Capstone Project - SuperFit gym in Paris

I. Introduction/Business Problem:

I.1 Background

SuperFit is a new gym company which would like to open its first gym in Paris. SuperFit is a low-cost gym and would like to open its gym where people live as more people work from home nowadays.

I.2 Problem

SuperFit would like to find a potential unmet market in Paris.

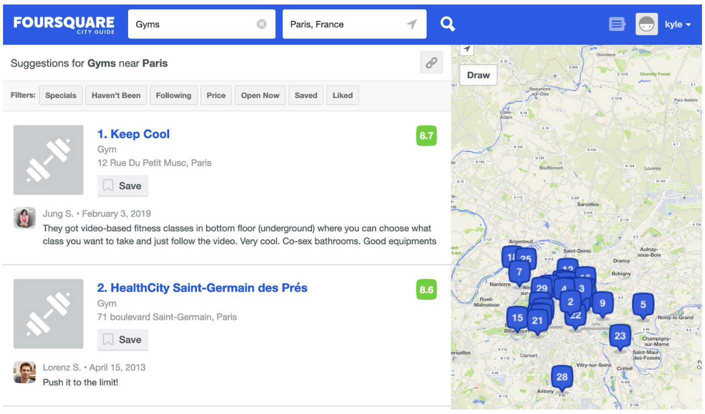
Paris is composed of 20 different districts. SuperFit would like to open its new gym in the district where the number of inhabitants per gym is the highest.

That way, SuperFit hopes to face less competition and grow faster.

II. Data:

II.1 Data sources

Competition: The existing gyms in Paris is retrieved using the Foursquare API



Gyms

Paris, France

Suggestions for **Gyms** near **Paris**

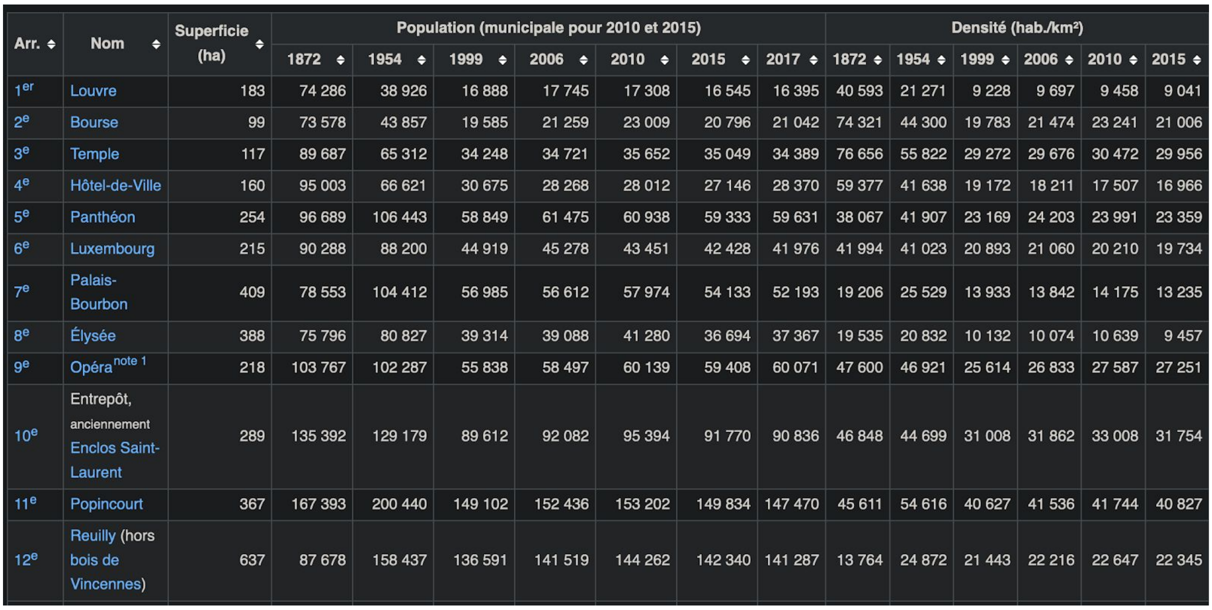
1. Keep Cool 8.7
Gym
12 Rue Du Petit Musc, Paris

Jung S. • February 3, 2019
They got video-based fitness classes in bottom floor (underground) where you can choose what class you want to take and just follow the video. Very cool. Co-sex bathrooms. Good equipments

2. HealthCity Saint-Germain des Prés 8.6
Gym
71 boulevard Saint-Germain, Paris

Lorenz S. • April 15, 2013
Push it to the limit!

Paris data: We are going to scrape data from the following wikipedia page (https://fr.wikipedia.org/wiki/Arrondissements_de_Paris) in order to have the number of inhabitants per district.



Arr.	Nom	Superficie (ha)	Population (municipale pour 2010 et 2015)							Densité (hab./km^2)					
			1872	1954	1999	2006	2010	2015	2017	1872	1954	1999	2006	2010	2015
1er	Louvre	183	74 286	38 926	16 888	17 745	17 308	16 545	16 395	40 593	21 271	9 228	9 697	9 458	9 041
2e	Bourse	99	73 578	43 857	19 585	21 259	23 009	20 796	21 042	74 321	44 300	19 783	21 474	23 241	21 006
3e	Temple	117	89 687	65 312	34 248	34 721	35 652	35 049	34 389	76 656	55 822	29 272	29 676	30 472	29 956
4e	Hôtel-de-Ville	160	95 003	66 621	30 675	28 268	28 012	27 146	28 370	59 377	41 638	19 172	18 211	17 507	16 966
5e	Panthéon	254	96 689	106 443	58 849	61 475	60 938	59 333	59 631	38 067	41 907	23 169	24 203	23 991	23 359
6e	Luxembourg	215	90 288	88 200	44 919	45 278	43 451	42 428	41 976	41 994	41 023	20 893	21 060	20 210	19 734
7e	Palais-Bourbon	409	78 553	104 412	56 985	56 612	57 974	54 133	52 193	19 206	25 529	13 933	13 842	14 175	13 235
8e	Élysée	388	75 796	80 827	39 314	39 088	41 280	36 694	37 367	19 535	20 832	10 132	10 074	10 639	9 457
9e	Opéra[note 1]	218	103 767	102 287	55 838	58 497	60 139	59 408	60 071	47 600	46 921	25 614	26 833	27 587	27 251
10e	Entrepôt, anciennement Enclos Saint-Laurent	289	135 392	129 179	89 612	92 082	95 394	91 770	90 836	46 848	44 699	31 008	31 862	33 008	31 754
11e	Popincourt	367	167 393	200 440	149 102	152 436	153 202	149 834	147 470	45 611	54 616	40 627	41 536	41 744	40 827
12e	Reuilly (hors bois de Vincennes)	637	87 678	158 437	136 591	141 519	144 262	142 340	141 287	13 764	24 872	21 443	22 216	22 647	22 345

Once we have all those data, we will be able to calculate the number of inhabitants per gym and we will be able to make a proposition to SuperFit about where it could be interesting to open the new gym.

II.2 Data cleaning

Paris data: We scraped the data from the wikipedia page and made a panda dataframe based on it. We had to make sure that the numbers were considered as integers and not strings. We simplified the table and we kept only the district numbers and the population in 2017.

FourSquare data: Our goal was to retrieve the list of gyms in Paris. As a free user, we could only retrieve 30 results per query which is lower than the number of gyms in Paris. In order to bypass this limitation we had to run the query 20 times for each district in Paris. We then combined the results into one dataframe and we dropped the duplicates and the columns which were not useful for this analysis. We tried to make a search by category but the results were not accurate. We settled on a search query for 'gym'.

II.3 Feature selection

Paris data:

Regarding the Paris data, we needed the most recent number of inhabitants per district in order to calculate later the number of inhabitants per gym in each district

Kept features	Dropped features
District number, Population in 2017	District name, district size, Population prior to 2017, population density

FourSquare data:

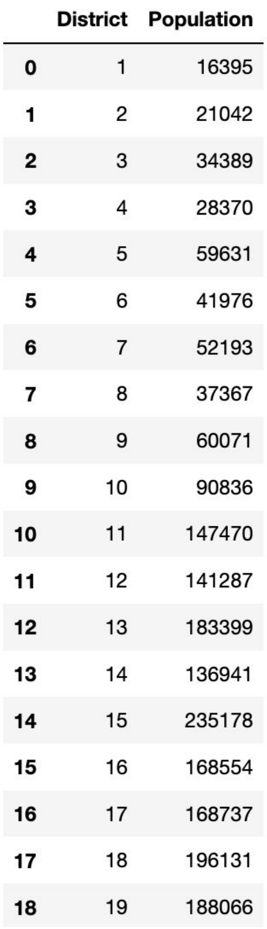
Regarding the FourSquare data, we wanted the list of gyms as well as the postal code in order to affect those gyms in their respective district. We kept as well the coordinates to map the gyms.

Kept features	Dropped features
postal code, name, id, address, latitude, longitude	category, perk, country code, city, country, street, labeledLatLngs, neighborhood, state, referralid, Page.id

III. Exploratory analysis:

III.1 Population per district:

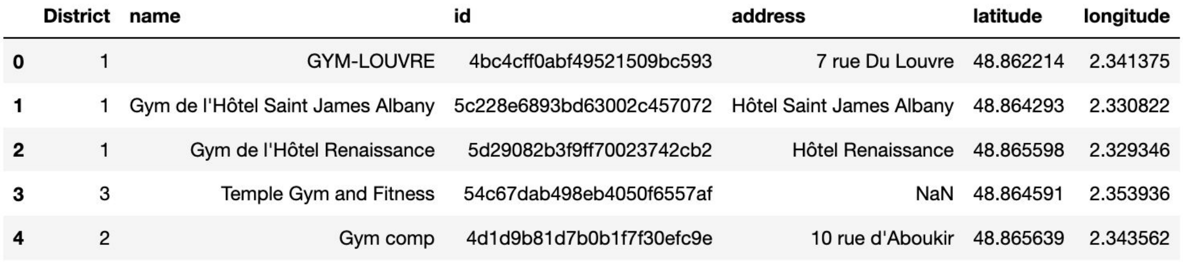
We used the Paris data for that. We kept the district number and the population per district in 2017. We converted the population as integers in order to make calculations later.



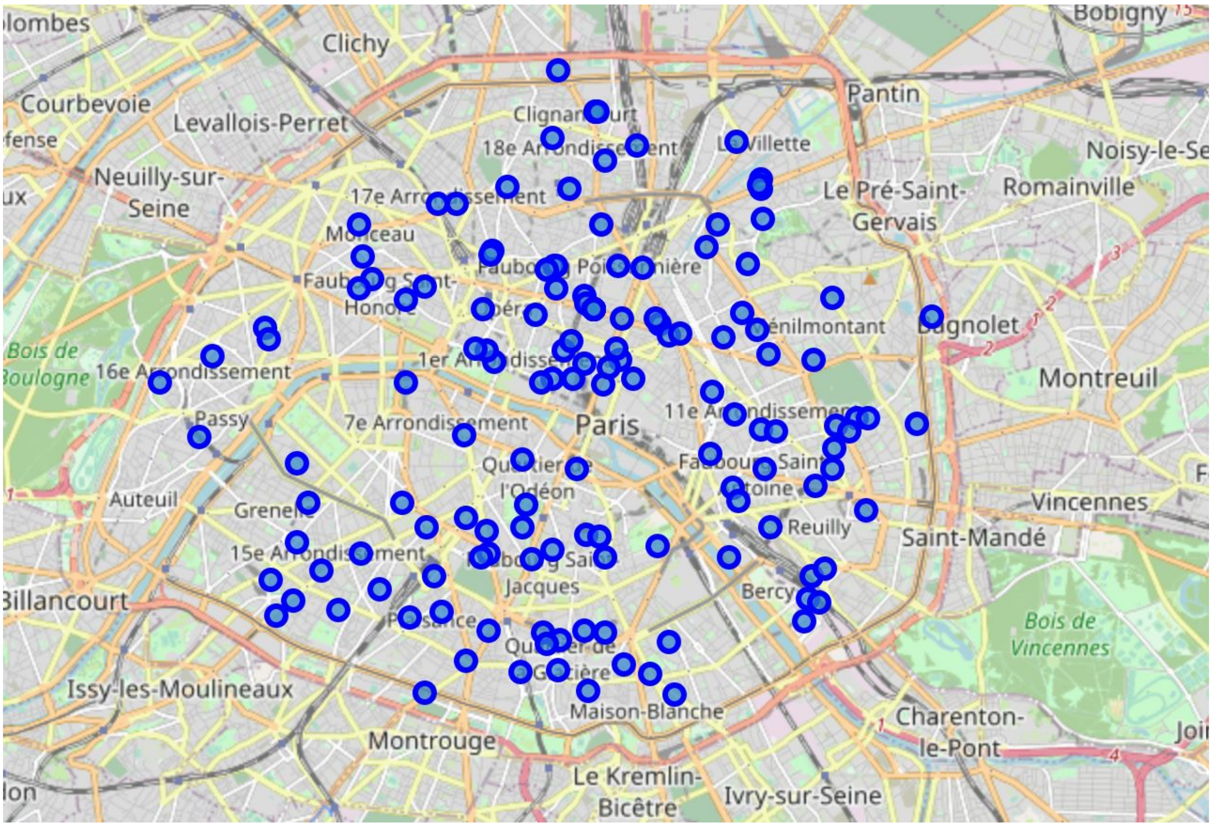
	District	Population
0	1	16395
1	2	21042
2	3	34389
3	4	28370
4	5	59631
5	6	41976
6	7	52193
7	8	37367
8	9	60071
9	10	90836
10	11	147470
11	12	141287
12	13	183399
13	14	136941
14	15	235178
15	16	168554
16	17	168737
17	18	196131
18	19	188066

III.2 Gyms in Paris:

We used the FourSquare API to get a list of gyms for each district in Paris. We then combined them and cleaned them. We kept the id column in order to count the gyms per district. We then mapped those gyms using Folium.

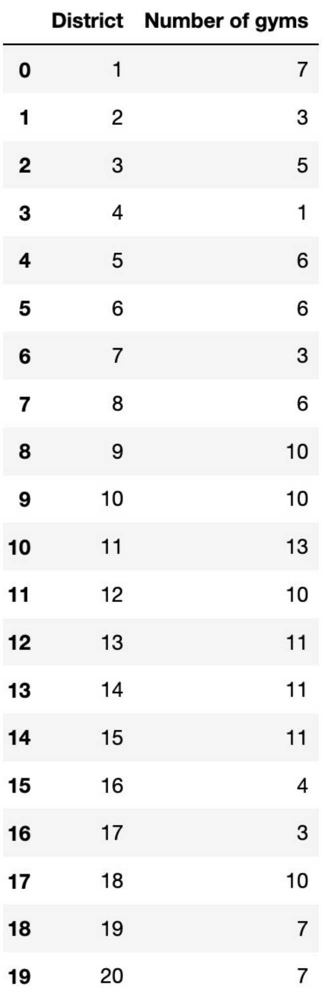


	District	name	id	address	latitude	longitude
0	1	GYM-LOUVRE	4bc4cff0abf49521509bc593	7 rue Du Louvre	48.862214	2.341375
1	1	Gym de l'Hôtel Saint James Albany	5c228e6893bd63002c457072	Hôtel Saint James Albany	48.864293	2.330822
2	1	Gym de l'Hôtel Renaissance	5d29082b3f9ff70023742cb2	Hôtel Renaissance	48.865598	2.329346
3	3	Temple Gym and Fitness	54c67dab498eb4050f6557af	NaN	48.864591	2.353936
4	2	Gym comp	4d1d9b81d7b0b1f7f30efc9e	10 rue d'Aboukir	48.865639	2.343562



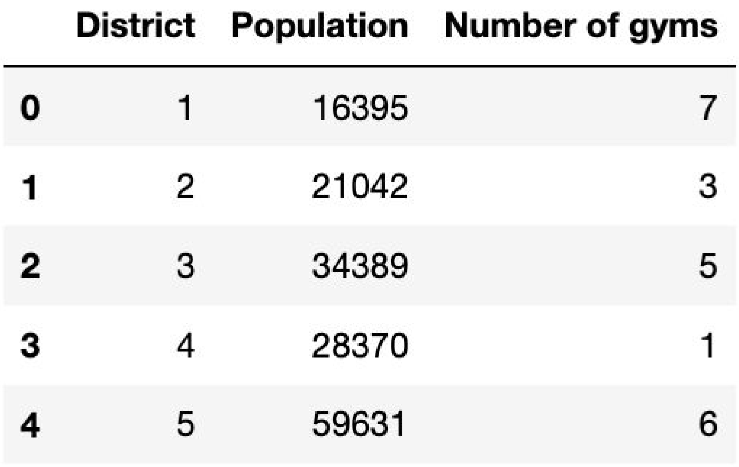
III.3 Inhabitants per gym per district in Paris:

We first counted the gyms in each district by grouping the gym dataframe by district and counting the ids.



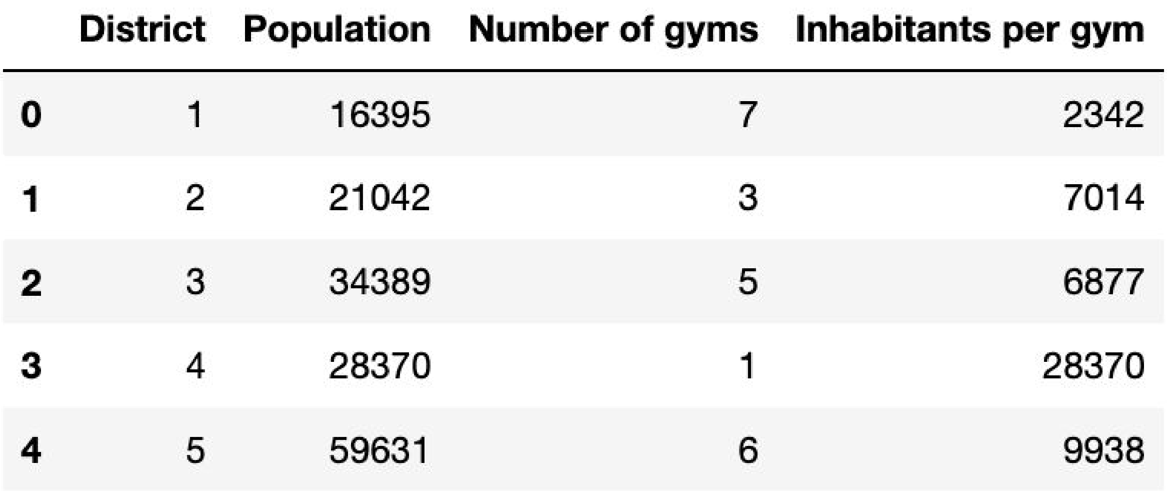
	District	Number of gyms
0	1	7
1	2	3
2	3	5
3	4	1
4	5	6
5	6	6
6	7	3
7	8	6
8	9	10
9	10	10
10	11	13
11	12	10
12	13	11
13	14	11
14	15	11
15	16	4
16	17	3
17	18	10
18	19	7
19	20	7

We then merged that table with the table containing the population per district. We used the district number as a key.



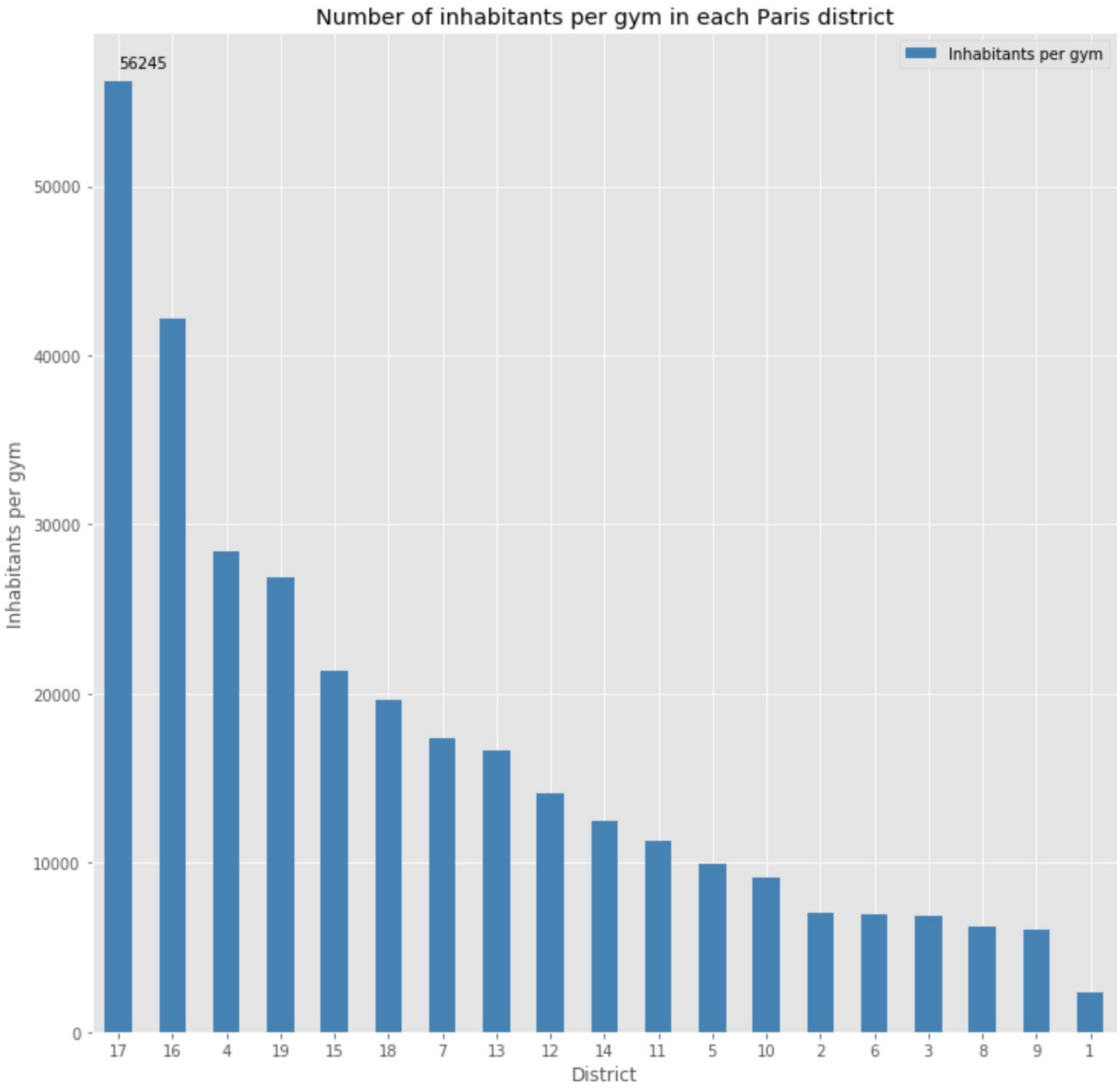
	District	Population	Number of gyms
0	1	16395	7
1	2	21042	3
2	3	34389	5
3	4	28370	1
4	5	59631	6

We created a new calculated column in order to have the number of inhabitants per gym.



	District	Population	Number of gyms	Inhabitants per gym
0	1	16395	7	2342
1	2	21042	3	7014
2	3	34389	5	6877
3	4	28370	1	28370
4	5	59631	6	9938

We then created a bar chart using matplotlib in order to have a visual representation of the preceding table



IV. Conclusion:

As a conclusion, this analysis showed us that the 17th district would be the more interesting district for SuperFit as it has the highest number of inhabitants per gym.

SuperFit should thus highly consider this district to open its new gym in Paris.