

Final Project - Battle of Neighborhoods - Paris Restaurants

Lucas DA RIN

1. INTRODUCTION

Paris is undoubtedly one of the most important cities in the western world when it comes to tourism, arts, fashion and culinary. As French cuisine expanded all across the globe, the contrary is also true: foreign culinary has set hundreds of restaurants all around Paris.

Since the city is known for many reasons, one of them its culinary, the idea of the project in a nutshell is to explore restaurants' data, like ratings and likes on Foursquare and, since the project is called Battle of Neighborhoods, this project is also going to compare neighborhoods in Paris, examining if french restaurants are the most common ones and, if not, what kind of restaurant is the most common in each of the 20 *arrondissements* of the city.

The idea is to start with Foursquare API in order to obtain information about venues in the city, given the geographic location of each neighborhood center. In order to filter the different types of venues found in the city, some criteria will be used for checking potential data to be treated as french restaurants or similar ones, and to classify bars and foreign cuisines as other types of restaurants.

Later on, the project will provide an overview of the best and most popular restaurants and bars and then focus on investigating the distribution of restaurant types per neighborhood. Tables and maps are going to be used as tools, in order to discover whether each neighborhood has more French Restaurants than bars and cafés, or if international cuisine is more common in a specific area.

The intention, by creating this project, is to **provide information for tourists, enthusiasts of the french cuisine and even locals about restaurants available in Paris.** Even though there are limitations in using a sample to solve the problems at hand, having an idea of which type of restaurant is the most common per area is an interesting outcome for me personally and treating the data, using foursquare API and adapting the problem to the city of Paris will be quite a challenge.

Finally, the idea of the project is mainly to be able to find out good restaurants in Paris and to get to know the most common kind of restaurant in each of the 20 neighborhoods of the city.

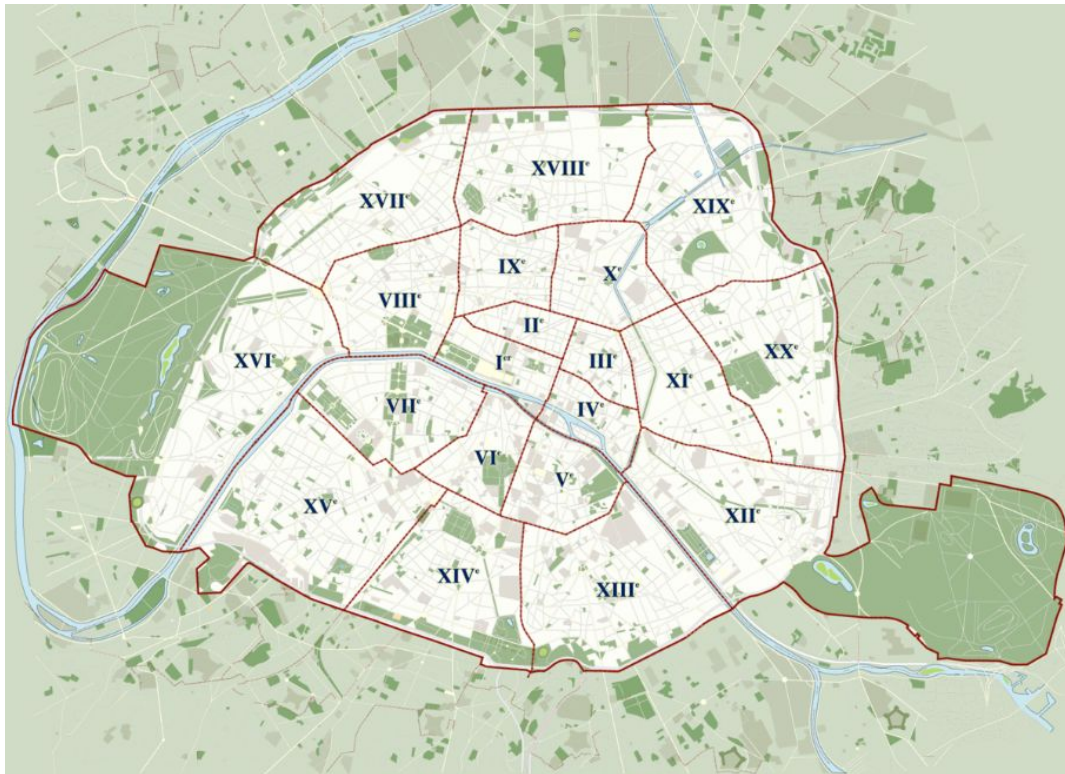
2. DATA ACQUISITION AND DATA CLEANING

2.1. DATA SOURCES

In order to have some initial information about the neighborhoods of Paris, a csv file will be used in this project. The csv of the neighborhoods of Paris and their geographic locations can be found at the French Government website (<https://www.data.gouv.fr/>). A path to my github account including the table will be used in the project, as the data is required. The data downloaded is yet to be treated for better use.

Since this work aims to identify the distribution of restaurants among specific locations, the Foursquare API is going to be the tool for finding our sample: venues near specific geographic locations. Given latitudes, longitudes and a radius, Foursquare API is useful for finding venues around a specific location.

Working with the map of Paris can be quite challenging, since the city has a 'snail-shell' shape, in which the most central neighborhoods have a much smaller surface than those on the outside. Looking for a solution for finding a great sample of venues per neighborhood without compromising the fact that the area varies a lot between the most central and most peripheral neighborhoods is the challenging task of this part.



Neighborhoods of Paris. Source: Wikipedia

The number of searches per day one can do in foursquare is limited. For that reason, the project may have functions to save dataframes as csv files for later use, avoiding the daily use of foursquare API.

Two functions are to be defined, the first to retrieve the list of venues given the locations and the second one for retrieve Foursquare information about the restaurants to go though further analysis, like ratings, tips and likes.

2.1. DATA CLEANING

The table provided for geographic locations of the neighborhoods comes with several columns not to be used in this project. The names of the remaining columns are also to be changed for better comprehension and use during the project. Except for the names of the columns and dropping the ones not needed, there's not much to be treated on the initial data, since it includes only the 20 neighborhoods of Paris and related data.

CAR		NAME	NSQAR	CAR.1	CARINSEE	LAR	NSQCO	SURFACE	PERIMETRE	Geometry_X	Geometry_Y
0	3	Temple	750000003	3	3	3eme Ardt	750001537	1170882828	4519264	48.862872	2.360001
1	19	Buttes-Chaumont	750000019	19	19	19eme Ardt	750001537	6792651129	11253182	48.887076	2.384821
2	14	Observatoire	750000014	14	14	14eme Ardt	750001537	5614877309	10317483	48.829245	2.326542
3	10	Entrepot	750000010	10	10	10eme Ardt	750001537	2891739442	6739375	48.876130	2.360728
4	12	Reuilly	750000012	12	12	12eme Ardt	750001537	16314782637	24089666	48.834974	2.421325
5	16	Passy	750000016	16	16	16eme Ardt	750001537	16372542129	17416110	48.860392	2.261971
6	11	Popincourt	750000011	11	11	11eme Ardt	750001537	3665441552	8282012	48.859059	2.380058
7	2	Bourse	750000002	2	2	2eme Ardt	750001537	991153745	4554104	48.868279	2.342803
8	4	Hotel-de-Ville	750000004	4	4	4eme Ardt	750001537	1600585632	5420908	48.854341	2.357630
9	17	Batignolles-Monceau	750000017	17	17	17eme Ardt	750001537	5668834504	10775580	48.887327	2.306777
10	18	Buttes-Montmartre	750000018	18	18	18eme Ardt	750001537	5996051308	9916464	48.892569	2.348161
11	1	Louvre	750000001	1	1	1er Ardt	750001537	1824612860	6054937	48.862563	2.336443
12	5	Pantheon	750000005	5	5	5eme Ardt	750001537	2539374623	6239195	48.844443	2.350715
13	7	Palais-Bourbon	750000007	7	7	7eme Ardt	750001537	4090057185	8099425	48.856174	2.312188
14	20	Menilmontant	750000020	20	20	20eme Ardt	750001537	5983446037	10704940	48.863461	2.401188
15	8	elysee	750000008	8	8	8eme Ardt	750001537	3880036397	7880533	48.872721	2.312554
16	9	Opera	750000009	9	9	9eme Ardt	750001537	2178303275	6471588	48.877164	2.337458
17	13	Gobelins	750000013	13	13	13eme Ardt	750001537	7149311091	11546547	48.828388	2.362272
18	15	Vaugirard	750000015	15	15	15eme Ardt	750001537	8494994081	13678798	48.840085	2.292826
19	6	Luxembourg	750000006	6	6	6eme Ardt	750001537	2153095586	6483687	48.849130	2.332898

Table before treatment

One last point to be treated in the data is that the neighborhoods are presented in a random order. So the data will be sorted in ascending order, as shown below

	ZIP Code	Neighborhood	SURFACE	Latitude	Longitude
11	75001	1er Ardt	1824612860	48.862563	2.336443
7	75002	2eme Ardt	991153745	48.868279	2.342803
0	75003	3eme Ardt	1170882828	48.862872	2.360001
8	75004	4eme Ardt	1600585632	48.854341	2.357630
12	75005	5eme Ardt	2539374623	48.844443	2.350715

Table after treatment

3. METHODOLOGY

This project consists of exploring venues around specific neighborhoods, using Foursquare API, investigating the ratings and popularity of venues and exploring the distribution of venues and their ratings across neighborhoods. The venues were found after the definition of a function for finding 100 results in a radius of, at first, 500 meters, as shown below.

```
def getNearbyVenues(zipcodes, names, latitudes, longitudes):
    radius=500
    LIMIT=100
    venues_list=[]
    for zipc, name, lat, lng in zip(zipcodes, names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([(
            zipc,
            name,
            lat,
            lng,
            v['venue']['id'],
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['ZIP Code',
                             'Neighborhood',
                             'Neighborhood Latitude',
                             'Neighborhood Longitude',
                             'Id',
                             'Venue',
                             'Venue Latitude',
                             'Venue Longitude',
                             'Venue Category']

    return(nearby_venues)
```

Function for extracting venues

The data about the neighborhoods comes in a csv(comma-separated values) file, consisting of the ZIP Code, Neighborhood Name and Geographical locations. By using the Foursquare API, the venues found were treated in order to keep the most likely of the occurrences in cases of duplicates.

A correlation analysis between ZIP Code and SURFACE was done in order to remove duplicates by descending order of likelihood of the venue being found in the neighborhood.

The data of the venues was then treated in order to keep only restaurants and bars. The criteria chosen was to check for 35 most common kinds of venues and filtering out those that were not related to restaurants and bars or that could not fit in “French Restaurants”, “Italian Restaurants”, “Cafés & Bars” and “Asian Restaurants” categories.

A function using Foursquare API was defined for extracting ‘Rating’, ‘Likes’ and ‘Tips’ fields. The ratings and popularity of restaurants were evaluated by two of the three criteria. A correlation analysis indicates ‘Likes’ and ‘Tips’ was done to find out if the fields were redundant.

```
def get_venue_details(venue_id):
    venue_details = []

    for ven in venue_id:
        #url to fetch data from foursquare api

        url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
            ven,
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION)
        # get all the data
        results = requests.get(url).json()
        print(results)
        venue_data = results["response"]

        try:
            venue_id=venue_data['venue']['id']
            venue_name=venue_data['venue']['name']
            venue_likes=venue_data['venue']['likes']['count']
            venue_rating=venue_data['venue']['rating']
            venue_tips=venue_data['venue']['tips']['count']
            venue_details.append([venue_id,venue_name,venue_likes,venue_rating,venue_tips])
        except KeyError:
            pass
        column_names=['ID','Name','Likes','Rating','Tips']
        df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

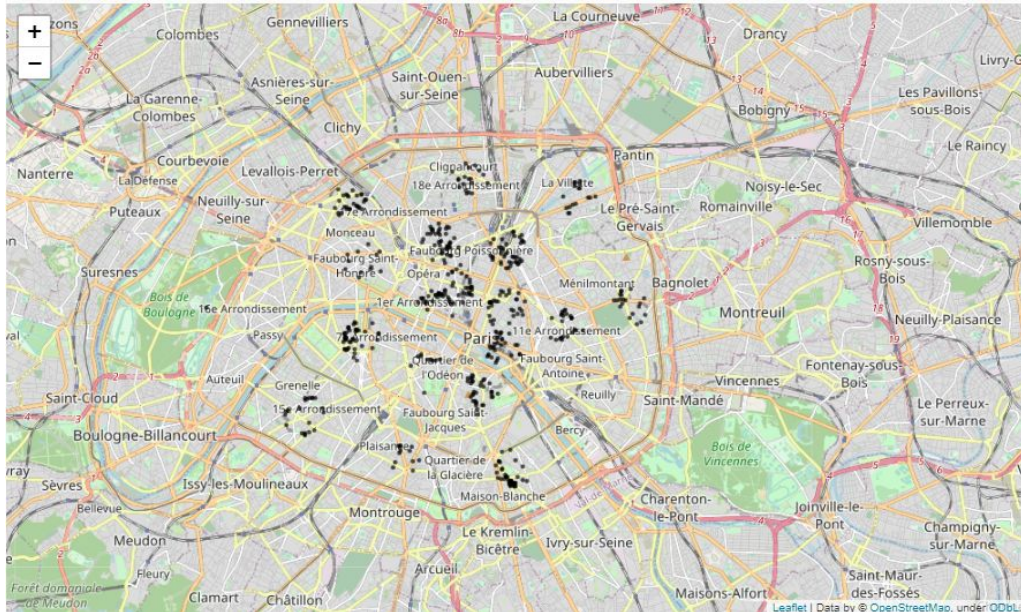
Function for finding Rating, Likes and Tips

The merging of the venues details table and the entire list of venues was used for the analysis of distribution or rates and geographical distribution of the venues. Finally, the results and discussion of the investigation are presented in the ‘Results’ and ‘Discussion’ sections of this report.

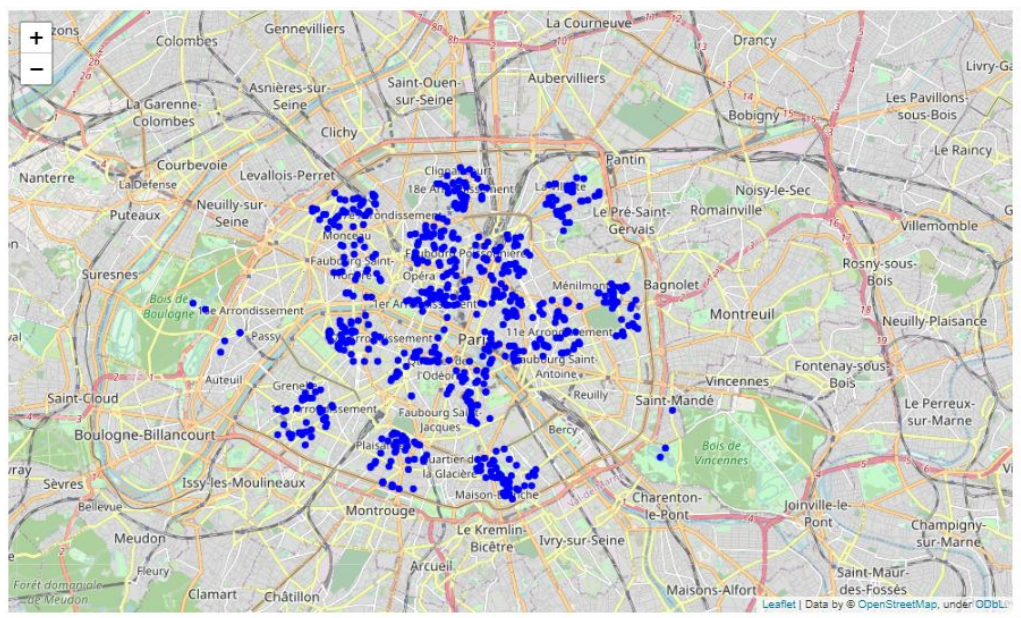
3.1. Refinement and Adjustments

After finding the venues in a 500 meters radius in the first try, the results were evaluated to check whether the radius was enough for covering a significant area of Paris. As the result was not satisfactory, a radius of 800 meters was later tried, and the

number of results was set up to 150. The final project consists a concatenation of the radius of 1000, 800 and 500, for a better representation of the reality. The final map can be found in the ipynb file.



Density of restaurants and bars with radius = 500 and limit = 100



Density of restaurants and bars with radius = 800 and limit = 150

4. RESULTS

The results show that French Restaurants are the most popular ones in 11 of the 20 neighborhoods of Paris, while Cafés & bars and Asian Restaurants account for 7 and 2 neighborhoods, respectively, of the most common venues of our list.

	ZIP Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	75010	Cafés and Bars	French Restaurant	Asian Restaurants
1	75011	Cafés and Bars	French Restaurant	Italian Restaurants
2	75012	French Restaurant	Cafés and Bars	Italian Restaurants
3	75013	Asian Restaurants	French Restaurant	Cafés and Bars
4	75014	French Restaurant	Cafés and Bars	Asian Restaurants
5	75015	French Restaurant	Cafés and Bars	Italian Restaurants
6	75016	French Restaurant	Italian Restaurants	Asian Restaurants
7	75017	French Restaurant	Italian Restaurants	Cafés and Bars
8	75018	Cafés and Bars	French Restaurant	Asian Restaurants
9	75019	Cafés and Bars	French Restaurant	Asian Restaurants
10	75001	Asian Restaurants	French Restaurant	Italian Restaurants
11	75020	Cafés and Bars	French Restaurant	Asian Restaurants
12	75002	Cafés and Bars	French Restaurant	Italian Restaurants
13	75003	Cafés and Bars	French Restaurant	Asian Restaurants
14	75004	French Restaurant	Cafés and Bars	Italian Restaurants
15	75005	French Restaurant	Cafés and Bars	Asian Restaurants
16	75006	French Restaurant	Cafés and Bars	Italian Restaurants
17	75007	French Restaurant	Cafés and Bars	Italian Restaurants
18	75008	French Restaurant	Cafés and Bars	Asian Restaurants
19	75009	French Restaurant	Cafés and Bars	Italian Restaurants

List of most common venue per neighborhood

The lists of most popular and best restaurants of our sample brings Café de Flore as the most popular one and La Cave de Septime as the best restaurant according to Foursquare users.

Finally, results suggest 3ème neighborhood is the one with the best ratings, with an average of 8.435, while 9ème, 4ème, 11ème, 2ème, 1er, 10ème and 5ème neighborhoods follow, all of them between 8.3 and 8.4, with pretty similar average grades. On the other hand, 12ème and 20ème have average ratings below 7.

5. DISCUSSION

5.1. Discussion of the results

The results showing that french restaurants, followed by cafés and bars are the most common venues in 18 of the 20 neighborhoods is no big surprise, as this project examines restaurants in Paris.

On the other hand, seeing that Asian Restaurants were the most common in our sample in two of the neighborhoods was pretty revealing. The 1st arrondissement has an entire street dedicated to japanese restaurants and culture. The 13th has an area known as *Quartier Asiatique*, which stands for Asian Neighborhood, what totally explains the results.

5.2. Limitations of the model

The model here presented uses a sample of the venues found in Paris and it does not intend to represent with perfect accuracy the reality.

The first limitation of the model is that, as it uses a sample of the venues and not all the possible ones, the analysis of the best restaurant and the most popular one may be somehow compromised, since the best restaurants may not be part of our sample. As the number of samples in the project is limited by the API quota, the number of restaurants and bars, 989, is the best one to be reached within the limits of the API quota and the 1-week-period suggested to the project.

6. CONCLUSION

This study concludes that french restaurants are most common ones in 11 of the 20 neighborhoods of Paris, followed by cafés and bars. Furthermore, french restaurants are by far the most common venue found in Paris. Another interesting point is that the 13th and the 1st shows more asian restaurants than any category of restaurant in our sample.

Finally, another interesting conclusion is that most neighborhoods have no significant difference in their average rating, while 12ème and 20ème classify by far as the least appealing neighborhoods in clients' opinion.