Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 1)

Introduction:

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 its close to transportation as well as office buildings and other entertainment avenues which provide more foot traffic.

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.

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.

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?

[ ]:

Analysis:

Install and Import Required Libraries

[127]:

```
#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/python/lib/python3.6/site-packages (4.9.3)
Requirement already satisfied: soupsieve>1.2; python_version >= "3.0" in /home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (from beautifulsoup4) (2.2.1)
Requirement already satisfied: lxml in /home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (4.6.3)
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

Folium installed
Libraries imported.
```

Get initial New York Data

[128]:

```
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']
    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)
    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']
        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]
        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_index=True)
    return new_york_data
```

[129]:

```
new_york_data = get_new_york_data()
new_york_data.head()
```

[129]:

Borough	Neighborhood	Latitude	Longitude

0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

[5]:
```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
        len(new_york_data['Borough'].unique()),
        new_york_data.shape[0]
    )
)
```

The dataframe has 5 boroughs and 306 neighborhoods.

[130]:
```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}.'.format(latitude, longitude))
```

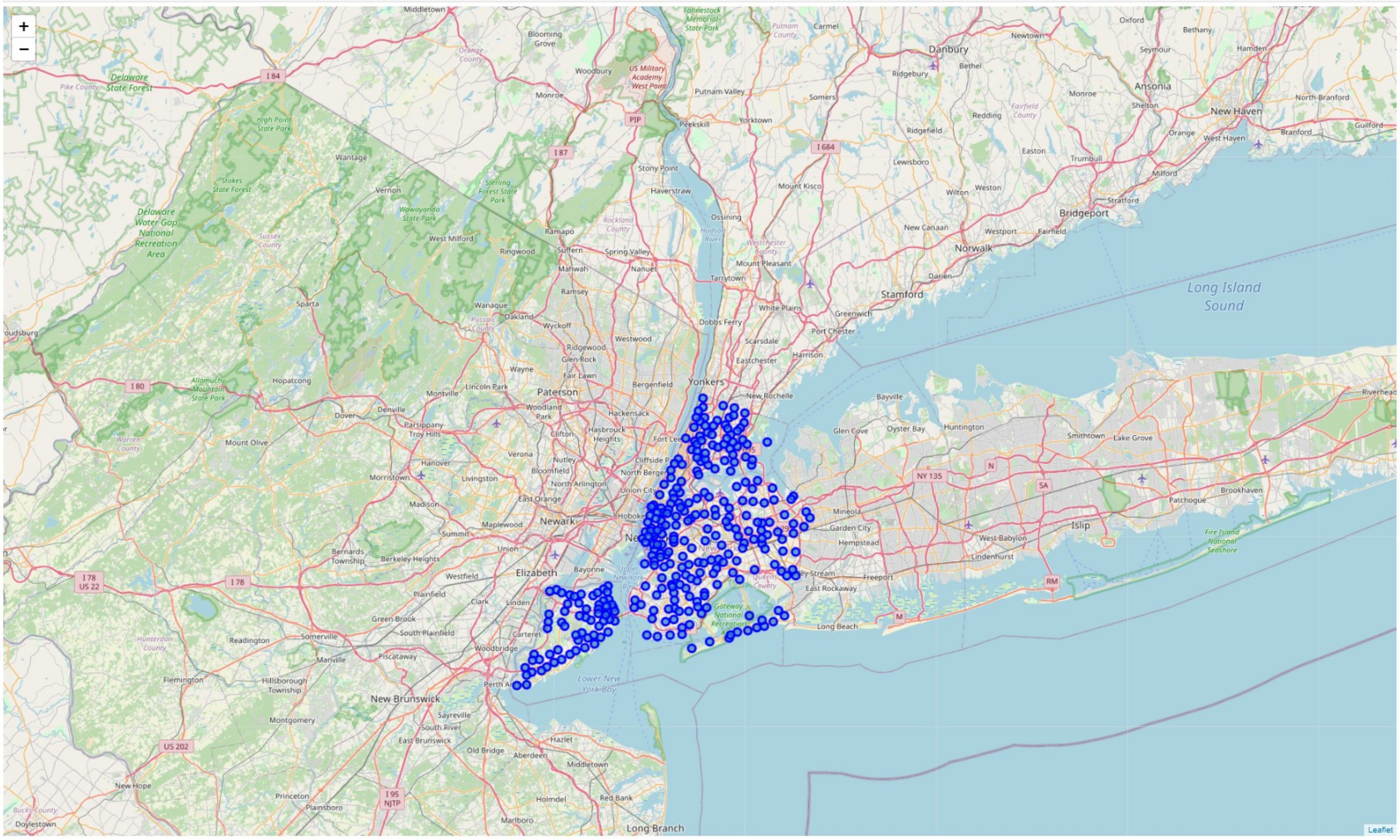
The geograpical coordinate of New York City are 40.7127281, -74.0060152.

[178]:
```
# create map of New York using latitude and longitude values
map_newyork = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(new_york_data['Latitude'], new_york_data['Longitude'], new_york_data['Borough'], new_york_data['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_newyork)

map_newyork
```

[178]:



Data Cleaning and Processing

Client wants to open Yoga Studio in Manhattan so lets create new Dataframe of Manahattan data and segment that neighborhood.

[132]:
```
manhattan_data = new_york_data[new_york_data['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head()
```

[132]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Getting geographical coordinates for Manahattan

[133]:
```
address = 'Manhattan, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Manhattan are {}, {}.'.format(latitude, longitude))
```

The geograpical coordinate of Manhattan are 40.7896239, -73.9598939.

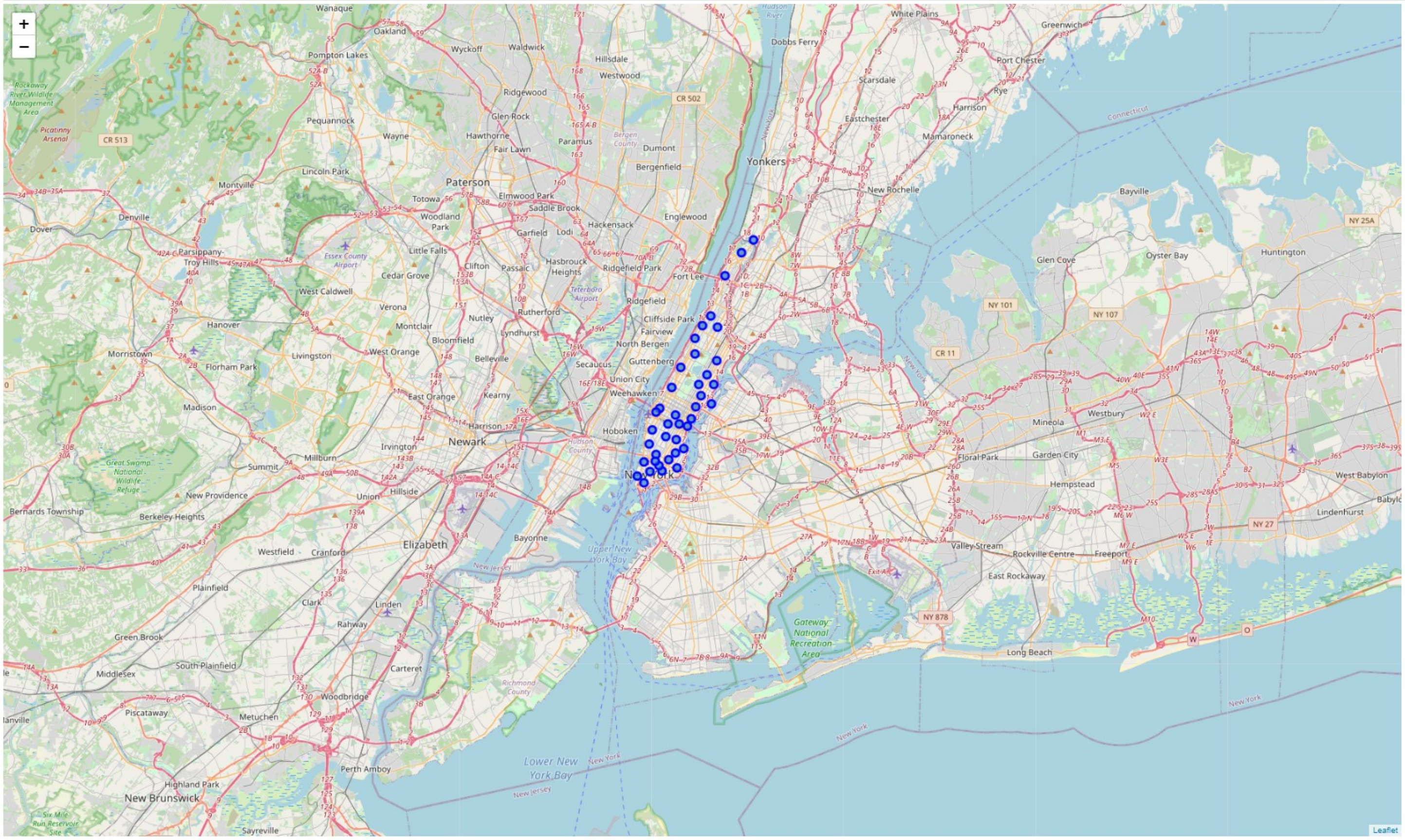
Creating map of Manhattan and the neighborhoods in it

[134]:
```
# create map of Manhattan using latitude and longitude values
map_manhattan = folium.Map(location=[latitude, longitude], zoom_start=11)
```

```python
# add markers to map
for lat, lng, label in zip(manhattan_data['Latitude'], manhattan_data['Longitude'], manhattan_data['Neighborhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_manhattan)

map_manhattan
```

[134]:



Start to use Foursquare API to explore the neighborhoods and segment them

Define Foursquare Credentials and Version

[11]:
```python
#hidden cell
CLIENT_ID = [redacted]' # your Foursquare ID
CLIENT_SECRET = '[redacted]' # your Foursquare Secret
ACCESS_TOKEN = '[redacted] # your FourSquare Access Token
```

[16]:
```python
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value
```

Exploring 'Marble Hill' as it is the first neighborhood

[12]:
```python
manhattan_data.loc[0, 'Neighborhood']
```

[12]: 'Marble Hill'

[13]:
```python
neighborhood_latitude = manhattan_data.loc[0, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = manhattan_data.loc[0, 'Longitude'] # neighborhood longitude value

neighborhood_name = manhattan_data.loc[0, 'Neighborhood'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_latitude,
                                                               neighborhood_longitude))
```

Latitude and longitude values of Marble Hill are 40.87655077879964, -73.91065965862981.

[14]:
```python
#search_query = 'Marble Hill'
radius = 500
LIMIT = 100
```

[166]:
```python
# type your answer here
#url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude,ACCESS_TOKEN, VE
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
#url # display URL
```

[ ]:
```python
results = requests.get(url).json()
results
```

[168]:
```python
# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

[169]:
```python
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:3: FutureWarning:

pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead
```

[169]:

	name	categories	lat	lng
0	Arturo's	Pizza Place	40.874412	-73.910271
1	Bikram Yoga	Yoga Studio	40.876844	-73.906204
2	Dunkin'	Donut Shop	40.877136	-73.906666
3	Tibbett Diner	Diner	40.880404	-73.908937
4	Starbucks	Coffee Shop	40.877531	-73.905582

[21]:
```python
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
23 venues were returned by Foursquare.
```

[22]:
```python
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

[23]:
```python
# type your answer here
manhattan_venues = getNearbyVenues(names=manhattan_data['Neighborhood'],
                                   latitudes=manhattan_data['Latitude'],
                                   longitudes=manhattan_data['Longitude']
                                  )
```

```
Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea
Greenwich Village
East Village
Lower East Side
Tribeca
Little Italy
Soho
West Village
Manhattan Valley
Morningside Heights
Gramercy
Battery Park City
Financial District
Carnegie Hill
Noho
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards
```

[24]:
```python
print(manhattan_venues.shape)
manhattan_venues.head()
```

```
(3252, 7)
```

[24]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
3	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

[62]:
```python
manhattan_venues.columns
```

[62]:
```
Index(['Neighborhood', 'Neighborhood Latitude', 'Neighborhood Longitude',
       'Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category'],
      dtype='object')
```

```
[63]: manhattan_venues[manhattan_venues['Neighborhood']=='Marble Hill']
```

[63]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
3	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
5	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
6	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium
7	Marble Hill	40.876551	-73.91066	Blink Fitness	40.877271	-73.905595	Gym
8	Marble Hill	40.876551	-73.91066	GameStop	40.874267	-73.909342	Video Game Store
9	Marble Hill	40.876551	-73.91066	T.J. Maxx	40.877232	-73.905042	Department Store
10	Marble Hill	40.876551	-73.91066	Vitamin Shoppe	40.877160	-73.905632	Supplement Shop
11	Marble Hill	40.876551	-73.91066	Rite Aid	40.875467	-73.908906	Pharmacy
12	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
13	Marble Hill	40.876551	-73.91066	Lot Less Closeouts	40.878270	-73.905265	Discount Store
14	Marble Hill	40.876551	-73.91066	Subway	40.874667	-73.909586	Sandwich Place
15	Marble Hill	40.876551	-73.91066	Baskin-Robbins	40.877132	-73.906678	Ice Cream Shop
16	Marble Hill	40.876551	-73.91066	Subway	40.878465	-73.905518	Sandwich Place
17	Marble Hill	40.876551	-73.91066	Parrilla Latina	40.877473	-73.906073	Steakhouse
18	Marble Hill	40.876551	-73.91066	Starbucks	40.873234	-73.908730	Coffee Shop
19	Marble Hill	40.876551	-73.91066	The Children's Place	40.873672	-73.908156	Kids Store
20	Marble Hill	40.876551	-73.91066	TD Bank	40.879496	-73.909286	Bank
21	Marble Hill	40.876551	-73.91066	Terrace View Delicatessen	40.875995	-73.913151	Deli / Bodega
22	Marble Hill	40.876551	-73.91066	Grill 26 at TCR	40.878802	-73.915672	American Restaurant

```
[65]: manhattan_venues.groupby('Neighborhood').count().sort_values(by='Venue',ascending=False)
```

[65]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Yorkville	100	100	100	100	100	100
Flatiron	100	100	100	100	100	100
Midtown	100	100	100	100	100	100
Noho	100	100	100	100	100	100
Soho	100	100	100	100	100	100
Sutton Place	100	100	100	100	100	100
Little Italy	100	100	100	100	100	100
Lenox Hill	100	100	100	100	100	100
Turtle Bay	100	100	100	100	100	100
Murray Hill	100	100	100	100	100	100
Greenwich Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100
East Village	100	100	100	100	100	100
Clinton	100	100	100	100	100	100
Civic Center	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Chelsea	100	100	100	100	100	100
West Village	100	100	100	100	100	100
Midtown South	100	100	100	100	100	100
Upper East Side	98	98	98	98	98	98
Lincoln Square	96	96	96	96	96	96
Upper West Side	95	95	95	95	95	95
Gramercy	94	94	94	94	94	94
Tribeca	93	93	93	93	93	93
Carnegie Hill	89	89	89	89	89	89
Battery Park City	87	87	87	87	87	87
Washington Heights	86	86	86	86	86	86
Tudor City	80	80	80	80	80	80
Hudson Yards	78	78	78	78	78	78
Hamilton Heights	60	60	60	60	60	60
Inwood	55	55	55	55	55	55
Lower East Side	49	49	49	49	49	49
Manhattanville	48	48	48	48	48	48
Central Harlem	46	46	46	46	46	46
Manhattan Valley	46	46	46	46	46	46
Morningside Heights	41	41	41	41	41	41
East Harlem	41	41	41	41	41	41
Roosevelt Island	28	28	28	28	28	28
Marble Hill	23	23	23	23	23	23
Stuyvesant Town	19	19	19	19	19	19

```
[27]: # one hot encoding
manhattan_onehot = pd.get_dummies(manhattan_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
manhattan_onehot['Neighborhood'] = manhattan_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [manhattan_onehot.columns[-1]] + list(manhattan_onehot.columns[:-1])
manhattan_onehot = manhattan_onehot[fixed_columns]

manhattan_onehot.head()
```

[27]:

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
2	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 339 columns

```
[155]: pd.set_option('display.max_columns', None)  # or 1000
pd.set_option('display.max_rows', None)  # or 1000
pd.set_option('display.max_colwidth', None)  # or 199
```

```
[83]: manhattan_onehot.shape
```

```
[83]: (3252, 339)
```

```
[156]: #pd.set_option('display.max_rows', df.shape[0]+1)
print(manhattan_onehot.columns)
```

```
Index(['Neighborhood', 'Accessories Store', 'Adult Boutique',
```

```
       'Afghan Restaurant', 'African Restaurant', 'American Restaurant',
       'Antique Shop', 'Argentinian Restaurant', 'Art Gallery', 'Art Museum',
       ...
       'Video Store', 'Vietnamese Restaurant', 'Volleyball Court',
       'Waterfront', 'Whisky Bar', 'Wine Bar', 'Wine Shop', 'Wings Joint',
       'Women's Store', 'Yoga Studio'],
      dtype='object', length=339)
```

[158]:
```
manhattan_grouped = manhattan_onehot.groupby('Neighborhood').mean().reset_index()
manhattan_grouped.head()
```

[158]:

ium	Australian Restaurant	Austrian Restaurant	BBQ Joint	Baby Store	Badminton Court	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Basketball Court	Beach Bar	Beer Bar	Beer Garden	Beer Store	Big Box Store	Bike Rental / Bike Share	Bike Trail	Bistro	Board Shop	Boat or Ferry	Bookstore	Boutique	Boxing Gym	Brazilian Restaurant	Breakfast Spot	Bridal Shop	Bridge	Bubble Tea Shop	Building	Burger Joint	Burrito Place
494	0.0	0.00	0.022989	0.0	0.0	0.000000	0.011494	0.0	0.011494	0.0	0.0	0.0	0.000000	0.022989	0.0	0.0	0.0	0.0	0.011494	0.0	0.022989	0.000000	0.011494	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.022989	0.011494
000	0.0	0.00	0.000000	0.0	0.0	0.011236	0.022472	0.0	0.033708	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.033708	0.000000	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.011236	0.000000
000	0.0	0.00	0.021739	0.0	0.0	0.021739	0.000000	0.0	0.043478	0.0	0.0	0.0	0.021739	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.021739	0.021739	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.000000	0.000000
000	0.0	0.00	0.000000	0.0	0.0	0.000000	0.050000	0.0	0.010000	0.0	0.0	0.0	0.010000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.020000	0.000000	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.010000	0.000000
000	0.0	0.01	0.000000	0.0	0.0	0.000000	0.080000	0.0	0.020000	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.010000	0.020000	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.000000	0.000000

[30]:
```
manhattan_grouped.shape
```

[30]: (40, 339)

[68]:
```
result_df = manhattan_grouped[manhattan_grouped['Yoga Studio'] > 0.0]
```

[90]:
```
gym_result_df = manhattan_grouped[manhattan_grouped['Gym'] > 0.0]
```

These are the neighborhoods with highest number of Yoga Studios in Manhattan

[75]:
```
result_df.sort_values(by='Yoga Studio',ascending=False)
```

[75]:

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
22	Marble Hill	0.000000	0.000000	0.0	0.0	0.043478	0.00	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.043478
1	Carnegie Hill	0.000000	0.000000	0.0	0.0	0.011236	0.00	0.011236	0.000000	0.022472	...	0.0	0.011236	0.000000	0.0	0.000000	0.011236	0.044944	0.000000	0.011236	0.033708
13	Hamilton Heights	0.000000	0.016667	0.0	0.0	0.000000	0.00	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.016667	0.000000	0.000000	0.000000	0.033333
35	Upper East Side	0.000000	0.000000	0.0	0.0	0.051020	0.00	0.000000	0.020408	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.030612	0.000000	0.020408	0.030612
5	Civic Center	0.000000	0.000000	0.0	0.0	0.040000	0.01	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.010000	0.020000	0.010000	0.000000	0.030000
10	Flatiron	0.000000	0.000000	0.0	0.0	0.040000	0.00	0.000000	0.010000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.030000	0.000000	0.010000	0.030000
20	Manhattan Valley	0.000000	0.000000	0.0	0.0	0.021739	0.00	0.000000	0.000000	0.000000	...	0.0	0.043478	0.000000	0.0	0.000000	0.000000	0.021739	0.021739	0.000000	0.021739
19	Lower East Side	0.000000	0.000000	0.0	0.0	0.000000	0.00	0.020408	0.061224	0.000000	...	0.0	0.020408	0.000000	0.0	0.000000	0.000000	0.020408	0.000000	0.020408	0.020408
15	Inwood	0.000000	0.000000	0.0	0.0	0.018182	0.00	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.036364	0.018182	0.000000	0.000000	0.018182
33	Tudor City	0.000000	0.000000	0.0	0.0	0.012500	0.00	0.000000	0.000000	0.000000	...	0.0	0.025000	0.000000	0.0	0.000000	0.000000	0.025000	0.000000	0.000000	0.012500
32	Tribeca	0.000000	0.000000	0.0	0.0	0.064516	0.00	0.010753	0.010753	0.000000	...	0.0	0.000000	0.010753	0.0	0.010753	0.043011	0.021505	0.000000	0.000000	0.010753
11	Gramercy	0.000000	0.000000	0.0	0.0	0.031915	0.00	0.000000	0.010638	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.000000	0.042553	0.000000	0.000000	0.010638
36	Upper West Side	0.010526	0.000000	0.0	0.0	0.021053	0.00	0.000000	0.000000	0.000000	...	0.0	0.010526	0.000000	0.0	0.000000	0.031579	0.010526	0.000000	0.000000	0.010526
17	Lincoln Square	0.000000	0.000000	0.0	0.0	0.020833	0.00	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.010417	0.031250	0.000000	0.000000	0.010417
12	Greenwich Village	0.000000	0.000000	0.0	0.0	0.020000	0.00	0.000000	0.000000	0.000000	...	0.0	0.020000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000
24	Midtown South	0.000000	0.000000	0.0	0.0	0.030000	0.00	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.0	0.000000	0.010000	0.000000	0.000000	0.000000	0.010000
27	Noho	0.000000	0.000000	0.0	0.0	0.010000	0.00	0.010000	0.030000	0.000000	...	0.0	0.010000	0.000000	0.0	0.010000	0.030000	0.020000	0.010000	0.000000	0.010000
31	Sutton Place	0.000000	0.010000	0.0	0.0	0.030000	0.00	0.000000	0.000000	0.000000	...	0.0	0.010000	0.000000	0.0	0.000000	0.010000	0.020000	0.000000	0.000000	0.010000

18 rows × 339 columns

These are the Neighborhoods with highest number of Yoga Studios, Gyms and Gym/Fitness centers

[159]:
```
gym_result_df = manhattan_grouped.loc[(manhattan_grouped['Yoga Studio'] > 0.0) | (manhattan_grouped['Gym'] > 0.0) | (manhattan_grouped['Gym / Fitness Center'] > 0.0) | (manhattan_grouped['Gymnastics Gym'] > 0.(
#gym_result_df['Neighborhood'], ['Gym']
```

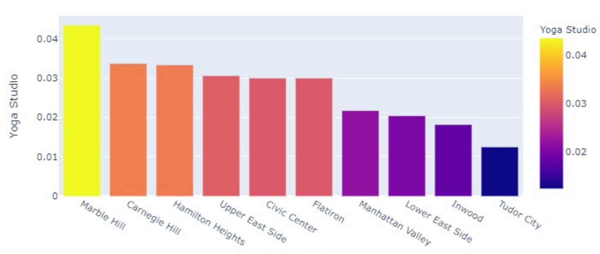
[161]:
```
show_df = gym_result_df[['Neighborhood', 'Yoga Studio','Gym','Gym / Fitness Center','Gymnastics Gym', 'Café']]
show_df_10 = show_df.sort_values(by='Yoga Studio',ascending=False).head(10)
show_df_10
```

[161]:

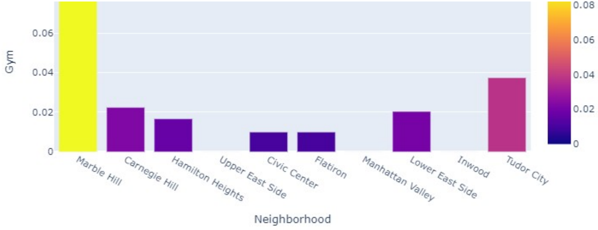
	Neighborhood	Yoga Studio	Gym	Gym / Fitness Center	Gymnastics Gym	Café
22	Marble Hill	0.043478	0.086957	0.000000	0.0	0.000000
1	Carnegie Hill	0.033708	0.022472	0.033708	0.0	0.067416
13	Hamilton Heights	0.033333	0.016667	0.000000	0.0	0.066667
35	Upper East Side	0.030612	0.000000	0.040816	0.0	0.000000
5	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

[122]:
```
import matplotlib.pyplot as plt
#import plotly.plotly as py
import plotly.express as px
```

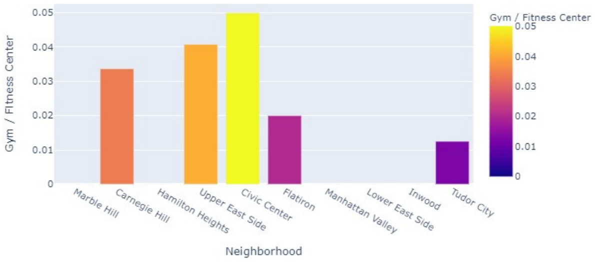
[174]:
```
fig = px.bar(show_df_10, x='Neighborhood', y='Yoga Studio', color='Yoga Studio', width=800, height=400)
fig.show()
```



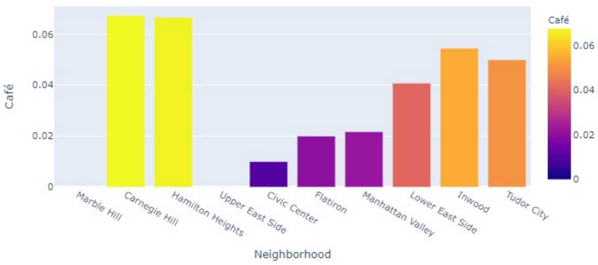
[175]:
```
fig = px.bar(show_df_10, x='Neighborhood', y='Gym', color='Gym', width=800, height=400)
fig.show()
```




```
[176]: fig = px.bar(show_df_10, x='Neighborhood', y='Gym / Fitness Center', color='Gym / Fitness Center', width=800, height=400)
fig.show()
```




```
[177]: fig = px.bar(show_df_10, x='Neighborhood', y='Café', color='Café', width=800, height=400)
fig.show()
```



Map to show the top 4 locations with potential to open and sustain a yoga studio

```
[150]: UES_data = new_york_data.loc[(new_york_data['Borough'] == 'Manhattan') & ((new_york_data['Neighborhood'] == 'Marble Hill') | (new_york_data['Neighborhood'] == 'Carnegie Hill') |(new_york_data['Neighborhood'] ==
UES_data.head()
```

[150]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Hamilton Heights	40.823604	-73.949688
2	Manhattan	Upper East Side	40.775639	-73.960508
3	Manhattan	Carnegie Hill	40.782683	-73.953256

```
[151]: address = 'Marble Hill, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Upper East Side are {}, {}.'.format(latitude, longitude))
```

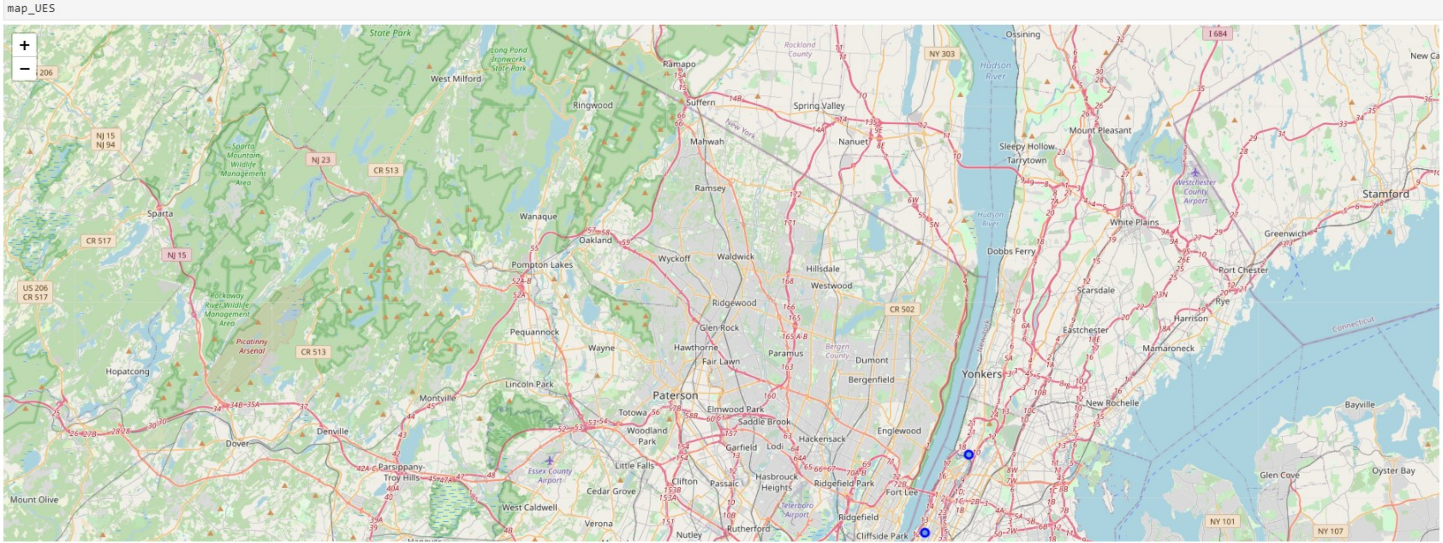
The geograpical coordinate of Upper East Side are 40.8762983, -73.9104292.

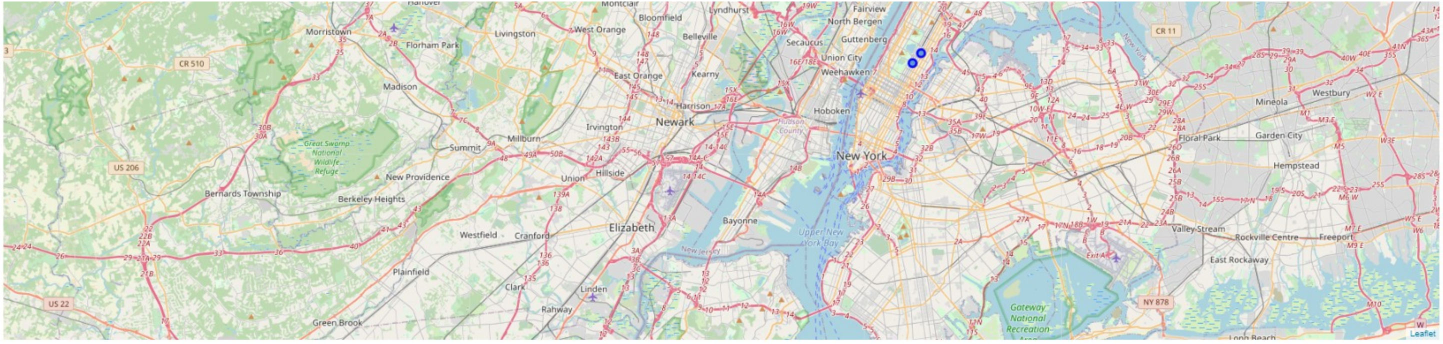
```
[152]: # create map of Manhattan using latitude and longitude values
map_UES = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, label in zip(UES_data['Latitude'], UES_data['Longitude'], UES_data['Neighborhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_UES)

map_UES
```

[152]:





Top Venues for each Neighborhood

[72]:
```python
num_top_venues = 5

for hood in manhattan_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = manhattan_grouped[manhattan_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Battery Park City----
                       venue  freq
0           Accessories Store  87.0
1                Optical Shop  87.0
2                 Pastry Shop  87.0
3                        Park  87.0
4  Paper / Office Supplies Store  87.0


----Carnegie Hill----
                       venue  freq
0           Accessories Store  89.0
1                Optical Shop  89.0
2                 Pastry Shop  89.0
3                        Park  89.0
4  Paper / Office Supplies Store  89.0


----Central Harlem----
                       venue  freq
0           Accessories Store  46.0
1                Optical Shop  46.0
2                 Pastry Shop  46.0
3                        Park  46.0
4  Paper / Office Supplies Store  46.0
```

```python
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = manhattan_grouped['Neighborhood']

for ind in np.arange(manhattan_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(manhattan_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted
```

[36]:

	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	Gym / Fitness Center	Spa	French Restaurant	American 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	Park	Cocktail Bar	Italian Restaurant
10	Flatiron	Italian Restaurant	Japanese Restaurant	New American Restaurant	American Restaurant	Mediterranean Restaurant
11	Gramercy	Italian Restaurant	Bar	Pizza Place	Bagel Shop	Wine Shop
12	Greenwich Village	Italian Restaurant	Clothing Store	Boutique	Sushi Restaurant	Dessert Shop
13	Hamilton Heights	Pizza Place	Café	Mexican Restaurant	Deli / Bodega	Coffee Shop
14	Hudson Yards	Gym / Fitness Center	Italian Restaurant	Hotel	American Restaurant	Thai Restaurant
15	Inwood	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant
16	Lenox Hill	Italian Restaurant	Coffee Shop	Sushi Restaurant	Pizza Place	Café
17	Lincoln Square	Plaza	Performing Arts Venue	Concert Hall	Café	Theater
18	Little Italy	Bakery	Café	Chinese Restaurant	Italian Restaurant	Cocktail Bar
19	Lower East Side	Chinese Restaurant	Art Gallery	Latin American Restaurant	Pharmacy	Pizza Place
20	Manhattan Valley	Mexican Restaurant	Bar	Coffee Shop	Vietnamese Restaurant	Playground
21	Manhattanville	Coffee Shop	Mexican Restaurant	Italian Restaurant	Bar	Seafood Restaurant
22	Marble Hill	Coffee Shop	Sandwich Place	Gym	Yoga Studio	Bank
23	Midtown	Hotel	Sporting Goods Shop	Clothing Store	Coffee Shop	Bookstore
24	Midtown South	Korean Restaurant	Hotel	Dessert Shop	Gym / Fitness Center	Hotel Bar
25	Morningside Heights	Coffee Shop	Park	Bookstore	American Restaurant	Burger Joint
26	Murray Hill	Hotel	Coffee Shop	American Restaurant	Sandwich Place	Japanese Restaurant
27	Noho	Italian Restaurant	Cocktail Bar	French Restaurant	Pizza Place	Hotel
28	Roosevelt Island	Deli / Bodega	Coffee Shop	Park	Bus Line	Bubble Tea Shop
29	Soho	Clothing Store	Boutique	Italian Restaurant	Mediterranean Restaurant	Shoe Store
30	Stuyvesant Town	Bar	Park	Coffee Shop	Cocktail Bar	Playground
31	Sutton Place	Italian Restaurant	Park	Pizza Place	Gym / Fitness Center	Indian Restaurant
32	Tribeca	American Restaurant	Park	Italian Restaurant	Wine Bar	Café
33	Tudor City	Park	Mexican Restaurant	Café	Pizza Place	Gym
34	Turtle Bay	Italian Restaurant	Coffee Shop	Sushi Restaurant	Deli / Bodega	Café
35	Upper East Side	Exhibit	Italian Restaurant	American Restaurant	Juice Bar	Gym / Fitness Center
36	Upper West Side	Italian Restaurant	Bakery	Wine Bar	Bar	Café
37	Washington Heights	Café	Bakery	Bank	Mobile Phone Shop	Pizza Place
38	West Village	Italian Restaurant	New American Restaurant	Cocktail Bar	American Restaurant	Park
39	Yorkville	Italian Restaurant	Gym	Coffee Shop	Deli / Bodega	Sushi Restaurant

[35]:
```python
neighborhoods_venues_sorted.shape
```

[35]: (40, 6)

Cluster Neighborhoods

Run K-means to cluster neighborhood into five clusters

[37]:
```python
# set number of clusters
kclusters = 5

manhattan_grouped_clustering = manhattan_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(manhattan_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

[37]:
```
array([1, 1, 1, 1, 1, 1, 1, 0, 1, 1], dtype=int32)
```

[38]:
```python
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

manhattan_merged = manhattan_data

# merge manhattan_grouped with manhattan_data to add latitude/longitude for each neighborhood
manhattan_merged = manhattan_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

manhattan_merged.head() # check the last columns!
```

[38]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Coffee Shop	Sandwich Place	Gym	Yoga Studio	Bank
1	Manhattan	Chinatown	40.715618	-73.994279	1	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Bank	Mobile Phone Shop	Pizza Place
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Café	Mexican Restaurant	Deli / Bodega	Coffee Shop

[39]:
```python
manhattan_merged.shape
```

[39]:
```
(40, 10)
```

[40]:
```python
manhattan_merged
```

[40]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Coffee Shop	Sandwich Place	Gym	Yoga Studio	Bank
1	Manhattan	Chinatown	40.715618	-73.994279	1	Chinese Restaurant	Bakery	Cocktail Bar	Dessert Shop	American Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Bank	Mobile Phone Shop	Pizza Place
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Café	Mexican Restaurant	Deli / Bodega	Coffee Shop
5	Manhattan	Manhattanville	40.816934	-73.957385	2	Coffee Shop	Mexican Restaurant	Italian Restaurant	Bar	Seafood Restaurant
6	Manhattan	Central Harlem	40.815976	-73.943211	1	African Restaurant	Chinese Restaurant	Public Art	French Restaurant	American Restaurant
7	Manhattan	East Harlem	40.792249	-73.944182	0	Bakery	Mexican Restaurant	Thai Restaurant	Sandwich Place	Latin American Restaurant
8	Manhattan	Upper East Side	40.775639	-73.960508	1	Exhibit	Italian Restaurant	American Restaurant	Juice Bar	Gym / Fitness Center
9	Manhattan	Yorkville	40.775930	-73.947118	2	Italian Restaurant	Gym	Coffee Shop	Deli / Bodega	Sushi Restaurant
10	Manhattan	Lenox Hill	40.768113	-73.958860	2	Italian Restaurant	Coffee Shop	Sushi Restaurant	Pizza Place	Café
11	Manhattan	Roosevelt Island	40.762160	-73.949168	2	Deli / Bodega	Coffee Shop	Park	Bus Line	Bubble Tea Shop
12	Manhattan	Upper West Side	40.787658	-73.977059	2	Italian Restaurant	Bakery	Wine Bar	Bar	Café
13	Manhattan	Lincoln Square	40.773529	-73.985338	2	Plaza	Performing Arts Venue	Concert Hall	Café	Theater
14	Manhattan	Clinton	40.759101	-73.996119	1	Italian Restaurant	Theater	Gym / Fitness Center	Sandwich Place	Coffee Shop
15	Manhattan	Midtown	40.754691	-73.981669	1	Hotel	Sporting Goods Shop	Clothing Store	Coffee Shop	Bookstore
16	Manhattan	Murray Hill	40.748303	-73.978332	1	Hotel	Coffee Shop	American Restaurant	Sandwich Place	Japanese Restaurant
17	Manhattan	Chelsea	40.744035	-74.003116	1	Coffee Shop	Bakery	Art Gallery	American Restaurant	Italian Restaurant
18	Manhattan	Greenwich Village	40.726933	-73.999914	2	Italian Restaurant	Clothing Store	Boutique	Sushi Restaurant	Dessert Shop
19	Manhattan	East Village	40.727847	-73.982226	1	Bar	Mexican Restaurant	Speakeasy	Pizza Place	Ice Cream Shop
20	Manhattan	Lower East Side	40.717807	-73.980890	0	Chinese Restaurant	Art Gallery	Latin American Restaurant	Pharmacy	Pizza Place
21	Manhattan	Tribeca	40.721522	-74.010683	2	American Restaurant	Park	Italian Restaurant	Wine Bar	Café
22	Manhattan	Little Italy	40.719324	-73.997305	1	Bakery	Café	Chinese Restaurant	Italian Restaurant	Cocktail Bar
23	Manhattan	Soho	40.722184	-74.000657	2	Clothing Store	Boutique	Italian Restaurant	Mediterranean Restaurant	Shoe Store
24	Manhattan	West Village	40.734434	-74.006180	2	Italian Restaurant	New American Restaurant	Cocktail Bar	American Restaurant	Park
25	Manhattan	Manhattan Valley	40.797307	-73.964286	1	Mexican Restaurant	Bar	Coffee Shop	Vietnamese Restaurant	Playground
26	Manhattan	Morningside Heights	40.808000	-73.963896	2	Coffee Shop	Park	Bookstore	American Restaurant	Burger Joint
27	Manhattan	Gramercy	40.737210	-73.981376	1	Italian Restaurant	Bar	Pizza Place	Bagel Shop	Wine Shop
28	Manhattan	Battery Park City	40.711932	-74.016869	1	Park	Coffee Shop	Hotel	Clothing Store	Women's Store
29	Manhattan	Financial District	40.707107	-74.010665	1	Coffee Shop	Pizza Place	Park	Cocktail Bar	Italian Restaurant
30	Manhattan	Carnegie Hill	40.782683	-73.953256	1	Coffee Shop	Café	Wine Shop	Yoga Studio	Bar
31	Manhattan	Noho	40.723259	-73.988434	2	Italian Restaurant	Cocktail Bar	French Restaurant	Pizza Place	Hotel
32	Manhattan	Civic Center	40.715229	-74.005415	1	Coffee Shop	Gym / Fitness Center	Spa	French Restaurant	American Restaurant
33	Manhattan	Midtown South	40.748510	-73.988713	1	Korean Restaurant	Hotel	Dessert Shop	Gym / Fitness Center	Hotel Bar
34	Manhattan	Sutton Place	40.760280	-73.963556	1	Italian Restaurant	Park	Pizza Place	Gym / Fitness Center	Indian Restaurant
35	Manhattan	Turtle Bay	40.752042	-73.967708	2	Italian Restaurant	Coffee Shop	Sushi Restaurant	Deli / Bodega	Café
36	Manhattan	Tudor City	40.746917	-73.971219	2	Park	Mexican Restaurant	Café	Pizza Place	Gym
37	Manhattan	Stuyvesant Town	40.731000	-73.974052	3	Bar	Park	Coffee Shop	Cocktail Bar	Playground
38	Manhattan	Flatiron	40.739673	-73.990947	1	Italian Restaurant	Japanese Restaurant	New American Restaurant	American Restaurant	Mediterranean Restaurant
39	Manhattan	Hudson Yards	40.756658	-74.000111	1	Gym / Fitness Center	Italian Restaurant	Hotel	American Restaurant	Thai Restaurant

Visualize the clusters

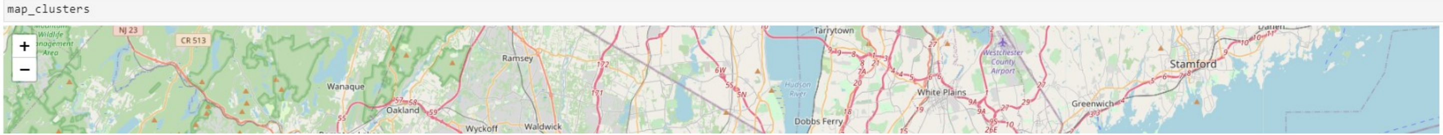
[43]:
```python
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)
```

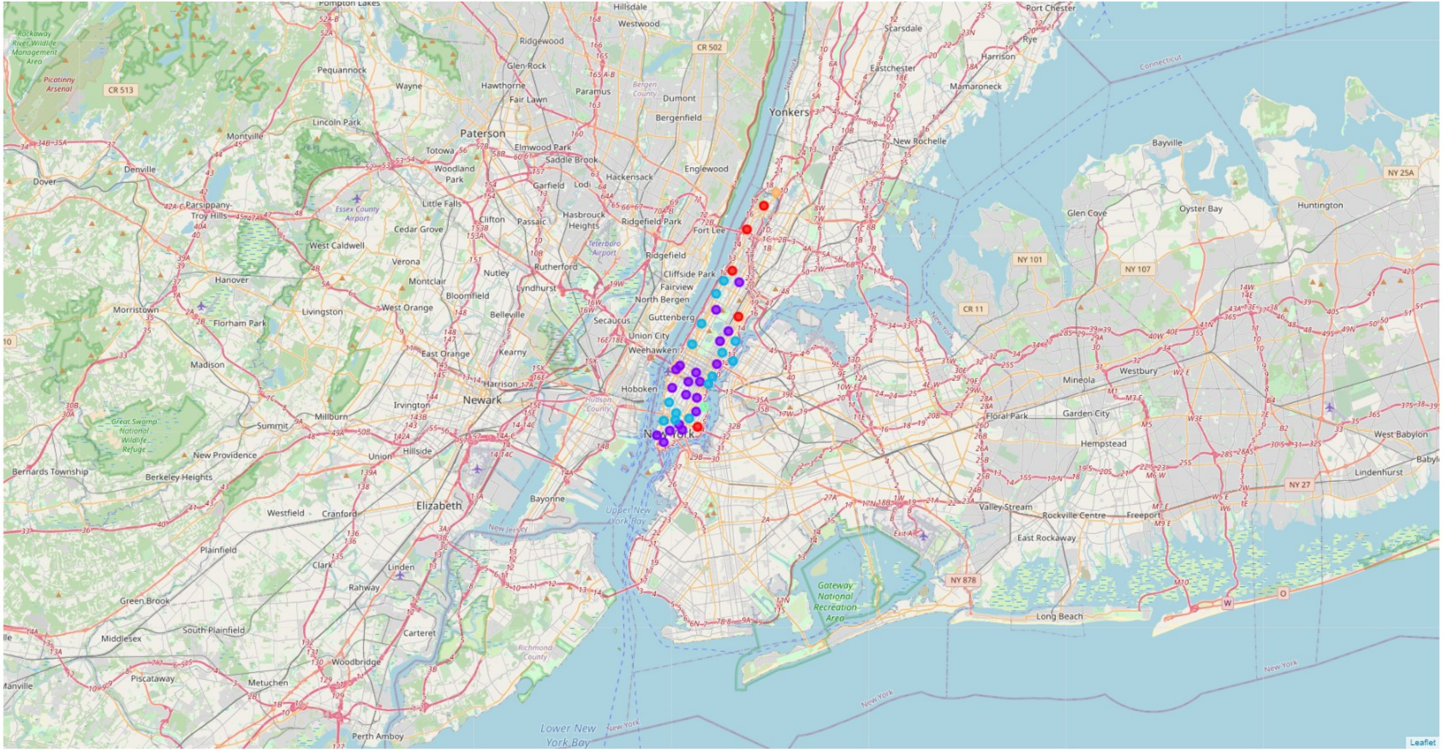
[44]:
```python
# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]
```

[45]:
```python
# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(manhattan_merged['Latitude'], manhattan_merged['Longitude'], manhattan_merged['Neighborhood'], manhattan_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)
```

[46]:
```python
map_clusters
```

[46]:





Examine Clusters

Cluster 1

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

[47]:

	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

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

[115]:

	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	Italian Restaurant	Cocktail Bar
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	Park	Cocktail Bar	Italian Restaurant
30	Carnegie Hill	Coffee Shop	Café	Wine Shop	Yoga Studio	Bar
32	Civic Center	Coffee Shop	Gym / Fitness Center	Spa	French Restaurant	American Restaurant
33	Midtown South	Korean Restaurant	Hotel	Dessert Shop	Gym / Fitness Center	Hotel Bar
34	Sutton Place	Italian Restaurant	Park	Pizza Place	Gym / Fitness Center	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

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

[116]:

	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

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

[117]:

	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

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

[118]:

	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

Results:

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!

[ ]: