

# Final Project: Data Exploration for Airbnb Listings and Neighborhood Similarities in NYC

In [89]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
import seaborn as sns

%matplotlib inline
```

## Part 1: Loading and Cleaning Data:

In [261]:

```
df = pd.read_csv('C:/Users/jry5/OneDrive/Desktop/listings.csv')
df.head()
```

Out[261]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude
0	2060	Modern NYC	2259	Jenny	Manhattan	Washington Heights	40.8572
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.7536
2	3831	Whole flr w/private bdrm, bath & kitchen(pls r...	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.6851
3	5099	Large Cozy 1 BR Apartment In Midtown East	7322	Chris	Manhattan	Murray Hill	40.7476
4	5121	BlissArtsSpace!	7356	Garon	Brooklyn	Bedford-Stuyvesant	40.6868

In [262]:

```
df.shape
```

Out[262]:

(49530, 16)

## Finding number of null values for each columns

In [263]:

```
df.isnull().sum()
```

Out[263]:

```
id                0
name             18
host_id          0
host_name        6
neighbourhood_group  0
neighbourhood    0
latitude         0
longitude        0
room_type        0
price            0
minimum_nights   0
number_of_reviews 0
last_review     11319
reviews_per_month 11319
calculated_host_listings_count 0
availability_365  0
dtype: int64
```

In our case, only column: "name", "last\_review", "reviews\_per\_month" contain missing data, and those three columns are not significant, meaning that we do not care whether it is null or it has some value. Therefore, I will not drop any rows by missing values. Instead, I will replace all missing values with 0.

In [264]:

```
# replace NA with 0
df.fillna('0', inplace=True)
df.isnull().sum()
```

Out[264]:

```
id                0
name              0
host_id          0
host_name        0
neighbourhood_group  0
neighbourhood    0
latitude         0
longitude        0
room_type        0
price            0
minimum_nights   0
number_of_reviews 0
last_review      0
reviews_per_month 0
calculated_host_listings_count 0
availability_365  0
dtype: int64
```

**Filter all listings that prices per night are less than 0 or greater than 500 USD**

In [265]:

```
df = df[(df.price <= 500) & (df.price > 0)]  
df.shape
```

Out[265]:

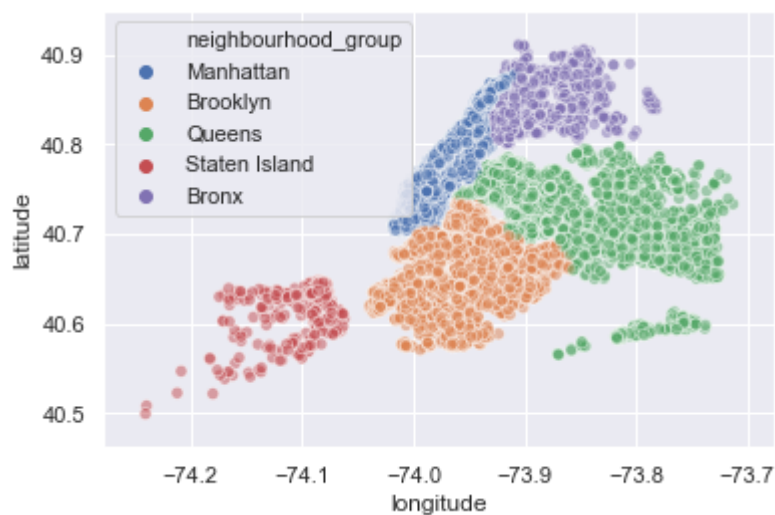
(48372, 16)

## Part 2: Data Visualization

### Scatterplot of Airbnb Listing Based on different Boroughs

In [266]:

```
sns.scatterplot(x='longitude', y='latitude', data=df, alpha=0.5, hue="neighbourhood_group");
```



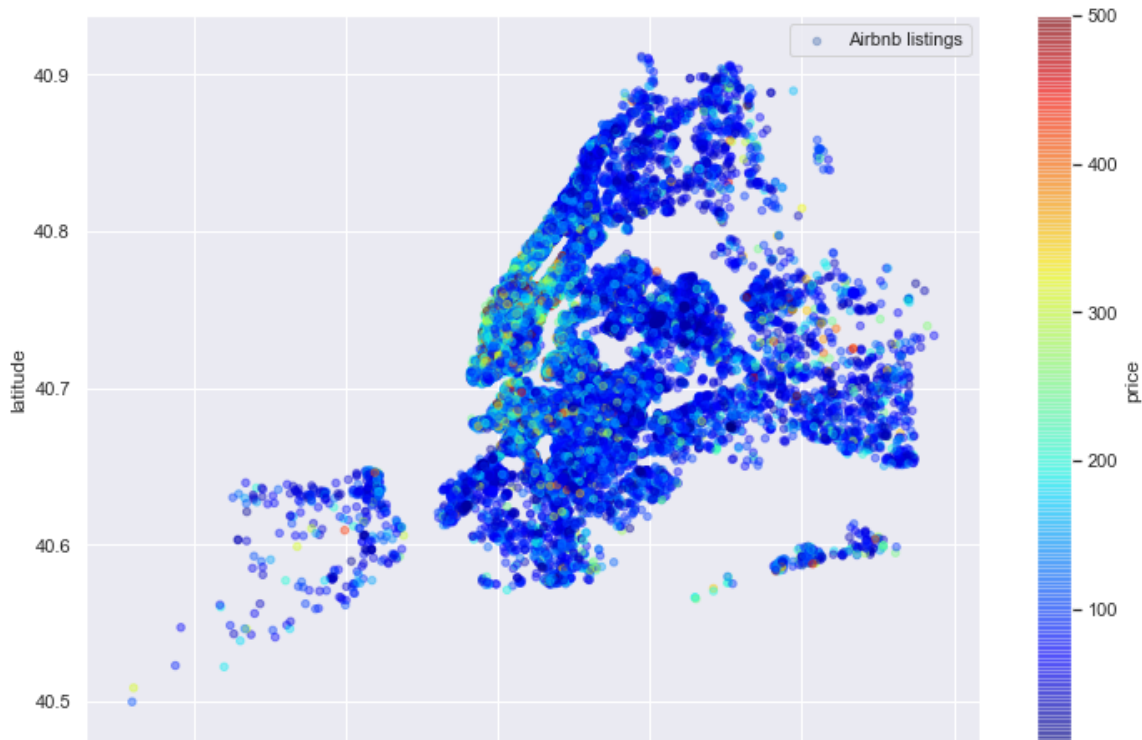
### Scatterplot of Airbnb Listing Based on Price

In [96]:

```
plot = df.plot(kind='scatter', x='longitude', y='latitude', c='price', label='Airbnb listings',  
                  cmap=plt.get_cmap('jet'), colorbar=True, alpha=0.4, figsize=(12,8))  
plot.legend()
```

Out[96]:

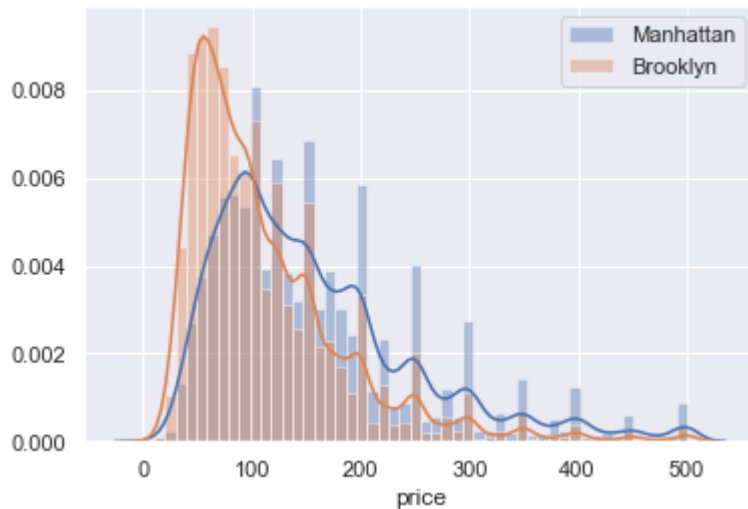
<matplotlib.legend.Legend at 0x2059d654188>



## Compare listing prices between Manhattan and Brooklyn

In [97]:

```
sns.distplot(df[df.neighbourhood_group == 'Manhattan'].price, label='Manhattan');
sns.distplot(df[df.neighbourhood_group == 'Brooklyn'].price, label='Brooklyn');
plt.legend();
```



## Summary Statistics of Price Group by Boroughs

In [98]:

```
#Brooklyn
sub_1=df.loc[airbnb['neighbourhood_group'] == 'Brooklyn']
price_sub1=sub_1[['price']]
#Manhattan
sub_2=df.loc[airbnb['neighbourhood_group'] == 'Manhattan']
price_sub2=sub_2[['price']]
#Queens
sub_3=df.loc[airbnb['neighbourhood_group'] == 'Queens']
price_sub3=sub_3[['price']]
#Staten Island
sub_4=df.loc[airbnb['neighbourhood_group'] == 'Staten Island']
price_sub4=sub_4[['price']]
#Bronx
sub_5=df.loc[airbnb['neighbourhood_group'] == 'Bronx']
price_sub5=sub_5[['price']]
#putting all the prices' dfs in the list
price_list_by_n=[price_sub1, price_sub2, price_sub3, price_sub4, price_sub5]
```

In [99]:

```

#creating an empty list that we will append later with price distributions for each neighbourhood_group
p_l_b_n_2=[]
#creating list with known values in neighbourhood_group column
nei_list=['Brooklyn', 'Manhattan', 'Queens', 'Staten Island', 'Bronx']
#creating a for loop to get statistics for price ranges and append it to our empty list
for x in price_list_by_n:
    i=x.describe(percentiles=[.25, .50, .75])
    i=i.iloc[3:]
    i.reset_index(inplace=True)
    i.rename(columns={'index':'Stats'}, inplace=True)
    p_l_b_n_2.append(i)
#changing names of the price column to the area name for easier reading of the table
p_l_b_n_2[0].rename(columns={'price':nei_list[0]}, inplace=True)
p_l_b_n_2[1].rename(columns={'price':nei_list[1]}, inplace=True)
p_l_b_n_2[2].rename(columns={'price':nei_list[2]}, inplace=True)
p_l_b_n_2[3].rename(columns={'price':nei_list[3]}, inplace=True)
p_l_b_n_2[4].rename(columns={'price':nei_list[4]}, inplace=True)
#finalizing our dataframe for final view
stat_df=p_l_b_n_2
stat_df=df.set_index('Stats')
stat_df=stat_df[0].join(stat_df[1:])
stat_df

```

Out[99]:

	Brooklyn	Manhattan	Queens	Staten Island	Bronx
Stats					
min	10.0	10.0	10.0	20.00	11.0
25%	60.0	90.0	50.0	50.00	49.0
50%	90.0	135.0	72.0	75.00	69.0
75%	145.0	200.0	109.0	109.75	100.0
max	500.0	500.0	500.0	429.00	500.0

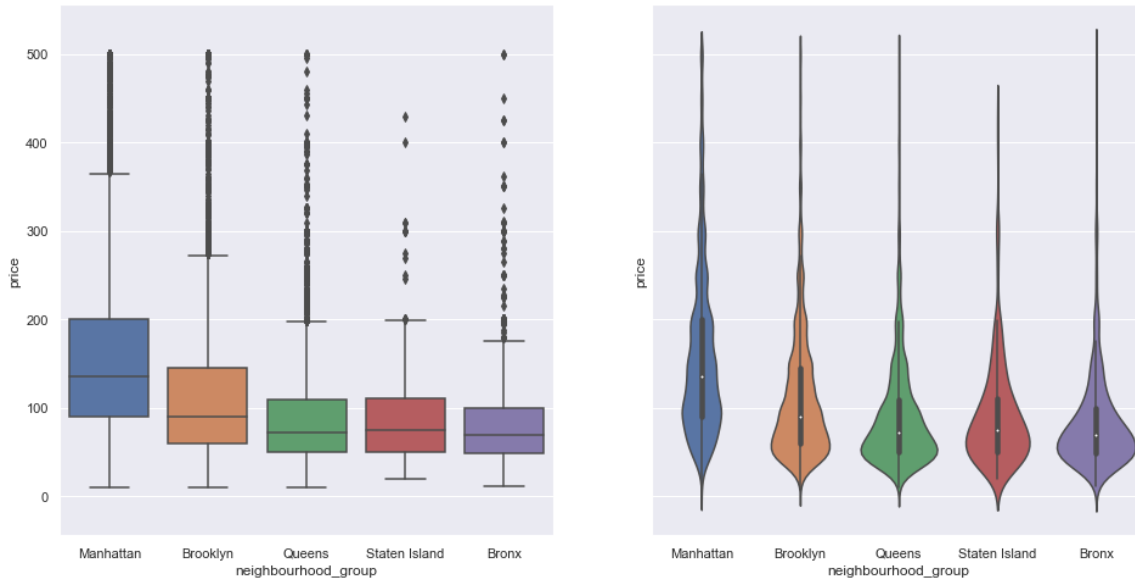
## Boxplot and Violinplot

In [100]:

```
fig, ax = plt.subplots(1, 2, figsize=(16, 8), sharey=True)
sns.boxplot(x='neighbourhood_group', y='price', data=df, ax=ax[0])
sns.violinplot(x='neighbourhood_group', y='price', data=df, ax=ax[1])
```

Out[100]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x205a0a79c08>

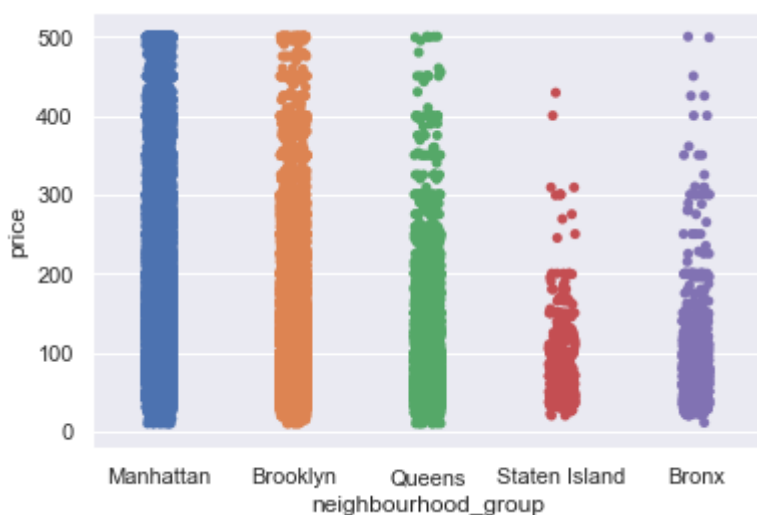


In [101]:

```
sns.stripplot(x='neighbourhood_group', y='price', data=df)
```

Out[101]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x205a05d1348>

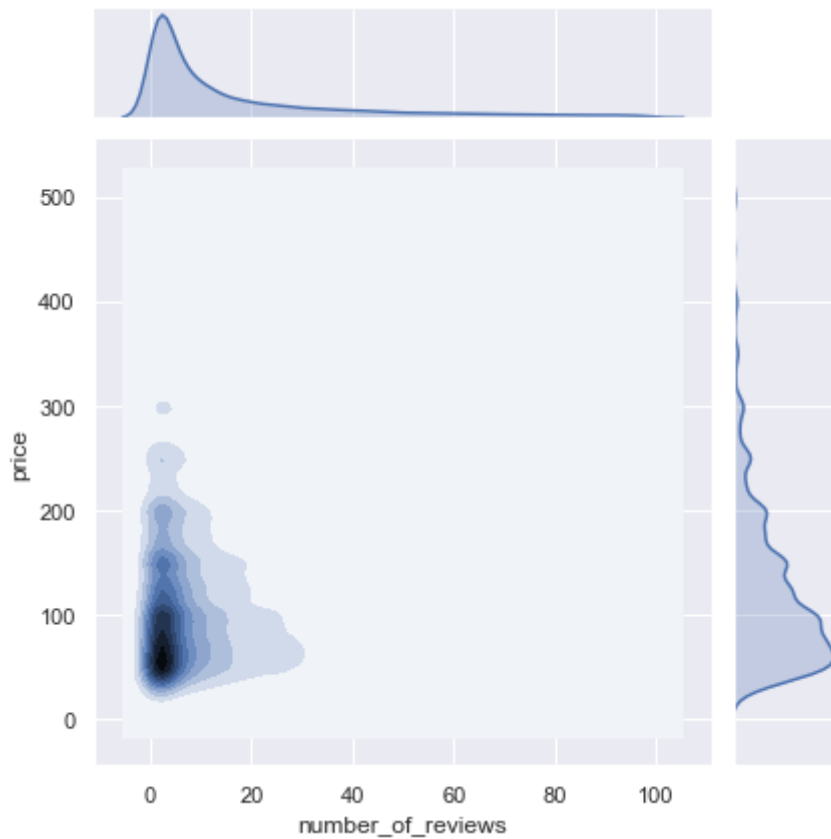


In [102]:

```
sns.jointplot(x='number_of_reviews', y='price',  
              data=df[(df.number_of_reviews < 100) & (df.number_of_reviews > 0)],  
              kind='kde')
```

Out[102]:

<seaborn.axisgrid.JointGrid at 0x205a05e8548>



## Part 3: Clustering and find suitable airbnb



In [103]:

```
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 geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
from geopy.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.

In [109]:

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

## Tranform the data into a pandas dataframe

In [112]:

```
neighborhoods_data = newyork_data['features']
```

In [113]:

```
# 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 [114]:

```
neighborhoods.head()
```

Out[114]:

	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

## Use geopy library to get the latitude and longitude values of Brooklyn

In [119]:

```
address = 'Brooklyn, NY'

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

The georapical coordinate of Brooklyn are 40.6501038, -73.9495823.

## Create a map of Brooklyn with neighborhoods superimposed on top

In [164]:

```
Brooklyn_data = neighborhoods[(neighborhoods['Borough'] == 'Brooklyn') | (neighborhoods['Borough'] == 'Manhattan')].reset_index(drop=True)
Brooklyn_data.head()
```

Out[164]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Brooklyn	Bay Ridge	40.625801	-74.030621
2	Brooklyn	Bensonhurst	40.611009	-73.995180
3	Brooklyn	Sunset Park	40.645103	-74.010316
4	Brooklyn	Greenpoint	40.730201	-73.954241

In [165]:

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

# add markers to map
for lat, lng, label in zip(Brooklyn_data['Latitude'], Brooklyn_data['Longitude'], Brooklyn_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_Brooklyn)

map_Brooklyn
```

Out[165]:

Make this Notebook Trusted to load map: File -> Trust Notebook

In [166]:

```
CLIENT_ID = '' # your Foursquare ID
CLIENT_SECRET = '' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
```

In [167]:

```
# create a function to get all the neighborhoods in Toronto
def getNearbyVenues(names, latitudes, longitudes, radius=500):
    venues_list=[]

    for name, lat, lng in zip(names, latitudes, longitudes):
        # print(name)

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

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

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

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

    return(nearby_venues)
```

In [168]:

```

LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 500 # define radius
df_venues = getNearbyVenues(names=Brooklyn_data['Neighborhood'],
                             latitudes=Brooklyn_data['Latitude'],
                             longitudes=Brooklyn_data['Longitude']
                             )
df_venues.head()

```

Out[168]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

In [169]:

```
# one hot encoding
df_venues_onehot = pd.get_dummies(df_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
df_venues_onehot['Neighborhood'] = df_venues['Neighborhood']

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

df_venues_onehot.head()
```

Out[169]:

	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Antique Shop	Re
0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	

In [170]:

```
df_venues_grouped = df_venues_onehot.groupby('Neighborhood').mean().reset_index()
df_venues_grouped.head()
```

Out[170]:

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	America Restaurant
0	Bath Beach	0.0	0.0	0.0	0.0	0.0	0.0	0.00000
1	Battery Park City	0.0	0.0	0.0	0.0	0.0	0.0	0.01492
2	Bay Ridge	0.0	0.0	0.0	0.0	0.0	0.0	0.03658
3	Bedford Stuyvesant	0.0	0.0	0.0	0.0	0.0	0.0	0.00000
4	Bensonhurst	0.0	0.0	0.0	0.0	0.0	0.0	0.00000

**Print each neighborhood along with the top 5 most common venues**

In [171]:

```
num_top_venues = 5

for hood in df_venues_grouped['Neighborhood']:
    print("-----"+hood+"-----")
    temp = df_venues_grouped[df_venues_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')
```



## ----Bath Beach----

	venue	freq
0	Pharmacy	0.07
1	Donut Shop	0.05
2	Chinese Restaurant	0.05
3	Pizza Place	0.05
4	Bubble Tea Shop	0.05

## ----Battery Park City----

	venue	freq
0	Park	0.09
1	Hotel	0.07
2	Gym	0.06
3	Coffee Shop	0.06
4	Memorial Site	0.04

## ----Bay Ridge----

	venue	freq
0	Italian Restaurant	0.07
1	Spa	0.06
2	Pizza Place	0.06
3	American Restaurant	0.04
4	Bagel Shop	0.04

## ----Bedford Stuyvesant----

	venue	freq
0	Coffee Shop	0.12
1	Bar	0.08
2	Café	0.08
3	Pizza Place	0.08
4	Deli / Bodega	0.04

## ----Bensonhurst----

	venue	freq
0	Chinese Restaurant	0.15
1	Italian Restaurant	0.08
2	Sushi Restaurant	0.08
3	Donut Shop	0.08
4	Ice Cream Shop	0.08

## ----Bergen Beach----

	venue	freq
0	Harbor / Marina	0.29
1	Baseball Field	0.14
2	Playground	0.14
3	Hockey Field	0.14
4	Athletics & Sports	0.14

## ----Boerum Hill----

	venue	freq
0	Bar	0.05
1	Coffee Shop	0.05
2	Dance Studio	0.05
3	Spa	0.03
4	Bakery	0.03

## ----Borough Park----

	venue	freq
0	Bank	0.22
1	Pizza Place	0.11
2	Deli / Bodega	0.11
3	Pharmacy	0.11
4	Fast Food Restaurant	0.11

## ----Brighton Beach----

	venue	freq
0	Restaurant	0.07
1	Beach	0.07
2	Russian Restaurant	0.07
3	Eastern European Restaurant	0.07
4	Sushi Restaurant	0.05

## ----Broadway Junction----

	venue	freq
0	Fried Chicken Joint	0.12
1	Donut Shop	0.12
2	Diner	0.12
3	Burger Joint	0.06
4	Nightclub	0.06

## ----Brooklyn Heights----

	venue	freq
0	Yoga Studio	0.04
1	Italian Restaurant	0.04
2	Deli / Bodega	0.04
3	Park	0.04
4	Pizza Place	0.03

## ----Brownsville----

	venue	freq
0	Fried Chicken Joint	0.11
1	Restaurant	0.11
2	Chinese Restaurant	0.11
3	Park	0.06
4	Performing Arts Venue	0.06

## ----Bushwick----

	venue	freq
0	Bar	0.10
1	Coffee Shop	0.07
2	Mexican Restaurant	0.07
3	Deli / Bodega	0.06
4	Discount Store	0.04

## ----Canarsie----

	venue	freq
0	Deli / Bodega	0.14
1	Asian Restaurant	0.14
2	Chinese Restaurant	0.14

3	Food	0.14
4	Gym	0.14

## ----Carnegie Hill----

	venue	freq
0	Coffee Shop	0.07
1	Café	0.04
2	Pizza Place	0.04
3	Bar	0.04
4	Bookstore	0.03

## ----Carroll Gardens----

	venue	freq
0	Italian Restaurant	0.11
1	Coffee Shop	0.07
2	Pizza Place	0.05
3	Bakery	0.05
4	Cocktail Bar	0.04

## ----Central Harlem----

	venue	freq
0	African Restaurant	0.07
1	French Restaurant	0.04
2	American Restaurant	0.04
3	Cosmetics Shop	0.04
4	Chinese Restaurant	0.04

## ----Chelsea----

	venue	freq
0	Coffee Shop	0.09
1	Art Gallery	0.05
2	Bakery	0.04
3	Seafood Restaurant	0.03
4	Italian Restaurant	0.03

## ----Chinatown----

	venue	freq
0	Chinese Restaurant	0.10
1	Cocktail Bar	0.05
2	Bakery	0.05
3	Vietnamese Restaurant	0.03
4	Salon / Barbershop	0.03

## ----City Line----

	venue	freq
0	Donut Shop	0.09
1	Pharmacy	0.06
2	Grocery Store	0.06
3	Mexican Restaurant	0.03
4	Bakery	0.03

## ----Civic Center----

	venue	freq
0	Coffee Shop	0.08

1	Cocktail Bar	0.05
2	Hotel	0.05
3	Italian Restaurant	0.04
4	Gym / Fitness Center	0.04

----Clinton----

	venue	freq
0	Italian Restaurant	0.06
1	Theater	0.06
2	Gym / Fitness Center	0.05
3	American Restaurant	0.04
4	Coffee Shop	0.04

----Clinton Hill----

	venue	freq
0	Italian Restaurant	0.05
1	Pizza Place	0.05
2	Wine Shop	0.04
3	Mexican Restaurant	0.04
4	Thai Restaurant	0.04

----Cobble Hill----

	venue	freq
0	Bar	0.04
1	Italian Restaurant	0.04
2	Cocktail Bar	0.04
3	Coffee Shop	0.04
4	Playground	0.04

----Coney Island----

	venue	freq
0	Baseball Stadium	0.12
1	Monument / Landmark	0.12
2	Intersection	0.06
3	Brewery	0.06
4	Beach	0.06

----Crown Heights----

	venue	freq
0	Pizza Place	0.14
1	Café	0.10
2	Museum	0.10
3	Burger Joint	0.05
4	Salon / Barbershop	0.05

----Cypress Hills----

	venue	freq
0	Latin American Restaurant	0.09
1	Fried Chicken Joint	0.06
2	Pizza Place	0.06
3	Spanish Restaurant	0.06
4	Donut Shop	0.06

----Ditmas Park----

	venue	freq
0	Chinese Restaurant	0.06
1	Deli / Bodega	0.06
2	Burger Joint	0.04
3	Donut Shop	0.04
4	Caribbean Restaurant	0.04

## ----Downtown----

	venue	freq
0	Burger Joint	0.05
1	Pizza Place	0.04
2	Coffee Shop	0.04
3	Wine Shop	0.03
4	Grocery Store	0.03

## ----Dumbo----

	venue	freq
0	Coffee Shop	0.08
1	Park	0.08
2	Bakery	0.05
3	Scenic Lookout	0.05
4	Yoga Studio	0.03

## ----Dyker Heights----

	venue	freq
0	Burger Joint	0.17
1	Golf Course	0.17
2	Bus Station	0.17
3	Bagel Shop	0.17
4	Plaza	0.17

## ----East Flatbush----

	venue	freq
0	Food & Drink Shop	0.08
1	Fast Food Restaurant	0.08
2	Liquor Store	0.08
3	Caribbean Restaurant	0.08
4	Park	0.08

## ----East Harlem----

	venue	freq
0	Mexican Restaurant	0.12
1	Thai Restaurant	0.07
2	Bakery	0.07
3	Deli / Bodega	0.05
4	Pizza Place	0.05

## ----East New York----

	venue	freq
0	Deli / Bodega	0.27
1	Pizza Place	0.07
2	Food Truck	0.07
3	Bus Stop	0.07
4	Music Venue	0.07

## ----East Village----

	venue	freq
0	Bar	0.07
1	Pizza Place	0.05
2	Ice Cream Shop	0.05
3	Cocktail Bar	0.04
4	Mexican Restaurant	0.04

## ----East Williamsburg----

	venue	freq
0	Bar	0.10
1	Deli / Bodega	0.07
2	Bakery	0.06
3	Coffee Shop	0.06
4	Cocktail Bar	0.06

## ----Erasmus----

	venue	freq
0	Caribbean Restaurant	0.17
1	Grocery Store	0.13
2	Yoga Studio	0.04
3	Bank	0.04
4	Playground	0.04

## ----Financial District----

	venue	freq
0	Coffee Shop	0.10
1	Bar	0.04
2	Gym	0.04
3	Cocktail Bar	0.04
4	Pizza Place	0.04

## ----Flatbush----

	venue	freq
0	Pharmacy	0.09
1	Caribbean Restaurant	0.09
2	Mexican Restaurant	0.09
3	Coffee Shop	0.09
4	Middle Eastern Restaurant	0.04

## ----Flatiron----

	venue	freq
0	Gym / Fitness Center	0.06
1	Italian Restaurant	0.04
2	New American Restaurant	0.04
3	Yoga Studio	0.03
4	Sporting Goods Shop	0.03

## ----Flatlands----

	venue	freq
0	Pharmacy	0.19
1	Fried Chicken Joint	0.12
2	Fast Food Restaurant	0.12
3	Caribbean Restaurant	0.12

4            Deli / Bodega   0.06

----Fort Greene----

	venue	freq
0	Coffee Shop	0.04
1	Italian Restaurant	0.04
2	Wine Shop	0.04
3	Pizza Place	0.04
4	Flower Shop	0.04

----Fort Hamilton----

	venue	freq
0	Italian Restaurant	0.06
1	Chinese Restaurant	0.05
2	Sandwich Place	0.05
3	Gym / Fitness Center	0.05
4	Deli / Bodega	0.03

----Fulton Ferry----

	venue	freq
0	Park	0.14
1	Scenic Lookout	0.05
2	American Restaurant	0.05
3	Bridge	0.03
4	Bakery	0.03

----Georgetown----

	venue	freq
0	Bank	0.10
1	Donut Shop	0.07
2	Pharmacy	0.07
3	Supermarket	0.03
4	Breakfast Spot	0.03

----Gerritsen Beach----

	venue	freq
0	Bar	0.13
1	Harbor / Marina	0.09
2	Ice Cream Shop	0.09
3	Pizza Place	0.09
4	Deli / Bodega	0.04

----Gowanus----

	venue	freq
0	Italian Restaurant	0.07
1	Bar	0.07
2	Furniture / Home Store	0.05
3	Coffee Shop	0.03
4	Pizza Place	0.03

----Gramercy----

	venue	freq
0	Bar	0.07
1	Coffee Shop	0.05

2	Pizza Place	0.05
3	Italian Restaurant	0.05
4	American Restaurant	0.05

## ----Gravesend----

	venue	freq
0	Bakery	0.12
1	Italian Restaurant	0.12
2	Lounge	0.12
3	Pizza Place	0.08
4	Music Venue	0.04

## ----Greenpoint----

	venue	freq
0	Bar	0.09
1	Pizza Place	0.07
2	Coffee Shop	0.06
3	Cocktail Bar	0.06
4	Yoga Studio	0.03

## ----Greenwich Village----

	venue	freq
0	Italian Restaurant	0.12
1	Sushi Restaurant	0.05
2	Café	0.04
3	Bubble Tea Shop	0.03
4	Indian Restaurant	0.03

## ----Hamilton Heights----

	venue	freq
0	Pizza Place	0.09
1	Coffee Shop	0.07
2	Café	0.07
3	Mexican Restaurant	0.05
4	Deli / Bodega	0.05

## ----Highland Park----

	venue	freq
0	Deli / Bodega	0.12
1	Grocery Store	0.12
2	Liquor Store	0.06
3	Pizza Place	0.06
4	Cosmetics Shop	0.06

## ----Homecrest----

	venue	freq
0	Bank	0.11
1	Donut Shop	0.08
2	Grocery Store	0.05
3	Mexican Restaurant	0.05
4	Chinese Restaurant	0.05

## ----Hudson Yards----

	venue	freq
--	-------	------



0	Hotel	0.07
1	American Restaurant	0.07
2	Gym / Fitness Center	0.07
3	Italian Restaurant	0.05
4	Café	0.03

----Inwood----

	venue	freq
0	Mexican Restaurant	0.07
1	Restaurant	0.06
2	Lounge	0.06
3	Café	0.06
4	Pizza Place	0.04

----Kensington----

	venue	freq
0	Grocery Store	0.09
1	Thai Restaurant	0.09
2	Ice Cream Shop	0.06
3	Pizza Place	0.06
4	Deli / Bodega	0.03

----Lenox Hill----

	venue	freq
0	Coffee Shop	0.07
1	Italian Restaurant	0.06
2	Sushi Restaurant	0.05
3	Café	0.04
4	Pizza Place	0.04

----Lincoln Square----

	venue	freq
0	Café	0.05
1	Plaza	0.05
2	Gym / Fitness Center	0.05
3	Theater	0.04
4	Italian Restaurant	0.04

----Little Italy----

	venue	freq
0	Bakery	0.06
1	Chinese Restaurant	0.05
2	Italian Restaurant	0.05
3	Café	0.05
4	Ice Cream Shop	0.04

----Lower East Side----

	venue	freq
0	Chinese Restaurant	0.06
1	Art Gallery	0.04
2	Ramen Restaurant	0.04
3	Bakery	0.04
4	Park	0.04

## ----Madison----

	venue	freq
0	Bagel Shop	0.22
1	Deli / Bodega	0.11
2	Spa	0.11
3	Pizza Place	0.11
4	Restaurant	0.11

## ----Manhattan Beach----

	venue	freq
0	Café	0.2
1	Beach	0.1
2	Harbor / Marina	0.1
3	Bus Stop	0.1
4	Ice Cream Shop	0.1

## ----Manhattan Terrace----

	venue	freq
0	Pizza Place	0.14
1	Donut Shop	0.09
2	Ice Cream Shop	0.09
3	Japanese Restaurant	0.05
4	Convenience Store	0.05

## ----Manhattan Valley----

	venue	freq
0	Bar	0.07
1	Yoga Studio	0.04
2	Coffee Shop	0.04
3	Thai Restaurant	0.04
4	Mexican Restaurant	0.04

## ----Manhattanville----

	venue	freq
0	Coffee Shop	0.10
1	Seafood Restaurant	0.05
2	Mexican Restaurant	0.05
3	Bar	0.05
4	Sushi Restaurant	0.05

## ----Marble Hill----

	venue	freq
0	Coffee Shop	0.08
1	Gym	0.08
2	Seafood Restaurant	0.04
3	Miscellaneous Shop	0.04
4	Supplement Shop	0.04

## ----Marine Park----

	venue	freq
0	Coffee Shop	0.08
1	Ice Cream Shop	0.08
2	Basketball Court	0.08
3	Baseball Field	0.08
4	Chinese Restaurant	0.08

## ----Midtown----

	venue	freq
0	Coffee Shop	0.07
1	Clothing Store	0.06
2	Hotel	0.05
3	Bakery	0.04
4	Theater	0.04

## ----Midtown South----

	venue	freq
0	Korean Restaurant	0.15
1	Hotel	0.06
2	Japanese Restaurant	0.04
3	Bakery	0.03
4	Cosmetics Shop	0.03

## ----Midwood----

	venue	freq
0	Pizza Place	0.36
1	Bakery	0.09
2	Ice Cream Shop	0.09
3	Convenience Store	0.09
4	Pharmacy	0.09

## ----Mill Basin----

	venue	freq
0	Chinese Restaurant	0.11
1	Pizza Place	0.08
2	Japanese Restaurant	0.08
3	Cosmetics Shop	0.06
4	Bank	0.06

## ----Mill Island----

	venue	freq
0	Locksmith	0.5
1	Pool	0.5
2	Yoga Studio	0.0
3	Other Nightlife	0.0
4	Pedestrian Plaza	0.0

## ----Morningside Heights----

	venue	freq
0	Park	0.10
1	American Restaurant	0.07
2	Coffee Shop	0.07
3	Bookstore	0.07
4	Food Truck	0.05

## ----Murray Hill----

	venue	freq
0	Sandwich Place	0.06
1	Japanese Restaurant	0.05
2	Coffee Shop	0.05

3	Pizza Place	0.03
4	Mediterranean Restaurant	0.03

## ----New Lots----

	venue	freq
0	Pizza Place	0.11
1	Fried Chicken Joint	0.11
2	Deli / Bodega	0.05
3	Bus Stop	0.05
4	Breakfast Spot	0.05

## ----Noho----

	venue	freq
0	Italian Restaurant	0.06
1	Coffee Shop	0.05
2	Pizza Place	0.04
3	French Restaurant	0.03
4	Rock Club	0.03

## ----North Side----

	venue	freq
0	Coffee Shop	0.10
1	Pizza Place	0.05
2	Wine Bar	0.05
3	American Restaurant	0.04
4	Bar	0.04

## ----Ocean Hill----

	venue	freq
0	Deli / Bodega	0.19
1	Southern / Soul Food Restaurant	0.07
2	Food	0.07
3	Playground	0.07
4	Convenience Store	0.04

## ----Ocean Parkway----

	venue	freq
0	Sushi Restaurant	0.18
1	Eastern European Restaurant	0.06
2	Pizza Place	0.06
3	Pharmacy	0.06
4	General Entertainment	0.06

## ----Paerdegat Basin----

	venue	freq
0	Harbor / Marina	0.25
1	Food	0.25
2	Home Service	0.25
3	Asian Restaurant	0.25
4	Yoga Studio	0.00

## ----Park Slope----

	venue	freq
0	Coffee Shop	0.08

1	Burger Joint	0.06
2	Italian Restaurant	0.05
3	Pizza Place	0.05
4	Bagel Shop	0.05

----Prospect Heights----

	venue	freq
0	Bar	0.11
1	Mexican Restaurant	0.07
2	Thai Restaurant	0.04
3	Bakery	0.04
4	Wine Shop	0.04

----Prospect Lefferts Gardens----

	venue	freq
0	Café	0.09
1	Bakery	0.09
2	Caribbean Restaurant	0.07
3	Pizza Place	0.07
4	Sushi Restaurant	0.04

----Prospect Park South----

	venue	freq
0	Caribbean Restaurant	0.08
1	Pizza Place	0.06
2	Mobile Phone Shop	0.06
3	Fast Food Restaurant	0.06
4	Grocery Store	0.06

----Red Hook----

	venue	freq
0	Seafood Restaurant	0.08
1	Art Gallery	0.06
2	Bar	0.06
3	American Restaurant	0.06
4	Park	0.06

----Remsen Village----

	venue	freq
0	Caribbean Restaurant	0.16
1	Fast Food Restaurant	0.11
2	Deli / Bodega	0.05
3	Construction & Landscaping	0.05
4	Gas Station	0.05

----Roosevelt Island----

	venue	freq
0	Park	0.07
1	Deli / Bodega	0.04
2	Coffee Shop	0.04
3	Scenic Lookout	0.04
4	School	0.04

----Rugby----

	venue	freq
0	Caribbean Restaurant	0.10
1	Grocery Store	0.10
2	Bank	0.10
3	Diner	0.05
4	Pharmacy	0.05

----Sea Gate----

	venue	freq
0	Sports Club	0.2
1	Beach	0.2
2	Bus Station	0.2
3	Spa	0.2
4	American Restaurant	0.2

----Sheepshead Bay----

	venue	freq
0	Dessert Shop	0.12
1	Turkish Restaurant	0.12
2	Sandwich Place	0.08
3	Playground	0.04
4	Pizza Place	0.04

----Soho----

	venue	freq
0	Italian Restaurant	0.09
1	Coffee Shop	0.07
2	Clothing Store	0.04
3	French Restaurant	0.04
4	Mediterranean Restaurant	0.04

----South Side----

	venue	freq
0	Bar	0.08
1	Coffee Shop	0.07
2	Pizza Place	0.05
3	Wine Bar	0.04
4	American Restaurant	0.04

----Starrett City----

	venue	freq
0	Bus Station	0.12
1	Bus Stop	0.12
2	American Restaurant	0.12
3	Shopping Mall	0.12
4	Pharmacy	0.12

----Stuyvesant Town----

	venue	freq
0	Boat or Ferry	0.16
1	Park	0.11
2	Playground	0.05
3	Farmers Market	0.05
4	Skating Rink	0.05

## ----Sunset Park----

	venue	freq
0	Mobile Phone Shop	0.09
1	Mexican Restaurant	0.09
2	Bank	0.09
3	Bakery	0.09
4	Pizza Place	0.09

## ----Sutton Place----

	venue	freq
0	Italian Restaurant	0.07
1	Gym / Fitness Center	0.05
2	Furniture / Home Store	0.04
3	Coffee Shop	0.04
4	Park	0.04

## ----Tribeca----

	venue	freq
0	Park	0.06
1	Italian Restaurant	0.06
2	Spa	0.04
3	Wine Bar	0.04
4	American Restaurant	0.04

## ----Tudor City----

	venue	freq
0	Park	0.06
1	Mexican Restaurant	0.06
2	Café	0.06
3	Sushi Restaurant	0.04
4	Greek Restaurant	0.04

## ----Turtle Bay----

	venue	freq
0	Coffee Shop	0.05
1	Sushi Restaurant	0.05
2	Italian Restaurant	0.05
3	Wine Bar	0.04
4	Park	0.04

## ----Upper East Side----

	venue	freq
0	Italian Restaurant	0.09
1	Bakery	0.05
2	Coffee Shop	0.05
3	Pizza Place	0.04
4	Gym / Fitness Center	0.04

## ----Upper West Side----

	venue	freq
0	Italian Restaurant	0.07
1	Coffee Shop	0.04
2	Bar	0.04
3	Wine Bar	0.04

4 Bakery 0.03

----Vinegar Hill----

	venue	freq
0	Food Truck	0.18
1	Coffee Shop	0.11
2	Café	0.07
3	Art Gallery	0.07
4	Ice Cream Shop	0.04

----Washington Heights----

	venue	freq
0	Café	0.06
1	Bakery	0.04
2	Chinese Restaurant	0.03
3	Grocery Store	0.03
4	Mobile Phone Shop	0.03

----Weeksville----

	venue	freq
0	Discount Store	0.12
1	Chinese Restaurant	0.12
2	Deli / Bodega	0.06
3	Liquor Store	0.06
4	Café	0.06

----West Village----

	venue	freq
0	Italian Restaurant	0.09
1	New American Restaurant	0.06
2	American Restaurant	0.05
3	Cocktail Bar	0.04
4	Park	0.04

----Williamsburg----

	venue	freq
0	Coffee Shop	0.09
1	Bar	0.09
2	Bagel Shop	0.06
3	Pizza Place	0.06
4	Lounge	0.03

----Windsor Terrace----

	venue	freq
0	Café	0.07
1	Plaza	0.07
2	Diner	0.07
3	Grocery Store	0.07
4	Park	0.07

----Wingate----

	venue	freq
0	Fried Chicken Joint	0.10
1	Deli / Bodega	0.05



2	Field	0.05
3	Discount Store	0.05
4	Donut Shop	0.05

----Yorkville----

	venue	freq
0	Italian Restaurant	0.08
1	Coffee Shop	0.06
2	Gym	0.06
3	Bar	0.06
4	Deli / Bodega	0.04

**Create the new dataframe and display the top 10 venues for each neighborhood.**

In [172]:

```
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 [216]:

```

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'] = df_venues_grouped['Neighborhood']

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

neighborhoods_venues_sorted.head()

```

Out[216]:

	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 M Comrn Ver
0	Bath Beach	Pharmacy	Chinese Restaurant	Pizza Place	Bubble Tea Shop	Fast Food Restaurant	Italian Restaurant	(
1	Battery Park City	Park	Hotel	Gym	Coffee Shop	Boat or Ferry	Memorial Site	Pl
2	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Bar	American Restaurant	Bagel Shop	Pharm
3	Bedford Stuyvesant	Coffee Shop	Pizza Place	Café	Bar	Deli / Bodega	Juice Bar	Gift SI
4	Bensonhurst	Chinese Restaurant	Ice Cream Shop	Italian Restaurant	Sushi Restaurant	Donut Shop	Noodle House	Ba SI

## Cluster Neighborhoods

In [217]:

```
# set number of clusters
kclusters = 10

df_venues_grouped_clustering = df_venues_grouped.drop('Neighborhood', 1)

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

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

Out[217]:

```
array([2, 1, 2, 1, 2, 4, 1, 9, 2, 9])
```

In [218]:

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

df__merged = Brooklyn_data

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
df__merged = df__merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

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

Out[218]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Gym	Coffee Shop	Seafood Restaurant
1	Brooklyn	Bay Ridge	40.625801	-74.030621	2	Italian Restaurant	Spa	Pizza Place
2	Brooklyn	Bensonhurst	40.611009	-73.995180	2	Chinese Restaurant	Ice Cream Shop	Italian Restaurant
3	Brooklyn	Sunset Park	40.645103	-74.010316	9	Pizza Place	Latin American Restaurant	Mobile Phone Shop
4	Brooklyn	Greenpoint	40.730201	-73.954241	2	Bar	Pizza Place	Coffee Shop

In [219]:

```
# 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(df__merged['Latitude'], df__merged['Longitude'], df__merged[
'Neighborhood'], df__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[219]:

Make this Notebook Trusted to load map: File -> Trust Notebook

In [223]:

```
df_merged.loc[(df_merged['Cluster Labels'] == 2) & (df_merged['Borough'] == 'Manhattan'), df_
_merged.columns[[1] + list(range(5, df_merged.shape[1]))]]
```

Out[223]:

	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 Common Venue
55	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Optical Shop	Spa	Ice Cream Shop	Bubb
56	Washington Heights	Café	Bakery	Mobile Phone Shop	Chinese Restaurant	Grocery Store	Deli / Bodega	Me
57	Inwood	Mexican Restaurant	Café	Lounge	Restaurant	Frozen Yogurt Shop	Deli / Bodega	Caril
58	Hamilton Heights	Pizza Place	Café	Coffee Shop	Deli / Bodega	Mexican Restaurant	Yoga Studio	San
60	Central Harlem	African Restaurant	Cosmetics Shop	American Restaurant	Seafood Restaurant	French Restaurant	Chinese Restaurant	
61	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Sandwich Place	Deli / Bodega	Pizza Place	Am
73	East Village	Bar	Pizza Place	Ice Cream Shop	Mexican Restaurant	Cocktail Bar	Italian Restaurant	Wir
74	Lower East Side	Chinese Restaurant	Japanese Restaurant	Park	Art Gallery	Ramen Restaurant	Bakery	
79	Manhattan Valley	Bar	Yoga Studio	Mexican Restaurant	Thai Restaurant	Pizza Place	Coffee Shop	Playg
81	Gramercy	Bar	Bagel Shop	Coffee Shop	Italian Restaurant	American Restaurant	Pizza Place	Gi

Result:

Assume I live in Kensington, Brooklyn. I want to find an Airbib in Manhattan:

- 1. I want to find an Airbnb in a neighborhood that has similar venues as Kensington, Brooklyn
- 1. The price should be 100-200 usd per day
- 1. Number of reviews should be greater than 50

1. Since I only want to spend 3 days in Manhattan, the minimum nights should not exceed 3 days

1. availability should be 365 days

In [267]:

```
df_filtered = df[(df.price >= 100) & (df.price <= 200)]
df_filtered = df_filtered[df_filtered.number_of_reviews > 50]
df_filtered = df_filtered[df_filtered.minimum_nights <= 3]
df_filtered = df_filtered[df_filtered.availability_365 == 365]
df_filtered.shape
```

Out[267]:

(140, 16)

In [289]:

```
# Find cluster of Kensington
cluster = list(df__merged.loc[df__merged.Neighborhood == 'Kensington', 'Cluster Labels'])
cluster
```

Out[289]:

[2]

In [290]:

```
# Find Airbnb listings that match the filter condition
selected_neighborhood = list(df__merged.loc[(df__merged['Cluster Labels'] == 2) & (df__merged['Borough'] == 'Manhattan')].Neighborhood)
select = df_filtered.neighbourhood.apply(lambda x: any(item for item in selected_neighborhood if item in x))
df_filtered = df_filtered[select]
df_filtered.shape
```

Out[290]:

(17, 16)

## Create map view for selected Airbnb listings

In [295]:

```
address = 'Manhattan, NY'

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

The georapical coordinate of Manhattan are 40.7896239, -73.9598939.

In [304]:

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

for lat, lng, label in zip(df_filtered['latitude'], df_filtered['longitude'], df_filtered['neighbourhood']):
    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_airbnb)

map_airbnb
```

Out[304]:

Make this Notebook Trusted to load map: File -> Trust Notebook

In [ ]: