Business Problem

The estate department of the government of Colombia is looking for the best location for its new consulate to serve Colombian nationals residing in the city of New York. In order to solve this problem, I will explore all the boroughs of NYC to find the areas most populated with Colombian people. I will do this by finding out which borough has the larger number of Colombian restaurants this will give me the right indication where the Colombian population resides within NYC and therefore where is the best location for the new Colombian consulate, which needs to be close to the people that it will serve.

Data

To solve this problem I will use the dataset provided in the lab https://cocl.us/new_york_dataset (https://cocl.us/new_york_dataset) which has the necessary information to segment the citie's neighborhoods. Also I will use the Fousquare API to explore the venues.

In [11]:
```
#!conda install -c conda-forge folium=0.5.0 --yes
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import requests
import sys
from bs4 import BeautifulSoup
import os
import folium # map rendering library
from geopy.geocoders import Nominatim
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
%matplotlib inline


print('Libraries imported.')
```

```
Libraries imported.
```

Now we define a function to get the geocodes i.e latitude and longitude of a given location using geopy.

In [12]:
```
def geo_location(address):
    # get geo location of address
    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    return latitude,longitude
```

We query the FourSquare API to get the top 300 venues within a radius of 2000 mts for a given latitude and longitude. The output will be venue id , venue name and category.

In [13]:
```
def get_venues(lat,lng):

    #set variables
    radius=2000
    LIMIT=300
    CLIENT_ID = 'ABKHZGTMF5MNAOGB0K2SXYLDJFHEJUSQPWOTDTSWTLIBXRX4' #Foursquare ID
    CLIENT_SECRET = 'FO50QPHK05HPVMOS3ZNASGMCSVFVWMP2IZIDZZMACXLQBZI2' #Foursquare
Secret
    VERSION = '20180605' # Foursquare API version

    #url to fetch data from foursquare api
    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)

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]['groups'][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

    column_names=['ID','Name','Category']
    df = pd.DataFrame(venue_details,columns=column_names)
    print("done")
    return df
```

Now we get venue details like name, rating , tips, likes etc.

In [14]:
```
def get_venue_details(venue_id):

    CLIENT_ID = 'ABKHZGTMF5MNAOGB0K2SXYLDJFHEJUSQPWOTDTSWTLIBXRX4' #Foursquare ID
    CLIENT_SECRET = 'FO50QPHK05HPVMOS3ZNASGMCSVFVWMP2IZIDZZMACXLQBZI2' #Foursquare
Secret
    VERSION = '20180605' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v
={}'.format(
            venue_id,
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION)

    # get all the data
    results = requests.get(url).json()
    venue_data=results['response']['venue']
    venue_details=[]
    try:
        venue_id=venue_data['id']
        venue_name=venue_data['name']
        venue_likes=venue_data['likes']['count']
        venue_rating=venue_data['rating']
        venue_tips=venue_data['tips']['count']
        venue_details.append([venue_id,venue_name,venue_likes,venue_rating,venue_ti
ps])
    except KeyError:
        pass

    column_names=['ID','Name','Likes','Rating','Tips']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

Now we get NYC details such as boroughs, neighborhoods and locations (latitude & longitude).

In [15]:
```
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

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

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

        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_in
dex=True)

    return new_york_data
```

We call the object to get NYC data.

In [16]:
```
# get new york data
new_york_data=get_new_york_data()
new_york_data.head()
```

Out[16]:

	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

In [17]:
```
new_york_data.shape
```

Out[17]:
```
(306, 4)
```

There are total of 306 different Neighborhoods in NYC

Now we collect Colombian restaurants for each Neighborhood

In [18]:
```
# prepare neighborhood list that contains Colombian restaurants
column_names=['Borough', 'Neighborhood', 'ID','Name']
colombian_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    colombian_restaurants=venues[venues['Category']=='Colombian Restaurant']
    print('(',count,'/',len(new_york_data),')','Colombian Restaurants in '+Neighbor
hood+', '+Borough+':'+str(len(colombian_restaurants)))
    for restaurant_detail in colombian_restaurants.values.tolist():
        id, name , category=restaurant_detail
        colombian_rest_ny = colombian_rest_ny.append({'Borough': Borough,
                                                      'Neighborhood': Neighborhood,
                                                      'ID': id,
                                                      'Name' : name
                                                     }, ignore_index=True)
    count+=1
```

```
done
( 1 / 306 ) Colombian Restaurants in Wakefield, Bronx:0
done
( 2 / 306 ) Colombian Restaurants in Co-op City, Bronx:0
done
( 3 / 306 ) Colombian Restaurants in Eastchester, Bronx:0
done
( 4 / 306 ) Colombian Restaurants in Fieldston, Bronx:0
done
( 5 / 306 ) Colombian Restaurants in Riverdale, Bronx:0
done
( 6 / 306 ) Colombian Restaurants in Kingsbridge, Bronx:0
done
( 7 / 306 ) Colombian Restaurants in Marble Hill, Manhattan:0
done
( 8 / 306 ) Colombian Restaurants in Woodlawn, Bronx:0
done
( 9 / 306 ) Colombian Restaurants in Norwood, Bronx:0
done
( 10 / 306 ) Colombian Restaurants in Williamsbridge, Bronx:0
done
( 11 / 306 ) Colombian Restaurants in Baychester, Bronx:0
done
( 12 / 306 ) Colombian Restaurants in Pelham Parkway, Bronx:0
done
( 13 / 306 ) Colombian Restaurants in City Island, Bronx:0
done
( 14 / 306 ) Colombian Restaurants in Bedford Park, Bronx:0
done
( 15 / 306 ) Colombian Restaurants in University Heights, Bronx:0
done
( 16 / 306 ) Colombian Restaurants in Morris Heights, Bronx:0
done
( 17 / 306 ) Colombian Restaurants in Fordham, Bronx:0
done
( 18 / 306 ) Colombian Restaurants in East Tremont, Bronx:0
done
( 19 / 306 ) Colombian Restaurants in West Farms, Bronx:0
done
( 20 / 306 ) Colombian Restaurants in High  Bridge, Bronx:0
done
( 21 / 306 ) Colombian Restaurants in Melrose, Bronx:0
done
( 22 / 306 ) Colombian Restaurants in Mott Haven, Bronx:0
done
( 23 / 306 ) Colombian Restaurants in Port Morris, Bronx:0
done
( 24 / 306 ) Colombian Restaurants in Longwood, Bronx:0
done
( 25 / 306 ) Colombian Restaurants in Hunts Point, Bronx:0
done
( 26 / 306 ) Colombian Restaurants in Morrisania, Bronx:0
done
( 27 / 306 ) Colombian Restaurants in Soundview, Bronx:0
done
( 28 / 306 ) Colombian Restaurants in Clason Point, Bronx:0
done
( 29 / 306 ) Colombian Restaurants in Throgs Neck, Bronx:0
done
( 30 / 306 ) Colombian Restaurants in Country Club, Bronx:0
done
( 31 / 306 ) Colombian Restaurants in Parkchester, Bronx:0
done
( 32 / 306 ) Colombian Restaurants in Westchester Square, Bronx:0
```

Now we got all Colombian restaurants in NYC

In [19]:
```
colombian_rest_ny.shape
```

Out[19]:
```
(10, 4)
```

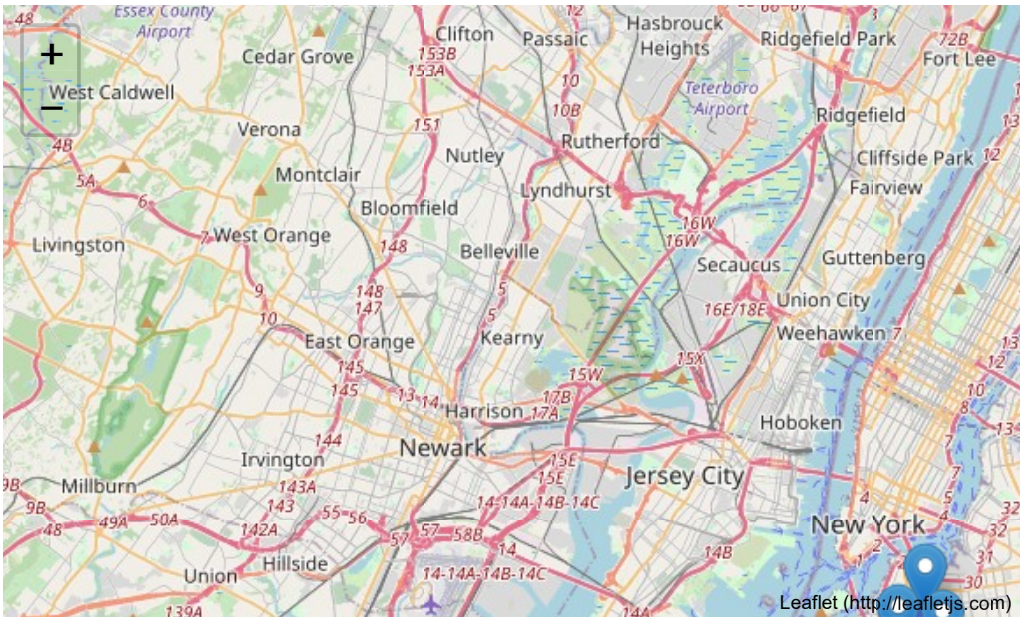
We found 10 Colombian restaurants in NYC. Let's Vizualize it

In [21]:
```
ny_map = folium.Map(location=geo_location('New York'), zoom_start=11)
incidents = folium.map.FeatureGroup()

ny_neighborhood_stats=pd.merge(colombian_rest_ny,new_york_data, on='Neighborhood')
for lat, lng, in ny_neighborhood_stats[['Latitude','Longitude']].values:
    incidents.add_child(
        folium.CircleMarker(
            [lat, lng],
            radius=10, # define how big you want the circle markers to be
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )

# add pop-up text to each marker on the map
for lat, lng in ny_neighborhood_stats[['Latitude','Longitude']].values:
    folium.Marker([lat, lng]).add_to(ny_map)
# add incidents to map
ny_map.add_child(incidents)
```

Out[21]:



Conclusion

There are clearly 2 clusters that seem to be relevant, Brooklyn & Queens. Further inspection on the size of these reveals that Queens is the borough where most Colombian people gather according to the data provided by FourSquare. Therefore, the new Colombian consulate should be located here.