The Battle of the Neighborhoods - Week 2 - Final

Problem Background

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Problem Description / Business Problem

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of New York is famous for its excelllent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens Italian immigrants - New York-style pizza and Italian cuisine Jewish immigrants and Irish immigrants - pastrami and corned beef Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city mobile food vendors - Some 4,000 licensed by the city Middle Eastern foods such as falafel and kebabs examples of modern New York street food It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants.The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. So it is evident that to survive in such competitive market it is very important to startegically plan. Various factors need to be studied inorder to decide on the Location such as :

New York Population New York City Demographics Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost? Are there any venues like Gyms, Entertainmnet zones, Parks etc nearby where floating population is high etc Who are the competitors in that location? Cuisine served / Menu of the competitors Segmentation of the Borough Untapped markets Saturated markets etc The list can go on...

Eventhough well funded we need to choose the correct location to start its first venture.If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

Data Description / How It Will Be Used

Explore Dataset Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement 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 the latitude and logitude 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

The description of the data and how it will be used to solve the problem are:

1. To identify the characteristics of our competitors' venues in Manhattan, we would first need to find out the number of sushi bars in Manhattan currently and their location.
2. We then used Google Map API to find their geographic coordinates based on their postal code addresses.
3. In Manhattan, there is 1763 sushi bars are currently operating.

Import Libraries

In this section we import the libraries that will be required to process the data.

In [63]:

```
# Import libraries
import numpy as np
import pandas as pd
```

```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

#!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim
import urllib.request
import json
from bs4 import BeautifulSoup
from urllib.request import urlopen
import requests
from pandas.io.json import json_normalize

import matplotlib.cm as cm
import matplotlib.colors as colors

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.pyplot as plt
import matplotlib.colors as colors
%matplotlib inline
from sklearn.cluster import KMeans

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

# To download url file
import urllib
```

Download and Read Dataset

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

After the dataset is downloaded, put the json file on our project directory.

In [64]:

```
# Read json file
with open('newyork_data.json') as json_data:
    newyork_data = json.load(json_data)
```

Transform Data into *Pandas* Dataframe

In [65]:

```
neighborhoods_data = newyork_data['features']
# Define the dataframe columns
column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']

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

for data in neighborhoods_data:
    borough = neighborhood_name = data['properties']['borough']
    neighborhood_name = data['properties']['name']

    neighborhood_latlon = data['geometry']['coordinates']
    neighborhood_lat = neighborhood_latlon[1]
    neighborhood_lon = neighborhood_latlon[0]

    neighborhoods = neighborhoods.append({'Borough': borough,
                                          'Neighborhood': neighborhood_name,
                                          'Latitude': neighborhood_lat,
                                          'Longitude': neighborhood_lon}, ignore_index=True)
```

In [66]:

```
neighborhoods.head()
```

Out[66]:

	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

Get Latitude and Longitude 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 [67]:

```
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.
```

In [68]:

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

Out[68]:

	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

Create a Map of New York

In [69]:

```
# 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(manhattan_data['Latitude'], manhattan_data['Longitude'],
manhattan_data['Borough'], manhattan_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
```

Out[69]:

Foursquare Venues

In [70]:

```
def getNearbyVenues(names, latitudes, longitudes, radius=5000, categoryIds=''):
    try:
        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/search?&client_id={}&client_secret={}&v={}&
ll={},{}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, VERSION, lat, lng, radius, LIMIT)

            if (categoryIds != ''):
                url = url + '&categoryId={}'
                url = url.format(categoryIds)

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

            # return only relevant information for each nearby venue
            for v in results:
                success = False
                try:
                    category = v['categories'][0]['name']
                    success = True
                except:
                    pass

                if success:
                    venues_list.append([(
                        name,
                        lat,
                        lng,
                        v['name'],
                        v['location']['lat'],
                        v['location']['lng'],
                        v['categories'][0]['name']
                    )])

        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']

    except:
        print(url)
        print(response)
        print(results)
        print(nearby_venues)

    return(nearby_venues)
```

In [71]:

```
# Foursquare ID & SECRET
LIMIT = 500
radius = 5000
CLIENT_ID = 'SNSELBPN5W3YRAETVYGCBSFGHU4NKLV1S3S5WXQW3CC54QM1' # Replace with your own Foursquare
APP CLIENT ID
CLIENT_SECRET = '53CIZG24GA5RMTVNXNJ54JNRSMQWNXGESGGKUZXCYRVPS0S0' # Replace with your own
Foursquare APP CLIENT SECRET
VERSION = '20200508' # Replace with your own Foursquare APP VERSION
```

In [72]:

```
# https://developer.foursquare.com/docs/resources/categories
# Sushi = 4bf58dd8d48988d1d2941735
neighborhoods = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
newyork_venues_sushi = getNearbyVenues(names=neighborhoods['Neighborhood'],
latitudes=neighborhoods['Latitude'], longitudes=neighborhoods['Longitude'], radius=1000,
categoryIds='4bf58dd8d48988d1d2941735')
newyork_venues_sushi.head()
```

Out[72]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Planet Tokyo	40.886233	-73.909479	Sushi Restaurant
1	Chinatown	40.715618	-73.994279	Shinsen	40.715608	-73.996611	Japanese Restaurant
2	Chinatown	40.715618	-73.994279	Sushumai Asian Fusion	40.721155	-73.987337	Sushi Restaurant
3	Chinatown	40.715618	-73.994279	Quan Sushi	40.720323	-73.996257	Sushi Restaurant
4	Chinatown	40.715618	-73.994279	Amano Sushi	40.716017	-73.992818	Sushi Restaurant

In [73]:

```
newyork_venues_sushi.shape
```

Out[73]:

```
(1087, 7)
```

In [74]:

```
def addToMap(df, color, existingMap):
    for lat, lng, local, venue, venueCat in zip(df['Venue Latitude'], df['Venue Longitude'], df['Ne
ighborhood'], df['Venue'], df['Venue Category']):
        label = '{} ({}) - {}'.format(venue, venueCat, local)
        label = folium.Popup(label, parse_html=True)
        folium.CircleMarker(
            [lat, lng],
            radius=5,
            popup=label,
            color=color,
            fill=True,
```

```
            fill_color=color,
            fill_opacity=0.7).add_to(existingMap)
```

In [75]:

```
map_newyork_sushi = folium.Map(location=[latitude, longitude], zoom_start=10)
addToMap(newyork_venues_sushi, 'red', map_newyork_sushi)

map_newyork_sushi
```

Out[75]:

In [76]:

```
def addColumn(startDf, columnTitle, dataDf):
    grouped = dataDf.groupby('Neighborhood').count()

    for n in startDf['Neighborhood']:
        try:
            startDf.loc[startDf['Neighborhood'] == n,columnTitle] = grouped.loc[n, 'Venue']
        except:
            startDf.loc[startDf['Neighborhood'] == n,columnTitle] = 0
```

In [77]:

```
manhattan_grouped = newyork_venues_sushi.groupby('Neighborhood').count()
manhattan_grouped
```

Out[77]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	21	21	21	21	21	21
Carnegie Hill	21	21	21	21	21	21
Central Harlem	3	3	3	3	3	3
Chelsea	41	41	41	41	41	41
Chinatown	26	26	26	26	26	26
Civic Center	30	30	30	30	30	30
Clinton	38	38	38	38	38	38

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
East Harlem	3	3	3	3	3	3
East Village	50	50	50	50	50	50
Financial District	21	21	21	21	21	21
Flatiron	50	50	50	50	50	50
Gramercy	47	47	47	47	47	47
Greenwich Village	47	47	47	47	47	47
Hamilton Heights	3	3	3	3	3	3
Hudson Yards	24	24	24	24	24	24
Inwood	4	4	4	4	4	4
Lenox Hill	34	34	34	34	34	34
Lincoln Square	23	23	23	23	23	23
Little Italy	39	39	39	39	39	39
Lower East Side	14	14	14	14	14	14
Manhattan Valley	9	9	9	9	9	9
Manhattanville	4	4	4	4	4	4
Marble Hill	1	1	1	1	1	1
Midtown	50	50	50	50	50	50
Midtown South	50	50	50	50	50	50
Morningside Heights	7	7	7	7	7	7
Murray Hill	50	50	50	50	50	50
Noho	50	50	50	50	50	50
Roosevelt Island	10	10	10	10	10	10
Soho	41	41	41	41	41	41
Stuyvesant Town	16	16	16	16	16	16
Sutton Place	42	42	42	42	42	42
Tribeca	18	18	18	18	18	18
Tudor City	36	36	36	36	36	36
Turtle Bay	44	44	44	44	44	44
Upper East Side	30	30	30	30	30	30
Upper West Side	16	16	16	16	16	16
Washington Heights	2	2	2	2	2	2
West Village	41	41	41	41	41	41
Yorkville	31	31	31	31	31	31

Analyze Each Neighborhood

In [78]:

```
# One hot encoding
manhattan_onehot = pd.get_dummies(newyork_venues_sushi[['Venue Category']], prefix="", prefix_sep="")

# Add neighborhood column back to dataframe
manhattan_onehot['Neighborhood'] = newyork_venues_sushi['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()
```

Out[78]:

	Neighborhood	Asian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant	Noodle House	Poke Place	Ramen Restaurant	Res

	Neighborhood	Asian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant	Noodle House	Poke Place	Ramen Restaurant	Res
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	
1	Chinatown	0	0	0	0	0	0	0	1	0	0	0	
2	Chinatown	0	0	0	0	0	0	0	0	0	0	0	
3	Chinatown	0	0	0	0	0	0	0	0	0	0	0	
4	Chinatown	0	0	0	0	0	0	0	0	0	0	0	

In [79]:

```
manhattan_grouped = manhattan_onehot.groupby('Neighborhood').mean().reset_index()
manhattan_grouped
```

Out[79]:

	Neighborhood	Asian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant	Noodle House	Poke Place	Rame Restaura
0	Battery Park City	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.095238	0.047619	0.000000	0.0
1	Carnegie Hill	0.047619	0.00	0.047619	0.000000	0.00000	0.000000	0.000000	0.095238	0.000000	0.000000	0.0
2	Central Harlem	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
3	Chelsea	0.048780	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.097561	0.000000	0.000000	0.0
4	Chinatown	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.192308	0.000000	0.000000	0.0
5	Civic Center	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.100000	0.033333	0.000000	0.0
6	Clinton	0.026316	0.00	0.026316	0.026316	0.00000	0.000000	0.000000	0.105263	0.000000	0.026316	0.0
7	East Harlem	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
8	East Village	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.120000	0.000000	0.000000	0.0
9	Financial District	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.190476	0.000000	0.000000	0.0
10	Flatiron	0.020000	0.00	0.020000	0.000000	0.00000	0.000000	0.000000	0.080000	0.000000	0.000000	0.0
11	Gramercy	0.000000	0.00	0.021277	0.000000	0.00000	0.000000	0.000000	0.148936	0.000000	0.000000	0.0
12	Greenwich Village	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.063830	0.000000	0.000000	0.0
13	Hamilton Heights	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
14	Hudson Yards	0.083333	0.00	0.000000	0.041667	0.00000	0.000000	0.000000	0.083333	0.000000	0.041667	0.0
15	Inwood	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
16	Lenox Hill	0.117647	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.088235	0.000000	0.000000	0.0
17	Lincoln Square	0.000000	0.00	0.043478	0.000000	0.00000	0.043478	0.000000	0.086957	0.000000	0.000000	0.0
18	Little Italy	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.102564	0.025641	0.000000	0.0
19	Lower East Side	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.285714	0.000000	0.000000	0.0
20	Manhattan Valley	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.111111	0.111111	0.000000	0.000000	0.0
21	Manhattanville	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
22	Marble Hill	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
23	Midtown	0.080000	0.02	0.000000	0.000000	0.00000	0.000000	0.000000	0.040000	0.000000	0.000000	0.0
24	Midtown South	0.060000	0.02	0.020000	0.000000	0.00000	0.000000	0.000000	0.040000	0.000000	0.000000	0.0
25	Morningside Heights	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.142857	0.000000	0.000000	0.000000	0.0
26	Murray Hill	0.040000	0.02	0.020000	0.000000	0.00000	0.000000	0.000000	0.080000	0.000000	0.000000	0.0
27	Noho	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.120000	0.000000	0.000000	0.0
28	Roosevelt Island	0.100000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
29	Soho	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.097561	0.024390	0.000000	0.0
30	Stuyvesant Town	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.125000	0.000000	0.000000	0.0
31	Sutton Place	0.071429	0.00	0.000000	0.000000	0.02381	0.000000	0.000000	0.071429	0.000000	0.000000	0.0

	Neighborhood	Asian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant	Noodle House	Poke Place	Ramen Restaurant
32	Tribeca	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.055556	0.055556	0.000000	0.0
33	Tudor City	0.055556	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.027778	0.000000	0.000000	0.0
34	Turtle Bay	0.022727	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.045455	0.000000	0.000000	0.0
35	Upper East Side	0.066667	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.100000	0.000000	0.000000	0.0
36	Upper West Side	0.062500	0.00	0.000000	0.000000	0.00000	0.062500	0.000000	0.250000	0.000000	0.000000	0.0
37	Washington Heights	0.000000	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
38	West Village	0.024390	0.00	0.000000	0.000000	0.00000	0.000000	0.000000	0.097561	0.000000	0.000000	0.0
39	Yorkville	0.032258	0.00	0.032258	0.000000	0.00000	0.000000	0.000000	0.129032	0.000000	0.000000	0.0

In [80]:

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [81]:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# Create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        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.head()
```

Out[81]:

	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	Battery Park City	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
1	Carnegie Hill	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
2	Central Harlem	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
3	Chelsea	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Smoothie Shop	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant	Ramen Restaurant
4	Chinatown	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant

Cluster Neighborhoods

In [82]:

```
# 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]
```

Out[82]:

```
array([3, 4, 1, 3, 3, 3, 4, 1, 3, 3])
```

In [83]:

```
# Add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

manhattan_merged = manhattan_data
manhattan_merged = manhattan_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on
='Neighborhood')

manhattan_merged.head()
```

Out[83]:

	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	6th Most Common Venue	7th M Comm Ver
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	D Bod
1	Manhattan	Chinatown	40.715618	-73.994279	3	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cock
2	Manhattan	Washington Heights	40.851903	-73.936900	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	D Bod
3	Manhattan	Inwood	40.867684	-73.921210	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	D Bod
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	D Bod

In [84]:

```
# Create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# 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]

# Add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(manhattan_merged['Latitude'], manhattan_merged['Longitude'], manh
attan_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)
```

```
map_clusters
```

Out[84]:

In [85]:

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

Out[85]:

	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
11	Roosevelt Island	Sushi Restaurant	Asian Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
18	Greenwich Village	Sushi Restaurant	Japanese Restaurant	Sake Bar	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
21	Tribeca	Sushi Restaurant	Noodle House	Japanese Restaurant	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
26	Morningside Heights	Sushi Restaurant	Hawaiian Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Japanese Restaurant
35	Turtle Bay	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Steakhouse	Seafood Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
36	Tudor City	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Steakhouse	Smoothie Shop	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant
38	Flatiron	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant	Ramen Restaurant

In [86]:

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

Out[86]:

	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	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
2	Washington Heights	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
3	Inwood	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
4	Hamilton Heights	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
5	Manhattanville	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
6	Central Harlem	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant
7	East Harlem	Sushi Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant

In [87]:

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

Out[87]:

	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
12	Upper West Side	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Grocery Store	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Hawaiian Restaurant
20	Lower East Side	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant

In [88]:

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

Out[88]:

	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
1	Chinatown	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
9	Yorkville	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
17	Chelsea	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Smoothie Shop	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant	Ramen Restaurant
19	East Village	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
22	Little Italy	Sushi Restaurant	Japanese Restaurant	Noodle House	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
23	Soho	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
24	West Village	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan	Sake Bar	Asian Restaurant	Seafood Restaurant	Sandwich Place	Smoothie Shop	Restaurant	Ramen Restaurant

	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
25	Manhattan Valley	Sushi Restaurant	Hawaiian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
27	Gramercy	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
28	Battery Park City	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
29	Financial District	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
31	Noho	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
32	Civic Center	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Vegetarian / Vegan Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store
37	Stuyvesant Town	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant

In [89]:

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

Out[89]:

	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
8	Upper East Side	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
10	Lenox Hill	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Noodle House	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
13	Lincoln Square	Sushi Restaurant	Japanese Restaurant	Smoothie Shop	Chinese Restaurant	Grocery Store	Vegetarian / Vegan Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bodega
14	Clinton	Sushi Restaurant	Japanese Restaurant	Poke Place	Chinese Restaurant	Cocktail Bar	Asian Restaurant	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant
15	Midtown	Sushi Restaurant	Asian Restaurant	Seafood Restaurant	Japanese Restaurant	Ramen Restaurant	Bakery	Vegetarian / Vegan Restaurant	Sandwich Place	Sake Bar	Restaurant
16	Murray Hill	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Restaurant	Bakery	Chinese Restaurant	Ramen Restaurant	Vegetarian / Vegan Restaurant	Sake Bar	Sandwich Place
30	Carnegie Hill	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Chinese Restaurant	Noodle House	Bakery	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant
33	Midtown South	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Restaurant	Bakery	Chinese Restaurant	Ramen Restaurant	Vegetarian / Vegan Restaurant	Sake Bar	Sandwich Place
34	Sutton Place	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Steakhouse	Seafood Restaurant	Deli / Bodega	Bakery	Chinese Restaurant	Cocktail Bar	Grocery Store
39	Hudson Yards	Sushi Restaurant	Asian Restaurant	Japanese Restaurant	Cocktail Bar	Poke Place	Seafood Restaurant	Sandwich Place	Sake Bar	Restaurant	Ramen Restaurant