Finding suitable locations to open a Gym in Athens, Greece

1. Introduction / Business Problem

The aim of this project is to find suitable locations to open a **gym** in the **Athens** greater metropolitan area [Athens and its suburbs], Greece.

The first requirement is that the new gym should be easily accessible by its prospective customers and more specifically it should be located **near a metro station**. The **number of gyms already existing in an area** should also be considered so that fierce competition be avoided if possible.

Apart from the obvious intended stakeholders, entrepreneurs looking to start a gym business, similar methodology could be used for other specific types of businesses. It can serve as an initial starting point of locations to consider to start their business.

For the project objectives to be achieved, python geolocation libraries were used, along with the Foursquare API. Also, in order to create clusters of similar candidate locations, the K-Means machine learning clustering algorithm was used.

2. Data

The necessary data for this project, based on the above stated requirements, are:

- The metro stations in the Athens greater metropolitan area
- Number of existing gyms near each station
- In addition, the distance to the nearest gym for every metro station will be used

In order to obtain the data, a combination of the **geopy** Python library and the **Foursquare API** were used:

1. *'Syntagma square'* was considered as the center of Athens. It is indeed one of the most central location in the city. I obtained its geospatial coordinates using the geopy library.
2. Having the coordinates of the 'center' of Athens, the Foursquare API was used to retrieve data for all the metro stations in Athens greater area in a radius of 15 km.
3. To find the existing gyms near the metro stations, the Foursquare API was again utilized for every station. I gathered data for all the gyms located in a radius of 750 meters of every metro station.

Using the collected data, I calculated the number of existing gyms near each station. **I was also able to determine the minimum distance to a gym for every metro station** from the 3rd step of the above process. This minimum distance to every metro station from a gym, along with the number of already existing gyms near the station were used as input to K-Means clustering algorithm to obtain the clusters of areas (metro stations).

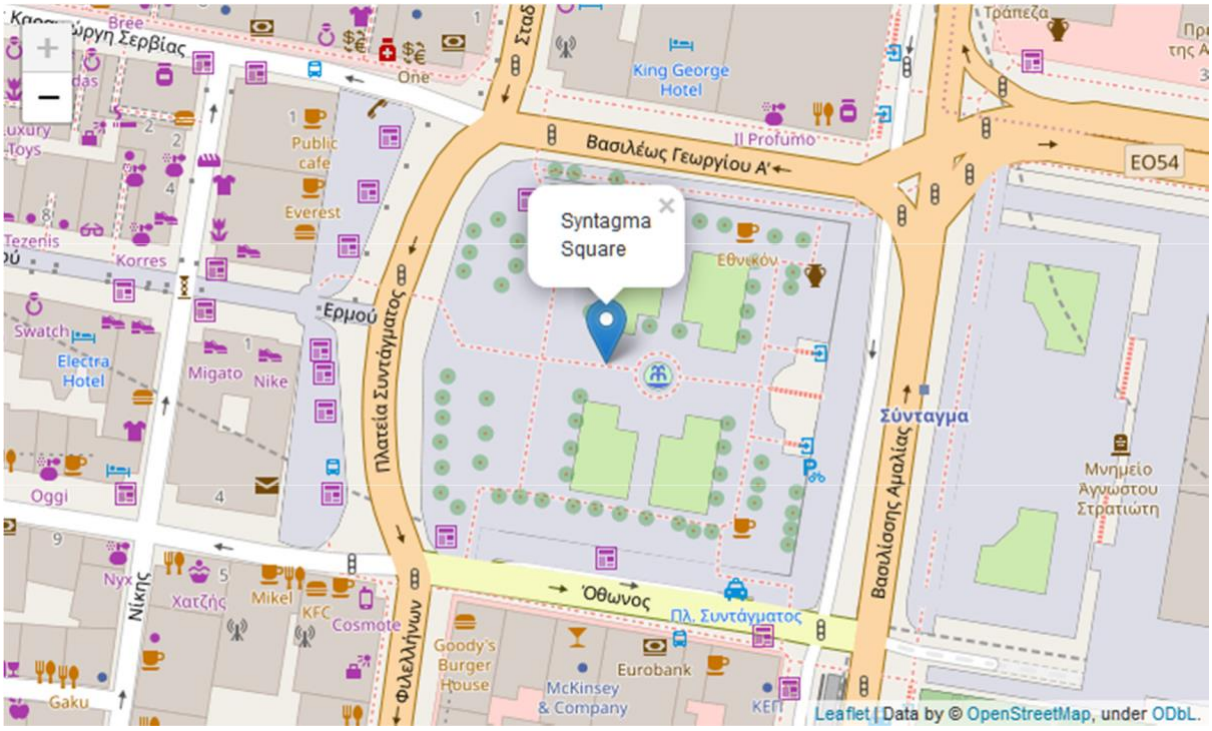
3. Methodology

The objective of this project is to obtain information about metro stations in the greater metropolitan Athens area with potential for opening a gym, and having as criteria:

- Low number of already existing gyms
- Minimum distance of each station to its nearby gyms

The steps I followed to identify potential areas (metro stations) were:

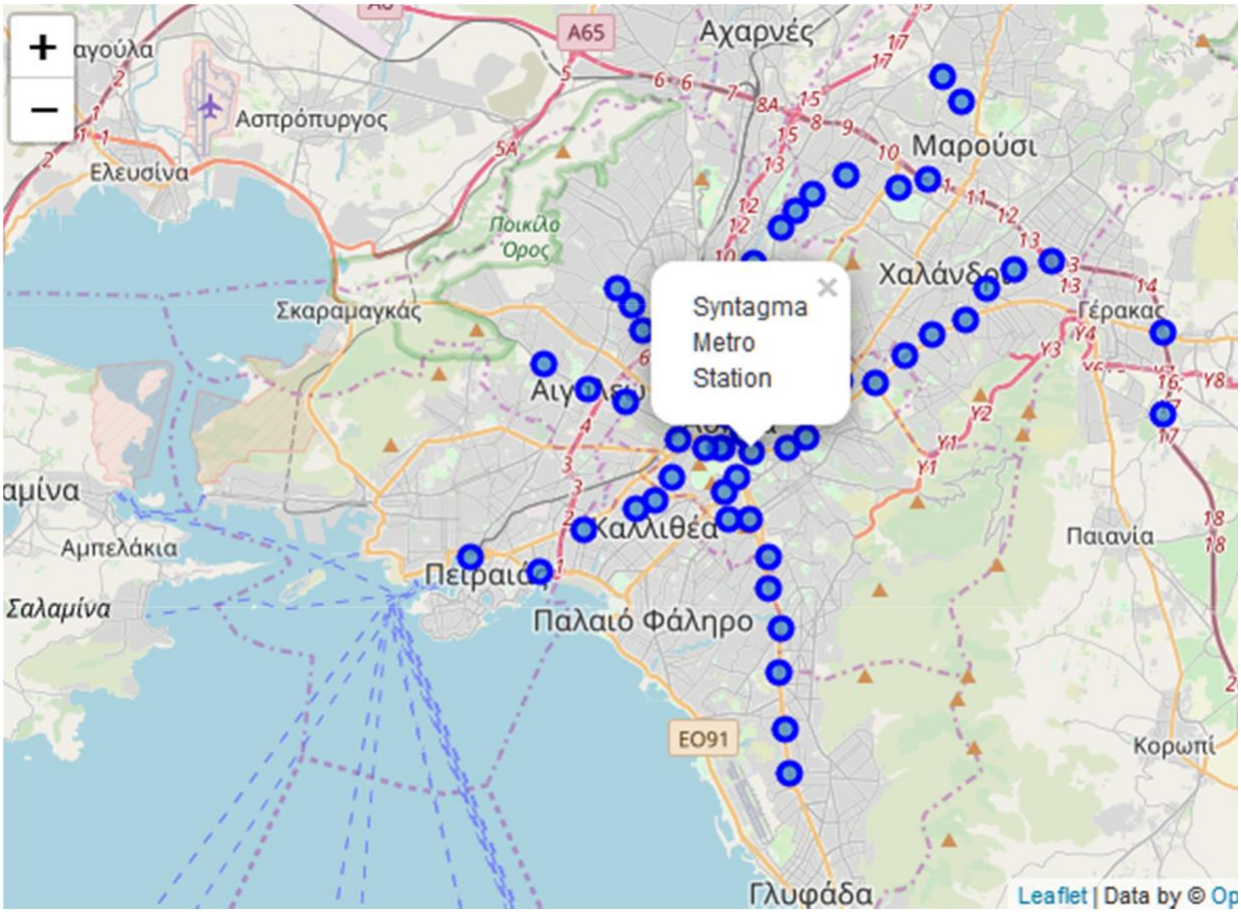
1. Considered Syntagma Square as the 'center' of Athens (indeed probably the most central location of the city) and acquired its latitude and longitude geospatial coordinates.



2. Based on the coordinates of Syntagma Square, I obtained information about metro stations in a radius of 15 km using the Foursquare API.
   At this stage I **removed from the above dataset 2 rows of data** that although they are identified as 'Metro stations' by the Foursquare API, they are only used as depots or maintenance gathering for the metro carriages.

	name	lat	lng	distance	postalCode	venue_type
40	Αμαξοστάσιο Μετρό Ελαιώνα	37.986554	23.686795	4378	NaN	Metro Station
61	Αμαξοστάσιο Μετρό Δουκίσης Πλακεντίας	38.022099	23.835587	10253	NaN	Metro Station

A visualization of the remaining metro stations on an Athens city map:



3. After the collection of metro stations information, I again utilized the Foursquare API to locate all the existing gyms in a radius of 750 meters from each station. The resulting subcategories of businesses found were:

Venue Category	station	lat	lng	Venue	Venue Latitude	Venue Longitude	Distance from Station
Athletics & Sports	2	2	2	2	2	2	2
Basketball Court	4	4	4	4	4	4	4
Boxing Gym	2	2	2	2	2	2	2
Climbing Gym	3	3	3	3	3	3	3
Cycle Studio	1	1	1	1	1	1	1
Dance Studio	7	7	7	7	7	7	7
Gym	156	156	156	156	156	156	156
Gym / Fitness Center	166	166	166	166	166	166	166
Gym Pool	3	3	3	3	3	3	3
Gymnastics Gym	7	7	7	7	7	7	7
Martial Arts Dojo	35	35	35	35	35	35	35
Massage Studio	3	3	3	3	3	3	3
Soccer Field	2	2	2	2	2	2	2
Spa	2	2	2	2	2	2	2
Track	5	5	5	5	5	5	5
Yoga Studio	93	93	93	93	93	93	93

- I kept as my data set the results that correspond **only** to **'Gym / Fitness Center'** and **'Gym'** subcategories. I removed the rest of the subcategories such as 'Dance Studio', 'Yoga Studio', 'Martial Arts Dojo' etc.
- I ignored for the purposes of clustering two metro stations that based on the results of the Foursquare API don't have any existing gyms in their vicinity. For those, there can either exist no data in the Foursquare database, or indeed there are no existing gyms near the corresponding stations.

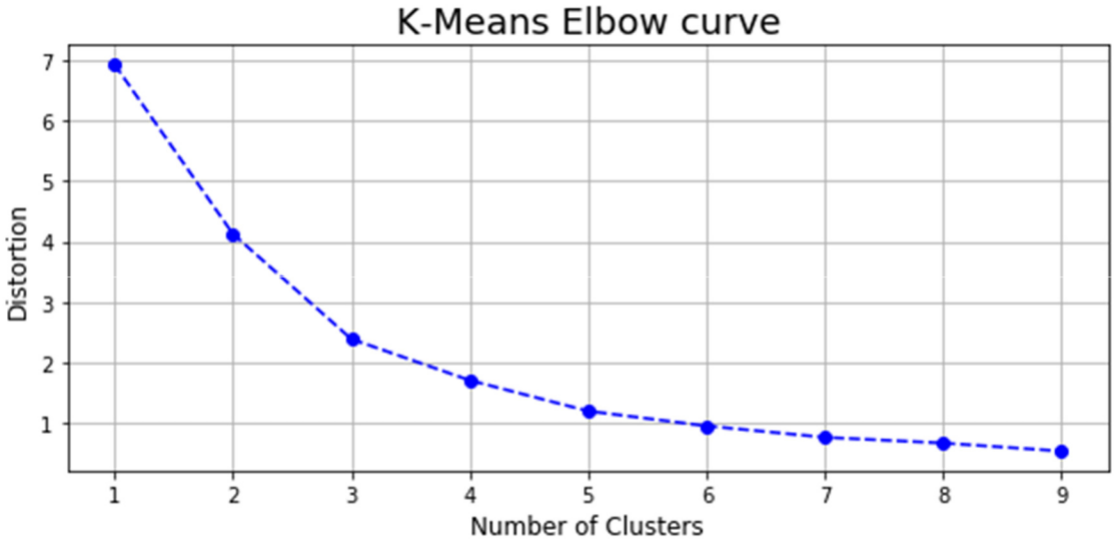
	station	lat_x	lng_x	lat_y	lng_y	Min Distance from Station	Gym Count
59	Paiania-Kantza Metro Station	37.984707	23.870084	NaN	NaN	NaN	NaN
60	Kifisia ISAP Station	38.071627	23.797488	NaN	NaN	NaN	NaN

The resulting data set will also contain the distance of each gym to the corresponding station.

4. Having the information about gyms around metro stations, I calculated the number of existing gyms near each station as well as the minimum distance from each station to a gym using available python statistical functions.
   Part of the data set containing the minimum distance and number of existing gyms for each station:

	station	lat	lng	Min Distance from Station	Gym Count
0	Syntagma Metro Station	37.975235	23.735298	112	13
13	Akropoli Metro Station	37.968516	23.730195	202	5
18	Evangelismos Metro Station	37.976173	23.747163	251	16
34	Monastiraki Metro Station	37.976067	23.725752	709	2
36	Panepistimio Metro Station	37.980004	23.732354	328	13

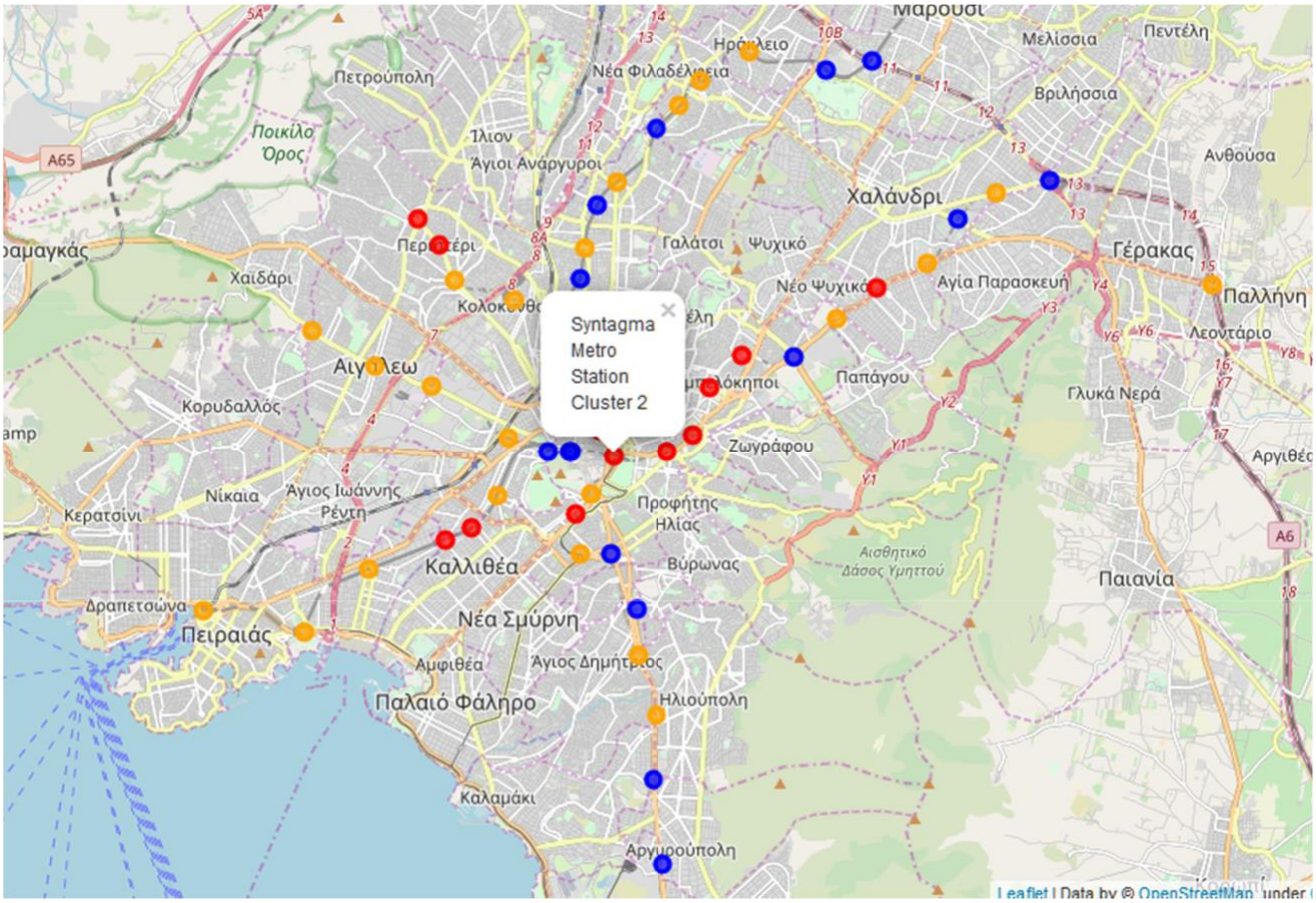
5. The data will be normalized so that both factors (minimum distance, number of existing gyms) will have equal weight when they will be used by a machine learning method.
6. The **K-Means Machine Learning clustering algorithm** will be used to divide the stations and gyms data set into clusters of similar locations.
   The elbow method will be used to find the most suitable number of clusters.



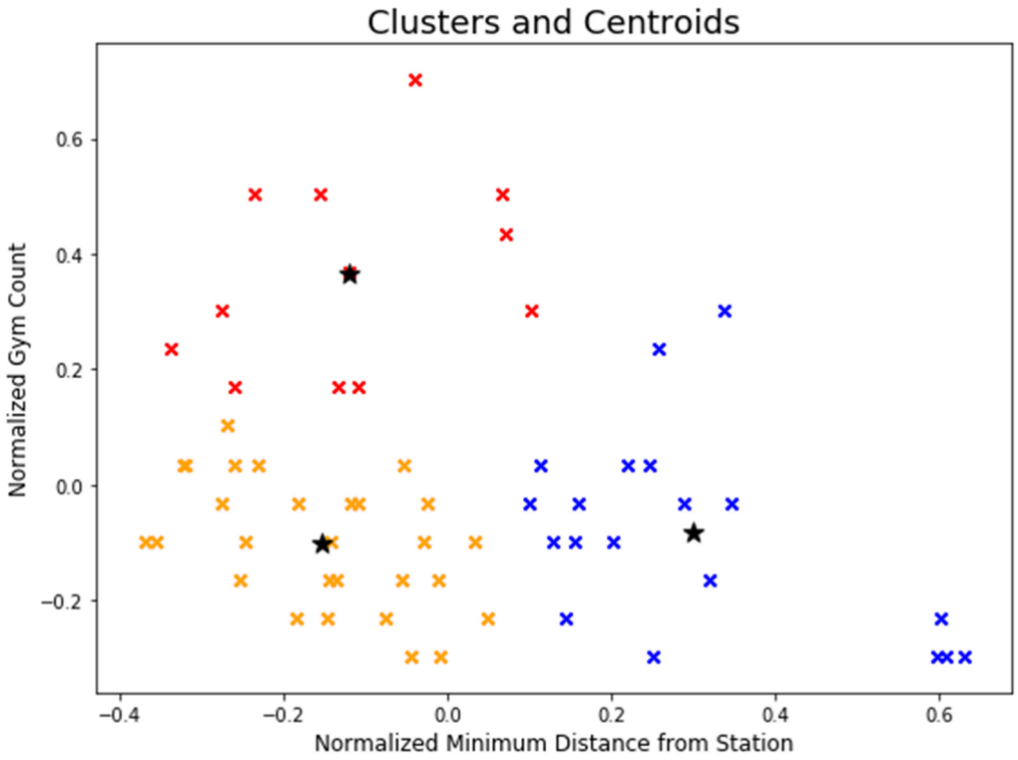
Although the elbow curve is not very steep, an elbow point of 3 clusters is clear, so this is the number of clusters that will be used for the **K-Means clustering algorithm**

4. Results

After executing the K-Means clustering algorithm three clusters of Metro stations were created, identified by their respective colors on the following map:



And a graph of the distribution of clusters and their final **centroids** (center points) in black, based on **normalized** values of minimum distance and number of existing gyms.



The three Metro stations clusters can be described as follows:

Cluster 1 [Cluster Label 0] – AVERAGE potential (color Orange on the map)

Although not a prohibitive metro station to open a gym in its vicinity, there is already a fair number of gyms in the area and the nearest one is not far from the metro station.

Examples:

station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
Akropoli Metro Station	0	37.968516	23.730195	202	5	-0.109479	-0.030508
Omonoia ISAP Station	0	37.984100	23.728071	262	5	-0.025209	-0.030508
Aghios Dimitrios Metro Station	0	37.940089	23.740915	183	3	-0.136165	-0.163842
Kerameikos Metro Station	0	37.978550	23.711564	52	6	-0.320153	0.036158
Attiki Metro Station	0	37.999452	23.722549	176	2	-0.145996	-0.230508
Neos Kosmos Metro Station	0	37.957934	23.727630	89	7	-0.268187	0.102825
Larisis Metro Station	0	37.991851	23.720942	242	6	-0.053299	0.036158
Petralona ISAP Station	0	37.968341	23.709012	315	2	0.049229	-0.230508
Egaleo Metro Station	0	37.991428	23.681813	115	6	-0.231670	0.036158
Attiki ISAP Station	0	37.999735	23.722709	148	2	-0.185322	-0.230508
Sepolia Metro Station	0	38.002892	23.713120	17	4	-0.369311	-0.097175
Eleonas Metro Station	0	37.987818	23.694272	248	1	-0.044872	-0.297175

Cluster 2 [Cluster Label 1] – LOW potential (color Red on the map)

There are already many existing gyms in the area and the nearest gym is in most cases in a relatively short distance from the station.

Examples:

	station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
0	Syntagma Metro Station	1	37.975235	23.735298	112	13	-0.235884	0.502825
2	Evangelismos Metro Station	1	37.976173	23.747163	251	16	-0.040659	0.702825
4	Panepistimio Metro Station	1	37.980004	23.732354	328	13	0.067487	0.502825
6	Sygrou-Fix Metro Station	1	37.964919	23.726618	185	8	-0.133356	0.169492
7	Panormou Metro Station	1	37.993298	23.764135	195	11	-0.119311	0.369492
14	Megaro Moussikis Metro Station	1	37.979014	23.753173	95	8	-0.259760	0.169492
20	Ampelokipi Metro Station	1	37.987439	23.757075	169	13	-0.155827	0.502825

Cluster 3 [Cluster Label 2] – HIGH potential (color Blue on the map)

There are not many already existing gyms in the area and the nearest gym is in most cases relatively not in a short distance to the metro station

Examples:

	station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
3	Monastiraki Metro Station	2	37.976067	23.725752	709	2	0.602600	-0.230508
5	Omonia Metro Station	2	37.984602	23.726186	424	4	0.202319	-0.097175
8	Monastiraki ISAP Station	2	37.976044	23.725204	705	1	0.596982	-0.297175
10	Thisseio ISAP Station	2	37.976166	23.720449	459	1	0.251476	-0.297175
13	Metaxourghio Metro Station	2	37.985549	23.720734	394	5	0.160184	-0.030508
16	Aghios Ioannis Metro Station	2	37.958003	23.734743	360	6	0.112431	0.036158
17	Victoria ISAP Station	2	37.993158	23.729811	508	3	0.320296	-0.163842
24	Dafni Metro Station	2	37.948246	23.740441	455	6	0.245858	0.036158
26	Agios Nikolaos ISAP Station	2	38.006748	23.727649	351	5	0.099791	-0.030508
30	Katehaki Metro Station	2	37.992993	23.775953	486	5	0.289397	-0.030508
33	Agios Eleftherios ISAP Station	2	38.019715	23.731599	391	4	0.155970	-0.097175
42	Alimos Metro Station	2	37.918175	23.744169	436	6	0.219173	0.036158
43	Perissos ISAP Station	2	38.033124	23.744857	527	5	0.346982	-0.030508
50	Argyroupoli Metro Station	2	37.903425	23.746266	372	4	0.129285	-0.097175
51	Agia Paraskevi Metro Station	2	38.017153	23.812426	383	2	0.144734	-0.230508

5. Discussion

Clusters of areas (in our case Metro stations) were identified as groups of similar in their potential locations for opening a gym.

Possible areas that were not in the Foursquare database should also be examined so that it can be determined if it is just lack of data about these stations or indeed there are no gyms in the vicinity of the stations.

A lot more factors can be considered when choosing an appropriate location. Some examples of extra factors can be:

- Population density in the area
- Number of businesses operating in the area (people may want to go to a gym close to work)
- Average age and household income in the area
- Property prices in the area

6. Conclusion

The above results can be **a good starting point** for a prospective businessman that is interested in opening a gym. Similar methodology can be used for other types of businesses probably with customized criteria.

With the availability of a number of different tools and Machine Learning algorithms, it is possible to find solutions (or possible solutions) to an ever increasing number of problems and queries.

And it is getting better and better!