Coursera Capstone Project

The Battle of Neighborhoods

October 23th 2020

Recommendation of borough in Paris to open a Restaurant

1. INTRODUCTION

Find the location of a new business like a restaurant is a difficult choice and an important one. Paris already has a lot of restaurants and the choice of the place is not an easy task: the number of potential customers around should be high enough, the rivals should not be too abundant, … A lot criterion should be examined to choose this place. Another point is the type of food which will be served.

Here we will focus on the choice of a borough of Paris taking into consideration the number of rivals in this borough and the density of residents in this area. For a 1st approach to the problem, we will not differentiate the kind of restaurant.

2. DATA

The first data used to solve this problem is the list of the borough of Paris that we find on Wikipedia with de Density of residents in each borough: Paris as 80 boroughs.

	Arrondissement[1],[n 1]	Quartiers	Quartiers.1	Population en1999 (hab.)[2]	Superficie(ha)[2]	Densitéhab/km2
0	1er arrondissementdit « du Louvre »	1er	Saint-Germain-l'Auxerrois	1 672	869	1 924
1	1er arrondissementdit « du Louvre »	2e	Halles	8 984	412	21 806
2	1er arrondissementdit « du Louvre »	3e	Palais-Royal	3 195	274	11 661
3	1er arrondissementdit « du Louvre »	4e	Place-Vendôme	3 044	269	11 316
4	2e arrondissementdit « de la Bourse »	5e	Gaillon	1 345	188	7 154
5	2e arrondissementdit « de la Bourse »	6e	Vivienne	2 917	244	11 955

Table 1-Borough of Paris (five first rows)

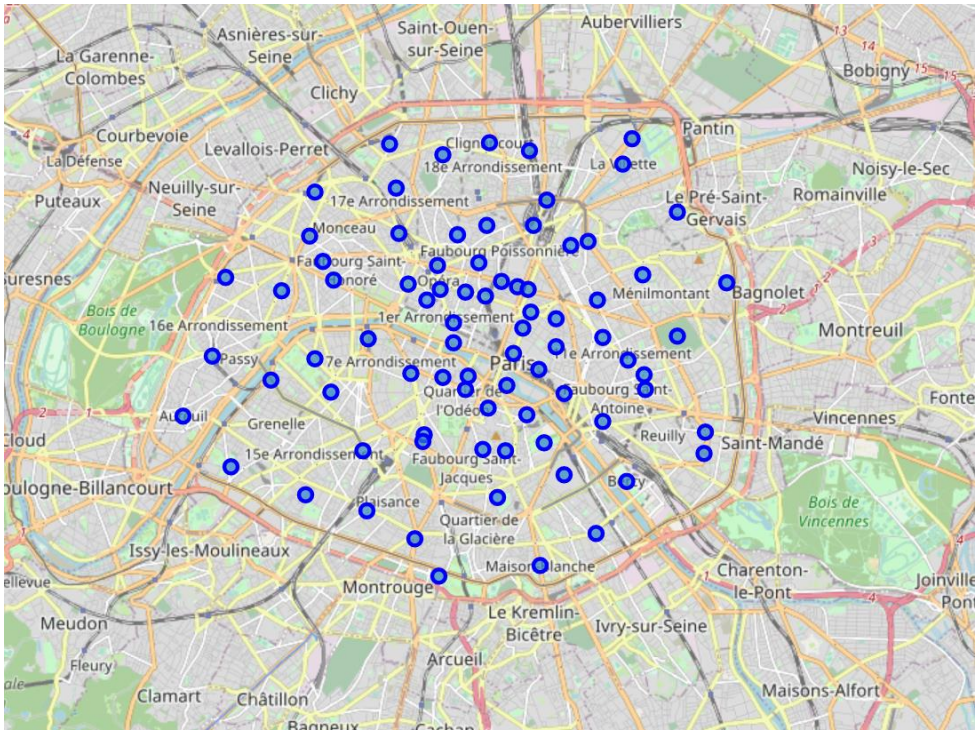
Then, our main data will be geolocation data collected from FourSquare. For each borough, we will request the 100 best venues in a 600m radius. After processing these data from FourSquare, we will obtain only the restaurants in this area and work on this data to analyze the area.

	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
17	Saint-Germain-l'Auxerrois	48.860211	2.336299	LouLou	48.862804	2.333500	Italian Restaurant
31	Saint-Germain-l'Auxerrois	48.860211	2.336299	Boutique yam'Tcha	48.861710	2.342380	Chinese Restaurant
40	Saint-Germain-l'Auxerrois	48.860211	2.336299	Sanukiya	48.864713	2.333805	Udon Restaurant
43	Saint-Germain-l'Auxerrois	48.860211	2.336299	Le Restaurant de L'Hôtel	48.856192	2.335133	French Restaurant
46	Saint-Germain-l'Auxerrois	48.860211	2.336299	Le Pot de Vins	48.864322	2.340193	Restaurant

Table 2 - Restaurants data

3. Methodology

Data scrapping from the Wikipedia page contains the borough with the density of resident in this area, after checking the data is complete, we must find the coordinates of each borough. We will use geocoder API for this request.



With the venues obtained with FourSquare, we will keep only restaurant locations and group them by borough.
We can now have a dataframe with the number of restaurants by borough and the density of residents in each on them.

	Borough	VenueCategory	Density
0	Amérique	3	30155
1	Archives	25	23394
2	Arsenal	25	19454
3	Arts-et-Métiers	36	30063
4	Auteuil	23	22431

After a standardization step, we use the K-Mean clustering technique to cluster each borough : we will try find 5 clusters.

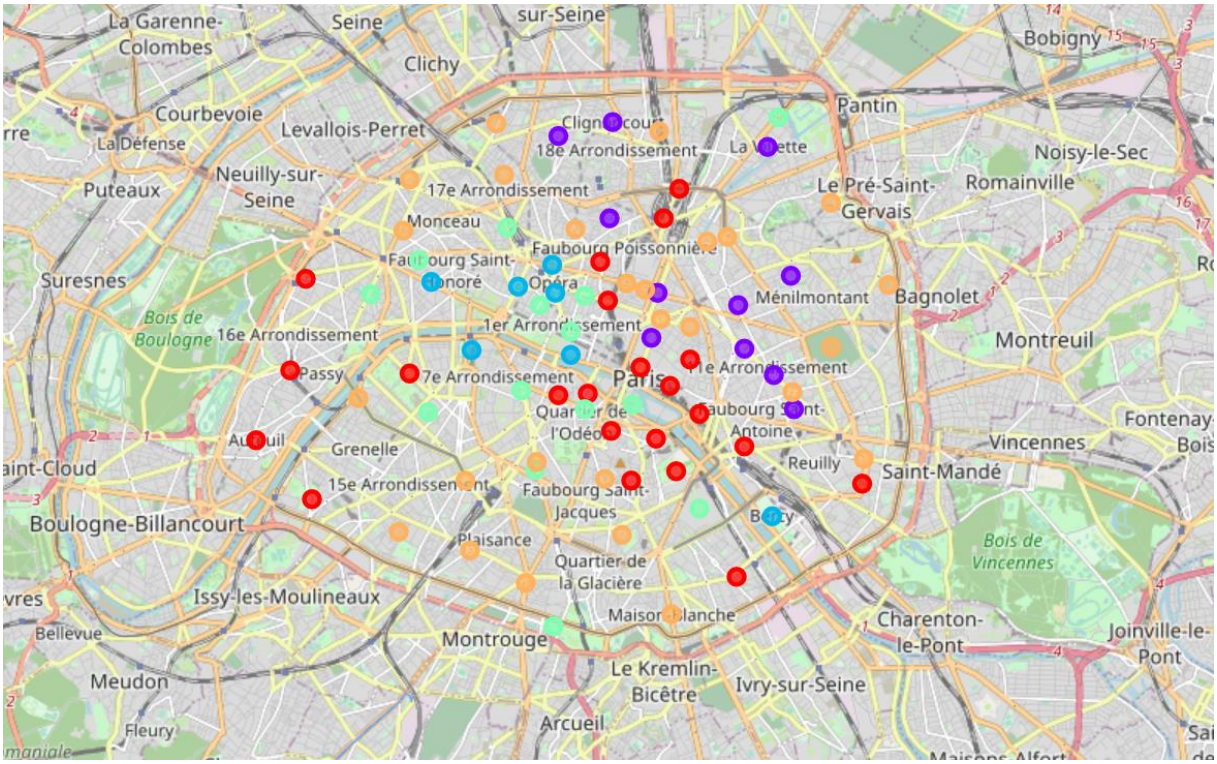
With this methodology we will be able to find the best recommendation to find a place for a restaurant in Paris.

4. Result

With the help of K-Mean clustering, we obtain 5 clusters of similar neighborhoods.

	Borough	VenueCategory	Density	Cluster Label
0	Amérique	3	30155	4
1	Archives	25	23394	0
2	Arsenal	25	19454	0
3	Arts-et-Métiers	36	30063	4
4	Auteuil	23	22431	0

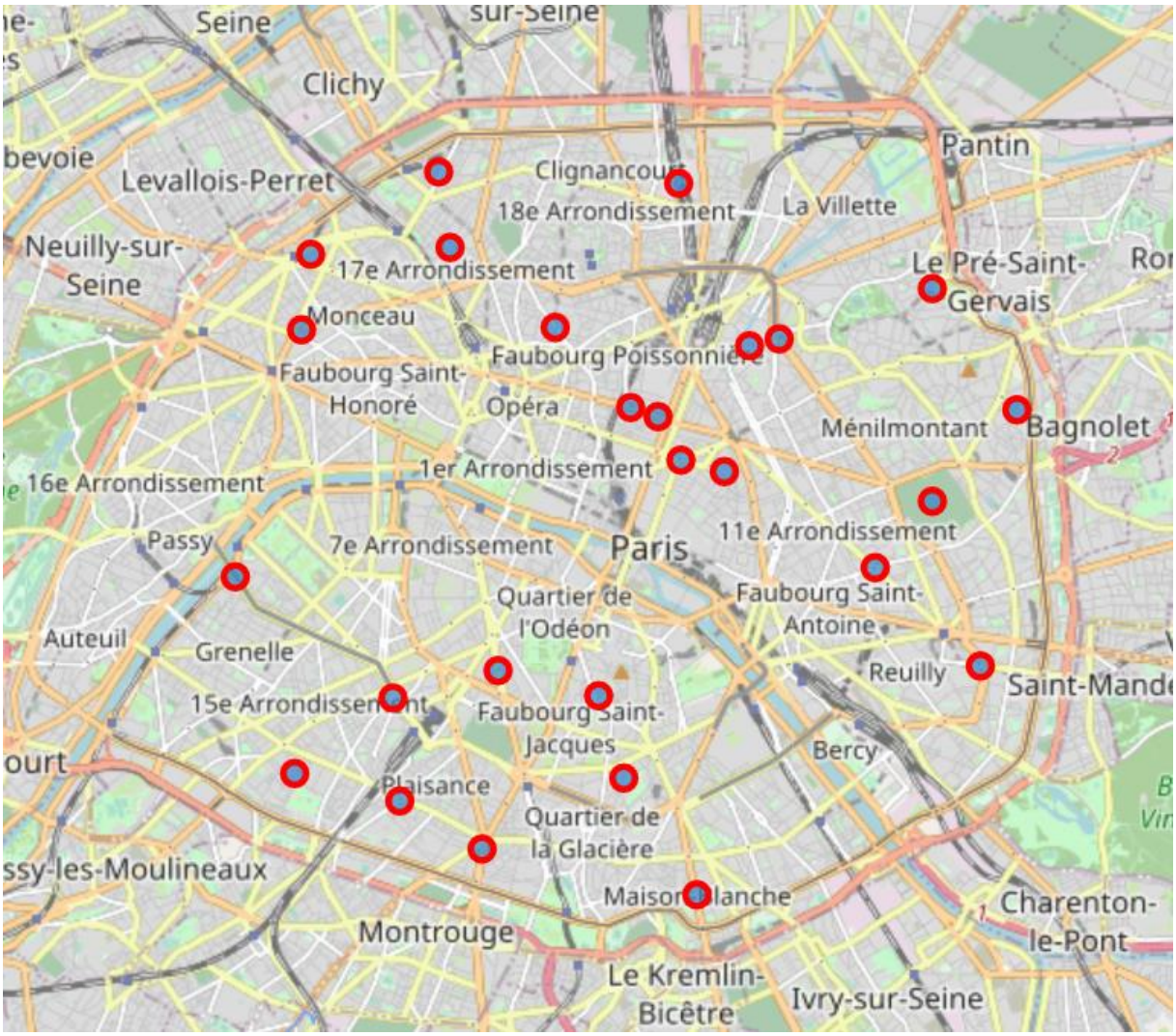
Now we can append the coordinates of each borough to map the clusters.



We will then study each cluster to understand which one would be the best.
According to our results, the cluster 4 is the best one, because of a lower number of restaurants regarding the density of residents.

	Borough	NumberRestaurants	Density	Cluster Label	Latitude	Longitude
65	Saint-Lambert	34	29330	4	48.834444	2.298216
61	Saint-Georges	34	29079	4	48.878447	2.337458
74	Ternes	47	26696	4	48.878286	2.299081
60	Saint-Fargeau	4	28303	4	48.870362	2.406736
75	Val-de-Grâce	24	27688	4	48.842213	2.343882
0	Amérique	3	30155	4	48.882424	2.394025
53	Porte-Saint-Denis	35	31919	4	48.869793	2.352686
3	Arts-et-Métiers	36	30063	4	48.865441	2.356132
5	Batignolles	50	26831	4	48.886355	2.321457
9	Bonne-Nouvelle	34	34514	4	48.870623	2.348750
13	Charonne	39	30082	4	48.854744	2.385356
16	Combat	36	30107	4	48.877421	2.371020
17	Croulebarbe	36	28217	4	48.833974	2.347634
18	Enfants-Rouges	29	31478	4	48.864332	2.362611
25	Goutte-d'Or	15	26169	4	48.892676	2.356040
27	Grenelle	15	32078	4	48.853890	2.289381
30	Hôpital-Saint-Louis	47	33790	4	48.876619	2.366364
36	Maison-Blanche	59	29031	4	48.822474	2.358504
40	Necker	23	29741	4	48.841966	2.312909
42	Notre-Dame-des-Champs	35	28724	4	48.844688	2.328831
46	Petit-Montrouge	22	27660	4	48.826938	2.326241
47	Picpus	18	33788	4	48.845075	2.401080
49	Plaine-Monceau	31	28149	4	48.885729	2.300543
50	Plaisance	20	32061	4	48.831780	2.314025
55	Père-Lachaise	25	26474	4	48.861217	2.393929
79	Épinettes	18	32186	4	48.893751	2.319856

Then we map only the cluster 4, to observe more clearly the best places.



5. Discussion

Before starting the business, some further analysis will be required. The choice of the kind of food is important to more accurate investigations, and also the drawing power of tourist in the borough should be analyze.

6. Conclusion

With this preliminary analyze, we can select 26 boroughs to start the business in Paris. We could first increase the number of clusters generated the restrain the number of areas, then use some other data to make more accurate.