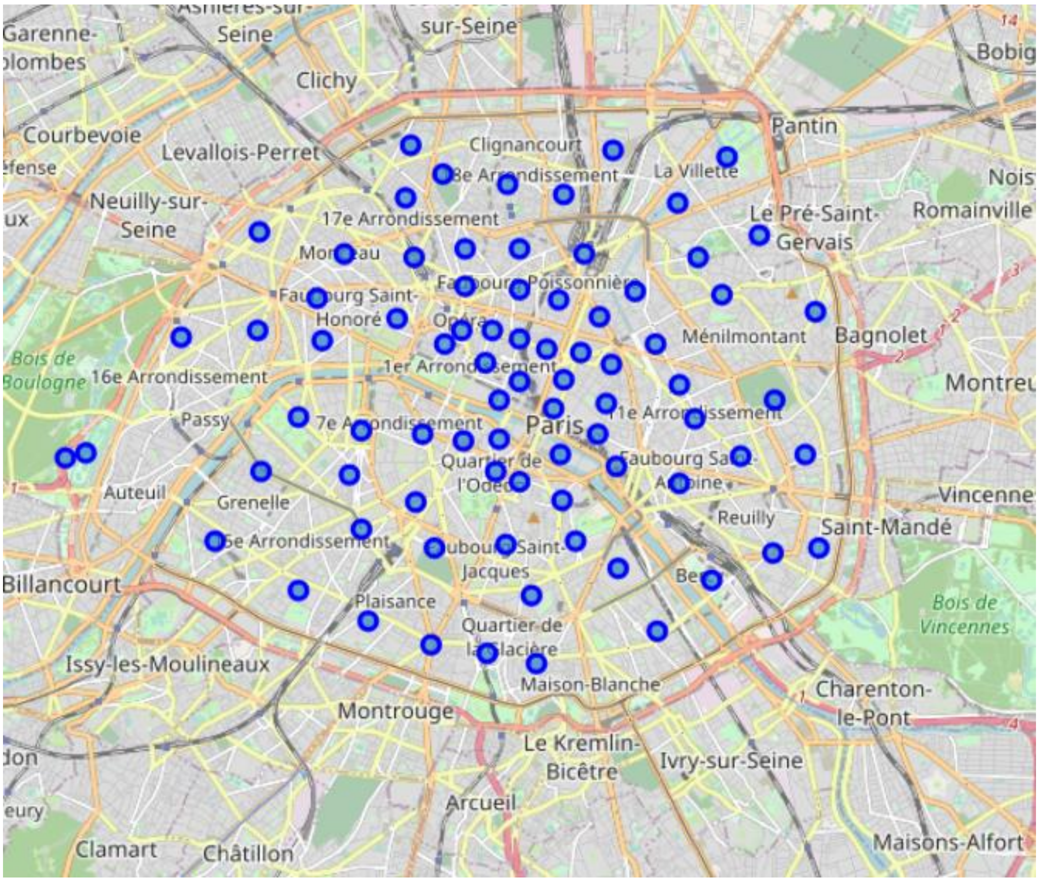


Figure 3 Folium map of Paris using the GPS coordinates collected for each neighbourhood

3. Methodology

3.1. Onehot Encoding for venue categories

The DataFrame that contains all the venues found by the Foursquare API have been onehot encoded in order to create a new DataFrame with numerical values for easier analysis.

This DataFrame has 294 columns, so we can say that the Foursquare API has retrieved venues from 293 categories. A grouping has also been made along the neighbourhood's name by mean value calculation. The higher the mean value is, the more common a venue category will be for a given neighbourhood. This allows to rank venue categories by their commonness easily.

3.2. Creation of features

In order to regroup places that will represent the competition for the new Bubble Tea Shop business, venue categories with 'Café', 'Coffee', 'Tea' in their names have been grouped into one venue category called 'Drink Place'.

To define the attractiveness of a neighbourhood for people interested in the eastern Asian culture, we will regroup venues categories with 'Asian', 'Japanese', 'Chinese' in their names into one venue category called 'Asian Venues'.

	Neighborhood	Drink Place	Asian Venues
0	Amérique	0.021978	0.054945
1	Archives	0.090000	0.040000
2	Arsenal	0.070000	0.020000
3	Arts-et-Métiers	0.100000	0.030000
4	Auteuil	0.014925	0.044776
...	...	...	...
75	Val-de-Grâce	0.060000	0.040000
76	Villette	0.062500	0.046875
77	Vivienne	0.030000	0.080000
78	École-Militaire	0.040000	0.010000
79	Épinettes	0.020000	0.020000

Figure 4 Visualization of 'Drink Place' and 'Asian Venues' features

3.3. Find the most suitable neighbourhood for the bubble tea shop business

The first step consisted in defining the 5 most commons venue categories for each neighbourhood using the DataFrame generated by onehot encoding and grouping. This ranking is stored in a new DataFrame.

```
neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Amérique	French Restaurant	Supermarket	Bar	Asian Venues	Bakery
1	Archives	Drink Place	French Restaurant	Clothing Store	Art Gallery	Coffee Shop
2	Arsenal	French Restaurant	Drink Place	Cocktail Bar	Plaza	Ice Cream Shop
3	Arts-et-Métiers	Drink Place	Coffee Shop	Restaurant	Bakery	Hotel
4	Auteuil	Tennis Court	French Restaurant	Asian Venues	Supermarket	Plaza

Figure 5 Visualization of the DataFrame storing the ranking for each candidate

The second step consisted in filtering the candidates with the requirements for the business implementation. At first was removed all candidates with a high density of Drink Place, i.e., Drink Place is on the top 5 of most common venue categories. Slightly more than half of the candidate were eliminated, leaving with 35 remaining.

As Paris is a large city in terms of population (at least for France) with several million people living, it is expected to see a lot of candidates with high density of drink places. If we did a similar analysis with restaurants (by grouping restaurants of all cuisines), it is expected to get a similar result.

During the assignment where we had the opportunity to analyse Toronto neighbourhoods, we could also see that a vast majority of them had a high density of drink places and restaurants.

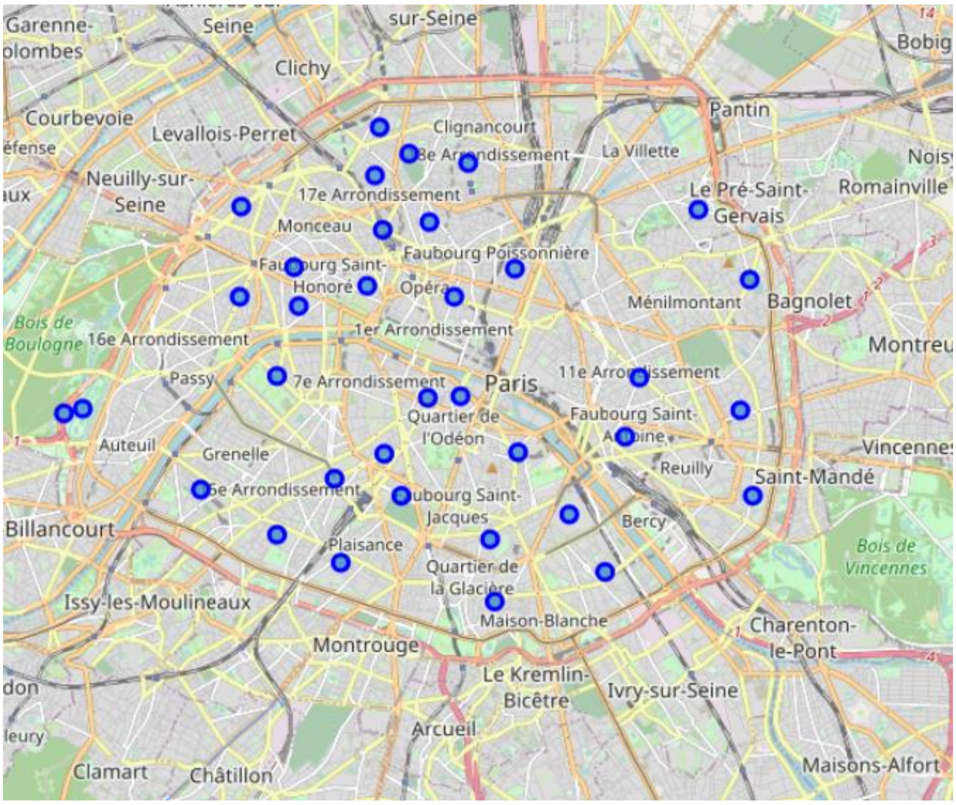


Figure 6 Map of Paris after removing the candidates with a high density of drink places

Then was included all candidates with a high density of ‘Asian Venues’, i.e., ‘Asian Venues’ is on the top 5 of most common venue categories. Slightly more than half of the candidate were eliminated, leaving with 20 remaining.

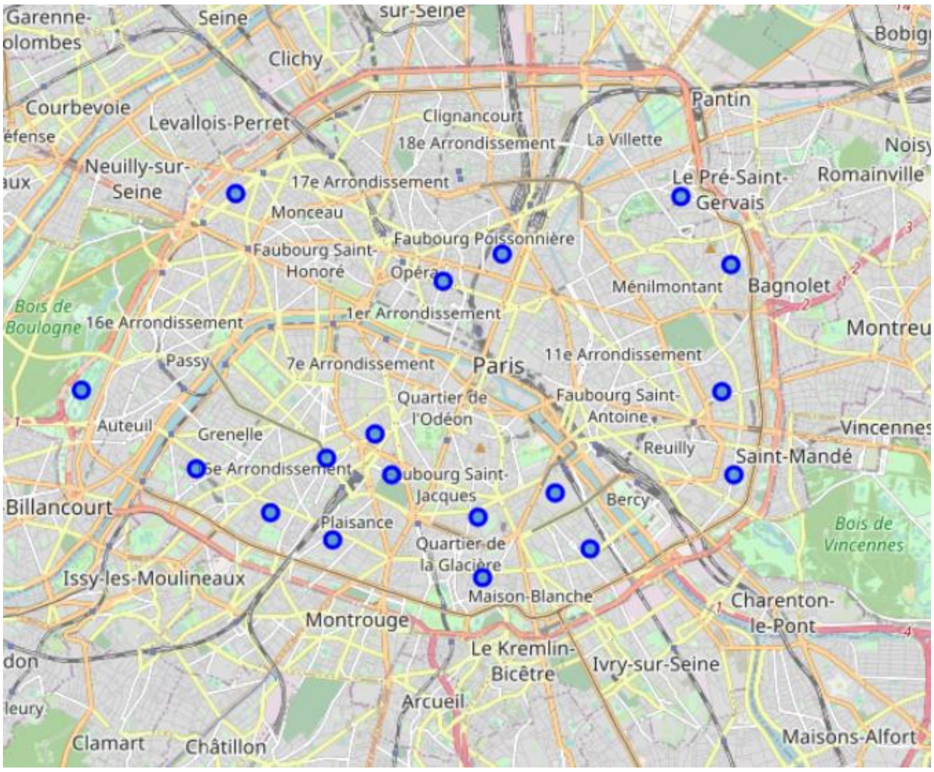


Figure 7 Map of Paris with remaining candidates with high density of 'Asian Venues'

As we still had 20 candidates left, we took a look at their attractiveness by checking the presence of the following venues:

- Park
- Garden
- Shopping Mall
- Shopping Plaza
- Monument / Landmark

	Neighborhood	Park	Garden	Shopping Mall	Shopping Plaza	Monument / Landmark
0	Amérique	0.022222	0.000000	0.00	0.000000	0.0
6	Bel-Air	0.020000	0.000000	0.00	0.000000	0.0
13	Charonne	0.010000	0.010000	0.00	0.000000	0.0
17	Croulebarbe	0.020000	0.000000	0.00	0.000000	0.0
24	Gare	0.020000	0.010000	0.00	0.000000	0.0
33	Javel	0.020000	0.000000	0.01	0.000000	0.0
36	Maison-Blanche	0.040000	0.000000	0.00	0.000000	0.0
38	Montparnasse	0.000000	0.010000	0.00	0.000000	0.0
39	Muette	0.011111	0.033333	0.00	0.011111	0.0
40	Necker	0.000000	0.000000	0.00	0.000000	0.0
42	Notre-Dame-des-Champs	0.000000	0.020000	0.00	0.000000	0.0
50	Plaisance	0.010000	0.000000	0.00	0.000000	0.0
53	Porte-Saint-Denis	0.000000	0.000000	0.00	0.000000	0.0
60	Saint-Fargeau	0.010417	0.010417	0.00	0.000000	0.0
65	Saint-Lambert	0.030000	0.000000	0.00	0.000000	0.0
72	Salpêtrière	0.000000	0.040000	0.00	0.000000	0.0
74	Ternes	0.010000	0.000000	0.00	0.000000	0.0
77	Vivienne	0.010000	0.010000	0.00	0.000000	0.0

Figure 8 Values of potential candidates for the presence of Park, Garden, Shopping Mall, Shopping Plaza and Monument/Landmark

It appears that only "Muette" has three of mentioned features (value superior to 0)

	Neighborhood	Park	Garden	Shopping Mall	Shopping Plaza	Monument / Landmark
39	Muette	0.011111	0.033333	0.00	0.011111	0.0

Figure 9 Values of the neighbourhood 'Muette' for the presence of Park, Garden, Shopping Mall, Shopping Plaza and Monument/Landmark

4. Results and Discussion

We first removed the neighbourhoods which have a high density of hot beverage selling points. As Paris is a large city, these places are extremely common so it allowed to remove approximatively half of the neighbourhoods.

We then searched for neighbourhoods with a strong influence from the Eastern Asian culture. So, we looked at venue categories which contain the words 'Asian', 'Japanese' or 'Chinese'. These venues have been summed into a new category, so we could look at neighbourhoods where these places are common.

These two filters allowed to remove 75% of the neighbourhoods (60 out of 80)

Figure 10 Map of Paris with all the neighbourhoods after analysis. Red: Not a potential candidate; Orange: Potential Candidates; Green: Best candidate so far

One interesting point is that all neighbourhoods of XIII District have been defined as potential candidates. This district is known to be heavily influenced by the eastern Asian culture. But it lacks of venues which attracts people to be in the top of our recommendations.

To pick among the remaining candidates, we evaluated their attractiveness by checking whether they have parks, gardens, shopping points or monuments. It appears that only "Muette" has three of mentioned features, which makes it our most promising candidate for our new bubble tea business. Of course, several other parameters have to be taken into account before taking a final decision.

You will find below a picture of the neighbourhood "Muette" as seen in Google Maps with several principal venues highlighted nearby. As the centroid determined by folium and the GPS coordinates found in Wikipedia is a bit off, we probably missed the Eiffel Tower by a few hundred meters in our analysis.

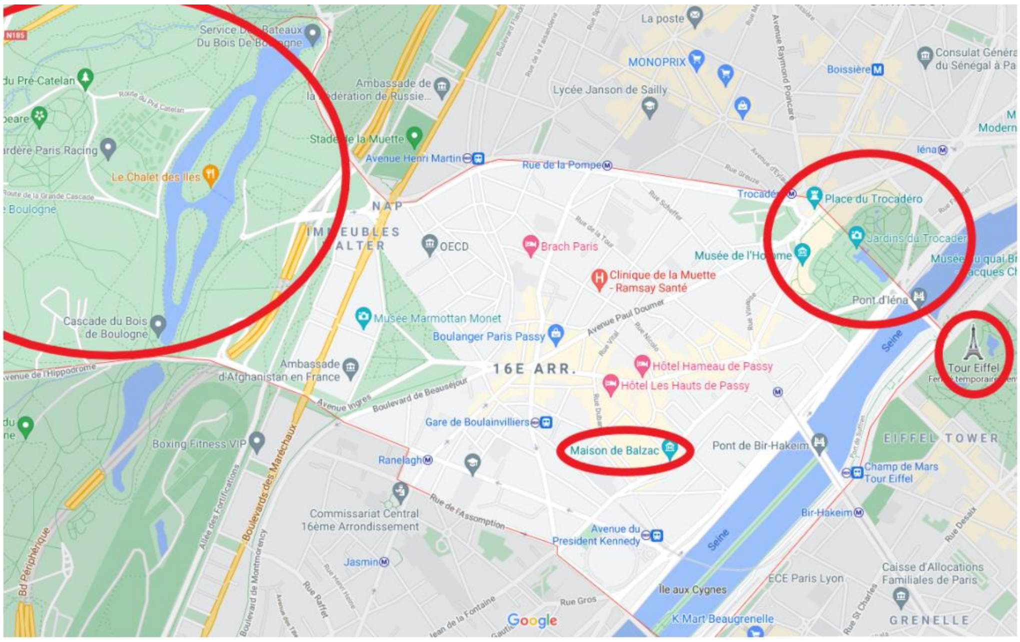


Figure 11 Map of 'Muette' neighbourhood with several venues of interest highlighted in red (Google Maps)

5. Conclusion

The purpose of this project was to give us an idea of neighbourhoods where it can be interesting to start a new business of bubble tea shop. We tried to avoid places where the competition is already strong or where the market target is not expected to go.

As we had still to deal with about 20 neighbourhoods, we tried to determine which neighbourhood is the most attractive by looking at parks, gardens, shopping points and monuments.

This analysis should only be used as a starting point as several parameters such as the availability and the price of the real-estate (especially when the Eiffel Tower is nearby) or social and economic dynamics should be taken into consideration before taking a final decision. The business strategy should also be addressed before taking a decision. For example, if we want to allow customers to be delivered at home, we should try to focus on neighbourhoods where the largest area possible in Paris downtown can be covered. For example, 'Notre-Dame-des-Champs' is interesting in a sense that it allows the delivery in a lot of neighbourhoods defined as potential candidates.

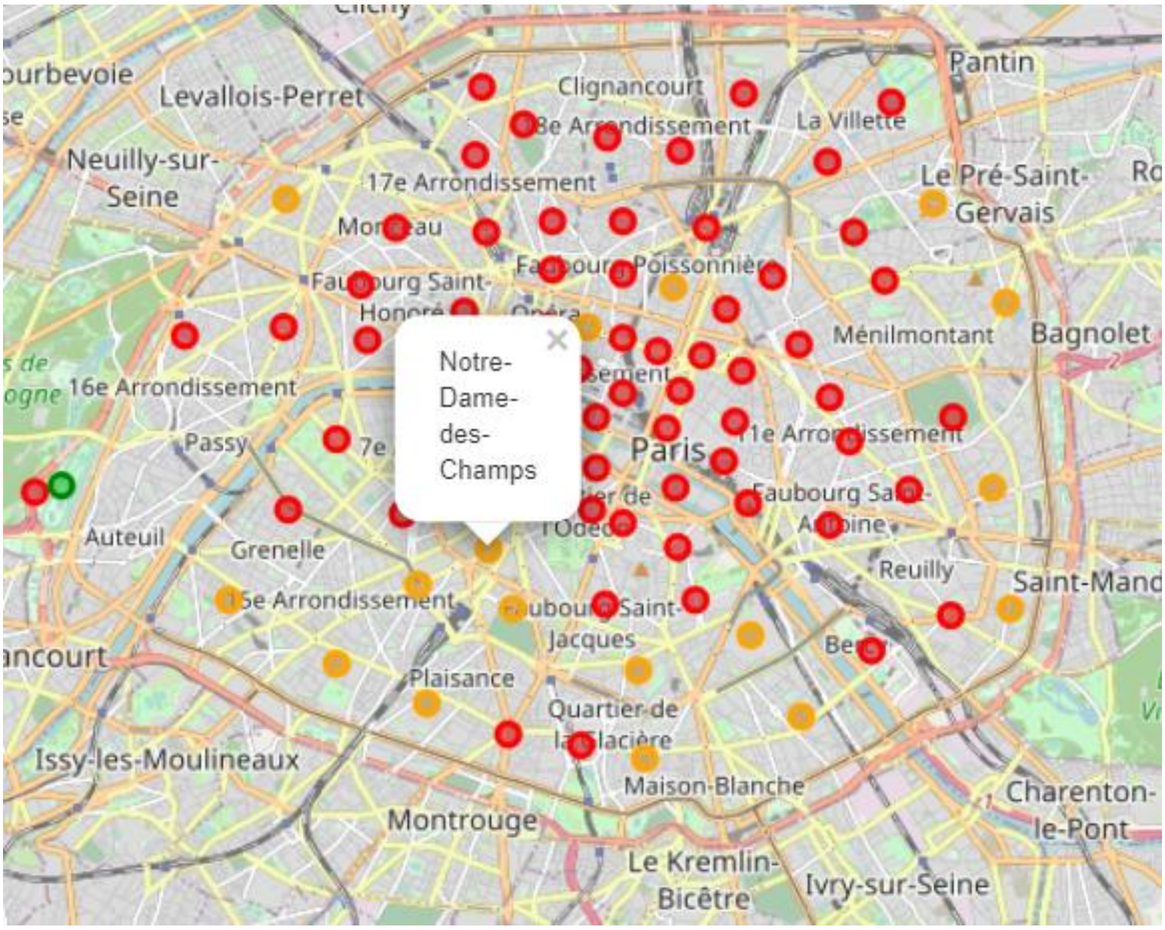


Figure 12 Map of Paris with the pinpoint of 'Notre-Dame-des-Champs' geographical position