Peer-graded Assignment: Capstone

Project - The Battle of Neighborhoods

(Week 1)

Dated: July 15, 2021

1. Introduction

1.1 Background:

New York City is not only the most populous city in the United States but also one of the world's largest mega cities with an estimated 2020 population of approximately 8.25 million spread over about 302.6 square miles including 5 boroughs. It is described as the cultural, financial and media capital of the world. It has population from all around the world and exerts significant influence on commerce, entertainment, research, technology, education, politics, art, tourism, sports etc. The pandemic beginning in March 2020 shut down most of the world including New York City. During this time, it has become very evident that some form of exercise is very useful for people not able to go out of the house. People took to doing yoga watching YouTube videos to stay healthy, unable to get to their regular yoga studio. Yoga has become an important part of people's lifestyle during this pandemic to stay healthy or even to recover from illnesses. Given these circumstances, our client would like to open a Yoga studio to teach face to face now that the city is opening back up again. So the basic question that we would like to answer is the best location for our client to open the Yoga studio. Client would like to stay in Manhattan where people are coming back again and offices are also opening up and demand would likely be high. The main criteria would be to find a location with easy access to subway, parking. Location also needs to be close to office buildings so people can step out during

lunch hours for their lessons or their way to and from work. So the question that this final project would like answer is where should our client open their Yoga studio in Manhattan? Main criteria is the location so it’s close to transportation as well as office buildings and other entertainment avenues which provide more foot traffic.

1.2 Business Problem

As part of the project, we will try to answer the following questions posed by our client:

1. Where in Manhattan is most optimum location for opening a Yoga studio which has the best chance of succeeding?
2. What location would provide ample traffic in surrounding areas that would also help bring clients to the Yoga studio?

2. Data

The data that will be used to order answer the business questions, is data on New York City neighborhoods, boroughs to include boundaries, latitude, longitude, restaurants, and other yoga studios. From the main data, the borough of Manhattan will be main focus since client wants to stay in Manhattan. New York City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source: https://cocl.us/new_york_dataset All data related to

locations and other entertainment venues and/or other office complexes etc will be obtained via the FourSquare API utilized via the Request library in Python.

3. Methodology

- Data will be collected from https://cocl.us/new_york_dataset and cleaned and processed into a Dataframe.
- Only Manhattan data will be used for the purpose answering the main questions.
- FourSquare app will be utilized via the Request library in Python to locate all venues and then filtered by optimum location by entertainment venues and added to the Dataframe.
- Data will be sorted based on rankings. We will explore the data using the data and ultimately visually assess the data using Python libraries.

4. Analysis

4.1 Install and Import Required Libraries

#Install and Import required libraries

!pip install beautifulsoup4

!pip install lxml

import requests # *library to handle requests*

```
import pandas as pd # library for data analsysis

import numpy as np # library to handle data in a vectorized manner

import random # library for random number generation

!conda install -c conda-forge geopy --yes

from geopy.geocoders import Nominatim # module to convert an address into

latitude and longitude values

# libraries for displaying images

from IPython.display import Image

from IPython.core.display import display, HTML



from IPython.display import display_html

import pandas as pd

import numpy as np

# tranforming json file into a pandas dataframe library

from pandas.io.json import json_normalize

!conda install -c conda-forge folium=0.5.0 --yes

import folium # plotting library

from bs4 import BeautifulSoup
```

```
from sklearn.cluster import KMeans
import matplotlib.cm as cm
import matplotlib.colors as colors

print('Folium installed')
print('Libraries imported.')
```

```
Requirement already satisfied: beautifulsoup4 in /home/jupyterlab/conda/envs/pyth
on/lib/python3.6/site-packages (4.9.3)
Requirement already satisfied: soupsieve>1.2; python_version >= "3.0" in /home/ju
pyterlab/conda/envs/python/lib/python3.6/site-packages (from beautifulsoup4) (2.2.
1)
Requirement already satisfied: lxml in /home/jupyterlab/conda/envs/python/lib/pytho
n3.6/site-packages (4.6.3)
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

4.2. Data acquisition and cleaning

The data that will be used to order answer the business questions, is data on New York City neighborhoods, boroughs to include boundaries, latitude, longitude, restaurants, and other yoga studios.

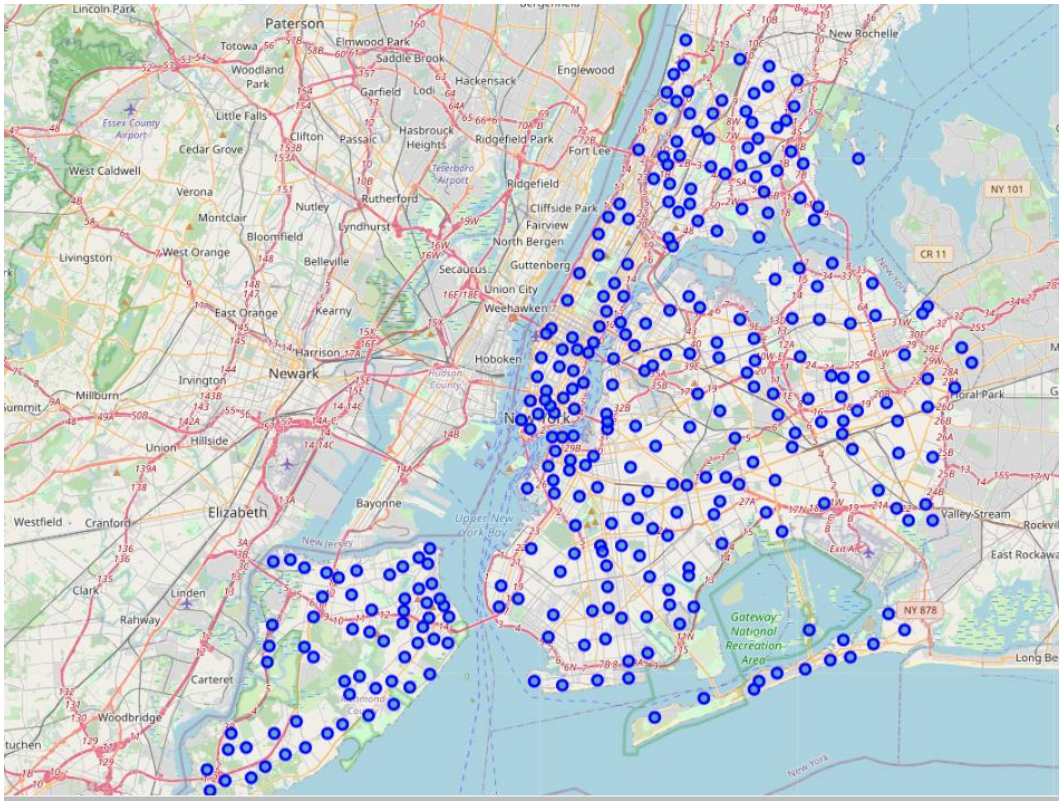
From the main data, the borough of Manhattan will be main focus since client wants to stay in Manhattan.

New York City data containing the neighborhoods and boroughs, latitudes, and longitudes was obtained from the data source: https://cocl.us/new_york_dataset

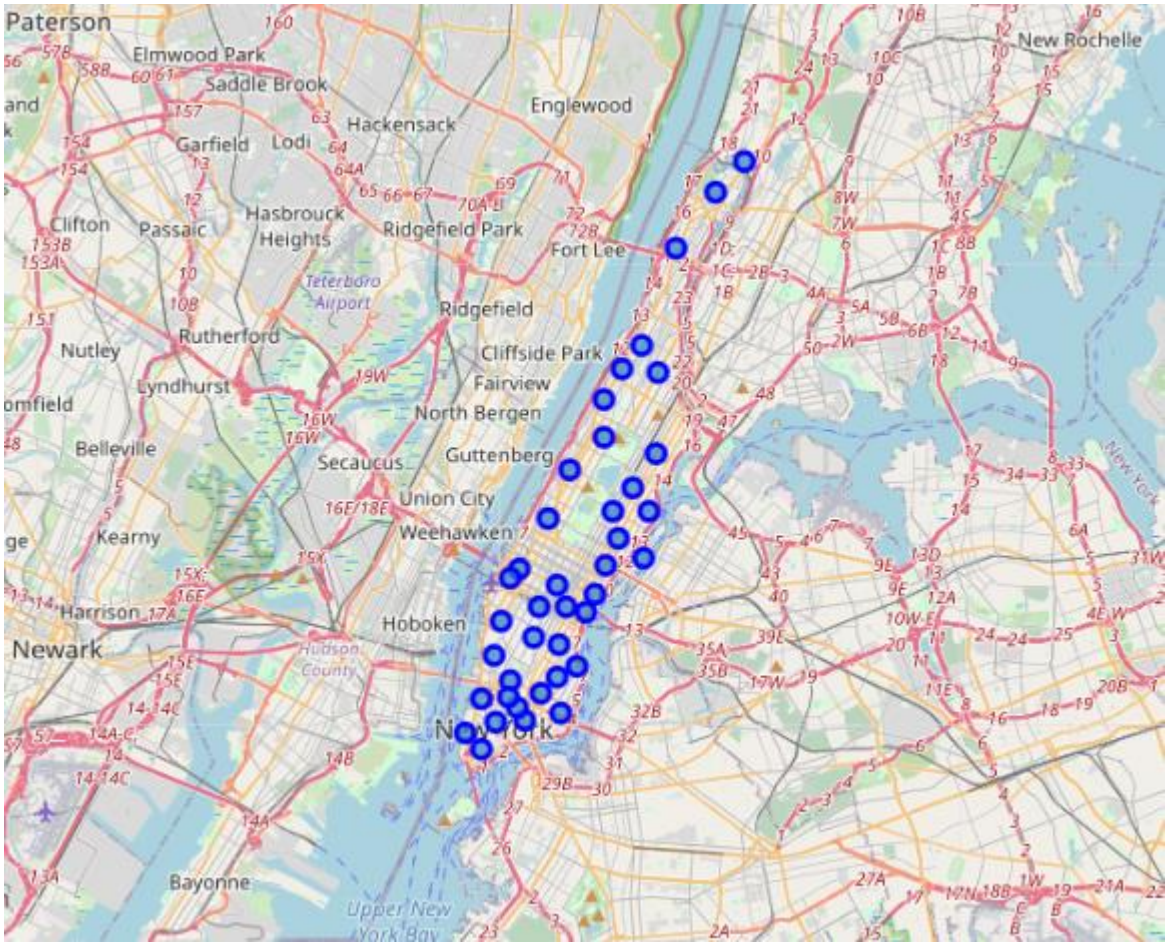
All data related to locations and other entertainment venues and/or other office complexes etc was obtained via the FourSquare API utilized via the Request library in Python.

4.3. Exploratory Data Analysis

Using geolocator and Folium packages of Python, we create a map of New York City showing all the neighborhoods.



Since client wants to open the studio in the borough of Manhattan, we subset the d ata to only show the borough of Manhattan.



Using FourSquare data we can create the top ten venues for each neighborhood. Using pandas, we create and explore the dataframe.

Data is grouped by their neighborhoods and their means are taken to calculate top neighborhoods for various Venue categories.

As can be seen from figure below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Clothing Store	Women's Store
1	Carnegie Hill	Coffee Shop	Café	Wine Shop	Yoga Studio	Bar
2	Central Harlem	African Restaurant	Chinese Restaurant	Public Art	French Restaurant	American Restaurant
3	Chelsea	Coffee Shop	Bakery	Art Gallery	American Restaurant	Italian Restaurant
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant
5	Civic Center	Coffee Shop	Spa	Gym / Fitness Center	American Restaurant	French Restaurant
6	Clinton	Italian Restaurant	Theater	Gym / Fitness Center	Sandwich Place	Coffee Shop
7	East Harlem	Bakery	Mexican Restaurant	Thai Restaurant	Sandwich Place	Latin American Restaurant
8	East Village	Bar	Mexican Restaurant	Speakeasy	Pizza Place	Ice Cream Shop
9	Financial District	Coffee Shop	Pizza Place	Cocktail Bar	Park	Bar
10	Flatiron	Italian Restaurant	Japanese Restaurant	New American Restaurant	American Restaurant	Mediterranean Restaurant

Grouping the data of Manhattan and after one hot encoding the top ten neighborhoods with maximum number of Yoga studios, gyms and cafes emerges.

Neighborhood	Yoga Studio	Gym	Gym / Fitness Center	Gymnastics Gym	Café	
22	Marble Hill	0.043478	0.086957	0.000000	0.0	0.000000
1	Carnegie Hill	0.033708	0.022472	0.033708	0.0	0.067416
13	Hamilton Heights	0.033333	0.016667	0.000000	0.0	0.066667
35	Upper East Side	0.030612	0.000000	0.040816	0.0	0.000000
5	Civic Center	0.030000	0.010000	0.050000	0.0	0.010000
10	Flatiron	0.030000	0.010000	0.020000	0.0	0.020000
20	Manhattan Valley	0.021739	0.000000	0.000000	0.0	0.021739
19	Lower East Side	0.020408	0.020408	0.000000	0.0	0.040816
15	Inwood	0.018182	0.000000	0.000000	0.0	0.054545
33	Tudor City	0.012500	0.037500	0.012500	0.0	0.050000

We plot the data to see the top neighborhoods better:

Figure 3. Bar plot showing highest number of Yoga studios in the neighborhoods of Manhattan

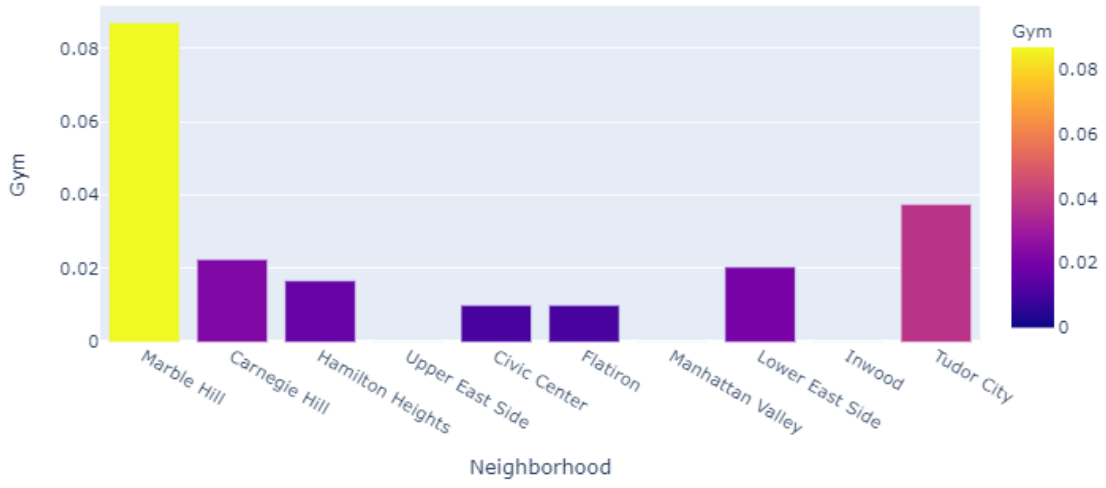


Figure 4. Bar plot showing highest number of Gyms in the neighborhoods of Manhattan

4.4 Cluster and Visualize the Neighborhoods

Using kmeans cluster, the neighborhoods of Manhattan were divided into 5 clusters.

Examine Clusters

Cluster 1

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Washington Heights	Café	Bakery	Bank	Mobile Phone Shop	Pizza Place
3	Inwood	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant
4	Hamilton Heights	Pizza Place	Café	Mexican Restaurant	Deli / Bodega	Coffee Shop
7	East Harlem	Bakery	Mexican Restaurant	Thai Restaurant	Sandwich Place	Latin American Restaurant
20	Lower East Side	Chinese Restaurant	Art Gallery	Latin American Restaurant	Pharmacy	Pizza Place

Cluster 2

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant
6	Central Harlem	African Restaurant	Chinese Restaurant	Public Art	French Restaurant	American Restaurant
8	Upper East Side	Exhibit	Italian Restaurant	American Restaurant	Juice Bar	Gym / Fitness Center
14	Clinton	Italian Restaurant	Theater	Gym / Fitness Center	Sandwich Place	Coffee Shop
15	Midtown	Hotel	Sporting Goods Shop	Clothing Store	Coffee Shop	Bookstore
16	Murray Hill	Hotel	Coffee Shop	American Restaurant	Sandwich Place	Japanese Restaurant
17	Chelsea	Coffee Shop	Bakery	Art Gallery	American Restaurant	Italian Restaurant
19	East Village	Bar	Mexican Restaurant	Speakeasy	Pizza Place	Ice Cream Shop
22	Little Italy	Bakery	Café	Chinese Restaurant	Mediterranean Restaurant	Hotel
25	Manhattan Valley	Mexican Restaurant	Bar	Coffee Shop	Vietnamese Restaurant	Playground
27	Gramercy	Italian Restaurant	Bar	Pizza Place	Bagel Shop	Wine Shop
28	Battery Park City	Park	Coffee Shop	Hotel	Clothing Store	Women's Store
29	Financial District	Coffee Shop	Pizza Place	Cocktail Bar	Park	Bar
30	Carnegie Hill	Coffee Shop	Café	Wine Shop	Yoga Studio	Bar
32	Civic Center	Coffee Shop	Spa	Gym / Fitness Center	American Restaurant	French Restaurant
33	Midtown South	Korean Restaurant	Hotel	Dessert Shop	Gym / Fitness Center	Hotel Bar
34	Sutton Place	Italian Restaurant	Pizza Place	Gym / Fitness Center	Coffee Shop	Indian Restaurant
38	Flatiron	Italian Restaurant	Japanese Restaurant	New American Restaurant	American Restaurant	Mediterranean Restaurant
39	Hudson Yards	Gym / Fitness Center	Italian Restaurant	Hotel	American Restaurant	Thai Restaurant

Cluster 3

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Manhattanville	Coffee Shop	Mexican Restaurant	Italian Restaurant	Bar	Seafood Restaurant
9	Yorkville	Italian Restaurant	Gym	Coffee Shop	Deli / Bodega	Sushi Restaurant
10	Lenox Hill	Italian Restaurant	Coffee Shop	Sushi Restaurant	Pizza Place	Café
11	Roosevelt Island	Deli / Bodega	Coffee Shop	Park	Bus Line	Bubble Tea Shop
12	Upper West Side	Italian Restaurant	Bakery	Wine Bar	Bar	Café
13	Lincoln Square	Plaza	Performing Arts Venue	Concert Hall	Café	Theater
18	Greenwich Village	Italian Restaurant	Clothing Store	Boutique	Sushi Restaurant	Dessert Shop
21	Tribeca	American Restaurant	Park	Italian Restaurant	Wine Bar	Café
23	Soho	Clothing Store	Boutique	Italian Restaurant	Mediterranean Restaurant	Shoe Store
24	West Village	Italian Restaurant	New American Restaurant	Cocktail Bar	American Restaurant	Park
26	Morningside Heights	Coffee Shop	Park	Bookstore	American Restaurant	Burger Joint
31	Noho	Italian Restaurant	Cocktail Bar	French Restaurant	Pizza Place	Hotel
35	Turtle Bay	Italian Restaurant	Coffee Shop	Sushi Restaurant	Deli / Bodega	Café
36	Tudor City	Park	Mexican Restaurant	Café	Pizza Place	Gym

Cluster 4

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

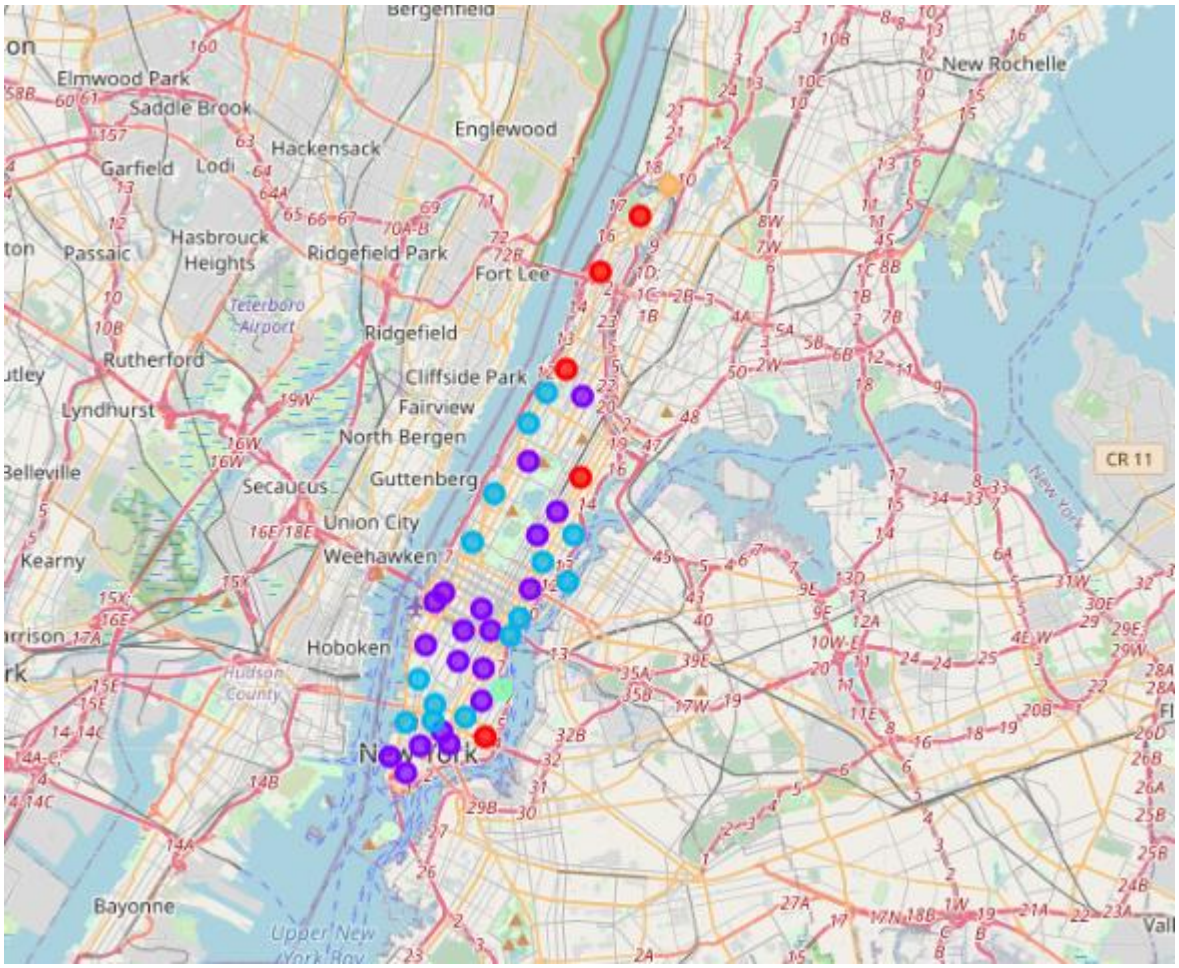
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
37	Stuyvesant Town	Bar	Park	Coffee Shop	Cocktail Bar	Playground

Cluster 5

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Marble Hill	Coffee Shop	Sandwich Place	Gym	Yoga Studio	Bank

4.5 Visualizing the clusters:



5. Results/Conclusions:

As can be seen from the above data processing and analysis that the most optimum locations for a new Yoga Studio would be the neighborhoods of Marble Hill, Carnegie Hill, Upper East Side and Hamilton Heights in Manhattan. The analysis has shown that these are the areas with most number of Gyms and Fitness centers. This shows that these areas are populated with people who are interested in health studios. These 4 areas also have many cafes situated in the area. This is also another avenue for foot traffic for the the new yoga studio. Of these four neighborhood areas

the most likely area would be Carnegie Hill. This neighborhood is adjacent to Upper East Side so it also has potential for more spill over traffic. Upper East Side also is also fairly residential and that increases the customer base.

This analysis needs more data for transportation issues. Any new business needs accessibility and it's proximity to subway stations, bus stations or parking areas is highly desired. So further comprehensive analysis with external transportation data would be an asset. This result is based on existing data showing existing gyms, fitness centers, cafes and restaurants in the area. The assumption being that since these businesses are established here there is scope for another fitness business to open and thrive. With the city opening back up again after the pandemic, people are eager to come out of homes and work out in the company of other people face to face. And this is where the new Yoga studio will make its mark. The pandemic had forced the closure of many small boutique businesses like yoga studios so with everything opening up there is definitely potential for a new yoga studio to thrive!