Analysis of Manhattan Neighborhoods –

Finding Optimal Location for opening a new Gym/Yoga Studio

1. Introduction

1.1. Background

Manhattan is the most densely populated of the five boroughs of New York City where over 16 million people live and it is also economic and administrative center of New York. It is 59 square km and is divided into 40 neighborhoods.

A large number of fitness centers are located in Manhattan – more than 95 fitness centers are present only within 1 km of each neighborhood center. Being a city of so many neighborhoods crowded with numerous fitness centers and people, it is difficult to decide the location for opening any new fitness center. The investors/businessmen prefer that neighborhood where there is the type of business they want to install is less intense and also within the reach of their client base.

1.2. Problem

The problem that needs to be solved here is - to find an optimal location for opening a new fitness center (Gym/Yoga Studio) in Manhattan, New York, USA.

In order to determine this optimal place, this project focuses on following data -

- To find locations with less crowded fitness centers (in comparison with other locations).
- To be particularly interested in areas with no Gym/Yoga Studio in vicinity.
- To prefer locations as close to city center as possible, provided that above two conditions are met.

1.3. Interest

Specifically, this project targets the stakeholders that are interested in opening either new gym or yoga studio. Investors would be interested in knowing the locations for fitness centers in Manhattan with less competition to make their business profitable.

Also, this project might help to determine which Manhattan population is more aware of the importance of fitness based on the density of fitness centers in neighborhoods, since more the fitness centers in area implies more customers are available there.

2. Data Acquisition and Cleaning

2.1. Data Sources

Based on definition of the problem, the factors influencing decision are:

- Number of existing fitness centers (of any type) in the neighborhood
- Number of Gyms/Yoga Studios already existing in the neighborhood and their distance from neighborhood center (if any)
- Distance of neighborhood from city center

This project uses all the 40 neighborhoods of Manhattan city located around the city center Times Square. Following data sources are used to extract/generate the required information:

- Neighborhoods of Manhattan are obtained in the form of JSON file from link - https://cocl.us/new_york_dataset
- Number of fitness centers, their type and location in every neighborhood are obtained using Foursquare API
- Coordinates of Manhattan center (Times Square – well known location) and distance from neighborhoods are obtained using Geocoders (of GeoPy library)

2.2. Data Cleaning

Data in the form of JSON is downloaded from above mentioned link and then only required fields such as borough, neighborhood, latitude and longitude are stored into 'pandas' dataframe. This downloaded data contains all the boroughs of New York City along with their neighborhoods. Only neighborhoods belonging to Manhattan and their geographical coordinates are retained and others are discarded.

Using 'geopy.distance' library, distances of all the neighborhoods from Times Square (city center) are calculated using geographical coordinates. The neighborhoods with the distance more than 9km - Washington Heights, Inwood, Marble Hill are ignored.

Another dataframe is populated with venues (up to 100) for each neighborhood present in Manhattan by calling Foursquare API with specific parameters. Only those venues belonging to 'Gym/Fitness center' category and located within 1km of neighborhood center are returned. This dataframe includes neighborhood name, neighborhood latitude and longitude, venue name, venue category id and name, venue address, venue distance and venue geographical coordinates. Since there are no NaN/ inappropriate values present, no further cleaning of data is required.

Total 3113 venues are returned for 37 neighborhoods of Manhattan. Among these venues, there are some that are not proper fitness centers such as Weight Loss Center,

Playground, Basketball court etc. They are not direct competitors so they need to be ignored. Hence only venues that have category name related to proper fitness centers are included, and the venues with specific category as Gym or Yoga Studio are determined and included. For that, two dictionaries are maintained – one with all proper fitness centers and the other one with all gyms/yoga studios. All these locations are plotted on Manhattan map to conclude data gathering phase.

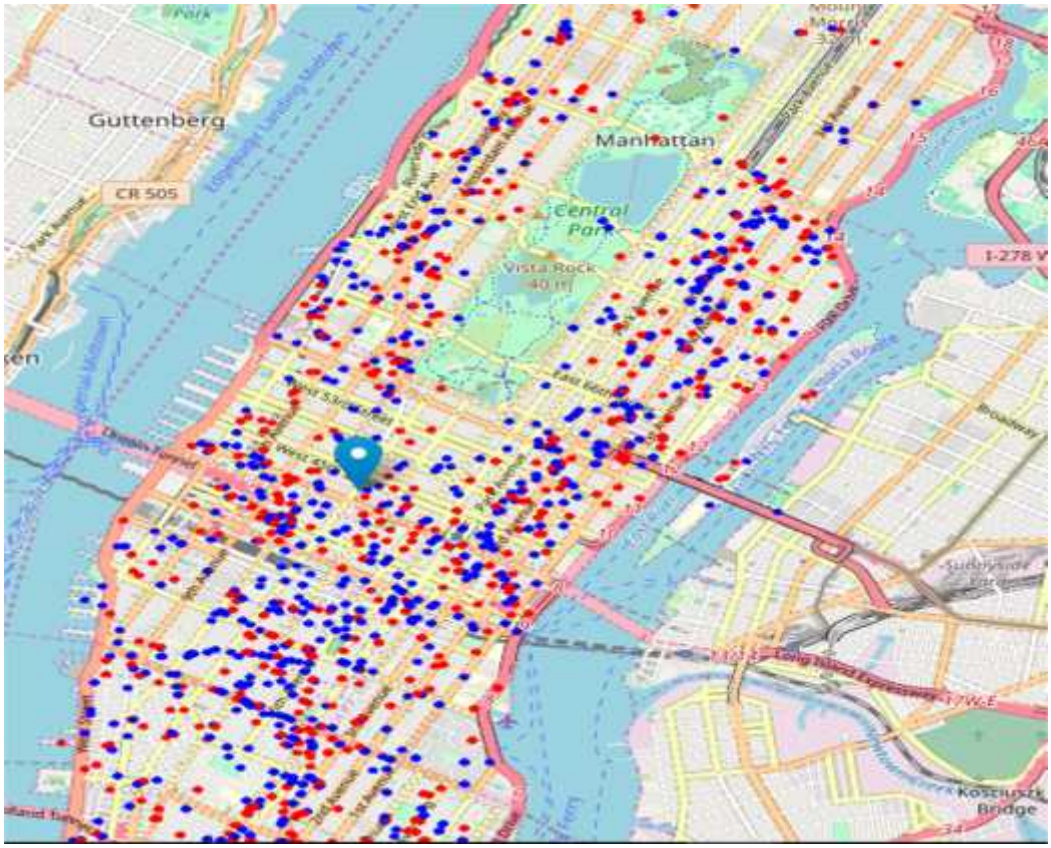

Figure 1 – Manhattan map with Fitness Centers (shown by blue circles) and Gym/Yoga Studio (shown by red circles).

3. Methodology

The focus is to detect areas of Manhattan that have low fitness center density, particularly those with low number of Gym/Yoga Studios. Here, the analysis is limited to area ~9km around city center.

It includes following steps –

i. Collecting required data which is :
   - Location and category of every fitness center, and
   - Determining Gym/Yoga Studio with the help of Foursquare categorization.

This is already done above.

ii. Calculation and exploration of 'fitness centers density' across different areas of Manhattan using heat maps in order to identify a few promising areas close to center with less number of fitness centers in general (and no Gym/Yoga Studio in vicinity)

iii. Finding most promising areas and creating clusters of the locations that meet some basic requirements :

- No more than 95 fitness centers within radius of 1 km, and
- Locations without Gym/Yoga Studios in radius of 150 meters.

iv. Plotting these locations on map and creating their clusters using k-means to identify neighborhoods/addresses for optimal venue location.

3.1. Exploratory Data Analysis

3.1.1. Density Distribution of all Fitness Centers

The density distribution of all fitness centers located in Manhattan neighborhoods is visualized with heat map (using Folium). A heat map is a graphical representation of data that uses a system of color-coding to represent different values.

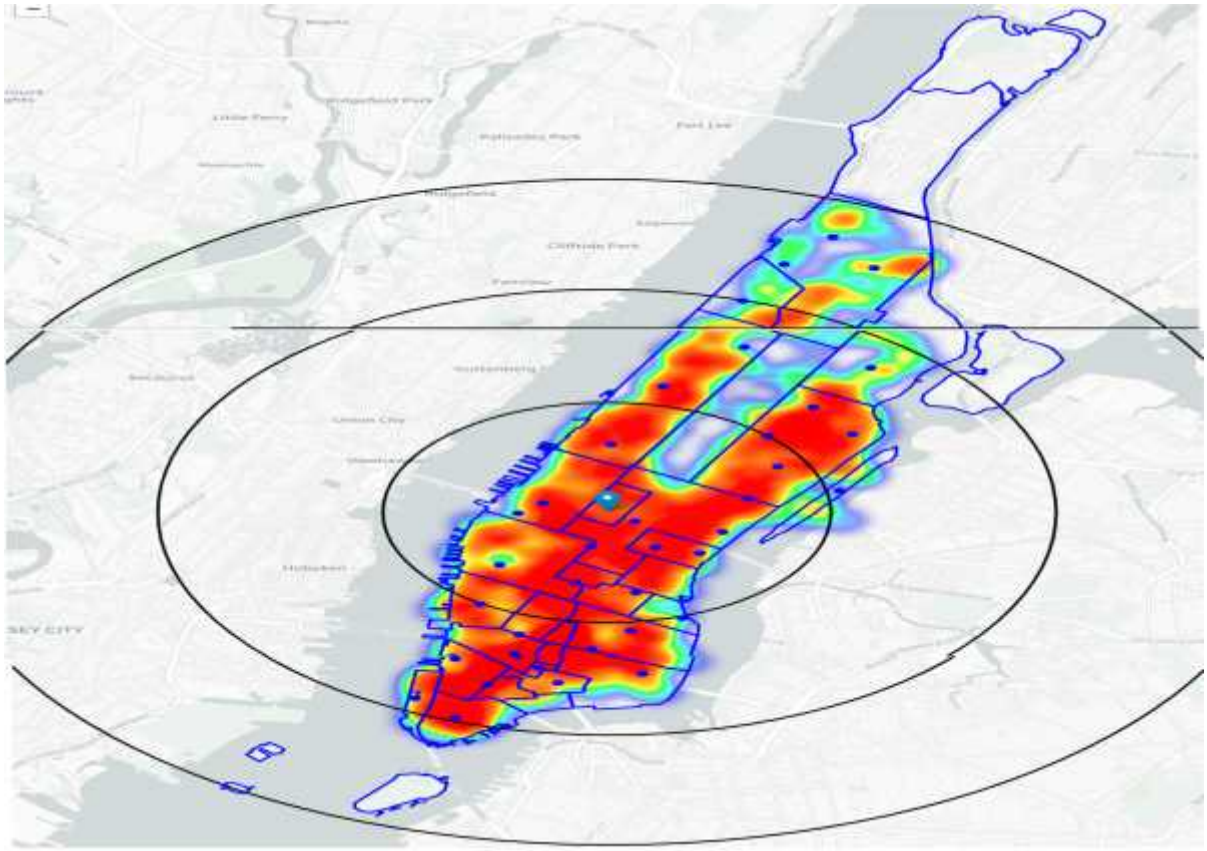

Figure 2 – Heat Map showing density distribution of fitness centers present in Manhattan

Above heat map shows a few pockets of low density fitness centers in north-east, east and south-east from Times Square (within 9 km).

3.1.2. Density Distribution of all Gym/Yoga Studios

The density distribution of all Gym/Yoga Studios located in Manhattan neighborhoods is visualized with heat map (using Folium).

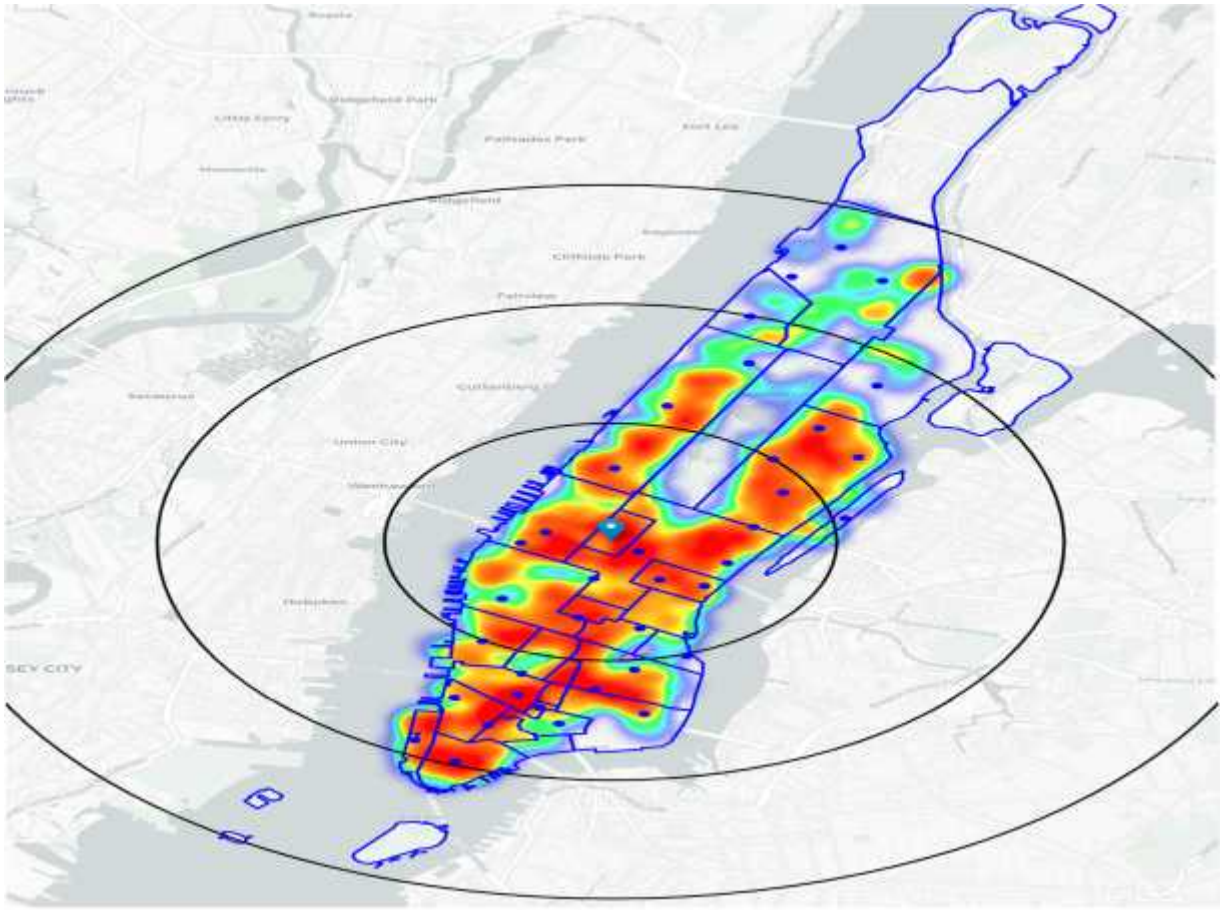

Figure 3 – Heat Map showing density distribution of Gym/Yoga Studios present in Manhattan

Above heat map is less 'hotter' than previous one, since the Gyms/Yoga Studios represent a subset of around 43% of all fitness centers in Manhattan. It also indicates higher density of Gym/Yoga Studios in south and west from Times Square whereas low density in north-east, east and south-east from city center.

3.1.3. Finding Promising Locations

For each neighborhood, total number of proper fitness centers (within 1 km) and distance to the closest Gym/Yoga Studio from neighborhood center are calculated. These neighborhoods are then filtered based on total number of fitness centers present (<=95) and distance to the closest Gym/Yoga Studio (>150m). Neighborhoods satisfying these two criteria are considered as good/optimal locations.

	Borough	Neighborhood	Latitude	Longitude	Distance from Center	Number of Fitness Centers	Distance to Gym/Yoga Studio
0	Manhattan	Chinatown	40.715618	-73.994279	4.627710	93	280.931123
1	Manhattan	Hamilton Heights	40.823604	-73.949688	8.026093	14	239.330304
2	Manhattan	Manhattanville	40.816934	-73.957385	7.100469	14	562.120618
3	Manhattan	Central Harlem	40.815976	-73.943211	7.498261	28	165.509915
4	Manhattan	East Harlem	40.792249	-73.944182	5.288173	28	343.028462

Figure 4 – Table with optimal locations with total count of fitness centers present and distance to closest gym/yoga studio.

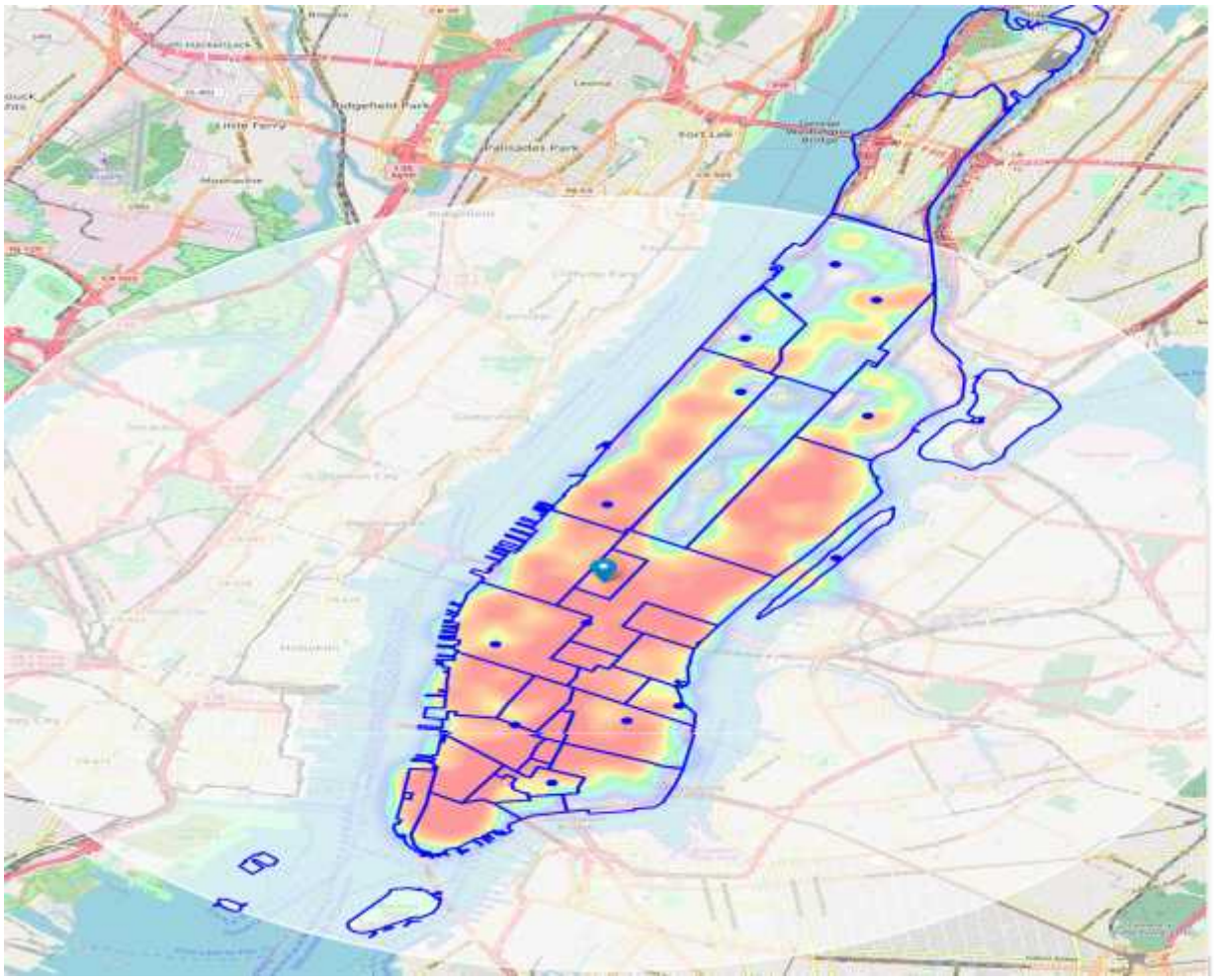

Figure 5 – Heat Map showing density distribution of fitness centers present in Manhattan along with optimal locations (shown by blue circles) within 9km from Times Square (shown by white circle)

A bunch of neighborhoods (blue circles) close to Times Square that satisfy the criteria of number of fitness centers and distance from existing Gym/Yoga Studio are obtained through this analysis. Any of these locations is a potential candidate for a new Gym/Yoga Studio, at least based on nearby competition.

3.2. Clustering

Clustering means finding clusters in dataset where cluster is group of objects similar to other objects in cluster and dissimilar to the data points in other clusters. **K-means** clustering is partitioning clustering and it is relatively efficient than the other algorithms. Here, K-means tries to group locations based on Euclidean distance between them and returns clusters of locations that are close to each other.

To apply K-means clustering, it is required to find optimal value of K i.e. number of clusters. It is determined using elbow method that includes plotting the sum of squared distances against the number of clusters.

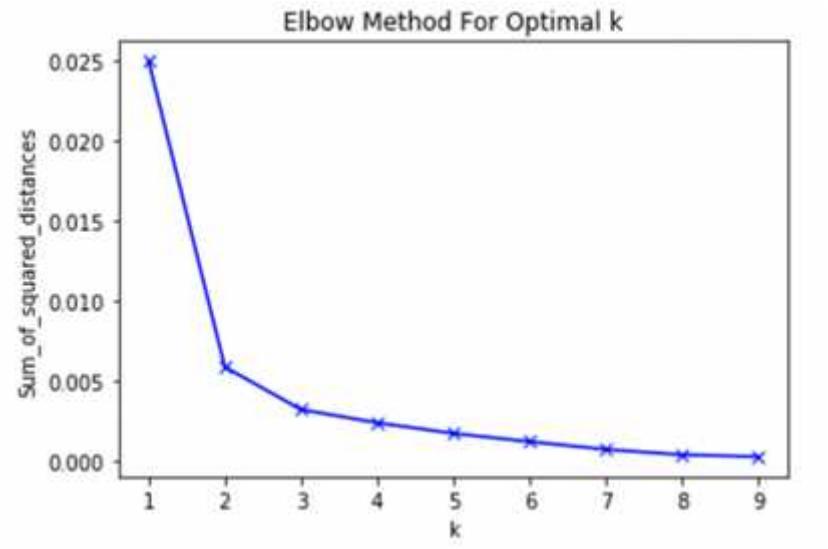

Figure 6 – Plot of sum of squared distances vs. k showing elbow at k=5

This elbow method gives optimal value of K as 5. Applying K-means clustering with K=5, cluster centers are returned which are the centers of zones containing good locations. These zones, their centers and addresses represent final result of analysis.

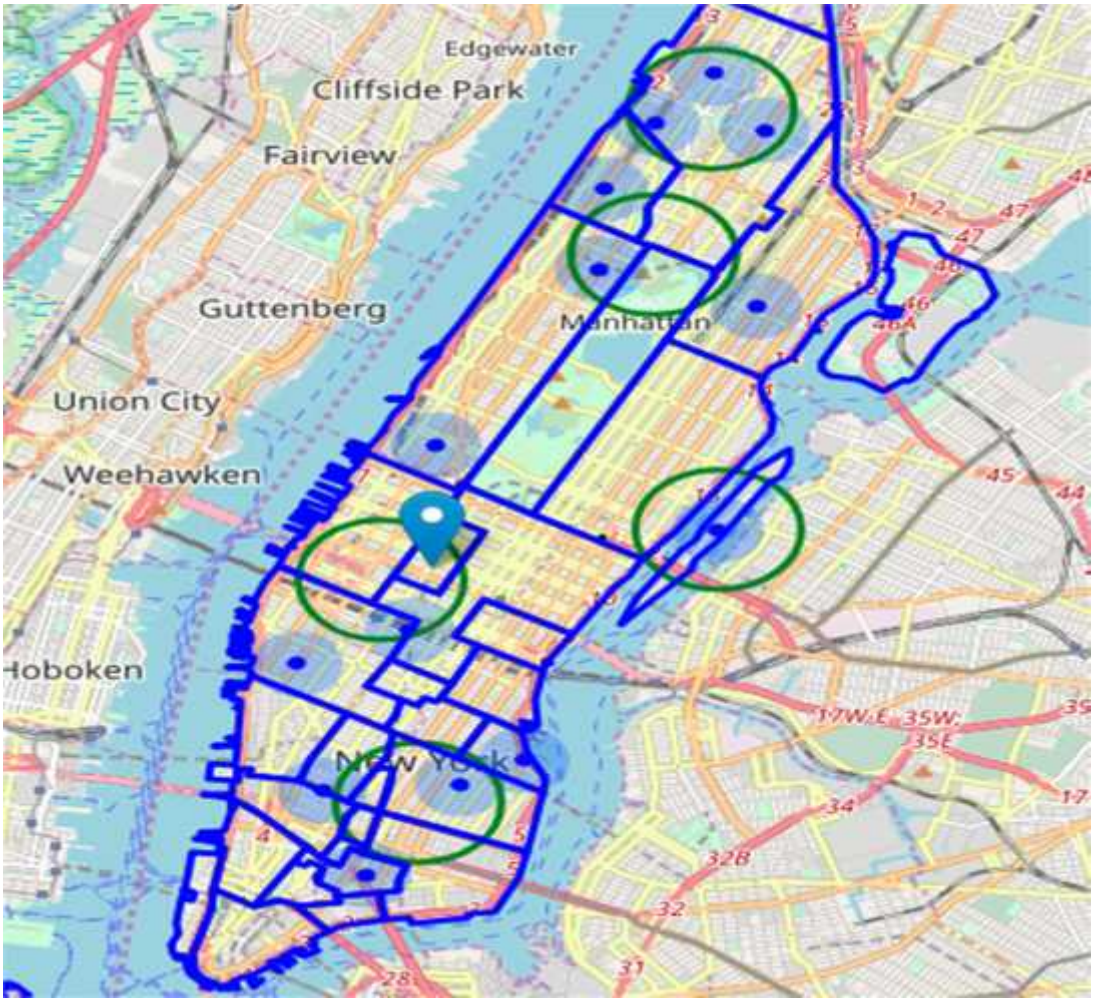

Figure 7 – Manhattan Map showing clusters (with green circles) and optimal neighborhoods (with blue circles)

The clusters represent groupings of most of the candidate locations and cluster centers are placed in the middle of these location candidates. Although clusters are shown on map with a radius of ~900 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search of potential gym/yoga studio locations.

Addresses of these cluster centers help to find the best possible locations/neighborhoods to open new gym/yoga studio. Hence, with the help of geocoders, addresses of these candidate centers are obtained which can be presented to stakeholders.

Addresses of centers of areas recommended for further analysis

Address	Distance in km
319, West 38th Street, Garment District, Manhattan, Manhattan Community Board 4, New York County, NYC, New York, 10018, USA	0.561
North Woods, East Drive, Central Park, Manhattan, New York County, NYC, New York, 10026, USA	5.287
A Philip Randolph Campus High School, West 135th Street, Hamilton Heights, Manhattan, Manhattan Community Board 9, New York County, NYC, New York, 10030, USA	7.526
126, East 4th Street, NoHo Historic District, NoHo, Manhattan, Manhattan Community Board 3, New York County, NYC, New York, 10003, USA	3.497
Roosevelt Landings, Main Street, Manhattan Community Board 8, New York County, NYC, New York, 10044, USA	3.169

Figure 8 – Table showing addresses of cluster centers and their distances from Times Square

These 5 addresses represent the centers of zones containing locations with low number of fitness centers and no Gym/Yoga Studio nearby and all zones are close to city center (within 8 km from Times Square). These addresses mainly belong to neighborhoods - Chelsea and Clinton, Central Harlem, Lower East Side and Upper East Side.

4. Results

The analysis shows that although there is a great number of fitness centers in Manhattan (within 1 km around each neighborhood), there are some neighborhoods (close to city center) with low density of fitness centers as compared to others.

Highest concentration of fitness centers is detected west and south from Times Square whereas lowest concentration is detected in north-east, east and south-east. Similarly for Gym/Yoga Studios, lowest concentration is detected in north-east, east and south-east. After filtering these locations based on total number of fitness centers and distance to closest gym/yoga studio, 14 neighborhoods are considered as possible optimal locations. Applying K-means clustering, five clusters are obtained and addresses of their centers are generated that are to be used as markers/starting points for more detailed local analysis based on other factors.

These five zones have the potential of opening new a fitness center only on the basis of the number of fitness centers and distance to existing venues of gym/yoga studios.

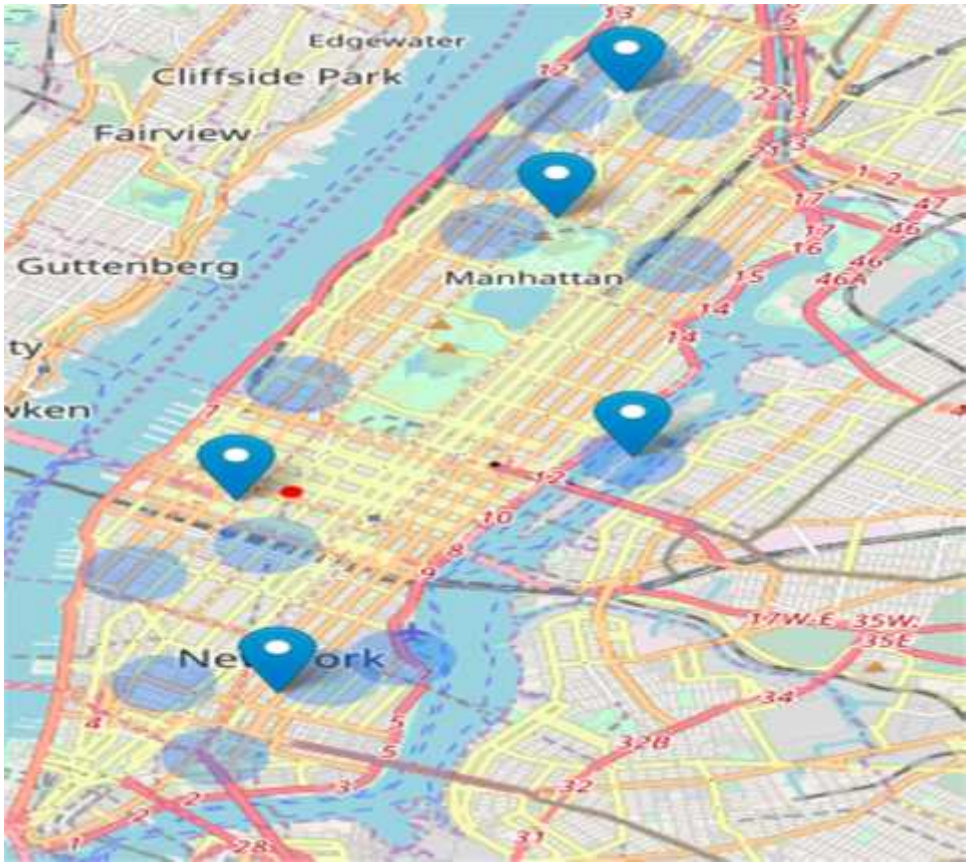

Figure 9 – Manhattan Map with cluster centers and possible optimal locations

5. Discussion

As mentioned before, Manhattan is a big city with a large number of fitness centers widely distributed among its neighborhoods. With such a complexity, many different approaches can be tried for analysis and clustering. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

Based on this analysis, possible optimal neighborhood to open new gym/yoga studios are – Harlem, Chelsea, Roosevelt Island, and Lower East Side.

The results retrieved with analysis above do not imply that those zones are actually optimal locations for opening a new gym/yoga studio. Purpose of this analysis is to only provide information on neighborhoods close to Manhattan center but not crowded with already existing gym/yoga studios - it is entirely possible that there is a very good reason for small

number of fitness centers in any of those neighborhoods, reasons which would make them unsuitable for a new gym/yoga studio regardless of lack of competition in the area.

Recommended areas should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6. Conclusion

Purpose of this project was to identify Manhattan locations close to center with low number of fitness centers, particularly Gym/Yoga studio, in order to aid stakeholders in narrowing down the search for optimal locations for a new Gym/Yoga Studio.

With calculation of fitness centers density distribution using Foursquare data, locations optimal for opening Gym/Yoga studio were identified. Clustering of those locations was then performed in order to create major zones of interest and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal fitness center location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like residential population, health awareness among residing people, real estate availability, prices, social and economic dynamics of every neighborhood etc.