Coursera - Applied Data Science Capstone Project

Best GYM in Manhattan NY

-Mahesh Goriwale

Coursera Capstone - REPORT CONTENT

1. Introduction Section :

Scenario:

Being a fitness freak, my first priority after shifting to a new place, is to find a good GYM.

So i will use this opportunity provided by coursera, to compare best GYM around Manhattan NY.

In order to make a comparison and evaluation of all fitness centers, i would set some restrictions, they are:

- Should be very popular with foursquare users.

- Should be in one of the neighborhood of Manhattan

Business Problem:

The challenge is to find a good GYM in Manhattan NY that aligns with the demands on rating, popularity and disagree percentage. The data required to resolve this challenge is described in the following section 2, below.

Interested Audience

I believe this is a relevant challenge with valid questions for anyone moving to other large city in US, EU or Asia and wanting to find good fitness centers. The same methodology can be applied in accordance to demands as applicable. This case is also applicable for anyone interested in exploring starting or locating a new business in any city. Lastly, it can also serve as a good practical exercise to develop Data Science skills.

2. Data Section:

Description of the Data:

The following data is required to answer the issues of the problem:

- Latitude and Longitutde of Manhattan NY.
- List of Fitness Centers around Manhattan NY.
- List of Fitness Centers with ratings, popularity count and agree/disagree count.
- NY Neighborhhod data
- Venues for each Fitness Gym (NY Neighborhood which can be clustered)

Source of the Data:

- Foursquare API.
- New York Neighborhood dataset.

How the data will be used to solve the problem?

The data will be used as follows:

- Use Foursquare to find top 10 venues for Manhattan NY,
- Use Foursquare and geopy data to map the location of Fitness Centers.
- Use Folium to create NY neighborhoods and clustered in groups (as per Course LAB)
- create a map that depicts, clusters of neighborhood and fitness gym in those area.

The procesing of these DATA will allow to answer the key questions to make a decision:

- which is the most popular Fitness Center in the neighborhood?
- What is the distance of Fitness Center to possible home in NY?
- How many Foursquare users agree with the review given for the Fitness Center?

3. Methodology section:

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

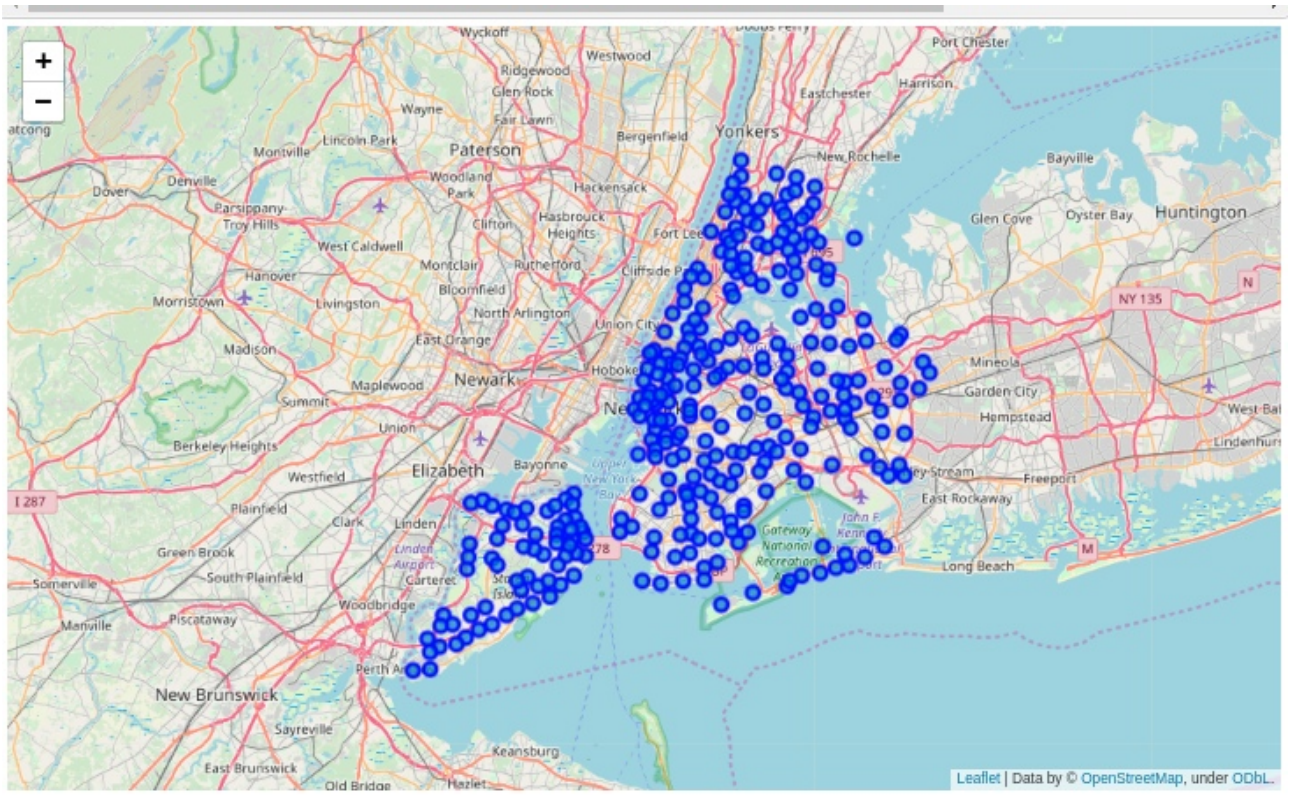
The Analysis and The Stragegy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places to join. The choice is made based on the demands imposed : distance near a propsed location, ratings and popularity amongst Foursquare users. This visual approach and maps with popups labels allow quick identification of location, name and ratings, thus making the selection very easy.

The Data Preparation Phase:

We follow below steps:

- Use the Manhattan data from previous course with Neighborhood and Borought.
- Extract venues from Foursquare API for all gyms near Manhattan NY.
- Filter venues having categories as GYM / Fitness Center.
- Search for venue with higher ratings.
- Calculate the percentage of users who agree with the review provided by the Foursquare user.

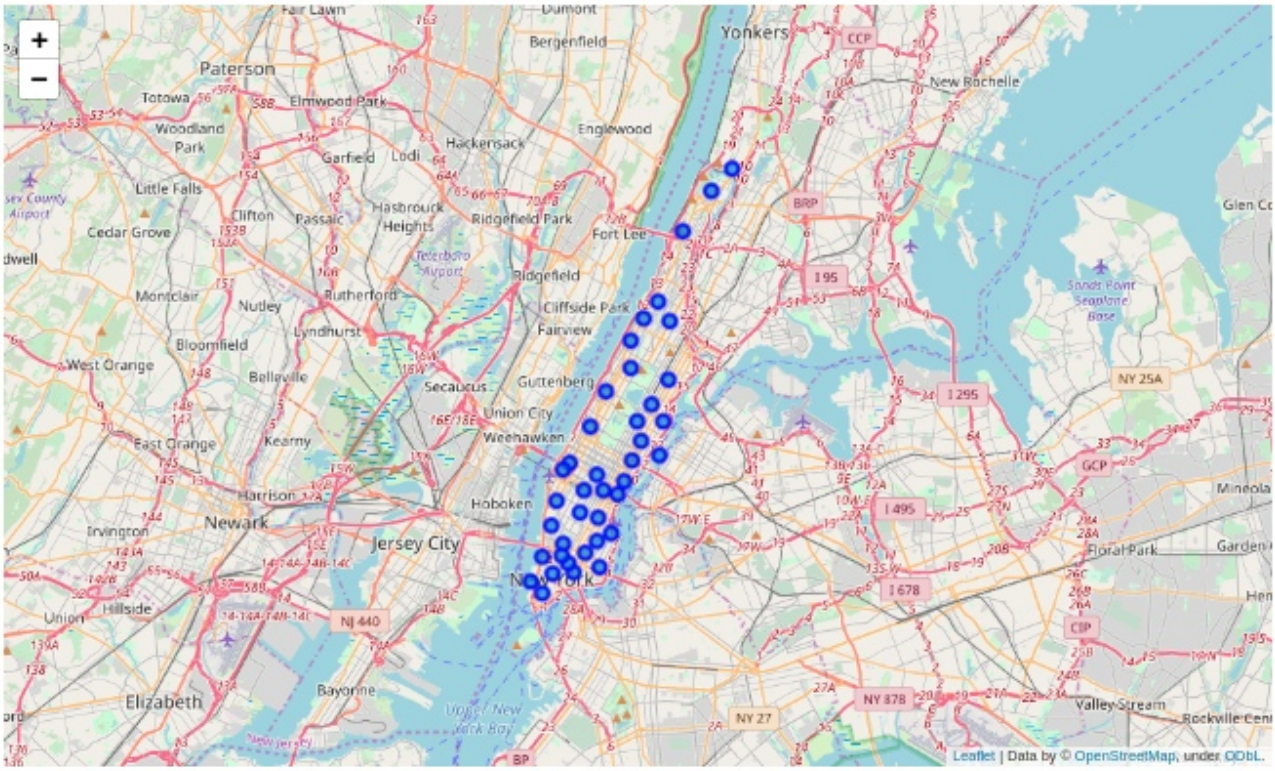


The Machine Learning and Visualization Phase:

We will create a cluster using kmeans for all neighborhood in Manhattan NY and also plot these cluster in map.

This would help us in visualizing different clusters in Manhattan NY.

Using both maps creating using kmeans and Foursquare api we can visualize in which cluster the top gym/fitness center belongs.



4. Results

Let's consolidate all the required inforamtion to make the gym / fitness center selection in one map

Map of Manhattan with names, ratings, distance and cluster of neighborhood.

Legends:

Cluster1: Red

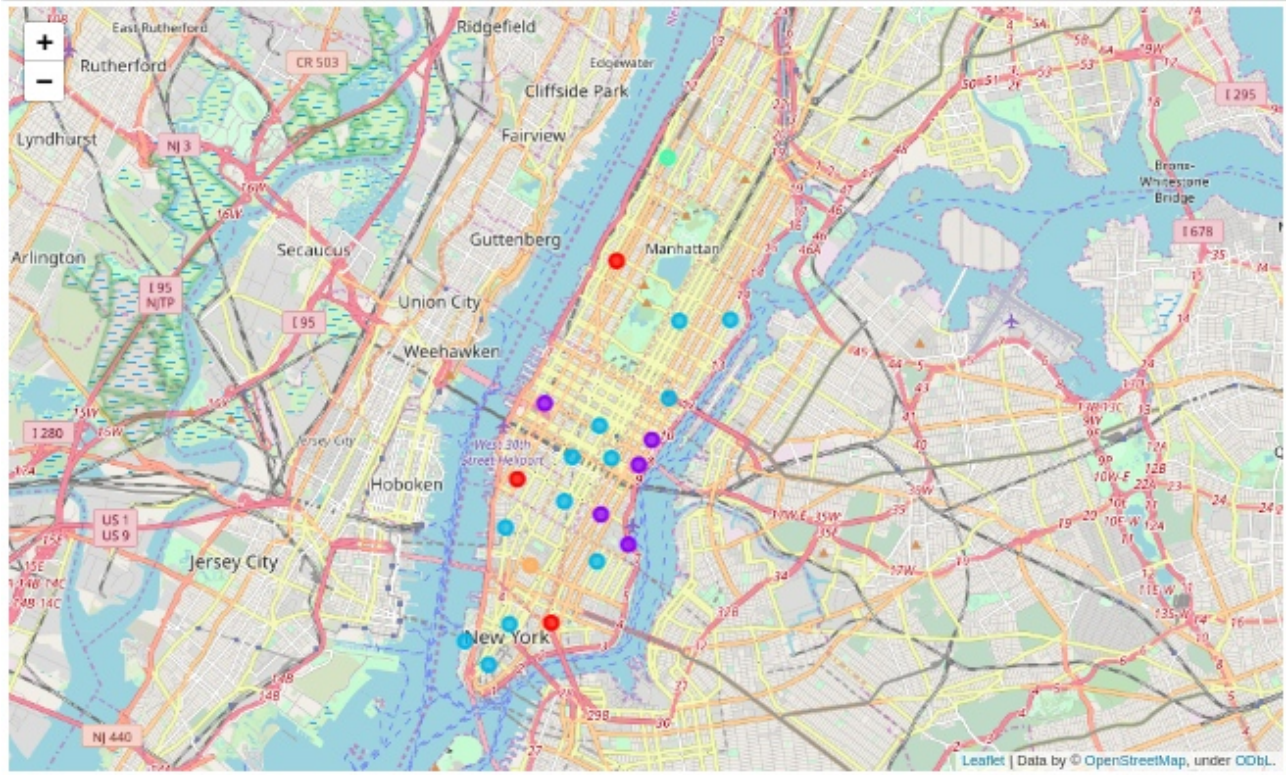
Cluster2: Purple

Cluster3: Blue

Cluster4: Teal

Cluster5: Orange

One Consodilated Map:



Selection

Using the "one map" above, I was able to explore all gyms across since the popups provide the information needed for a good decision.I was able to decide a good and popular gym.

5. DISCUSSION

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course

I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.

I have created a good project that I can present as an example to show my potential.

I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

6. Conclusion

I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.

This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.

The Foursquare API is also a great tool which can be used to extract data for exploring our neighborhood which is very important for anyone moved to a new location.

One must keep abreast of new tools for DS that continue to appear for application in several business fields.