Finding suitable locations to open a Gym in Athens, Greece

M.Manikanta Reddy (April 2020)

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 3 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
11	Θησείο	37.977394	23.720287	1365	NaN	Metro Station
42	Αμαξοστάσιο Μετρό Ελαιώνα	37.986554	23.686795	4461	NaN	Metro Station
62	Αμαξοστάσιο Μετρό Δουκίσης Πλακεντίας	38.022099	23.835587	10185	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:
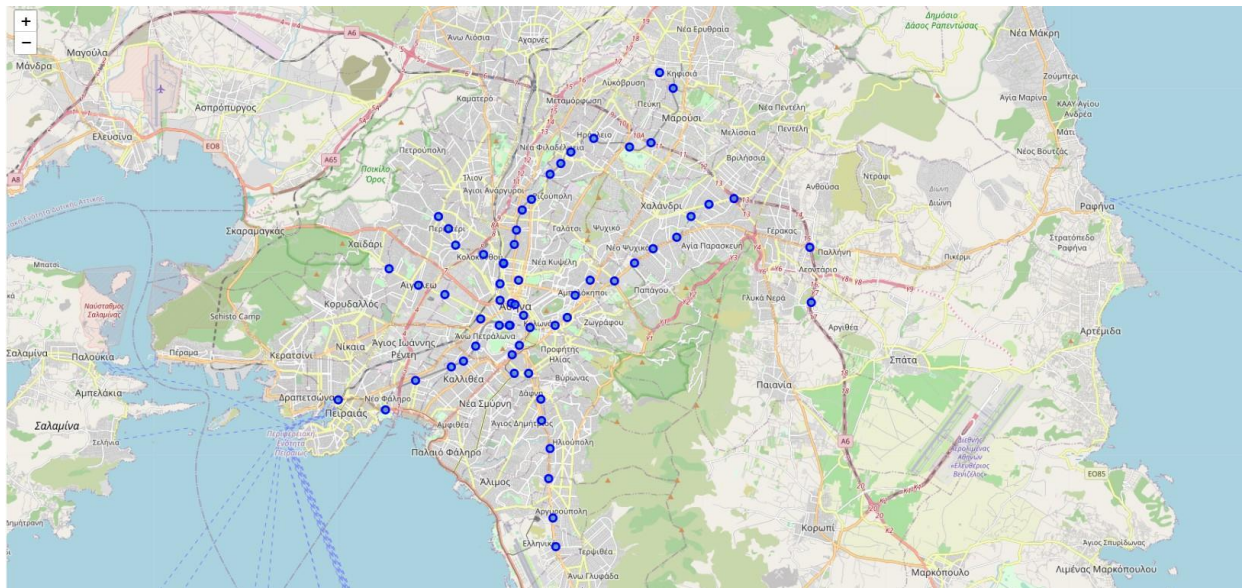
Venue Category	station	lat	lng	Venue	Venue Latitude	Venue Longitude	Distance from Station
Athletics & Sports	2	2	2	2	2	2	2
Basketball Court	4	4	4	4	4	4	4
Boxing Gym	3	3	3	3	3	3	3
Climbing Gym	3	3	3	3	3	3	3
Cycle Studio	1	1	1	1	1	1	1
Dance Studio	5	5	5	5	5	5	5
Gym	156	156	156	156	156	156	156
Gym / Fitness 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	94	94	94	94	94	94	94

- 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	15
33	Monastiraki Metro Station	37.976067	23.725752	709	2
35	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.
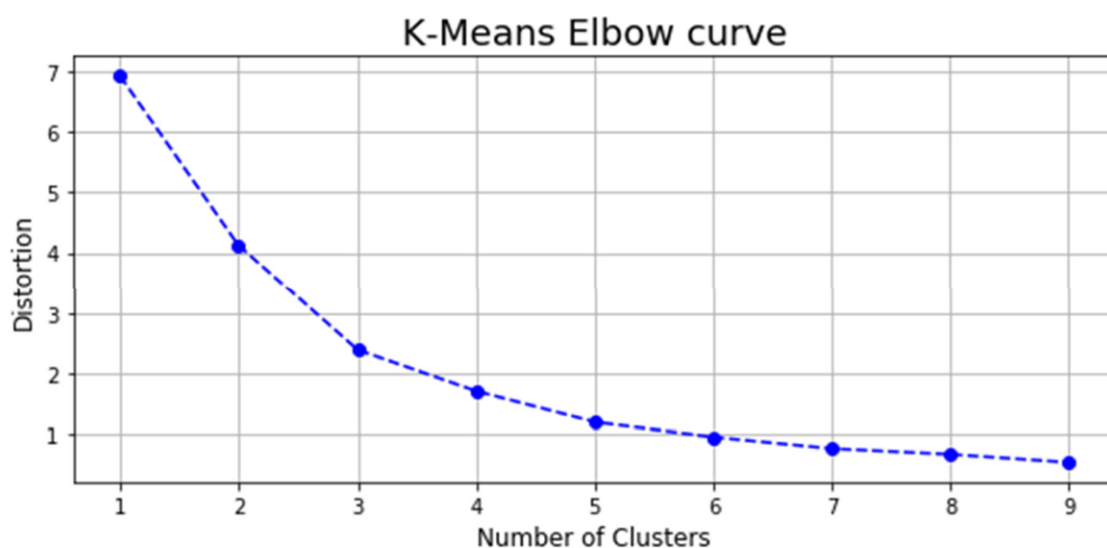

K-Means Elbow curve

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:
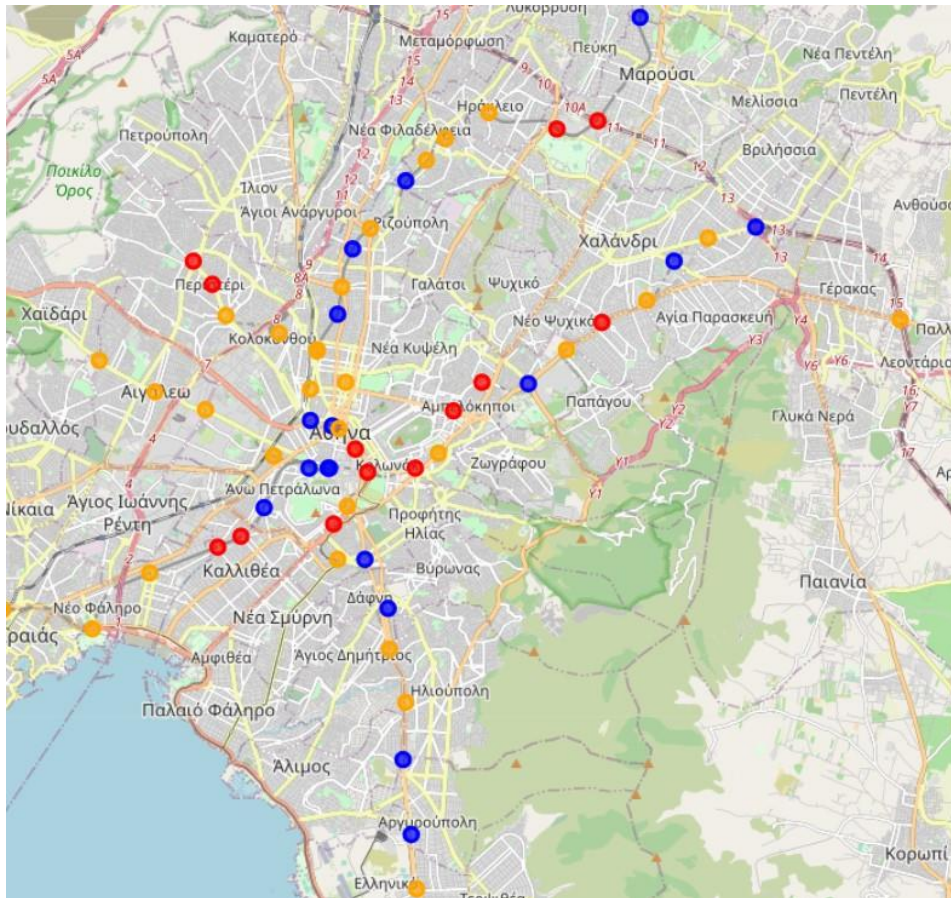
And a graph of the distribution of clusters and their final **centroids** (center points) in black, based on
normalized values of minimum distance and number of existing gyms.
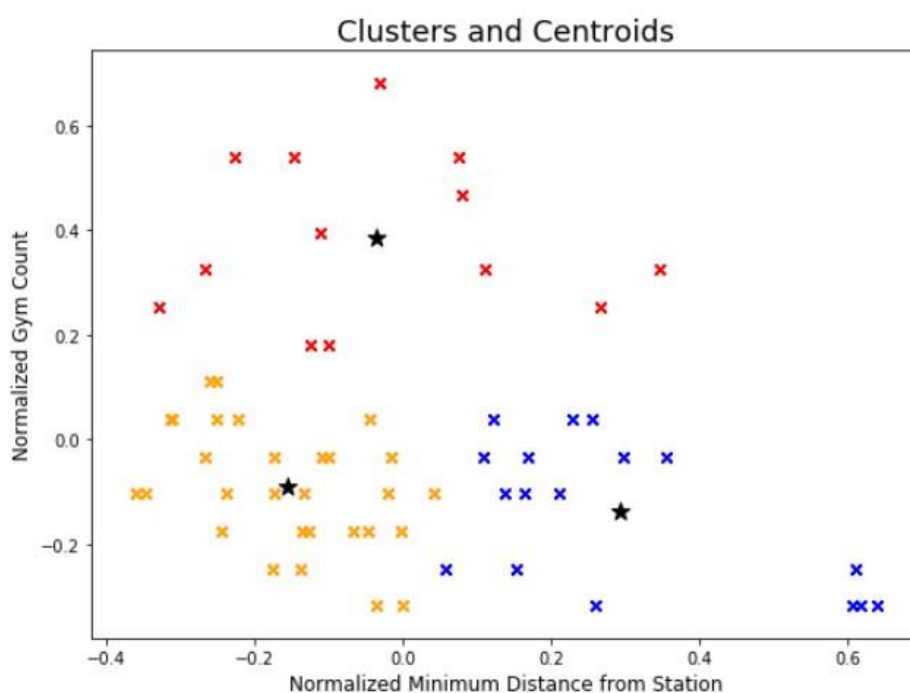
The three Metro stations clusters can be described as follows:

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
1	Akropoli Metro Station	0	37.968516	23.730195	202	5	-0.100957	-0.032688
9	Omonoia ISAP Station	0	37.984100	23.728071	262	5	-0.016687	-0.032688
11	Aghios Dimitrios Metro Station	0	37.940089	23.740915	183	3	-0.127642	-0.175545
12	Kerameikos Metro Station	0	37.978550	23.711564	52	6	-0.311631	0.038741
13	Megaro Moussikis Metro Station	0	37.979014	23.753173	95	7	-0.251238	0.110169
15	Attiki Metro Station	0	37.999452	23.722549	176	2	-0.137474	-0.246973
17	Victoria ISAP Station	0	37.993158	23.729811	150	4	-0.173991	-0.104116
18	Neos Kosmos Metro Station	0	37.957934	23.727630	89	7	-0.259665	0.110169
19	Larisis Metro Station	0	37.991851	23.720942	242	6	-0.044777	0.038741
22	Egaleo Metro Station	0	37.991428	23.681813	115	6	-0.223148	0.038741
23	Attiki ISAP Station	0	37.999735	23.722709	148	2	-0.176800	-0.246973
27	Sepolia Metro Station	0	38.002892	23.713120	17	4	-0.360788	-0.104116
29	Eleonas Metro Station	0	37.987818	23.694272	248	1	-0.036350	-0.318402
31	Kato Patisia ISAP Station	0	38.012168	23.728596	27	4	-0.346743	-0.104116
32	Aghios Antonios Metro Station	0	38.006380	23.699421	304	4	0.042301	-0.104116
34	Ethniki Amyna Metro Station	0	37.999722	23.785535	84	5	-0.266687	-0.032688
35	Ilioupoli Metro Station	0	37.929474	23.744744	179	4	-0.133260	-0.104116
36	Moschato ISAP Station	0	37.955221	23.680209	272	3	-0.002642	-0.175545
37	Ano Patisia ISAP Station	0	38.023857	23.735795	104	4	-0.238597	-0.104116
43	Agia Marina Metro Station	0	37.997628	23.667484	226	3	-0.067249	-0.175545
44	Halandri Metro Station	0	38.021967	23.820975	273	1	-0.001238	-0.318402
45	Pefkakia ISAP Station	0	38.037224	23.749974	50	6	-0.314440	0.038741
46	Neo Faliro ISAP Station	0	37.944256	23.666055	241	3	-0.046182	-0.175545
47	Nomismatokopio Metro Station	0	38.009402	23.805635	95	6	-0.251238	0.038741
48	Nea Ionia ISAP Station	0	38.041531	23.755007	196	5	-0.109384	-0.032688
49	Piraeus ISAP Station	0	37.948087	23.643211	178	3	-0.134665	-0.175545
52	Irakleio ISAP Station	0	38.046674	23.765953	151	5	-0.172586	-0.032688
55	Elliniko Metro Station	0	37.892466	23.747613	99	3	-0.245620	-0.175545
57	Pallini Metro Station	0	38.005699	23.869600	259	4	-0.020901	-0.104116

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	-0.227361	0.538741
2	Evangelismos Metro Station	1	37.976173	23.747163	251	15	-0.032137	0.681598
4	Panepistimio Metro Station	1	37.980004	23.732354	328	13	0.076009	0.538741
6	Sygrou-Fix Metro Station	1	37.964919	23.726618	185	8	-0.124833	0.181598
7	Panormou Metro Station	1	37.993298	23.764135	195	11	-0.110788	0.395884
20	Ampelokipi Metro Station	1	37.987439	23.757075	169	13	-0.147305	0.538741
25	Tavros ISAP Station	1	37.962494	23.703367	330	12	0.078818	0.467312
28	Kallithea ISAP Station	1	37.960339	23.697373	39	9	-0.329890	0.253027
38	Peristeri Metro Station	1	38.012749	23.696161	202	8	-0.100957	0.181598
39	Anthoupoli Metro Station	1	38.017143	23.691265	353	10	0.111122	0.324455
40	Holargos Metro Station	1	38.005225	23.794296	84	10	-0.266687	0.324455
53	Eirini ISAP Station	1	38.043475	23.783093	463	9	0.265616	0.253027
54	Neratziotissa ISAP Station	1	38.045224	23.793136	521	10	0.347077	0.324455

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	0.611122	-0.246973
5	Omonia Metro Station	2	37.984602	23.726186	424	4	0.210841	-0.104116
8	Monastiraki ISAP Station	2	37.976044	23.725204	705	1	0.605504	-0.318402
10	Thisseio ISAP Station	2	37.976166	23.720449	459	1	0.259998	-0.318402
14	Metaxourghio Metro Station	2	37.985549	23.720734	394	5	0.168706	-0.032688
16	Aghios Ioannis Metro Station	2	37.958003	23.734743	360	6	0.120953	0.038741
21	Petralona ISAP Station	2	37.968341	23.709012	315	2	0.057751	-0.246973
24	Dafni Metro Station	2	37.948246	23.740441	455	6	0.254380	0.038741
26	Agios Nikolaos ISAP Station	2	38.006748	23.727649	351	5	0.108313	-0.032688
30	Katehaki Metro Station	2	37.992993	23.775953	486	5	0.297919	-0.032688
33	Agios Eleftherios ISAP Station	2	38.019715	23.731599	391	4	0.164492	-0.104116
41	Alimos Metro Station	2	37.918175	23.744169	436	6	0.227695	0.038741
42	Perissos ISAP Station	2	38.033124	23.744857	527	5	0.355504	-0.032688
50	Argyroupoli Metro Station	2	37.903425	23.746266	372	4	0.137807	-0.104116
51	Agia Paraskevi Metro Station	2	38.017153	23.812426	383	2	0.153257	-0.246973
56	Doukissis Plakentias Metro Station	2	38.024105	23.833275	729	1	0.639212	-0.318402
58	KAT ISAP Station	2	38.065601	23.803997	714	1	0.618144	-0.318402

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!