Gym market in San Francisco

IBM APPLIED DATA SCIENTE CAPSTONE PROJECT – COURSERA

GIAMPAOLO TUMMINELLO

Problem

Awareness about health issues has resulted in an increase in expenditure on healthy lifestyle and fitness activities.

Revenue of the fitness industry is expected to show an annual growth rate of 0.6%, by 2024.

For entrepreneurs who wants to enter the market, **where** to open a new gym **is crucial for success** of the investment.

The distribution of gyms in the city of San Francisco is analyzed, to **see where the "good spots" are hidden** in the city.

Good spot's characteristics:

- Great potential client base.
- No strong direct competition.
- Good geographical position overall.

Data

- Neighborhoods of San Francisco, from government websites.
- Coordinates of each neighborhood, using the Geocoder library.
- Venues for each neighborhood, retrieved using Foursquare API.

	name	Latitude	Longitude
0	Seacliff	37.788540	-122.486920
1	Lake Street	37.785060	-122.464040
2	Presidio National Park	37.799930	-122.463580
3	Presidio Terrace	37.788260	-122.460800
4	Inner Richmond	37.780900	-122.465560
...	...	...	...
112	Corona Heights	37.763640	-122.440390
113	Ashbury Heights	37.764670	-122.445870
114	Eureka Valley	37.757501	-122.437941
115	St. Francis Wood	37.734650	-122.468030
116	Sherwood Forest	37.737510	-122.460060

Methodology

The process followed these steps:

1. Create the data frame of neighborhoods and coordinates;
2. Getting venues data and aggregating with the above data frame;
3. Identify the different kinds of gym and aggregate them;
4. Apply the k-means cluster algorithm;
5. Identify the cluster which represents the “good spot”.

Results

The k-means cluster algorithm applied consisted in the division of all the neighborhoods of San Francisco into 10 clusters. After, for each cluster, the number of gyms is considered.

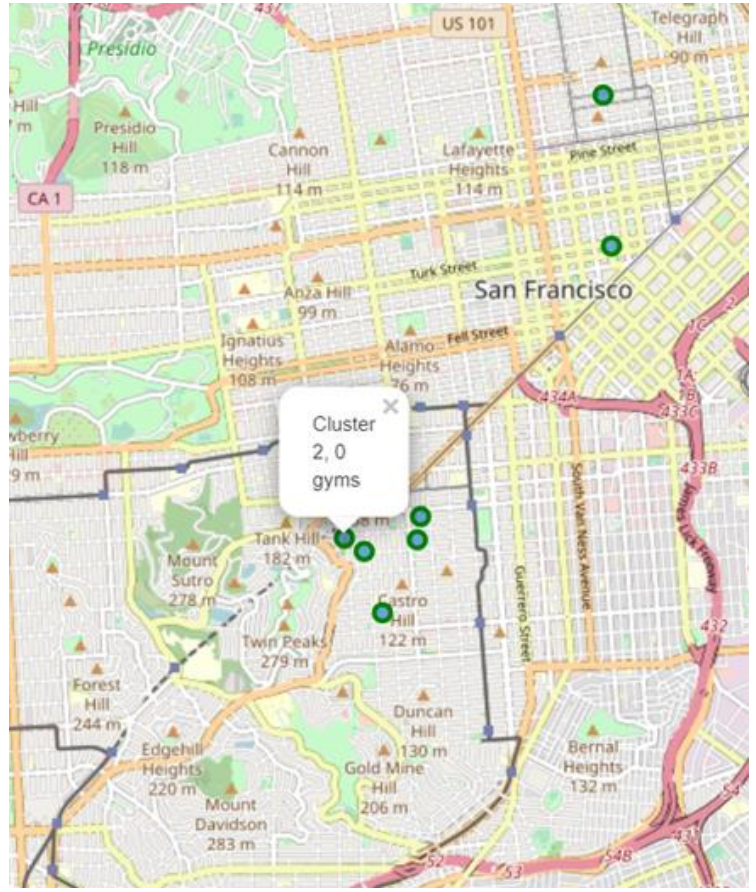
Cluster 2 and cluster 9 have no gyms, but cluster 9 is outside the city centre, so **cluster 2 is the good spot.**

	Cluster	Latitude	Longitude	Gyms
0	0	37.757576	-122.439539	38
1	1	37.782505	-122.413980	36
2	2	37.758625	-122.441638	0
3	3	37.764866	-122.442029	52
4	4	37.758477	-122.434021	25
5	5	37.752511	-122.437581	42
6	6	37.778950	-122.398707	16
7	7	37.760353	-122.433764	25
8	8	37.794835	-122.414810	14
9	9	37.965865	-121.722447	0

Conclusions

Cluster 2 has 0 gyms in it, even if it is surrounded by other clusters containing respectively 38 gyms (cluster 0) and 52 gyms (cluster 3).

It represents a great opportunity for entrepreneurs who wants to enter in the fitness industry in San Francisco.



Thanks for your attention