Open a new climbing gym in France

Préparé par : Rudy M.

31 mars 2020

INTRODUCTION

Background

Climbing gym is a place where the mountain lovers like to go to train when the wether is too bad to climb on the rock. For a long time, these place has reserved to the people living near of a mountain because we didn't talk about climbing elsewhere. This is not still true today. The interest for this sport is growing quickly everywhere in France and some climbing gym open in cities faraway from the mountains.

Problem

We want to take the opportunity and open a climbing gym in a city. We can place your business everywhere is France, but where is the best city and the best place to be ? May the data available on internet can help us to choose.
Obviously, the goal is to create the most proficient company, this analysis will help us to make the good chose.

DATA ACQUISITION AND CLEANING

Data sources

For this project, I used two sources of data. The first one is dataset available on gouvernement's web site, it contain informations about all the cities in France like localisation, name, population. The second one is the Foursquare API, it'll help me to get informations about places and competitors (others climbing gym).

Data cleaning and feature selection

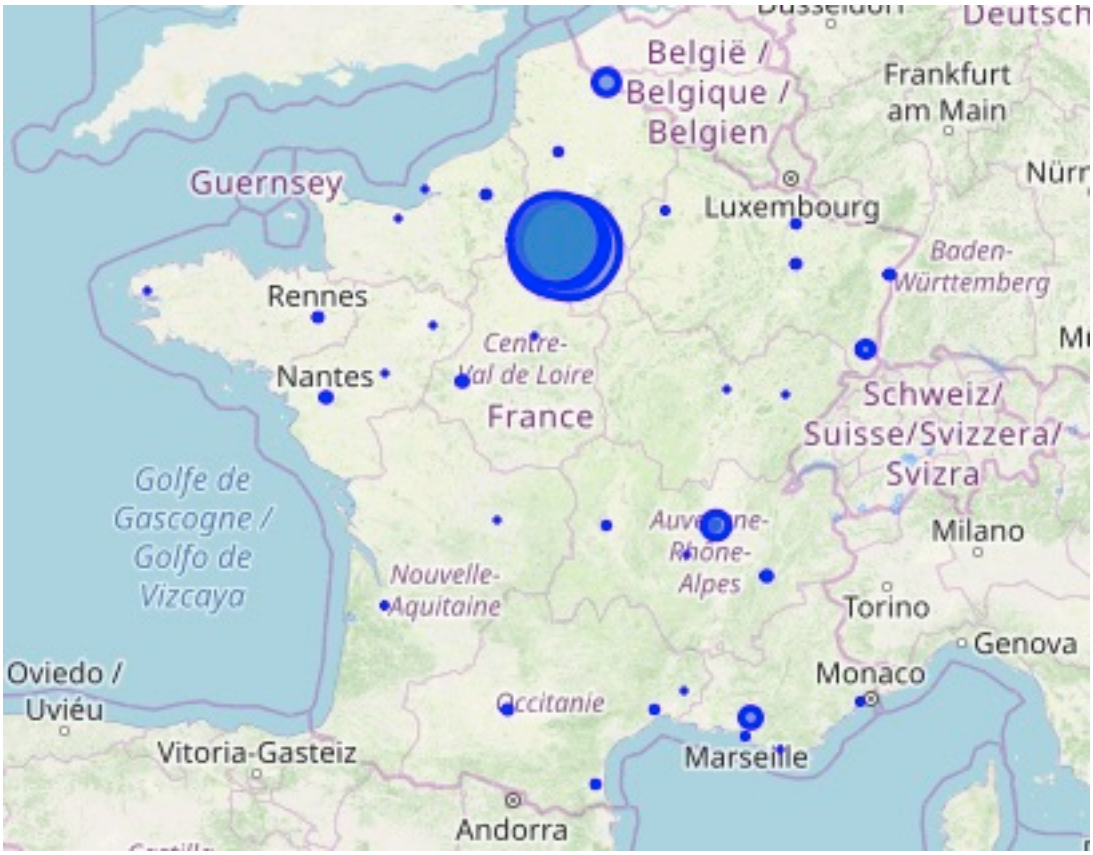
I checked about missing value or other error in the dataset but I found nothing. The dataset is really well made.

After a quick analysis of the dataset, I selected four column to keep: Name, Longitude, Latitude, 2010 Population. The others, like old population data, are not necessary for the project. Also I removed the cities with less than 100,000 inhabitant, the target is not the small cities. I also needed to add some data about climbing gym, so I used the Foursquare API to get the number of this kind of venues in each city. Then I calculate the number of climbing gym by inhabitant of each one. There is a part of the result :

	com_nom	long	lat	pop_2010	number_of_gym	gym_by_inhabitant
0	PARIS	2.352222	48.856614	2.240213e+06	19	0.000008
969	BOULOGNE-BILLANCOURT	2.237803	48.843250	1.139234e+05	19	0.000167
1020	MONTREUIL	2.448451	48.863812	1.063122e+05	20	0.000188
1034	SAINT-DENIS	2.357443	48.936181	1.097270e+05	18	0.000164
1117	ARGENTEUIL	2.248950	48.947411	1.053462e+05	18	0.000171

DATA ANALYSIS

Find the best city

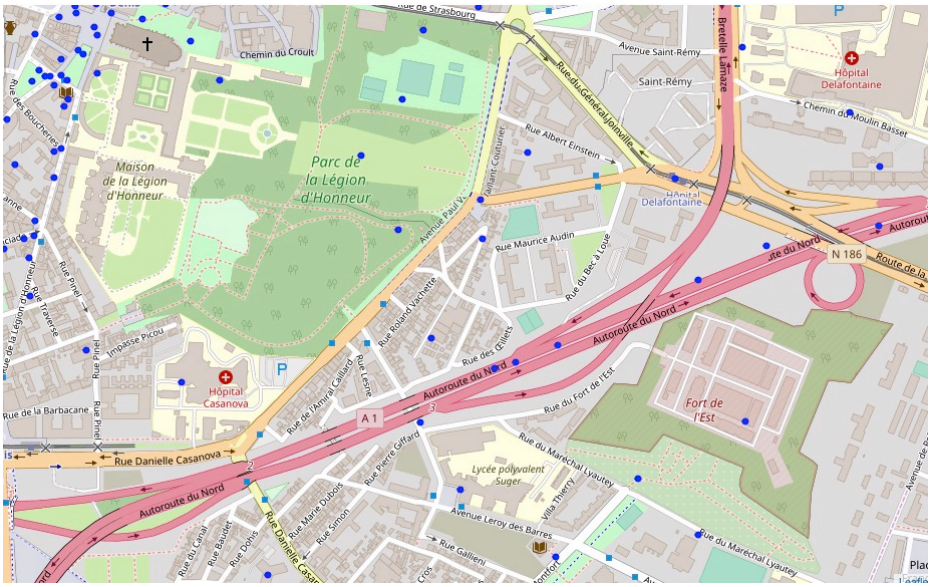


Here is a map of the climbing gym in French biggest cities. The radius each point is defined by the number by inhabitant.

The localisation does not mater anymore to develop interests in climbing so the best city seems to be the one with the lowest climbing gym by inhabitant. After sorting data, that was « Limoges » with 0. That's a good news, we'll be the first gym of this kind here.

Create neighborhood

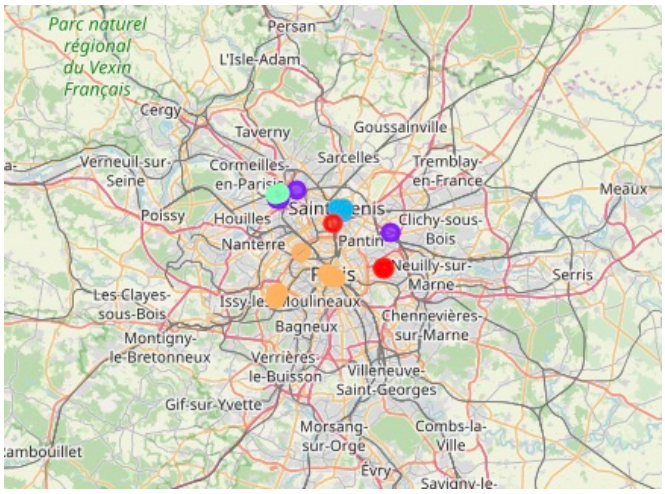
I didn't find any dataset about neighborhood in France, so I decided to create some virtual one using clustering. First I use Foursquare API to get about 500 venues for each city. Here is an example of city :



Next, I'm group them with k-means clustering on their latitude and longitude. This way I can create virtual neighborhood distributed over the city.

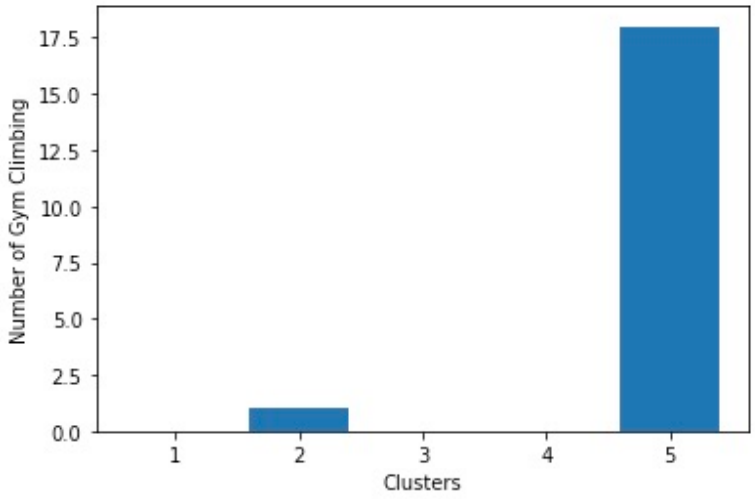
Find the best neighborhood

The first step to find the best is to get informations about our newly created neighborhood. I used Foursquare again an ask for the venues around each one. I calculated the proportion of each type of venue in the neighborhood. Next, I used k-means again but this time on the proportion just calculated. I got 5 group of neighborhood.



The last question is : Is there a kind of neighborhood where Climbing Gym is preferably placed ?

The answer is in the distribution of the climbing gym between the groups, let's check it :



As we can see, the cluster 5 is really plebiscite than the other. I'm surprised by the clearness of this result !

So we must open your business is Limoges and in a neighborhood of the 5th cluster. Let's find one :

	com_nom	lat	long	PostalCode	Cluster Labels
125	LIMOGES	45.830455	1.258095	125	4
126	LIMOGES	45.831846	1.256621	126	4
128	LIMOGES	45.831340	1.259499	128	4
129	LIMOGES	45.832217	1.258238	129	4

There are four lines which perfectly match our needs, so let's find an area for sale there.

CONCLUSIONS

In this study, I tried to find the best place to open a Climbing Gym with unsupervised learning. I found that there is a very high proportion of those place in only one kind of neighborhood and that there are still some city like Limoges without any climbing gym. The situation is clear, we need to chose a place in Limoges and in the plebiscite kind of neighborhood.

DISCUSSION AND FUTURE DIRECTIONS

Obviously, in a real project this study will not be enough because I didn't explore many way to classify and rank competitors and neighborhoods because of time and expensiveness of Foursquare premium plan. However, I think that study is a very good start point for a real project because I succeed to show that there is city without this kind of service and I also found a clear relation between the neighborhood and the number of climbing gym.

There are many future directions to explore like the closeness of the mountains, the ranking of the competitors, the salary of peoples in the cities (Climbing Gym attract mainly the well paid people), and so on.