**Recommending where to construct gyms in Manhattan, New York**

**May 26th 2019**

1. **Introduction / Business Problem**

In this project we will try to find the best suited location for opening gyms. This report is targeted and will be a help for stakeholders who want to open a gym in Manhattan, New York.

**Problem:**

The population of Manhattan is significantly high as compared to the other neighborhoods in the city. Hence, demand for gyms has increased significantly. Plus, some areas are completely devoid of gyms. People living in these areas have to travel far just for such a basic facility. Since, gyms are very important from the perspective of staying fit, new gyms need to be constructed. This report will help in identifying areas which have no facility of gym nearby. Hence new gym can be created in this vicinity.

1. **Data**

This dataset is from New York University Libraries of New York from Spatial Data Repository of NYU. This dataset exists for free on the web. The .json file has coordinates of the city of New York. This is the link to the dataset <https://geo.nyu.edu/catalog/nyu_2451_34572>.

This file is downloaded and cleaned. The data is placed on the server, so that we can simply use a `wget` command and access the data. The dataset contains 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. We only need 1 borough – Manhattan, hence we need to narrow down the data and only process Manhattan’s neighborhoods.

Based on definition of our problem, factors that will influence our decision are:

1. Number of existing gyms in the neighbourhood

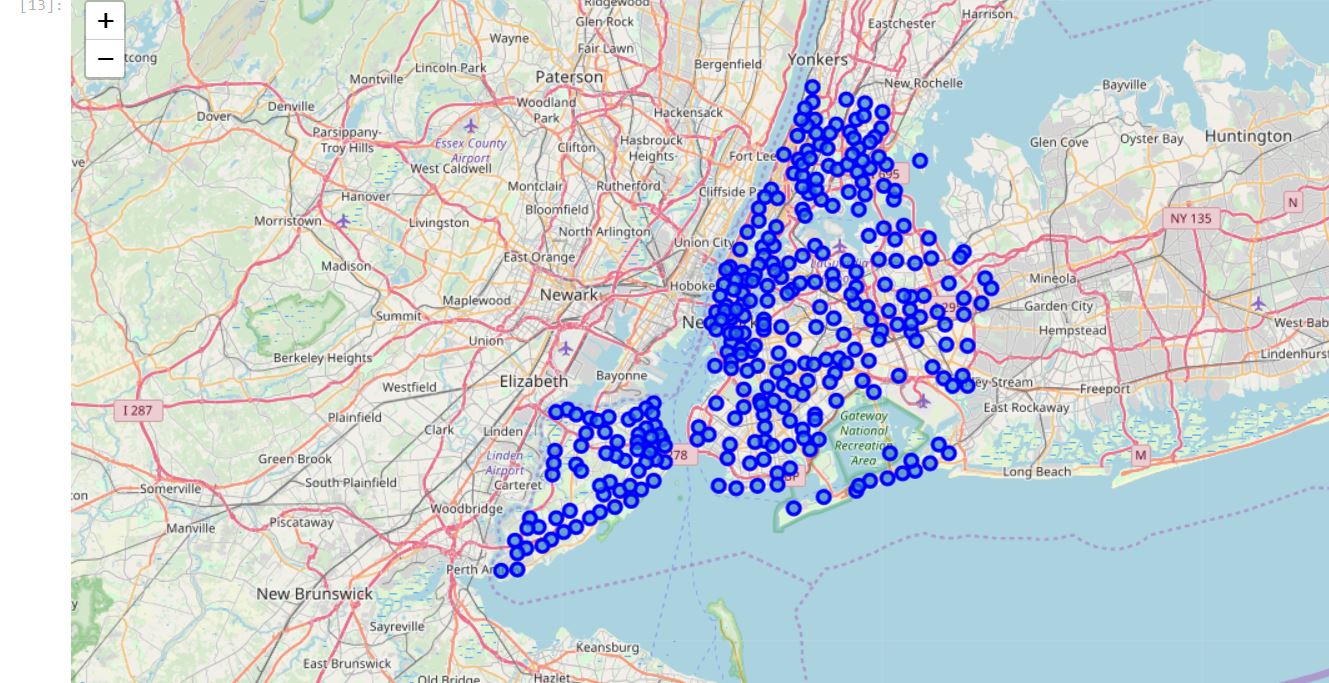
2. Number of neighbourhood devoid of gyms

Following data sources will be needed to extract/generate the required information:

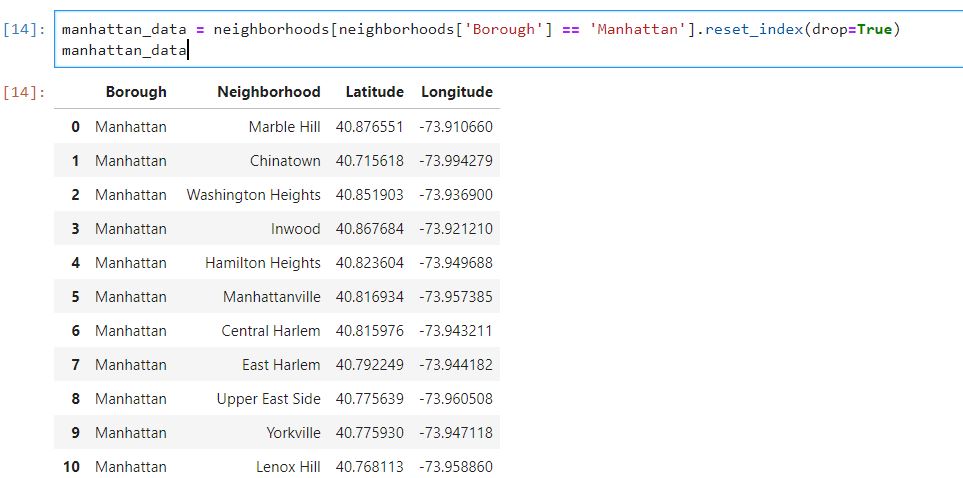
* **Foursquare API** will be needed to get the most common venues of Manhattan Borough of New York. The venue that we are interested in is “Gym”.
* Coordinate of Manhattan centre will be obtained using **Google Maps API geocoding** of well known location Manhattan.

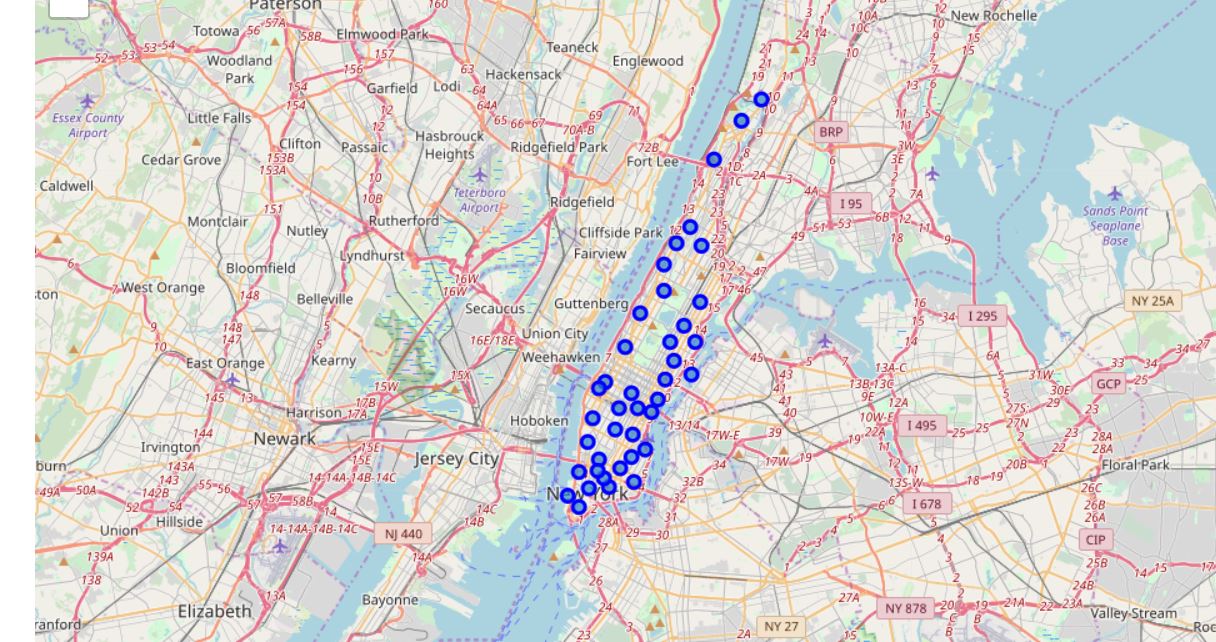
Data pre-processing

We acquire data from <https://geo.nyu.edu/catalog/nyu_2451_34572> link. Then the data is cleaned. This data is in unstructured format. So, we make it in structured form and store it in a dataframe. After examining the dataframe we find that the data has 5 boroughs and 306 neighbourhoods. Then we use “geopy library” to get longitude and latitudes of values of New York. After this, we create a map of New York with all the neighbourhoods. We will use python **folium** library to visualize all the maps in this project.



Our interest is only in Manhattan. Hence, we simply the above map. After simplifying the dataframe with respect to Manhattan, we visualize it.





1. **Methodology**

We utilize the Foursquare API to explore the neighbourhoods in Manhattan and segment them. Our focus is to find gyms in Manhattan. Hence, we need to explore venues in Manhattan. For this we use Foursquare API which returns us venues. We set the Foursquare API so that it returns us the top 100 venues within the radius of 500 meters. Then we create a new dataframe **manhattan\_venues**.



Now, we have 2 goals –

1. To find neighbourhoods with very few gyms.
2. To find neighbourhoods with no gyms

Our main goal is to construct gyms in the areas which have no gyms. We need these neighbourhoods because these neighbourhoods are areas where gyms can be constructed. So stakeholders can construct gyms in these areas.

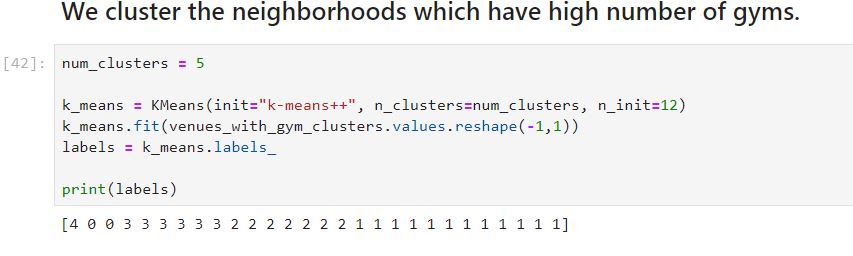
First, let’s find neighbourhoods which have very few gyms. For this first we need to filter out the neighbourhoods which have gyms from the original dataframe. We store this data in dataframe **venues\_with\_gym**. This dataframe has columns of latitude and longitude corresponding to each neighbourhood, venues, venue’s latitude and longitudes and venue category.



Then, we analyze how many gyms are present for each neighbourhood in dataframe **venues\_with\_gym**. We find that certain places already have high amount of gyms. These places do not need construction of new gyms.



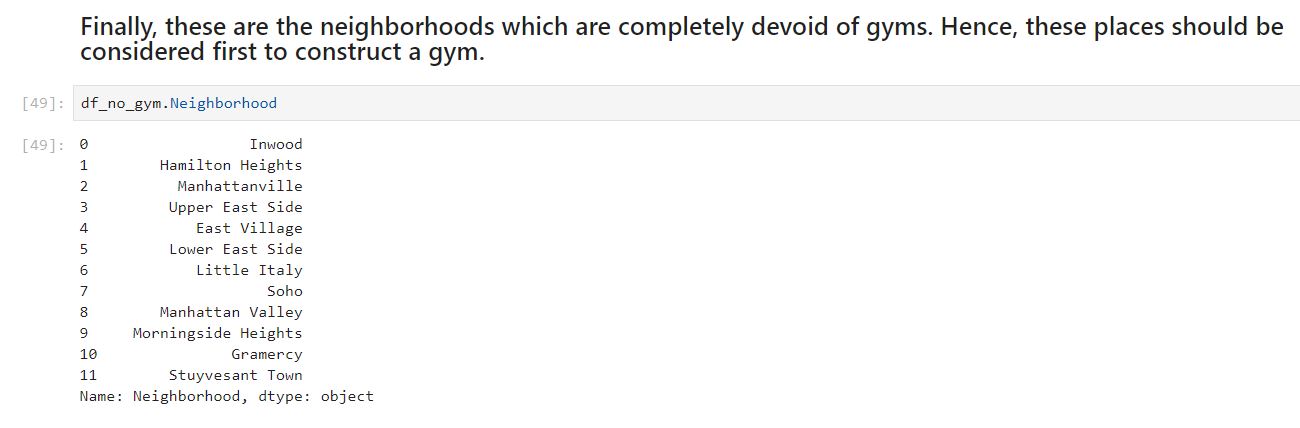
Now, we will use clustering to distinguish between neighbourhoods that have little number of gyms and neighbourhoods which have high number of gyms.



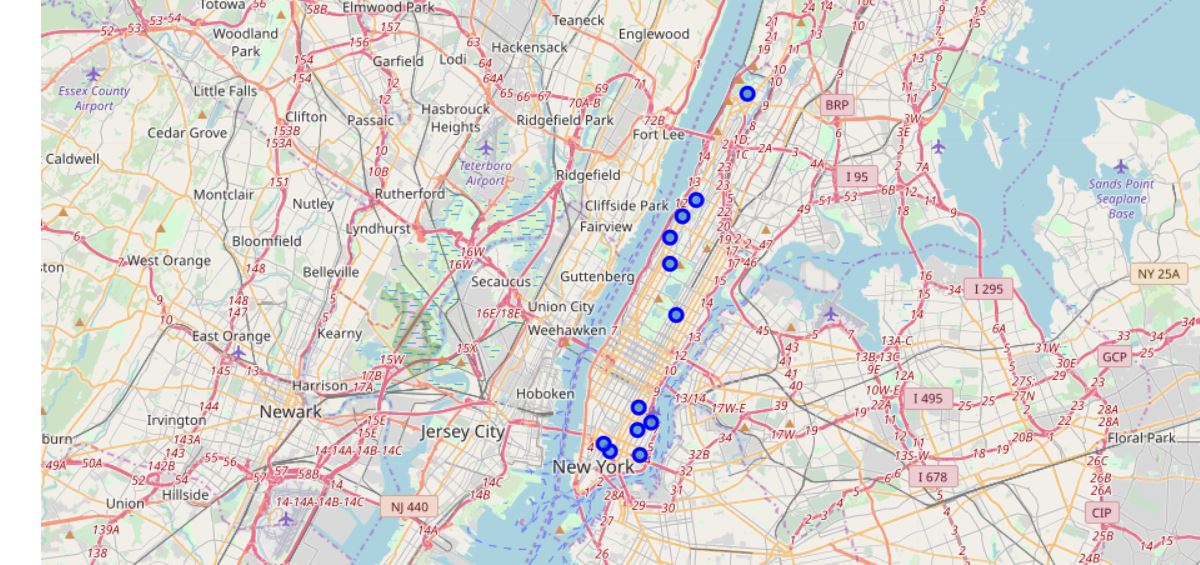
Now, let’s find neighbourhoods with no gyms. This is also our main goal. For this we construct a new dataframe naming **df\_no\_gyms**.



This dataframe still has many redundant rows and columns. So, we process to get ultimately all the neighbourhoods with no gyms.



We visualize the neighbourhoods which have no gyms using Folium.



**4. Results**

After all the processing, we observe the following results -:

1. There are very few neighbourhoods which have sufficient amount of gyms. And these neighbourhoods are :-

|  |  |
| --- | --- |
| **Neighbourhoods** | **Number of gyms** |
| Yorkville | 6 |
| Flatiron | 5 |
| Financial District, Battery Park City | 4 |
| Tribeca, Murray Hill, Lenox Hill, Carnegie Hill, Sutton Place | 3 |

1. Excluding the above places, rest all the neighbourhoods are good options to create gyms.
2. The best options for creating gyms are neighbourhoods with **no gym**. These are –

* Inwood
* Hamilton Heights
* Manhattanville
* Upper East Side
* East Village
* Lower East Side
* Little Italy
* Soho
* Manhattan Valley
* Morningside Heights
* Gramercy
* Stuyvesant Town

**5. Discussion**

Gyms are very important, especially from the perspective of fitness. All age groups should have access to gyms. Gyms help in increasing bone density if followed by proper diet. It’s a basic facility which is neglected a lot. Stakeholders who are seeking profits in fitness industry should start with creating new gyms or any fitness school. New gyms can open up even employment for those who pursue fitness as their careers, for example gym trainers. This report helps stakeholders to target neighbourhoods in Manhattan through which they can attain maximum profits. Plus new gyms will be constructed for the people and benefit people.

The best strategy would be start your business related to gym or construction of gyms from the neighbourhoods which do not have gyms. Even in these neighbourhoods we can start from neighbourhoods which are more popular and dense like Manhattan valley. If opting only to promote gyming equipments and stakeholders are not interested in construction of gyms, then opting neighbourhoods which have few gyms would be a good option. Neighbourhood like Chinatown which have only one gym but is a neighbourhood which is fairly popular, there could be a lot of load on a single gym. Hence chances are that they will need renewal of their equipments more.

**6. Conclusion**

Gyms and fitness are very important aspects of human life. Enhancing and increasing quality of them is hence very necessary. This report is beneficial to people who promote fitness or want to construct gyms and who aim to make peoples’ lives fitter and healthier.