New Fitness center in Paris

Author: Rodica Militaru

Date: November 22, 2019

Introduction/Business Problem

Paris is the capital of France and its largest city. Well known for its cuisine, art and fashion, its economic and cultural life is flourishing.

It is the most crowded city in Europe and one of the most important political and educational centers in EU.

People are living well, but what are they doing in order to maintain their health?

The question this project aims to answer is: where, in Paris, would be the best place to open a fitness center?

Data

Paris has 20 boroughs, each with 4 neighborhoods. We will try to identify the place for a new gym based on the density of population and the number of gyms in each area.

This will be our starting point to finally identify the most appropriate place to open a gym.

The data for this project will be taken from 2 sources:

- a wikipedia page: https://en.wikipedia.org/wiki/Quarters_of_Paris with info about the population in each neighborhood;

- a file from the french government with latitude and longitude of each neighborhood: https://www.data.gouv.fr/en/datasets/quartiers-administratifs

Methodology

We will use pandas, folium, heatmaps and foursquare to identify the already existing gyms.

First, we will process the two files, in order to be able to merge them.

Secondly, we will get the information about the gyms from foursquare and place it into a new dataframe.

Thirdly, we will merge this dataframe with the previous one, keeping only the necessary information.

The next step is to use the describe function to find the mean for population and gyms.

Once this is done, we will drop from our dataframe the values: less than mean for population and greater than mean for gyms and we will generate a heatmap showing the best places to open a gym.

Data preparation

For this project, we need to know the density of the population and the number of gyms in each neighborhood of France, in order to be able to offer solution.

The first table comes from Wikipedia and contains information regarding the population:

	Arrondissement(Districts)	Quartiers(Quarters)	Quartiers(Quarters).1	Population in1999[3]	Area(hectares)[3]	Map
0	1st arrondissement(Called "du Louvre")	1st	Saint-Germain-l'Auxerrois	1672	86.9	NaN
1	1st arrondissement(Called "du Louvre")	2nd	Les Halles	8984	41.2	NaN
2	1st arrondissement(Called "du Louvre")	3rd	Palais-Royal	3195	27.4	NaN
3	1st arrondissement(Called "du Louvre")	4th	Place-Vendôme	3044	26.9	NaN
4	2nd arrondissement(Called "de la Bourse")	5th	Gaillon	1345	18.8	NaN

The second table comes from the French government website and contains the latitude and longitude of the 80 neighborhoods:

	N_SQ_QU	C_QU	C_QUINSEE	L_QU	C_AR	N_SQ_AR	PERIMETRE	SURFACE	Geometry X Y	Geometry
0	750000008	8	7510204	Bonne-Nouvelle	2	750000002	2233.976030	2.814482e+05	48.867150	2.350080
1	750000018	18	7510502	Jardin-des-Plantes	5	750000005	4052.729521	7.983894e+05	48.841940	2.356894
2	750000030	30	7510802	Faubourg-du-Roule	8	750000008	3773.673073	7.965891e+05	48.874136	2.304119
3	750000052	52	7511304	Croulebarbe	13	750000013	3289.230480	6.920677e+05	48.833734	2.347673
4	750000061	61	7511601	Auteuil	16	750000016	12452.253930	6.383888e+06	48.850622	2.252277

We need to merge these tables, in order to have all the info in one place. As the name of the neighborhoods is written differently in the two tables, I decided to use the neighborhoods number instead (Quartiers(Quarters) and C_QU); I changed the Quartiers(Quarters) variable from string to integer and I merged the 2 tables into a new dataframe:

	N_SQ_QU	C_QU	C_QUINSEE	Neighborhood_x	C_AR	N_SQ_AR	PERIMETRE	SURFACE	Latitude	Longitude	Borough	Neighborhood_
0	750000008	8	7510204	Bonne-Nouvelle	2	750000002	2233.976030	2.814482e+05	48.867150	2.350080	2nd arrondissement(Called "de la Bourse")	Bonne-Nouvell
1	750000018	18	7510502	Jardin-des-Plantes	5	750000005	4052.729521	7.983894e+05	48.841940	2.356894	5th arrondissement(Called "du Panthéon")	Jardin-des Plante
2	750000030	30	7510802	Faubourg-du-Roule	8	750000008	3773.673073	7.965891e+05	48.874136	2.304119	8th arrondissement(Called "de l'Élysée")	Faubourg-du Roul
3	750000052	52	7511304	Croulebarbe	13	750000013	3289.230480	6.920677e+05	48.833734	2.347673	13th arrondissement(Called "des Gobelins")	Croulebarb
4	750000061	61	7511601	Auteuil	16	750000016	12452.253930	6.383888e+06	48.850622	2.252277	16th arrondissement(Called "de Passy")	Auteu

Next, I cleaned the new dataframe and only kept the necessary info:

	N_SQ_AR	Borough	N_SQ_QU	C_QU	Neighborhood	Latitude	Longitude	Population
0	750000002	2nd arrondissement(Called "de la Bourse")	750000008	8	Bonne-Nouvelle	48.867150	2.350080	9595
1	750000005	5th arrondissement(Called "du Panthéon")	750000018	18	Jardin-des-Plantes	48.841940	2.356894	18005
2	750000008	8th arrondissement(Called "de l'Élysée")	750000030	30	Faubourg-du-Roule	48.874136	2.304119	10038
3	750000013	13th arrondissement(Called "des Gobelins")	750000052	52	Croulebarbe	48.833734	2.347673	19526
4	750000016	16th arrondissement(Called "de Passy")	750000061	61	Auteuil	48.850622	2.252277	67967

I used this table to get the information regarding the number of gyms from foursquare.

The first info I get is not very encouraging or it is very encouraging, depending on how you look at the situation: gyms are not too present in top 10 venues of Paris's neighborhoods. This could mean that either French people are not very fond of sport or that there are not enough investments in this area yet.

	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	Amérique	French Restaurant	Café	Supermarket	Bistro	Street Art	Bakery	Bed & Breakfast	Tram Station	Park	Plaza
1	Archives	Clothing Store	French Restaurant	Art Gallery	Japanese Restaurant	Burger Joint	Coffee Shop	Café	Tea Room	Bookstore	Bakery
2	Arsenal	French Restaurant	Hotel	Plaza	Park	Italian Restaurant	Tapas Restaurant	Seafood Restaurant	Gastropub	Bakery	Pedestrian Plaza
3	Arts-et-Métiers	French Restaurant	Hotel	Italian Restaurant	Cocktail Bar	Bar	Chinese Restaurant	Wine Bar	Vietnamese Restaurant	Coffee Shop	Restaurant
4	Auteuil	Tennis Court	Stadium	Sporting Goods Shop	French Restaurant	Botanical Garden	Garden	Office	Outdoors & Recreation	Museum	Racecourse
5	Batignolles	French Restaurant	Hotel	Italian Restaurant	Bar	Bistro	Café	Restaurant	Japanese Restaurant	Pizza Place	Park
6	Bel-Air	Recreation Center	Plaza	French Restaurant	Café	Playground	Sports Club	Cupcake Shop	Ethiopian Restaurant	Cosmetics Shop	Fountain

As I found two types of venues referring to gym (Gym and Gym/ Fitness center), I decided to sum them, to make sure I cover all areas.

Finally, I created a new dataframe with all the info, merging the previous table with the information regarding the gyms from foursquare:

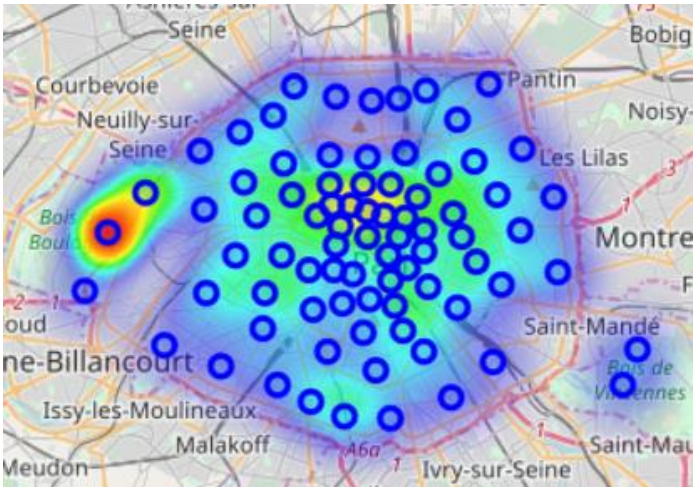
	N_SQ_AR	Borough	N_SQ_QU	C_QU	Neighborhood	Latitude	Longitude	Population	gymall
0	750000002	2nd arrondissement(Called "de la Bourse")	750000008	8	Bonne-Nouvelle	48.867150	2.350080	9595	0.020000
1	750000005	5th arrondissement(Called "du Panthéon")	750000018	18	Jardin-des-Plantes	48.841940	2.356894	18005	0.014286
2	750000008	8th arrondissement(Called "de l'Élysée")	750000030	30	Faubourg-du-Roule	48.874136	2.304119	10038	0.010000
3	750000013	13th arrondissement(Called "des Gobelins")	750000052	52	Croulebarbe	48.833734	2.347673	19526	0.000000
4	750000016	16th arrondissement(Called "de Passy")	750000061	61	Auteuil	48.850622	2.252277	67967	0.000000

Now, we are able to see the general statistics in the dataframe:

	N_SQ_AR	N_SQ_QU	C_QU	Latitude	Longitude	Population	Gym
count	8.000000e+01	8.000000e+01	80.0000	80.000000	80.000000	80.000000	80.000000
mean	7.500000e+08	7.500000e+08	40.5000	48.860991	2.343915	26573.137500	0.013354
std	5.802662e+00	2.323790e+01	23.2379	0.018429	0.035163	20235.804847	0.021321
min	7.500000e+08	7.500000e+08	1.0000	48.823128	2.252277	1345.000000	0.000000
25%	7.500000e+08	7.500000e+08	20.7500	48.848734	2.324470	9413.750000	0.000000
50%	7.500000e+08	7.500000e+08	40.5000	48.862451	2.344880	21418.500000	0.010000
75%	7.500000e+08	7.500001e+08	60.2500	48.873565	2.363586	39025.250000	0.020000
max	7.500000e+08	7.500001e+08	80.0000	48.895556	2.433178	82032.000000	0.125000

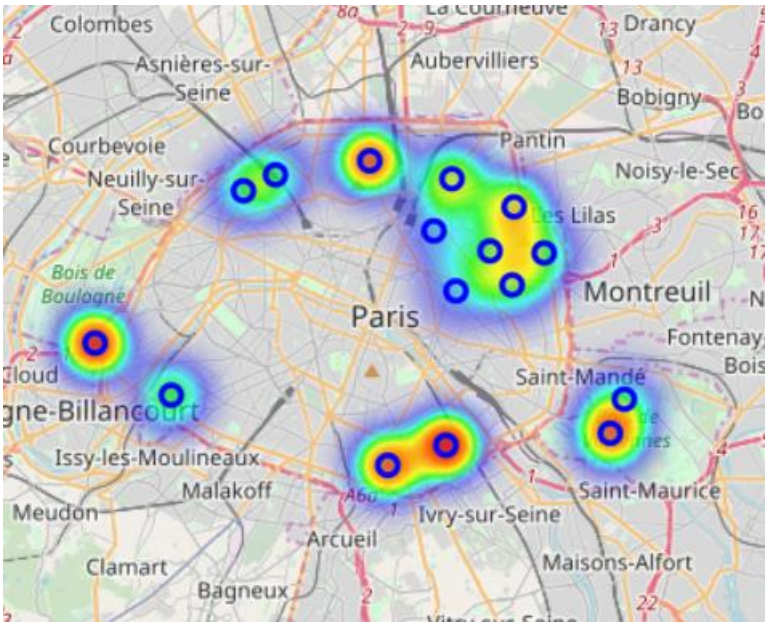
It is now clear that our area of interest in Paris will be reduced to those neighborhoods with a population of more than 26573 (the mean) and a Gym value of less than 0.013 (the mean).

The initial heatmap of Paris was this:



Results

After eliminating the neighborhoods with a small number of inhabitants and a high number of gyms, the new map was this:



According to the map, we have 16 potential neighborhoods were new gyms can be opened, and the top 5 neighborhoods that could be good options are in the boroughs 12,13,16,18:

	N_SQ_AR	Borough	N_SQ_QU	C_QU	Neighborhood	Latitude	Longitude	Population	Gym
28	750000013	13th arrondissement(Called "des Gobelins")	750000050	50	Gare	48.827527	2.372398	69008	0.0
4	750000016	16th arrondissement(Called "de Passy")	750000061	61	Auteuil	48.850622	2.252277	67967	0.0
39	750000018	18th arrondissement(Called "des Buttes-Montmar...	750000070	70	Clignancourt	48.891668	2.345979	64868	0.0
68	750000013	13th arrondissement(Called "des Gobelins")	750000051	51	Maison-Blanche	48.823128	2.352433	64797	0.0
58	750000012	12th arrondissement(Called "de Reuilly")	750000046	46	Picpus	48.830359	2.428827	62947	0.0

Discussion

Based on the number of population and number of gyms, we identified 16 potential neighborhoods were gym can be opened, out of which 5 seem to be best options.

However, this research is very limited, as the information for population is old (1999) and it only takes into account 2 parameters. For the necessary space for the gym, a rent will have to be paid, so the average rent in every neighborhood is definitely an important factor in making a decision, as many areas of Paris are very expensive. At the time of this research, we didn't have any information regarding the average rent.

Also, the small number of gyms is a red flag for me; maybe a deeper analysis is needed to identify why this happens. Maybe the French people don't train too much, so a gym could be a bad investment.

Conclusion

We identified 5 best options for a new gym in Paris: Gare, Auteuil, Clignancourt, Maison-Blanche and Picpus. They have a large number of inhabitants and no gyms, so the potential is unlimited.

A final decision should be taken, however, after consulting the average rent in every neighborhood.