

# Prospective Areas to Set-up a Gym in New York

## ❖ Introduction / Business Problem

The aim of this project is to find suitable locations to open a gym in New York, USA. One of the main requirements to open a new gym is that it should be easily accessible by customers and it should be located near a metro station. As New York is known as city of dreams, it is very important to eliminate any kind of competition in order to flourish the business. So, we also look into the number of gyms already existing in an area before setting up Gym. This project can be utilized by any stakeholders/entrepreneur who would want to set up their business in a new city. It can serve as stepping stone to start their business. To achieve the above mentioned tasks, python geolocation libraries will be 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.

## ❖ Data

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

- The dataset for New York, which consists of 5 boroughs and 306 neighborhoods, is readily available on the web and the link is: [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572). However, as we already have this data downloaded from our previous lab session (JSON file) [https://cocl.us/new\\_york\\_dataset/n](https://cocl.us/new_york_dataset/n), we will be using the same.
- The metro stations in New York City 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:

- New York City will be considered as the main center as it is indeed one of the most central location in the city. Its geospatial coordinates are obtained using the geopy library.
- Having the coordinates of the New York, the Foursquare API will be used to retrieve data for all the metro stations in this area in a radius of 15 km.
- Data for all the gyms located in a radius of 750 meters of every metro station are obtained using the Foursquare API.

Using the collected data, the number of existing gyms near each station are calculated and also the minimum distance of a gym for every metro station is determined. 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).

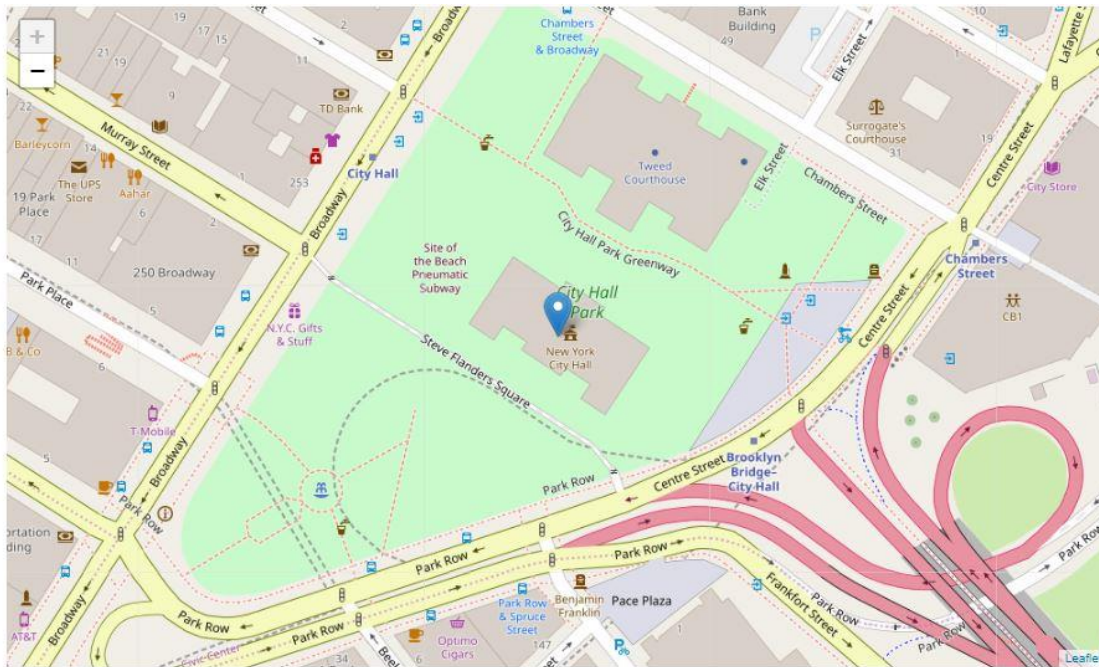
## ❖ Methodology

The objective of this project is to find an optimum place to set up a gym that is close to a metro station and having following criteria:

- Number of existing gyms near that metro station must be **less**.
- Distance from the metro station must be **less**.

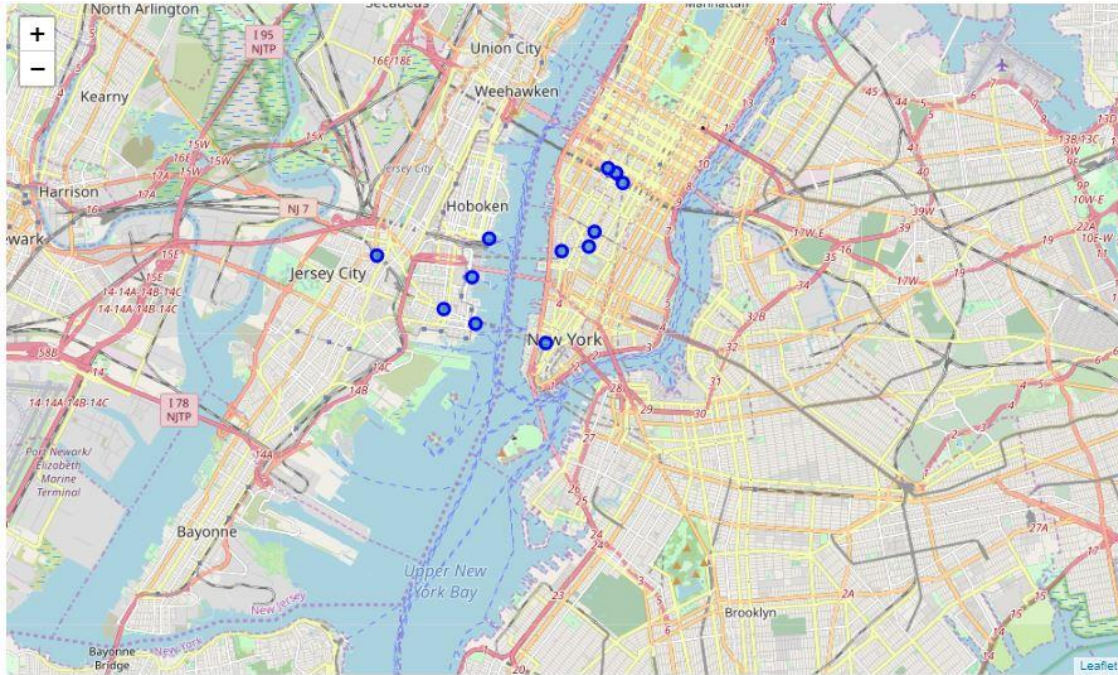
The steps to identify potential areas (metro stations) are:

- Acquiring the latitude and longitude geospatial coordinates of New York City.



**Fig 1 : Map of New York City**

- Based on the coordinates of New York City, obtain the information about metro stations in a radius of 15 km using the Foursquare API. Here, we retained only the most prominent stations (**stations that has the word ‘station’ in its name**) because not all trains halts at every station.



**Fig 2: Map of New York showcasing Metro stations**

- After the collecting information on metro stations information, the next step is to find the gyms around these stations. This was achieved by 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:

	station	lat	lng	Venue	Venue Latitude	Venue Longitude	Distance from Station
Venue Category							
Advertising Agency	1	1	1	1	1	1	1
Boxing Gym	12	12	12	12	12	12	12
Building	3	3	3	3	3	3	3
Construction & Landscaping	1	1	1	1	1	1	1
Cycle Studio	9	9	9	9	9	9	9
Dance Studio	3	3	3	3	3	3	3
Gym	220	220	220	220	220	220	220
Gym / Fitness Center	430	430	430	430	430	430	430
Gym Pool	2	2	2	2	2	2	2
Martial Arts Dojo	34	34	34	34	34	34	34
Medical Center	6	6	6	6	6	6	6
Park	1	1	1	1	1	1	1
Pilates Studio	13	13	13	13	13	13	13
Pool	1	1	1	1	1	1	1
Residential Building (Apartment / Condo)	2	2	2	2	2	2	2
Spa	1	1	1	1	1	1	1
Tennis Court	1	1	1	1	1	1	1
Track	4	4	4	4	4	4	4
Yoga Studio	80	80	80	80	80	80	80

**Fig 3: Venue Categories in New York**

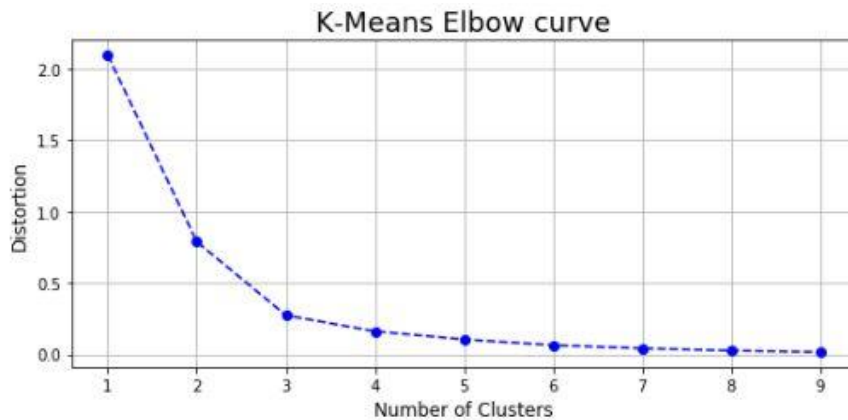
- Data that correspond only to ‘**Gym / Fitness Center**’ and ‘**Gym**’ subcategories are retained. Rest of the subcategories such as ‘Dance Studio’, ‘Yoga Studio’, ‘Martial Arts Dojo’ and so on are omitted.
- With the above information on Gym and metro stations, the number of existing gyms near each station as well as the minimum distance from each station to a gym are calculated 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	Venue	Venue Latitude	Venue Longitude	Distance from Station	Venue Category	Min Distance from Station	Gym Count
0	World Trade Center PATH Station	40.711807	-74.011687	The Class by Taryn Toomey	40.712753	-74.008734	270	Gym / Fitness Center	47	95
1	World Trade Center PATH Station	40.711807	-74.011687	Equinox Brookfield Place	40.712704	-74.014995	296	Gym / Fitness Center	47	95
2	World Trade Center PATH Station	40.711807	-74.011687	Crunch - FIDi	40.708614	-74.010013	382	Gym / Fitness Center	47	95
3	World Trade Center PATH Station	40.711807	-74.011687	Equinox Tribeca	40.714099	-74.009686	305	Gym / Fitness Center	47	95
4	World Trade Center PATH Station	40.711807	-74.011687	Exceed Physical Culture	40.715629	-74.007992	527	Gym / Fitness Center	47	95
5	World Trade Center PATH Station	40.711807	-74.011687	Physique 57	40.706844	-74.013014	563	Gym / Fitness Center	47	95
6	World Trade Center PATH Station	40.711807	-74.011687	Equinox Wall Street	40.707273	-74.010543	513	Gym	47	95

**Fig 4: Gym count and minimum distance from Metro Station**

- The data then 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.
- 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.



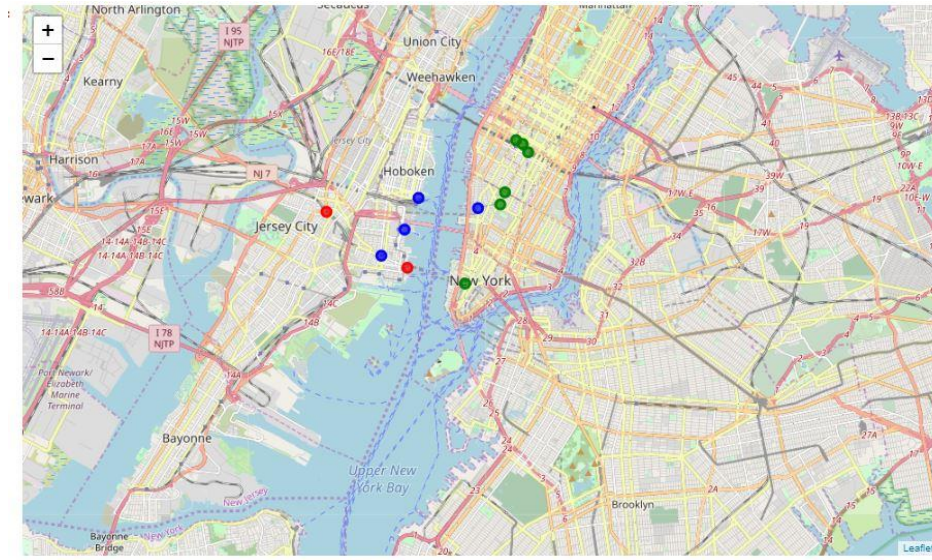
**Fig 5: Elbow Curve**



- 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**.

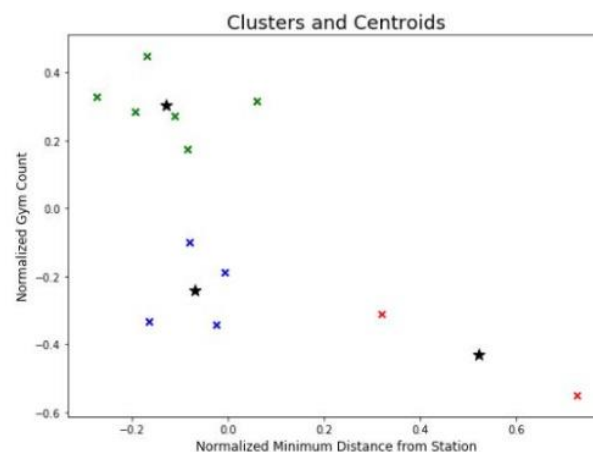
## ❖ Results

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



**Fig 6: Three Clusters of Metro Stations**

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.



**Fig 7: Clusters and Centroid**

The three Metro stations clusters can be described as follows:

➤ **Cluster 1 [Cluster Label 0] (Green color 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.

	station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
0	World Trade Center PATH Station	0	40.711807	-74.011687	47	95	-0.169683	0.448718
3	9th Street PATH Station	0	40.734119	-73.998916	66	70	-0.083710	0.173993
5	14th Street PATH Station	0	40.737438	-73.996931	42	80	-0.192308	0.283883
8	33rd St PATH Station	0	40.748613	-73.988417	60	79	-0.110860	0.272894
9	MTA Subway - 34th St/Penn Station (1/2/3)	0	40.750996	-73.990688	98	83	0.061086	0.316850
10	MTA Subway - 34th St/Penn Station (A/C/E)	0	40.751995	-73.992888	24	84	-0.273756	0.327839

**Fig 8: Clusters 1**

➤ **Cluster 2 [Cluster Label 1] (Red color 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.

	station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
2	Exchange Place PATH Station	1	40.716249	-74.033207	155	26	0.319005	-0.309524
11	Journal Square PATH Station	1	40.732037	-74.062827	245	4	0.726244	-0.551282

**Fig 9: Clusters 2**

➤ **Cluster 3 [Cluster Label 2] (Blue color 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.

	station	Cluster Label	lat	lng	Min Distance from Station	Gym Count	Norm Min Distance from Station	Norm Gym Count
1	Christopher Street PATH Station	2	40.732937	-74.006905	83	37	-0.006787	-0.188645
4	Newport PATH Station	2	40.726838	-74.034201	48	24	-0.165158	-0.331502
6	Grove Street PATH Station	2	40.719785	-74.042707	67	45	-0.079186	-0.100733
7	Hoboken PATH Station	2	40.735691	-74.029110	79	23	-0.024887	-0.342491

**Fig 10: Clusters 3**

## ❖ 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

## ❖ Conclusion

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