Finding suitable locations to open a Gym in Athens, Greece

Vasilis Kokkinos (June 2019)

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, entrepreneur 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 will used, along with the Foursquare API. Also, in order to create clusters of similar candidate locations, the K-Means machine learning clustering algorithm will be 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** will be used:

1. *'Syntagma square'* will be considered as the center of Athens. It is indeed one of the most central location in the city. I will obtain its geospatial coordinates using the geopy library
2. Having the coordinates of the 'center' of Athens, the Foursquare API will be 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 will again be used for every station. I will obtain data for all the gyms located in a radius of 750 meters of every metro station

Using the collected data, I will calculate the number of existing gyms near each station. **I will also be 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 will be used as input to K-Means clustering algorithm to obtain the clusters of areas (metro stations).