Diving equipment store in Haute-Savoie

Table of contents

1. Introduction : Business Problem

In this project we will try to find an optimal location for a diving equipment store. Specifically, this report will be targeted to stakeholders interested in opening a **diving & watersport equipment store** in **Haute-Savoie**, France.

Since the store is dedicated to scuba diving and watersport activities, we will first try to detect **locations with *water access* in vicinity**. In order to limit the effect of competition, it will also be necessary to ensure that **no such sports store is already established near the selected areas**. Then, assuming that these two conditions are met, we would prefer **frequented places** with a fairly high population density.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed, so that the stakeholders can easily make a choice.

2. Data

Based on our business problem, we will have to gather data such as :

- Name and localization of all the cities in Haute-Savoie
- Population of each city
- Number and type of water access in the neighborhood of each city, if any
- Number of existing sports store in the neighborhood of each city

Following data sources will be needed :

- Names and populations of the cities in Haute-Savoie will be obtained using an available table of **Wikipedia**
- Localization of the cities will be read from a local csv file
- Venues such as water access and sports store will be extracted using **Foursquare API**

a. Cities and their population in Haute-Savoie

Let's extract the available table of Wikipedia where all the cities ('communes' in french) of Haute-Savoie are listed with also the associated population.

Here are the first lines of this table :

	Nom	CodeInsee	Code postal	Arrondissement	Canton	Intercommunalité	Superficie(km2)	Population(dernière pop. légale)	Densité(hab./km2)	Modifier
0	Annecy(préfecture)	74010	74000743707460074940749 60	Annecy	Annecy-1Annecy-2Annecy-le-VieuxSeynod	CA du Grand Annecy	6694	126 924 (2017)	1 896	NaN
1	Abondance	74001	74360	Thonon-les-Bains	Évian-les-Bains	CC Pays d'Évian Vallée d'Abondance	5884	1 439 (2017)	24	NaN
2	Alby-sur-Chéran	74002	74540	Annecy	Rumilly	CA du Grand Annecy	656	2 579 (2017)	393	NaN
3	Alex	74003	74290	Annecy	Faverges	CC des vallées de Thônes	1702	1 072 (2017)	63	NaN
4	Allèves	74004	74540	Annecy	Rumilly	CA du Grand Annecy	881	411 (2017)	47	NaN

We only need the columns with the name and the population, so let's clean the table.

Here's the first lines of the cleaned table :

	Commune	Population
0	Annecy	126 924
1	Abondance	1 439
2	Alby-sur-Chéran	2 579
3	Alex	1 072
4	Allèves	411

b. Localization of the cities

Now let's collect also the latitude and longitude of all these cities. For that we have to read a local csv file.

Here are the first lines of this table :

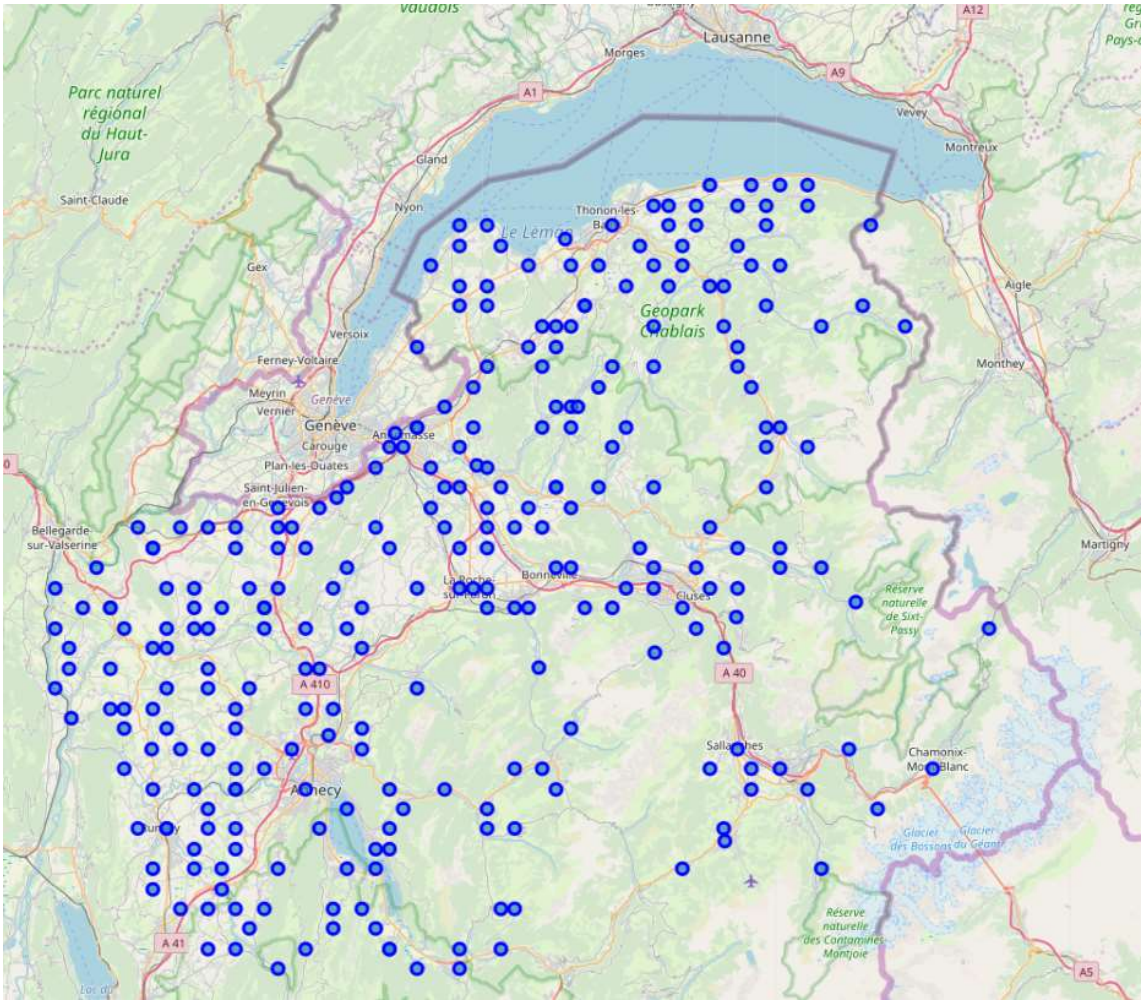
	Commune	Code Postal	Latitude	Longitude
0	Annecy	74000\n74370\n74600\n74940\n74960	45.900002	6.11667
1	Abondance	74360	46.283329	6.73333
2	Alby-sur-Chéran	74540	45.816700	6.01670
3	Alex	74290	45.883331	6.23333
4	Allèves	74540	45.750000	6.08333

Let's first remove the postal code from the previous table and then combine the 2 databases. We obtain a list of **279 cities**.

Here are the first lines of the combined database :

	Commune	Population	Latitude	Longitude
0	Annecy	126924	45.900002	6.11667
1	Abondance	1439	46.283329	6.73333
2	Alby-sur-Chéran	2579	45.816700	6.01670
3	Alex	1072	45.883331	6.23333
4	Allèves	411	45.750000	6.08333

In order to visualize on a map the repartition of the cities, we use folium and obtain the following map :



c. Venues in the neighborhoods

Now that we have a good overview of the cities in Haute-Savoie, let's find out where are the 'water access' and sports stores around them. We will use Foursquare API to gather this information.

Let's now go over our neighborhood locations and get nearby venues. We look for a hundred of venues in a radius of 2500 m from each city. We obtain a list of **2366 venues**.

Here are the first lines of this list :

	Commune	Commune Latitude	Commune Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Id
0	Annecy	45.900002	6.11667	Brumes	45.899517	6.123535	Coffee Shop	4bf58dd8d48988d1e0931735
1	Annecy	45.900002	6.11667	Beer O'Clock	45.897427	6.123039	Bar	4bf58dd8d48988d116941735
2	Annecy	45.900002	6.11667	Une Autre Histoire	45.899761	6.121999	Tea Room	4bf58dd8d48988d1dc931735
3	Annecy	45.900002	6.11667	Chez Pen	45.904349	6.121446	Bar	4bf58dd8d48988d116941735
4	Annecy	45.900002	6.11667	Le Barista Café	45.900824	6.124986	Coffee Shop	4bf58dd8d48988d1e0931735

By regrouping the venues by category, we find out that there are **228 unique categories** in this list.

We have to define the list of relevant venues. We can find all the needed documentation on Foursquare website *(https://developer.foursquare.com/docs/resources/categories)*.

Let's extract only the relevant venues for our research, that is to say all the venues related to water access and sports shops.

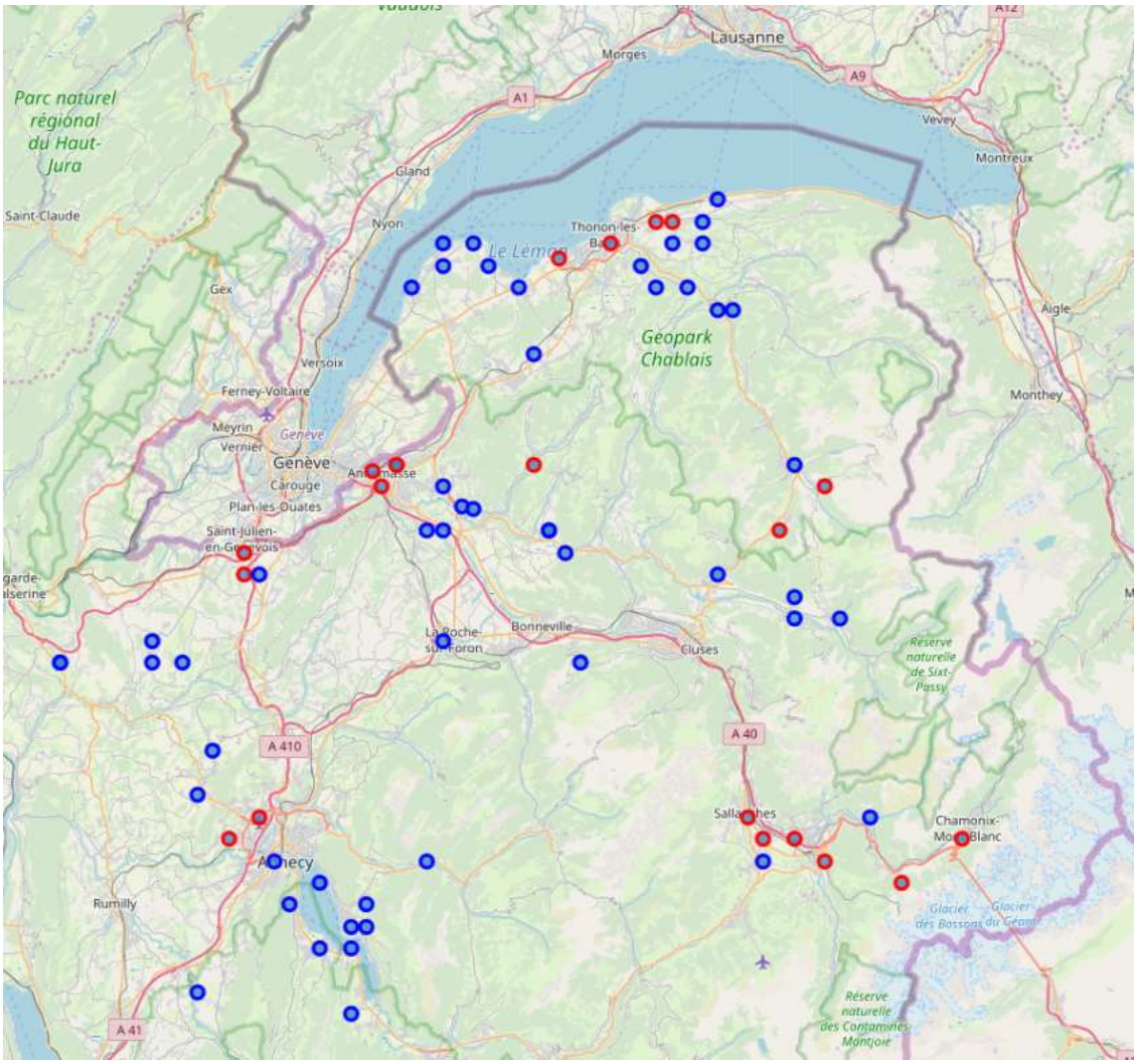
The results is a list of **113 venues** of **14 unique categories**.

Here are the first lines of this list :

	Commune	Bay	Beach	Dive Spot	Harbor / Marina	Hot Spring	Lake	Outdoor Supply Store	Pool	Rafting	Reservoir	River	Spa	Sporting Goods Shop	Water Park	Waterfall
46	Annecy	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
50	Annecy	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
77	Annecy	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
134	Ambilly	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
194	Annemasse	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Let's group by city and add the population and localization. We have then a list of **70 cities** in Haute-Savoie.

Let's now see all the collected data on a map. All the cities that have water access are shown with a blue dot, and if there's a sports shop at the same time we add a red circle around :



This concludes the data gathering phase - we're now ready to use this data for analysis to find an optimal location for a new dive equipment store !

3. Methodology and data analysis

Now that we have the needed data, let's go back to our criteria to define the optimal location :

- locations with water access in vicinity
- no sports store already established
- frequented places

a. Types of water access

As we saw above, there are different types of water access. To help the stakeholders in their final decision, it can be interesting to group them as follow :

- Open water access such as beach or lake
- Confined water access such as pool, spa or water park
- Rapids access such as rafting or waterfall

Indeed these 3 types could involve specific sports equipment. So according to the preferential target aimed by the articles of the store, it could be helpful to have this distinction.

So let's create a new table with these categories. At the same time we can clean the list by removing all the cities that have only a sports store and no water access.

Here's the list of the 58 cities that correspond to our criteria, with their population, localization and number of sports store and water accesses by type :

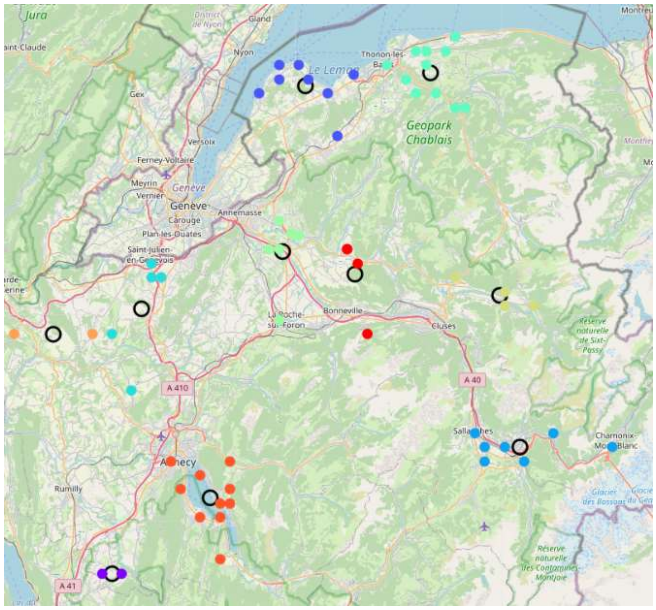
	Commune	Sporting Goods Shop	Population	Latitude	Longitude	Open water	Confined water	Rapids
0	Annecy	0	126924	45.900002	6.11667	3	0	0
1	Anthy-sur-Léman	0	2171	46.355300	6.42730	1	0	0
2	Armoy	0	1295	46.349998	6.51667	0	0	1
3	Arthaz-Pont-Notre-Dame	0	1577	46.150002	6.28333	0	1	0
4	Bluffy	0	391	45.866669	6.21667	2	0	0
5	Bonne	0	3218	46.168200	6.32150	0	1	0
6	Brenthonne	0	1037	46.283329	6.40000	0	1	0
7	Brizon	0	485	46.049999	6.45000	0	1	0
8	Chainaz-les-Frasses	0	728	45.766670	6.00000	0	1	0
9	Chamonix-Mont-Blanc	1	8611	45.916672	6.86667	0	2	0
10	Champanges	0	1015	46.366669	6.55000	0	0	1
11	Chavannaz	0	236	46.049999	6.01667	0	1	0
12	Chens-sur-Léman	0	2776	46.333328	6.26667	1	0	0
13	Choisy	0	1604	45.983330	6.05000	1	0	0
14	Chêne-en-Semine	0	496	46.049999	5.85000	1	0	0
15	Combloux	0	2106	45.900002	6.65000	1	0	0
16	Cranves-Sales	0	6793	46.183331	6.30000	0	1	0
17	Cusy	0	1850	45.766670	6.03333	0	1	0
18	Dingy-Saint-Clair	0	1433	45.900002	6.21667	0	0	1
19	Domancy	0	2113	45.916672	6.65000	2	0	0
20	Duingt	0	972	45.833328	6.20000	5	0	1
21	Etaux	0	1979	46.066669	6.30000	0	1	0
22	Excenevex	0	1104	46.349998	6.35000	1	0	0
23	Feigères	1	1682	46.116669	6.08333	0	1	0
24	Fillinges	0	3413	46.166672	6.33333	0	1	0
25	La Forclaz	0	233	46.316669	6.61667	0	0	1
26	La Tour	0	1275	46.133331	6.43333	1	0	0
27	La Vernaz	0	337	46.316669	6.60000	1	0	0
28	Lathuile	0	1016	45.783329	6.20000	1	0	0
29	Lyaud	0	1729	46.333328	6.53333	0	0	1

30	Marin	1	1789	46.383331	6.53333	1	0	0
31	Menthon-Saint-Bernard	0	1889	45.849998	6.20000	6	0	1
32	Messery	0	2134	46.349998	6.30000	2	0	0
33	Minzier	0	1023	46.049999	5.98333	0	1	0
34	Morillon	0	662	46.083328	6.68333	2	0	0
35	Nangy	0	1618	46.150002	6.30000	0	1	0
36	Nernier	0	380	46.366669	6.30000	2	0	0
37	Neuvecelle	0	3048	46.400002	6.60000	0	1	0
38	Neydens	1	1915	46.116669	6.10000	0	1	0
39	Passy	1	10902	45.916672	6.68333	1	1	0
40	Publier	1	7148	46.383331	6.55000	1	0	0
41	Reyvroz	0	510	46.333328	6.56667	0	0	1
42	Saint-Gervais-les-Bains	1	5573	45.900002	6.71667	0	1	0
43	Saint-Jorioz	0	5738	45.833328	6.16667	2	0	0
44	Saint-Julien-en-Genevois	1	14258	46.133331	6.08333	0	1	0
45	Sallanches	1	16088	45.933331	6.63333	2	0	0
46	Samoëns	0	2458	46.083328	6.73333	1	0	0
47	Sciez	0	6033	46.333328	6.38333	3	0	0
48	Servoz	0	999	45.933331	6.76667	1	0	0
49	Sevrier	0	4161	45.866669	6.13333	4	0	0
50	Talloires-Montmin	0	1996	45.849998	6.21667	2	0	1
51	Taninges	0	3443	46.116669	6.60000	1	0	0
52	Thonon-les-Bains	0	34756	46.366669	6.48333	1	0	0
53	Verchaix	0	752	46.099998	6.68333	1	0	0
54	Veyrier-du-Lac	0	2251	45.883331	6.16667	5	0	0
55	Ville-en-Sallaz	0	905	46.150002	6.41667	1	0	0
56	Viuz-en-Sallaz	0	4374	46.150002	6.41667	1	0	0
57	Yvoire	0	993	46.366669	6.33333	2	0	0
58	Évian-les-Bains	0	9098	46.383331	6.58333	0	1	0

b. Clustering

Let's now cluster those cities to create centers of zones containing good locations.
We will use **k-means** method with a number of **10 clusters** to do so, and the clusters will be based on the localizations of the 58 cities.

Let's see on a map where are the centers of the 10 defined clusters. They are represented with a black circle and the clusters of cities are shown with 10 different colors.



Now that we have defined 10 zones that could be good candidates, let's compare them more in details.

Let's have a look at each cluster. For each of the zones, we can resume their population and numbers of sport shops and types of water access. The table below shows the results :

	Cluster Labels	Sporting Goods Shop	Population	Open water	Confined water	Rapids
0	0	0	7039	3	1	0
1	1	0	2578	0	2	0
2	2	0	16628	12	1	0
3	3	4	46392	7	4	0
4	4	3	19695	1	4	0
5	5	2	60958	4	2	5
6	6	0	18598	0	6	0
7	7	0	7315	5	0	0
8	8	0	1519	1	1	0
9	9	0	146771	30	0	4

c. Finding the optimal location

Our criteria to define the optimal locations were :

- locations with water access in vicinity
- no sports store already established
- frequented places

So let's compare each cluster, that is to say each potential zone with these criteria. On the below table the total number of water accesses for each cluster is written :

	Cluster Labels	Sporting Goods Shop	Population	Total number of water access
0	0	0	7039	4
1	1	0	2578	2
2	2	0	16628	13
3	3	4	46392	11
4	4	3	19695	5
5	5	2	60958	11
6	6	0	18598	6
7	7	0	7315	5
8	8	0	1519	2
9	9	0	146771	34

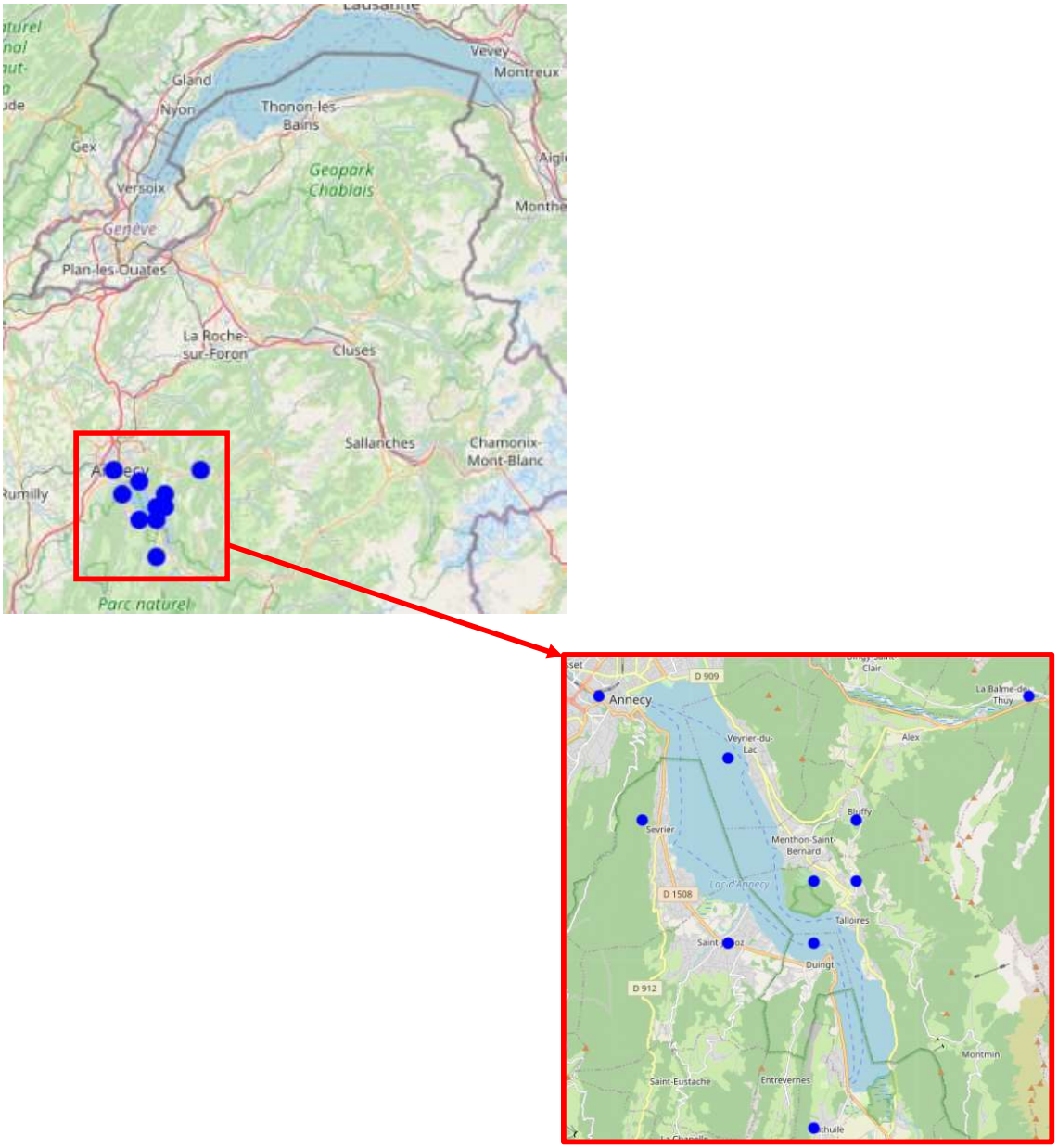
It appears clearly that **cluster n°9** is the best candidate with the maximum number of water accesses, the largest population and no sport shop in the area.

4. Results

The table below corresponds to the list of the cities linked to the optimal cluster. It resumes also the population of each city, their number of water accesses by type and their localization :

	Commune	Sporting Goods Shop	Population	Open water	Confined water	Rapids	Latitude	Longitude
0	Annecy	0	126924	3	1	0	45.900002	6.11667
1	Bluffy	0	391	2	0	0	45.866669	6.21667
2	Duingt	0	972	6	0	1	45.833328	6.20000
3	La Balme-de-Thuy	0	462	0	0	1	45.900002	6.28333
4	Lathuile	0	1016	1	0	0	45.783329	6.20000
5	Menthon-Saint-Bernard	0	1889	7	0	1	45.849998	6.20000
6	Saint-Jorioz	0	5738	1	0	0	45.833328	6.16667
7	Sevrier	0	4161	3	0	0	45.866669	6.13333
8	Talloires-Montmin	0	1996	2	0	1	45.849998	6.21667
9	Veyrier-du-Lac	0	2251	5	0	0	45.883331	6.16667

Let's see on a map the cities of the optimal zone found :



The best zone in Haute-Savoie to create a diving & watersport equipment store, according to our criteria, seems to be around the famous *'Lake of Annecy'*.

5. Discussion

Our analysis shows that although there is a great number of water accesses in Haute-Savoie, the best zone to create a new diving & watersport equipment store is **around the lake of Annecy**. It is where the 3 criteria of our research are clearly satisfied : 34 water accesses, no existing sport shop in the area and more than 145 000 inhabitants.

Nevertheless, we've also seen that water access could be of 3 different types : open water, confined water or rapids. These 3 types could involve specific sports equipment. So according to the preferential target aimed by the articles of the store, another analysis could be useful to defined the most appropriate location in the defined area.

Moreover the surroundings of the lake of Annecy is quite a big area and this study should be considered as **a starting point to a more detailed analysis** which could also consider **new criteria**, such as **real estate availability** in the selected areas and prices, **proximity to major roads**...

6. Conclusion

Purpose of this project was to identify an optimal location in Haute-Savoie to open a new diving & watersport equipment store. The criteria for this analysis where :

- water access in the vicinity
- no already existing sports store
- freqented place

The surroundings of the Lake of Annecy was clearly identified as the best location.

A list of the cities in this area was etablished with, for each of them, the number of water access separated in 3 different types which could be used as a starting point for final exploration by stakeholders.

Final decision on optimal location will be made by stakeholders based on specific characteristics of neighborhoods and locations in this recommended zone, taking into consideration additional factors like attractiveness of each location, proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood...