

Segmenting and Clustering Neighborhoods in New York City

Introduction

In this lab, you will learn how to convert addresses into their equivalent latitude and longitude values. Also, you will use the Foursquare API to explore neighborhoods in New York City. You will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. You will use the k-means clustering algorithm to complete this task. Finally, you will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters.

Table of Contents

1. Download and Explore Dataset
2. Explore Neighborhoods in New York City
3. Analyze Each Neighborhood
4. Cluster Neighborhoods
5. Examine Clusters

Before we get the data and start exploring it, let's download all the dependencies that we will need.

```
In [1]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#conda install -c conda-forge geoppy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geoppy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

Libraries imported.

1. Download and Explore Dataset

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

For your convenience, I downloaded the files and placed it on the server, so you can simply run a `wget` command and access the data. So let's go ahead and do that.

```
In [2]: # wget -q -O 'newyork_data.json' https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IB
# Move iperskillsnewwork-cop701ism-SkillsNetwork/labs/newyork_data.json
# print('Data downloaded!')
```

Load and explore the data

Next, let's load the data.

```
In [3]: with open('newyork_data.json') as json_data:
newyork_data = json.load(json_data)
```

Let's take a quick look at the data.

[illegible]

[illegible]


```
"stacked": 2,  
"name": "Beadows",  
"annolinz": "Beaowes",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.7960319588289, 40.7046573606817],  
"id": "myuu_2451_34572.161",  
"geomtry_name": "geom",  
"properties": {"name": "Jamica Center"},  
"stacked": 2,  
"annolinz": "Jamica",  
"annolinz2": "Center",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.7866750567424, 40.7046573606817],  
"id": "myuu_2451_34572.162",  
"geomtry_name": "geom",  
"properties": {"name": "Oakland Gardens"},  
"stacked": 2,  
"annolinz": "Oakland",  
"annolinz2": "Gardens",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.7592500393594, 40.7124344193104],  
"id": "myuu_2451_34572.163",  
"geomtry_name": "geom",  
"properties": {"name": "South Jamaica"},  
"stacked": 2,  
"annolinz": "South Jamaica",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.7592500393594, 40.6961123793898],  
"id": "myuu_2451_34572.164",  
"geomtry_name": "geom",  
"properties": {"name": "St Albans"},  
"stacked": 2,  
"annolinz": "St Albans",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.758676327717, 40.694453852359],  
"id": "myuu_2451_34572.165",  
"geomtry_name": "geom",  
"properties": {"name": "Springfield Gardens"},  
"stacked": 2,  
"annolinz": "Springfield",  
"annolinz2": "Gardens",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6662304936858],  
"id": "myuu_2451_34572.166",  
"geomtry_name": "geom",  
"properties": {"name": "Camria Heights"},  
"stacked": 2,  
"annolinz": "Camria",  
"annolinz2": "Heights",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.167",  
"geomtry_name": "geom",  
"properties": {"name": "Brookdale"},  
"stacked": 2,  
"annolinz": "Brookdale",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.168",  
"geomtry_name": "geom",  
"properties": {"name": "Broad Channel"},  
"stacked": 2,  
"annolinz": "Broad",  
"annolinz2": "Channel",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.169",  
"geomtry_name": "geom",  
"properties": {"name": "Breezy Point"},  
"stacked": 2,  
"annolinz": "Breezy",  
"annolinz2": "Point",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.170",  
"geomtry_name": "geom",  
"properties": {"name": "Edgemere"},  
"stacked": 2,  
"annolinz": "Edgemere",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.171",  
"geomtry_name": "geom",  
"properties": {"name": "Forest Hills"},  
"stacked": 2,  
"annolinz": "Forest",  
"annolinz2": "Hills",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.172",  
"geomtry_name": "geom",  
"properties": {"name": "Leffrak City"},  
"stacked": 2,  
"annolinz": "Leffrak",  
"annolinz2": "City",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.173",  
"geomtry_name": "geom",  
"properties": {"name": "Bellevue"},  
"stacked": 2,  
"annolinz": "Bellevue",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.174",  
"geomtry_name": "geom",  
"properties": {"name": "Brighton"},  
"stacked": 2,  
"annolinz": "Brighton",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.175",  
"geomtry_name": "geom",  
"properties": {"name": "Richmond"},  
"stacked": 2,  
"annolinz": "Richmond",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.176",  
"geomtry_name": "geom",  
"properties": {"name": "Somerville"},  
"stacked": 2,  
"annolinz": "Somerville",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.177",  
"geomtry_name": "geom",  
"properties": {"name": "Brooklynville"},  
"stacked": 2,  
"annolinz": "Brooklyn",  
"annolinz2": "ville",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.178",  
"geomtry_name": "geom",  
"properties": {"name": "Staten Island"},  
"stacked": 2,  
"annolinz": "Staten",  
"annolinz2": "Island",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.179",  
"geomtry_name": "geom",  
"properties": {"name": "Rosedale"},  
"stacked": 2,  
"annolinz": "Rosedale",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.180",  
"geomtry_name": "geom",  
"properties": {"name": "Tottus"},  
"stacked": 2,  
"annolinz": "Tottus",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.181",  
"geomtry_name": "geom",  
"properties": {"name": "North Corona"},  
"stacked": 2,  
"annolinz": "North",  
"annolinz2": "Corona",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.182",  
"geomtry_name": "geom",  
"properties": {"name": "Forest Hills Gardens"},  
"stacked": 2,  
"annolinz": "Forest",  
"annolinz2": "Hills Gardens",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.183",  
"geomtry_name": "geom",  
"properties": {"name": "Forest Hills"},  
"stacked": 2,  
"annolinz": "Forest",  
"annolinz2": "Hills",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.184",  
"geomtry_name": "geom",  
"properties": {"name": "Brooklynville"},  
"stacked": 2,  
"annolinz": "Brooklyn",  
"annolinz2": "ville",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.185",  
"geomtry_name": "geom",  
"properties": {"name": "Staten Island"},  
"stacked": 2,  
"annolinz": "Staten",  
"annolinz2": "Island",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.186",  
"geomtry_name": "geom",  
"properties": {"name": "Rosedale"},  
"stacked": 2,  
"annolinz": "Rosedale",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.187",  
"geomtry_name": "geom",  
"properties": {"name": "Tottus"},  
"stacked": 2,  
"annolinz": "Tottus",  
"annolinz2": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.188",  
"geomtry_name": "geom",  
"properties": {"name": "North Corona"},  
"stacked": 2,  
"annolinz": "North",  
"annolinz2": "Corona",  
"annolinz": "None",  
"annangle": 0.0,  
"bbox": [-73.752500393594, 40.6927463916045],  
"id": "myuu_2451_34572.189",  
"
```


[illegible]


```

"geometry": {"type": "Point",
             "coordinates": [-73.9671678028349, 40.733500254297573]},
"geometry_name": "geom",
"properties": {"name": "Utopia",
               "annolinel1": "Utopia",
               "annolinel2": "Utopia",
               "annolinel3": None,
               "annolinel4": None,
               "annolinel5": None,
               "borough": "Queens",
               "bbox": [-73.79671678028349,
                        -73.79671678028349,
                        40.733500254297573]},
('type': 'Feature',
 'id': 'nyu_2451_34572.266',
 'geometry': {'type': 'Point',
              'coordinates': [-73.80486120040537, 40.734936180754781]},
 'geometry_name': 'geom',
 'properties': {'name': 'Fomonok',
               'annolinel1': "Fomonok",
               'annolinel2': "Fomonok",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.80486120040537,
                        40.734936180754781,
                        40.734936180754781]},
('type': 'Feature',
 'id': 'nyu_2451_34572.267',
 'geometry': {'type': 'Point',
              'coordinates': [-73.89467996270574, 40.77031739299822]},
 'geometry_name': 'geom',
 'properties': {'name': 'Astoria Heights',
               'annolinel1': "Astoria",
               'annolinel2': "Astoria",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.89467996270574,
                        40.77031739299822,
                        40.77031739299822]},
('type': 'Feature',
 'id': 'nyu_2451_34572.268',
 'geometry': {'type': 'Point',
              'coordinates': [-73.90119903387667, 40.831428341615481]},
 'geometry_name': 'geom',
 'properties': {'name': 'Claremont Village',
               'annolinel1': "Claremont",
               'annolinel2': "Claremont",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.90119903387667,
                        40.831428341615481,
                        40.831428341615481]},
('type': 'Feature',
 'id': 'nyu_2451_34572.269',
 'geometry': {'type': 'Point',
              'coordinates': [-73.91584652759009, 40.824780490842905]},
 'geometry_name': 'geom',
 'properties': {'name': 'Concourse Village',
               'annolinel1': "Concourse",
               'annolinel2': "Village",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.91584652759009,
                        40.824780490842905,
                        40.824780490842905]},
('type': 'Feature',
 'id': 'nyu_2451_34572.270',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9165551964419, 40.848362176716541]},
 'geometry_name': 'geom',
 'properties': {'name': 'Mount Eden',
               'annolinel1': "Mount",
               'annolinel2': "Eden",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9165551964419,
                        40.848362176716541,
                        40.848362176716541]},
('type': 'Feature',
 'id': 'nyu_2451_34572.271',
 'geometry': {'type': 'Point',
              'coordinates': [-73.90829930881988, 40.848841607246651]},
 'geometry_name': 'geom',
 'properties': {'name': 'Mount Hope',
               'annolinel1': "Mount",
               'annolinel2': "Hope",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.90829930881988,
                        40.848841607246651,
                        40.848841607246651]},
('type': 'Feature',
 'id': 'nyu_2451_34572.272',
 'geometry': {'type': 'Point',
              'coordinates': [-73.91894796251049303, 40.760280331313741]},
 'geometry_name': 'geom',
 'properties': {'name': 'Sutton Place',
               'annolinel1': "Sutton",
               'annolinel2': "Place",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.91894796251049303,
                        40.760280331313741,
                        40.760280331313741]},
('type': 'Feature',
 'id': 'nyu_2451_34572.273',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9538678230745, 40.7434140900735361]},
 'geometry_name': 'geom',
 'properties': {'name': 'Hunters Point',
               'annolinel1': "Hunters",
               'annolinel2': "Point",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9538678230745,
                        40.7434140900735361,
                        40.7434140900735361]},
('type': 'Feature',
 'id': 'nyu_2451_34572.274',
 'geometry': {'type': 'Point',
              'coordinates': [-73.96770824581834, 40.752042369507221]},
 'geometry_name': 'geom',
 'properties': {'name': 'Turtle Bay',
               'annolinel1': "Turtle",
               'annolinel2': "Bay",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.96770824581834,
                        40.752042369507221,
                        40.752042369507221]},
('type': 'Feature',
 'id': 'nyu_2451_34572.275',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9712192872265, 40.7469174107401951]},
 'geometry_name': 'geom',
 'properties': {'name': 'Tudor City',
               'annolinel1': "Tudor",
               'annolinel2': "City",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9712192872265,
                        40.7469174107401951,
                        40.7469174107401951]},
('type': 'Feature',
 'id': 'nyu_2451_34572.276',
 'geometry': {'type': 'Point',
              'coordinates': [-73.97405170469203, 40.730999554770611]},
 'geometry_name': 'geom',
 'properties': {'name': 'Stuyvesant Town',
               'annolinel1': "Stuyvesant",
               'annolinel2': "Town",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.97405170469203,
                        40.730999554770611,
                        40.730999554770611]},
('type': 'Feature',
 'id': 'nyu_2451_34572.277',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9909471052826, 40.7396730476384261]},
 'geometry_name': 'geom',
 'properties': {'name': 'Flatiron',
               'annolinel1': "Flatiron",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9909471052826,
                        40.7396730476384261,
                        40.7396730476384261]},
('type': 'Feature',
 'id': 'nyu_2451_34572.278',
 'geometry': {'type': 'Point',
              'coordinates': [-73.91819286431682, 40.745651806080761]},
 'geometry_name': 'geom',
 'properties': {'name': 'Sunnyside Gardens',
               'annolinel1': "Sunnyside",
               'annolinel2': "Gardens",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.91819286431682,
                        40.745651806080761,
                        40.745651806080761]},
('type': 'Feature',
 'id': 'nyu_2451_34572.279',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9324425260178, 40.737250716944971]},
 'geometry_name': 'geom',
 'properties': {'name': 'Blissville',
               'annolinel1': "Blissville",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9324425260178,
                        40.737250716944971,
                        40.737250716944971]},
('type': 'Feature',
 'id': 'nyu_2451_34572.280',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9857528074504, 40.70317632826921]},
 'geometry_name': 'geom',
 'properties': {'name': 'Fulton Ferry',
               'annolinel1': "Fulton",
               'annolinel2': "Ferry",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9857528074504,
                        40.70317632826921,
                        40.70317632826921]},
('type': 'Feature',
 'id': 'nyu_2451_34572.281',
 'geometry': {'type': 'Point',
              'coordinates': [-73.9811603592393, 40.703321498828741]},
 'geometry_name': 'geom',
 'properties': {'name': 'Vinegar Hill',
               'annolinel1': "Vinegar",
               'annolinel2': "Hill",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.9811603592393,
                        40.703321498828741,
                        40.703321498828741]},
('type': 'Feature',
 'id': 'nyu_2451_34572.282',
 'geometry': {'type': 'Point',
              'coordinates': [-73.93053108817338, 40.675039865032371]},
 'geometry_name': 'geom',
 'properties': {'name': 'Weeksville',
               'annolinel1': "Weeksville",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.93053108817338,
                        40.675039865032371,
                        40.675039865032371]},
('type': 'Feature',
 'id': 'nyu_2451_34572.283',
 'geometry': {'type': 'Point',
              'coordinates': [-73.90331684852599, 40.677861047695311]},
 'geometry_name': 'geom',
 'properties': {'name': 'Broadway Junction',
               'annolinel1': "Broadway",
               'annolinel2': "Junction",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.90331684852599,
                        40.677861047695311,
                        40.677861047695311]},
('type': 'Feature',
 'id': 'nyu_2451_34572.284',
 'geometry': {'type': 'Point',
              'coordinates': [-73.987528074504, 40.70317632826921]},
 'geometry_name': 'geom',
 'properties': {'name': 'Dumbo',
               'annolinel1': "Dumbo",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.987528074504,
                        40.70317632826921,
                        40.70317632826921]},
('type': 'Feature',
 'id': 'nyu_2451_34572.285',
 'geometry': {'type': 'Point',
              'coordinates': [-74.12059399718001, 40.6018095976314441]},
 'geometry_name': 'geom',
 'properties': {'name': 'Manor Heights',
               'annolinel1': "Manor",
               'annolinel2': "Heights",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.12059399718001,
                        40.6018095976314441,
                        40.6018095976314441]},
('type': 'Feature',
 'id': 'nyu_2451_34572.286',
 'geometry': {'type': 'Point',
              'coordinates': [-74.13208447484298, 40.603706926273711]},
 'geometry_name': 'geom',
 'properties': {'name': 'Willowbrook',
               'annolinel1': "Willowbrook",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.13208447484298,
                        40.603706926273711,
                        40.603706926273711]},
('type': 'Feature',
 'id': 'nyu_2451_34572.287',
 'geometry': {'type': 'Point',
              'coordinates': [-74.217766306968567, 40.5411399220917661]},
 'geometry_name': 'geom',
 'properties': {'name': 'Sandy Ground',
               'annolinel1': "Sandy",
               'annolinel2': "Ground",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.217766306968567,
                        40.5411399220917661,
                        40.5411399220917661]},
('type': 'Feature',
 'id': 'nyu_2451_34572.288',
 'geometry': {'type': 'Point',
              'coordinates': [-74.12727240604946, 40.5791187429612141]},
 'geometry_name': 'geom',
 'properties': {'name': 'Egbertville',
               'annolinel1': "Egbertville",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.12727240604946,
                        40.5791187429612141,
                        40.5791187429612141]},
('type': 'Feature',
 'id': 'nyu_2451_34572.289',
 'geometry': {'type': 'Point',
              'coordinates': [-73.89213760232822, 40.567375889570321]},
 'geometry_name': 'geom',
 'properties': {'name': 'Roxbury',
               'annolinel1': "Roxbury",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.89213760232822,
                        40.567375889570321,
                        40.567375889570321]},
('type': 'Feature',
 'id': 'nyu_2451_34572.290',
 'geometry': {'type': 'Point',
              'coordinates': [-73.958520951372551, 40.598520951372551]},
 'geometry_name': 'geom',
 'properties': {'name': 'Homestead',
               'annolinel1': "Homestead",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.958520951372551,
                        40.598520951372551,
                        40.598520951372551]},
('type': 'Feature',
 'id': 'nyu_2451_34572.291',
 'geometry': {'type': 'Point',
              'coordinates': [-73.8814319200604, 40.716414511581851]},
 'geometry_name': 'geom',
 'properties': {'name': 'Middle Village',
               'annolinel1': "Middle",
               'annolinel2': "Village",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.8814319200604,
                        40.716414511581851,
                        40.716414511581851]},
('type': 'Feature',
 'id': 'nyu_2451_34572.292',
 'geometry': {'type': 'Point',
              'coordinates': [-74.2015255657658, 40.526264067348121]},
 'geometry_name': 'geom',
 'properties': {'name': "Prince's Bay",
               'annolinel1': "Bay",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.2015255657658,
                        40.526264067348121,
                        40.526264067348121]},
('type': 'Feature',
 'id': 'nyu_2451_34572.293',
 'geometry': {'type': 'Point',
              'coordinates': [-74.1379266371568, 40.576506293794891]},
 'geometry_name': 'geom',
 'properties': {'name': 'Lighthouse Hill',
               'annolinel1': "Lighthouse",
               'annolinel2': "Hill",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.1379266371568,
                        40.576506293794891,
                        40.576506293794891]},
('type': 'Feature',
 'id': 'nyu_2451_34572.294',
 'geometry': {'type': 'Point',
              'coordinates': [-74.22957080626941, 40.519541457488091]},
 'geometry_name': 'geom',
 'properties': {'name': 'Richmond Valley',
               'annolinel1': "Richmond",
               'annolinel2': "Valley",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.22957080626941,
                        40.519541457488091,
                        40.519541457488091]},
('type': 'Feature',
 'id': 'nyu_2451_34572.295',
 'geometry': {'type': 'Point',
              'coordinates': [-73.82667757138641, 40.790601556701481]},
 'geometry_name': 'geom',
 'properties': {'name': 'Malba',
               'annolinel1': "Malba",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.82667757138641,
                        40.790601556701481,
                        40.790601556701481]},
('type': 'Feature',
 'id': 'nyu_2451_34572.296',
 'geometry': {'type': 'Point',
              'coordinates': [-73.890345709872, 40.6819934517311]},
 'geometry_name': 'geom',
 'properties': {'name': 'Highland Park',
               'annolinel1': "Highland",
               'annolinel2': "Park",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.890345709872,
                        40.6819934517311,
                        40.6819934517311]},
('type': 'Feature',
 'id': 'nyu_2451_34572.298',
 'geometry': {'type': 'Point',
              'coordinates': [-73.86172577555115, 40.852722976330171]},
 'geometry_name': 'geom',
 'properties': {'name': 'Bronxdale',
               'annolinel1': "Bronxdale",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.86172577555115,
                        40.852722976330171,
                        40.852722976330171]},
('type': 'Feature',
 'id': 'nyu_2451_34572.299',
 'geometry': {'type': 'Point',
              'coordinates': [-73.85931863221647, 40.865787878029821]},
 'geometry_name': 'geom',
 'properties': {'name': 'Allerton',
               'annolinel1': "Allerton",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.85931863221647,
                        40.865787878029821,
                        40.865787878029821]},
('type': 'Feature',
 'id': 'nyu_2451_34572.300',
 'geometry': {'type': 'Point',
              'coordinates': [-73.90152264513144, 40.8703293914171]},
 'geometry_name': 'geom',
 'properties': {'name': 'Kingsbridge Heights',
               'annolinel1': "Kingsbridge",
               'annolinel2': "Heights",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.90152264513144,
                        40.8703293914171,
                        40.8703293914171]},
('type': 'Feature',
 'id': 'nyu_2451_34572.301',
 'geometry': {'type': 'Point',
              'coordinates': [-73.94817709920184, 40.64692606585791]},
 'geometry_name': 'geom',
 'properties': {'name': 'Baywater',
               'annolinel1': "Baywater",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.94817709920184,
                        40.64692606585791,
                        40.64692606585791]},
('type': 'Feature',
 'id': 'nyu_2451_34572.302',
 'geometry': {'type': 'Point',
              'coordinates': [-74.08179992211962, 40.613216912838341]},
 'geometry_name': 'geom',
 'properties': {'name': 'Fox Hills',
               'annolinel1': "Fox",
               'annolinel2': "Hills",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.08179992211962,
                        40.613216912838341,
                        40.613216912838341]},
('type': 'Feature',
 'id': 'nyu_2451_34572.303',
 'geometry': {'type': 'Point',
              'coordinates': [-73.8472052054902, 40.8947051766111]},
 'geometry_name': 'geom',
 'properties': {'name': "Wakefield",
               'annolinel1': "Wakefield",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.8472052054902,
                        40.8947051766111,
                        40.8947051766111]},
('type': 'Feature',
 'id': 'nyu_2451_34572.304',
 'geometry': {'type': 'Point',
              'coordinates': [-73.76596781445627, 40.6113216912838341]},
 'geometry_name': 'geom',
 'properties': {'name': 'Baywater',
               'annolinel1': "Baywater",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.76596781445627,
                        40.6113216912838341,
                        40.6113216912838341]},
('type': 'Feature',
 'id': 'nyu_2451_34572.305',
 'geometry': {'type': 'Point',
              'coordinates': [-73.890345709872, 40.6819934517311]},
 'geometry_name': 'geom',
 'properties': {'name': 'Queensbridge',
               'annolinel1': "Queensbridge",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-73.890345709872,
                        40.6819934517311,
                        40.6819934517311]},
('type': 'Feature',
 'id': 'nyu_2451_34572.306',
 'geometry': {'type': 'Point',
              'coordinates': [-74.08179992211962, 40.613216912838341]},
 'geometry_name': 'geom',
 'properties': {'name': 'Fox Hills',
               'annolinel1': "Fox",
               'annolinel2': "Hills",
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.08179992211962,
                        40.613216912838341,
                        40.613216912838341]},
('type': 'Feature',
 'id': 'nyu_2451_34572.307',
 'geometry': {'type': 'Point',
              'coordinates': [-74.249259487305, 40.5033187866211]},
 'geometry_name': 'geom',
 'properties': {'name': "urn:ogc:def:crs:EPSG::4326",
               'annolinel1': "urn:ogc:def:crs:EPSG::4326",
               'annolinel2': None,
               'annolinel3': None,
               'annolinel4': None,
               'annolinel5': None,
               "bbox": [-74.249259487305,
                        40.5033187866211,
                        40.5033187866211]}
}

```

Notice how all the relevant data is in the `features` key, which is basically a list of the neighborhoods. So, let's define a new variable that includes this data.

```
In [5]: neighborhoods_data = newycork_data["features"]
```

Let's take a look at the first item in this list.

```
In [6]: neighborhoods_data[0]
```

```
Out [6]: {'type': 'Feature',
'id': 'nyu_2451_34572.1',
'geometry': {'type': 'Point',
'coordinates': [-73.8472052054902, 40.8947051766111]},
'properties': {'name': 'Wakefield',
'annolinel1': None,
'annolinel2': None,
'annolinel3': None,
'annolinel4': None,
'annolinel5': None,
'borough': 'Bronx',
'bbox': [-73.8472052054902,
40.8947051766111,
40.8947051766111]}}
```

Transform the data into a pandas dataframe

The next task is essentially transforming this data of nested Python dictionaries into a pandas dataframe. So let's start by creating an empty dataframe.

```
In [7]: # define the dataframe columns
column_names = ['borough', 'Neighborhood', 'Latitude', 'Longitude']

# instantiate the dataframe
neighborhoods = pd.DataFrame(columns=column_names)
```

Take a look at the empty dataframe to confirm that the columns are as intended.

```
In [8]: neighborhoods

Out [8]:
```

Borough Neighborhood Latitude Longitude

Then let's loop through the data and fill the dataframe one row at a time.

```
In [9]: for data in neighborhoods_data:
borough = neighborhoods_data["properties"]["borough"]
neighborhood_name = data["geometry"]["coordinates"]
neighborhood_lat = data["geometry"]["coordinates"]
neighborhood_lon = data["geometry"]["coordinates"]
neighborhoods = neighborhoods.append({'borough': borough,
'Neighborhood': neighborhood_name,
'Latitude': neighborhood_lat,
'Longitude': neighborhood_lon}, ignore_index=True)
```

Quickly examine the resulting dataframe.

```
In [10]: neighborhoods

Out [10]:
```

```
Borough Neighborhood Latitude Longitude
0 Bronx Wakefield 40.894705 -73.847201
1 Bronx Co-op City 40.874294 -73.829399
2 Bronx Eastchester 40.887556 -73.827806
3 Bronx Fairview 40.895437 -73.805643
4 Bronx Riverdale 40.890834 -73.912885
```

And make sure that the dataset has all 5 boroughs and 306 neighborhoods.

```
In [11]: print(len(neighborhoods_data))
print(len(neighborhoods_data["properties"].format(neighborhoods_data["properties"].unique(),)))
print(len(neighborhoods_data["properties"].format(neighborhoods_data["properties"].unique(),)))
```

The dataframe has 5 boroughs and 306 neighborhoods.

Use `geopy` library to get the latitude and longitude values of New York City.

In order to define an instance of the `geocoder`, we need to define a user agent. We will name our agent `ny_explorer`, as shown below.

```
In [12]: address = 'New York City, NY'
geolocator = Nominatim(user_agent='ny_explorer')
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geopy geographic coordinate of New York City are {}, {}'.format(latitude, longitude))
```

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

Create a map of New York with neighborhoods superimposed on top.

```
In [13]: # 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(neighborhoods['Latitude'], neighborhoods['Longitude'], neighborhoods['borough'], neighborhoods['Neighborhood']):
    label = '{} {}'.format(borough, neighborhood)
    folium.CircleMarker(
        location=(lat, lng),
        radius=5,
        popup=label,
        fill=True,
        fill_color='blue',
        parse_html=True).add_to(map_newyork)
```

```
Out [13]:
```

New York City map showing neighborhoods superimposed on top.

Folium is a great visualization library. Feel free to zoom into the above map, and click on each circle mark to reveal the name of the neighborhood and its respective borough.

However, for illustration purposes, let's simplify the above map and segment and cluster only the neighborhoods in Manhattan. So let's slice the original dataframe and create a new dataframe of the Manhattan data.

```
In [14]: manhattan_data = wghborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head()
```

```
Out[14]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910680
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

Let's get the geographical coordinates of Manhattan.

```
In [15]: address = 'Manhattan, NY'

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

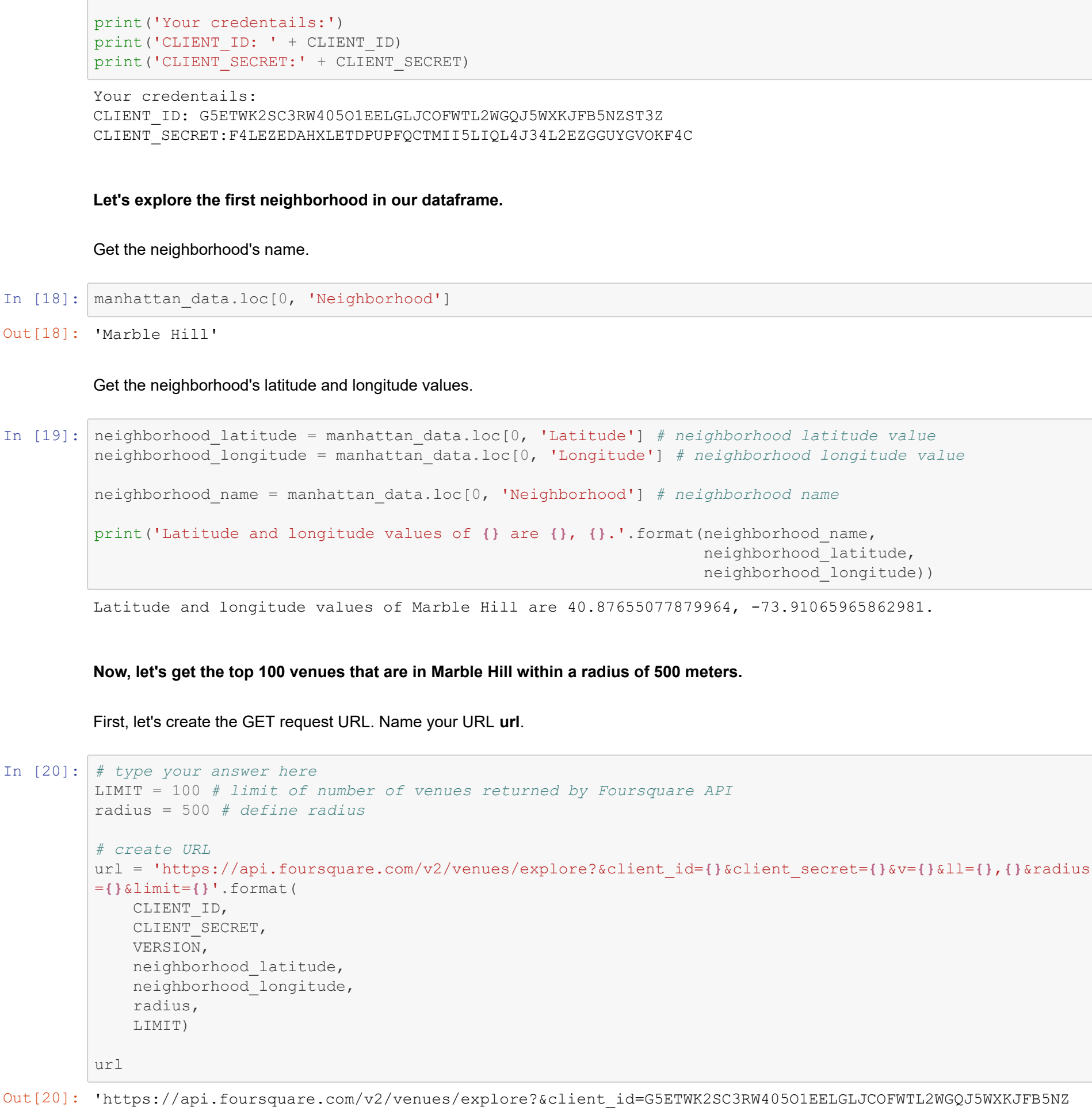
The geographical coordinate of Manhattan are 40.7896239, -73.9598939.
```

As we did with all of New York City, let's visualizat Manhattan the neighborhoods in it.

```
In [16]: # create map of Manhattan using latitude and longitude values
map_manhattan = folium.Map(location=[latitude, longitude], zoom_start=11)

# 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
```



Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

Define Foursquare Credentials and Version

```
In [17]: CLIENT_ID = 'G5ETWR28C3RW405O1EELGLJCOFWTL2WQJ5WKKJFB5NZ8T3E' # your Foursquare ID
CLIENT_SECRET = 'F41E2EDAHKLETDPUFFQCTMI15L1QL4J34L2E2GGUYGVORF4C' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentials:')
print('CLIENT ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

Your credentials:
CLIENT_ID: G5ETWR28C3RW405O1EELGLJCOFWTL2WQJ5WKKJFB5NZ8T3E
CLIENT_SECRET: F41E2EDAHKLETDPUFFQCTMI15L1QL4J34L2E2GGUYGVORF4C

Let's explore the first neighborhood in our dataframe.

Get the neighborhood's name.

```
In [18]: manhattan_data.loc[0, 'Neighborhood']
```

```
Out[18]: 'Marble Hill'
```

Get the neighborhood's latitude and longitude values.

```
In [19]: 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, -73.9106595862981.
```

Now, let's get the top 100 venues that are in Marble Hill within a radius of 500 meters.

First, let's create the GET request URL. Name your URL url.

```
In [20]: # type your answer here
LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 500 # define radius

# create URL
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
```

```
Out[20]: 'https://api.foursquare.com/v2/venues/explore?client_id=G5ETWR28C3RW405O1EELGLJCOFWTL2WQJ5WKKJFB5NZ8T3E&client_secret=F41E2EDAHKLETDPUFFQCTMI15L1QL4J34L2E2GGUYGVORF4C&v=20180605&ll=40.87655077879964,-73.9106595862981&radius=500&limit=100'
```

Double-click here for the solution.

Send the GET request and examine the results


```
In [27]: manhattan_venues = pd.NearbyVenues(name=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
Financial District
Carnegie Hill
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards

Double-click here for the solution.

Let's check the size of the resulting dataframe

```
In [28]: print(manhattan_venues.shape)
manhattan_venues.head()
```

	Neighborhood	Neighborhood	Latitude	Neighborhood	Longitude	Venue	Venue	Latitude	Venue	Longitude	Venue	Category
0	Marble Hill		40.876551		-73.91066	Arturo's		40.87442		-73.910271		Pizza Place
1	Marble Hill		40.876551		-73.91066	Bkram Yoga		40.87884		-73.906204		Yoga Studio
2	Marble Hill		40.876551		-73.91066	Tibbet Diner		40.880404		-73.908937		Diner
3	Marble Hill		40.876551		-73.91066	Dunkin'		40.877136		-73.906666		Donut Shop
4	Marble Hill		40.876551		-73.91066	Starbucks		40.877531		-73.905582		Coffee Shop

Let's check how many venues were returned for each neighborhood

```
In [29]: manhattan_venues.groupby('Neighborhood').count()
```

	Neighborhood	Neighborhood	Latitude	Neighborhood	Longitude	Venue	Venue	Latitude	Venue	Longitude	Venue	Category
0	Battery Park City		66		66	66		66		66		66
1	Carnegie Hill		86		86	86		86		86		86
2	Central Harlem		45		45	45		45		45		45
3	Chinatown		100		100	100		100		100		100
4	Civic Center		100		100	100		100		100		100
5	Civic Center		100		100	100		100		100		100
6	Clinton		100		100	100		100		100		100
7	East Harlem		40		40	40		40		40		40
8	East Village		100		100	100		100		100		100
9	Financial District		100		100	100		100		100		100
10	Flatiron		100		100	100		100		100		100
11	Gramercy		94		94	94		94		94		94
12	Greenwich Village		100		100	100		100		100		100
13	Hamilton Heights		63		63	63		63		63		63
14	Hudson Yards		62		62	62		62		62		62
15	Inwood		57		57	57		57		57		57
16	Lenox Hill		100		100	100		100		100		100
17	Lincoln Square		92		92	92		92		92		92
18	Little Italy		100		100	100		100		100		100
19	Lower East Side		47		47	47		47		47		47
20	Manhattan Valley		49		49	49		49		49		49
21	Manhattanville		47		47	47		47		47		47
22	Marble Hill		22		22	22		22		22		22
23	Noho		100		100	100		100		100		100
24	Midtown South		100		100	100		100		100		100
25	Morningside Heights		43		43	43		43		43		43
26	Murray Hill		100		100	100		100		100		100
27	Noho		100		100	100		100		100		100
28	Roosevelt Island		26		26	26		26		26		26
29	Soho		100		100	100		100		100		100
30	Stuyvesant Town		17		17	17		17		17		17
31	Sutton Place		100		100	100		100		100		100
32	Tribeca		88		88	88		88		88		88
33	Tudor City		81		81	81		81		81		81
34	Turtle Bay		100		100	100		100		100		100
35	Upper East Side		97		97	97		97		97		97
36	Upper West Side		98		98	98		98		98		98
37	Washington Heights		84		84	84		84		84		84
38	West Village		100		100	100		100		100		100
39	Yorkville		100		100	100		100		100		100

Let's find out how many unique categories can be curated from all the returned venues

```
In [30]: print('There are {} unique categories.'.format(len(manhattan_venues['Venue Category'].unique())))
There are 322 unique categories.
```

3. Analyze Each Neighborhood

```
In [31]: # 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()
```

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arropa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Marble Hill	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
2	Marble Hill	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
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0

And let's examine the new dataframe size.

```
In [32]: manhattan_onehot.shape
Out[32]: (3204, 323)
```

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

```
In [33]: manhattan_grouped = manhattan_onehot.groupby('Neighborhood').mean().reset_index()
manhattan_grouped
```

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arropa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Battery Park City	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
1	Carnegie Hill	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.011628	0.000000	0.011628	0.000000	0.00
2	Central Harlem	0.000000	0.00	0.006667	0.044444	0.00	0.000000	0.000000	0.022222	0.000000	0.000000	0.00
3	Chelsea	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.050000	0.000000	0.000000	0.00
4	Chinatown	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.02
5	Civic Center	0.000000	0.00	0.000000	0.030000	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
6	Clinton	0.000000	0.00	0.000000	0.050000	0.00	0.000000	0.000000	0.010000	0.000000	0.000000	0.00
7	East Harlem	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
8	East Village	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.010000	0.010000	0.000000	0.010000	0.00
9	Financial District	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.020000	0.000000	0.000000	0.00
10	Flatiron	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.020000	0.000000	0.010000	0.00
11	Gramercy	0.000000	0.00	0.000000	0.042553	0.00	0.000000	0.000000	0.010638	0.000000	0.000000	0.00
12	Greenwich Village	0.010000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.010638	0.000000	0.000000	0.00
13	Hamilton Heights	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
14	Hudson Yards	0.000000	0.00	0.000000	0.048387	0.00	0.000000	0.000000	0.016129	0.000000	0.000000	0.00
15	Inwood	0.000000	0.00	0.000000	0.035088	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
16	Lenox Hill	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.020000	0.000000	0.000000	0.00
17	Lincoln Square	0.000000	0.00	0.000000	0.021739	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
18	Little Italy	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
19	Lower East Side	0.000000	0.00	0.000000	0.021277	0.00	0.000000	0.021277	0.042553	0.000000	0.000000	0.00
20	Manhattan Valley	0.000000	0.00	0.000000	0.020408	0.00	0.000000	0.000000	0.000000	0.000000	0.020408	0.00
21	Manhattanville	0.000000	0.00	0.000000	0.021277	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
22	Marble Hill	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
23	Midtown	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.010000	0.000000	0.000000	0.00
24	Midtown South	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
25	Morningside Heights	0.000000	0.00	0.000000	0.069767	0.00	0.000000	0.000000	0.000000	0.000000	0.023256	0.00
26	Murray Hill	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.020000	0.000000	0.000000	0.00
27	Noho	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.010000	0.040000	0.000000	0.000000	0.01
28	Roosevelt Island	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
29	Soho	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.000000	0.010000	0.000000	0.010000	0.00
30	Stuyvesant Town	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
31	Sutton Place	0.000000	0.01	0.000000	0.020000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.01
32	Tribeca	0.000000	0.00	0.000000	0.056818	0.00	0.000000	0.000000	0.011364	0.027227	0.000000	0.000000
33	Tudor City	0.000000	0.00	0.000000	0.012346	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
34	Turtle Bay	0.000000	0.00	0.000000	0.020000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.03
35	Upper East Side	0.000000	0.00	0.000000	0.020619	0.00	0.000000	0.000000	0.000000	0.010309	0.000000	0.00
36	Upper West Side	0.010204	0.00	0.000000	0.020408	0.00	0.000000	0.000000	0.000000	0.000000	0.010204	0.01
37	Washington Heights	0.011905	0.00	0.000000	0.011905	0.00	0.011905	0.000000	0.000000	0.000000	0.000000	0.00
38	West Village	0.010000	0.00	0.000000	0.060000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
39	Yorkville	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.01

Let's confirm the new size

```
In [34]: manhattan_grouped.shape
Out[34]: (40, 323)
```

Let's print each neighborhood along with the top 5 most common venues

```
In [35]: 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 Park 0.09
1 Hotel 0.08
2 Gym 0.06
3 Coffee Shop 0.04
```


In [41]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]

Out [41]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Marble Hill	Sandwich Place	Coffee Shop	Gym	Discount Store	Supplement Shop	Donut Shop	Seafood Restaurant	Tennis Stadium	Pharmacy	Kids Store
2	Washington Heights	Café	Bakery	Grocery Store	Bank	Sandwich Place	Coffee Shop	Park	Spanish Restaurant	Deli / Bodega	New American Restaurant
5	Manhattanville	Seafood Restaurant	Coffee Shop	Deli / Bodega	Italian Restaurant	Mexican Restaurant	Indian Restaurant	Lounge	Boutique	Sushi Restaurant	Supermarket
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Bar	Chinese Restaurant	Bookstore	Boutique	Gym
9	Yorkville	Italian Restaurant	Coffee Shop	Gym	Bar	Sushi Restaurant	Deli / Bodega	Japanese Restaurant	Wine Shop	Diner	Mexican Restaurant
11	Roosevelt Island	Deli / Bodega	Japanese Restaurant	Coffee Shop	Greek Restaurant	Supermarket	Bubble Tea Shop	Playground	Food & Drink Shop	Soccer Field	Farmers Market
14	Clinton	Theater	American Restaurant	Gym / Fitness Center	Coffee Shop	Sandwich Place	Gym	Cocktail Bar	Italian Restaurant	Spa	Hotel
16	Murray Hill	Japanese Restaurant	Coffee Shop	Hotel	Gym / Fitness Center	Sandwich Place	American Restaurant	Bar	Restaurant	Italian Restaurant	Grocery Store
17	Chelsea	Coffee Shop	Art Gallery	Bakery	American Restaurant	Ice Cream Shop	Italian Restaurant	Japanese Restaurant	Park	Cycle Studio	Cupcake Shop
21	Tribeca	Park	American Restaurant	Wine Bar	Italian Restaurant	Coffee Shop	Café	Spa	Greek Restaurant	Poke Place	Playground
26	Morningside Heights	Park	American Restaurant	Coffee Shop	Bookstore	Café	Deli / Bodega	Burger Joint	Indian Restaurant	Seafood Restaurant	Grocery Store
28	Battery Park City	Park	Hotel	Gym	Coffee Shop	Memorial Site	Playground	Gourmet Shop	Boat or Ferry	Burger Joint	Plaza
29	Financial District	Coffee Shop	Pizza Place	Cocktail Bar	American Restaurant	Hotel	Park	Café	Sandwich Place	Salad Place	Steakhouse
34	Sutton Place	Italian Restaurant	Park	Furniture / Home Store	Gym / Fitness Center	Coffee Shop	Pizza Place	Gym	Latin American Restaurant	Beer Garden	Mediterranean Restaurant
35	Turtle Bay	Italian Restaurant	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Park	Ramen Restaurant	Seafood Restaurant	Deli / Bodega	Hotel	Pharmacy
36	Tudor City	Café	Park	Mexican Restaurant	Asian Restaurant	Diner	Deli / Bodega	Coffee Shop	Restaurant	Garden	Sushi Restaurant
39	Hudson Yards	Hotel	Gym / Fitness Center	Italian Restaurant	American Restaurant	Café	Coffee Shop	Boat or Ferry	Park	Dog Run	Gym

Cluster 2

In [42]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]

Out [42]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
33	Midtown South	Korean Restaurant	Hotel	Japanese Restaurant	Cosmetics Shop	Burger Joint	Coffee Shop	Dessert Shop	Gym / Fitness Center	Bakery	American Restaurant

Cluster 3

In [43]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]

Out [43]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 C
1	Chinatown	Chinese Restaurant	Cocktail Bar	Bakery	Dessert Shop	American Restaurant	Salon / Barbershop	Optical Shop	Noodle House	Hotpot Restaurant	
8	Upper East Side	Exhibit	Italian Restaurant	Coffee Shop	Bakery	Gym / Fitness Center	Yoga Studio	Juice Bar	Cosmetics Shop	French Restaurant	
10	Lenox Hill	Italian Restaurant	Sushi Restaurant	Coffee Shop	Pizza Place	Cocktail Bar	Burger Joint	Gym	Gym / Fitness Center	Deli / Bodega	
12	Upper West Side	Italian Restaurant	Bar	Café	Indian Restaurant	Coffee Shop	Wine Bar	Pizza Place	Bakery	Ice Cream Shop	Mediterranean Restaurant
13	Lincoln Square	Plaza	Performing Arts Venue	Café	Theater	Concert Hall	Italian Restaurant	Wine Shop	French Restaurant	Gym / Fitness Center	Indian Restaurant
15	Midtown	Hotel	Coffee Shop	Bakery	Steakhouse	Clothing Store	Sporting Goods Shop	Theater	Pizza Place	Bookstore	
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	Clothing Store	Café	Indian Restaurant	American Restaurant	Gym	Boutique	Bubble Tea Shop	
19	East Village	Bar	Ice Cream Shop	Pizza Place	Mexican Restaurant	Wine Bar	Korean Restaurant	Italian Restaurant	Coffee Shop	Cocktail Bar	Vegan Restaurant
22	Little Italy	Bakery	Café	Italian Restaurant	Bubble Tea Shop	Chinese Restaurant	Mediterranean Restaurant	Ice Cream Shop	Cocktail Bar	Pizza Place	Coffee Shop
23	SoHo	Clothing Store	Italian Restaurant	Coffee Shop	Boutique	Mediterranean Restaurant	Shoe Store	Bakery	Café	French Restaurant	Pizzeria
24	West Village	Italian Restaurant	American Restaurant	New American Restaurant	Cocktail Bar	Park	Wine Bar	Theater	Jazz Club	Coffee Shop	Sushi Restaurant
25	Manhattan Valley	Bar	Coffee Shop	Yoga Studio	Pizza Place	Playground	Thai Restaurant	Mexican Restaurant	French Restaurant	Szechuan Restaurant	Reception
27	Gramercy	Bar	Italian Restaurant	Coffee Shop	American Restaurant	Thai Restaurant	Pizza Place	Bagel Shop	Ice Cream Shop	Cocktail Bar	Thrift / Vintage Store
30	Carnegie Hill	Coffee Shop	Café	Bookstore	Italian Restaurant	Gym / Fitness Center	Gym	French Restaurant	Yoga Studio	Wine Shop	Vietnamese Restaurant
31	Noho	Italian Restaurant	American Restaurant	Hotel	Art Gallery	Yoga Studio	Bookstore	Coffee Shop	Cocktail Bar	French Restaurant	Pizzeria
32	Civic Center	Spa	Gym / Fitness Center	Coffee Shop	Hotel	French Restaurant	Cocktail Bar	Yoga Studio	American Restaurant	Sushi Restaurant	
38	Flatiron	Italian Restaurant	Japanese Restaurant	New American Restaurant	Mediterranean Restaurant	Cycle Studio	Gym	Gym / Fitness Center	Furniture / Home Store	Sporting Goods Shop	

Cluster 4

In [44]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]

Out [44]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37	Stuyvesant Town	Park	Coffee Shop	Cocktail Bar	Gym / Fitness Center	Baseball Field	Bar	Bistro	Heliprot	Farmers Market	Boat or Ferry

Cluster 5

In [45]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]

Out [45]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Inwood	Mexican Restaurant	Café	Restaurant	Lounge	Frozen Yogurt Shop	Bakery	Pizza Place	Park	Chinese Restaurant	Deli / Bodega
4	Hamilton Heights	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Cocktail Bar	Indian Restaurant	Liquor Store	Sushi Restaurant	Park	Deli / Bodega
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Deli / Bodega	Latin American Restaurant	Sandwich Place	Taco Place	Beer Bar	Liquor Store	Grocery Store
20	Lower East Side	Chinese Restaurant	Pharmacy	Coffee Shop	Café	Bakery	Japanese Restaurant	Art Gallery	Pizza Place	Ramen Restaurant	Pet Café

Thank you for completing this lab!

This notebook was created by [Alex Aklonis](#) and [Polong Lin](#). I hope you found this lab interesting and educational. Feel free to contact us if you have any questions!

This notebook is part of a course on [Coursera](#) called [Applied Data Science Capstone](#). If you accessed this notebook outside the course, you can take this course online by clicking [here](#).