

▼ Introduction/Business Problem

One of the social inequities that has come to light during the COVID-19 pandemic is the shortage of affordable daycare options and the stress that childcare puts on working families.

In this project, I use Foursquare location data to identify neighborhoods in Manhattan that are underserved with Daycares. One of the assumptions I have made here is that Manhattan is densely populated across the entire island, so that all neighborhoods need daycares—maybe not an equal number of daycares in each neighborhood, but still several to choose from.

Data

I will be using the NYU neighborhood location data set found at

https://geo.nyu.edu/catalog/nyu_2451_34572

to divide Manhattan into neighborhoods. I will use Foursquare's venue categories, found here

<https://developer.foursquare.com/docs/build-with-foursquare/categories/>

to count Daycares.

I will then merge the two datasets to find neighborhoods that have many daycares, and neighborhoods that have few.

▼ Step 0: 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

print('Libraries imported.')
```

Libraries imported.

▼ Step 1: Download and Explore Dataset

NYU's Furman Center maintains a free database of 306 NYC neighborhoods and the coordinates of the neighborhoods' centers. You can find the database in geojson format here:

<https://geo.nyu.edu/download/file/nyu-2451-34572-geojson.json>

```
import urllib.request, json
with urllib.request.urlopen("https://geo.nyu.edu/download/file/nyu-2451-34572-geojson.json") as url:
    newyork_data = json.loads(url.read().decode())
    print(newyork_data)
```

```
{'type': 'FeatureCollection', 'totalFeatures': 306, 'features': [{'type': 'Feature', 'id': 'nyu_2451_34572.1
```

▼ Step 2: Tranform the data into a pandas dataframe

```
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)
```

▼ Step 3: Examine the Dataframe

```
neighborhoods.head()
```

	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

▼ Step 4: Use Folium to create a map of New York with neighborhoods superimposed on top.

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

	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

Check the geolocator functionality.

```
address = 'New York City, NY'

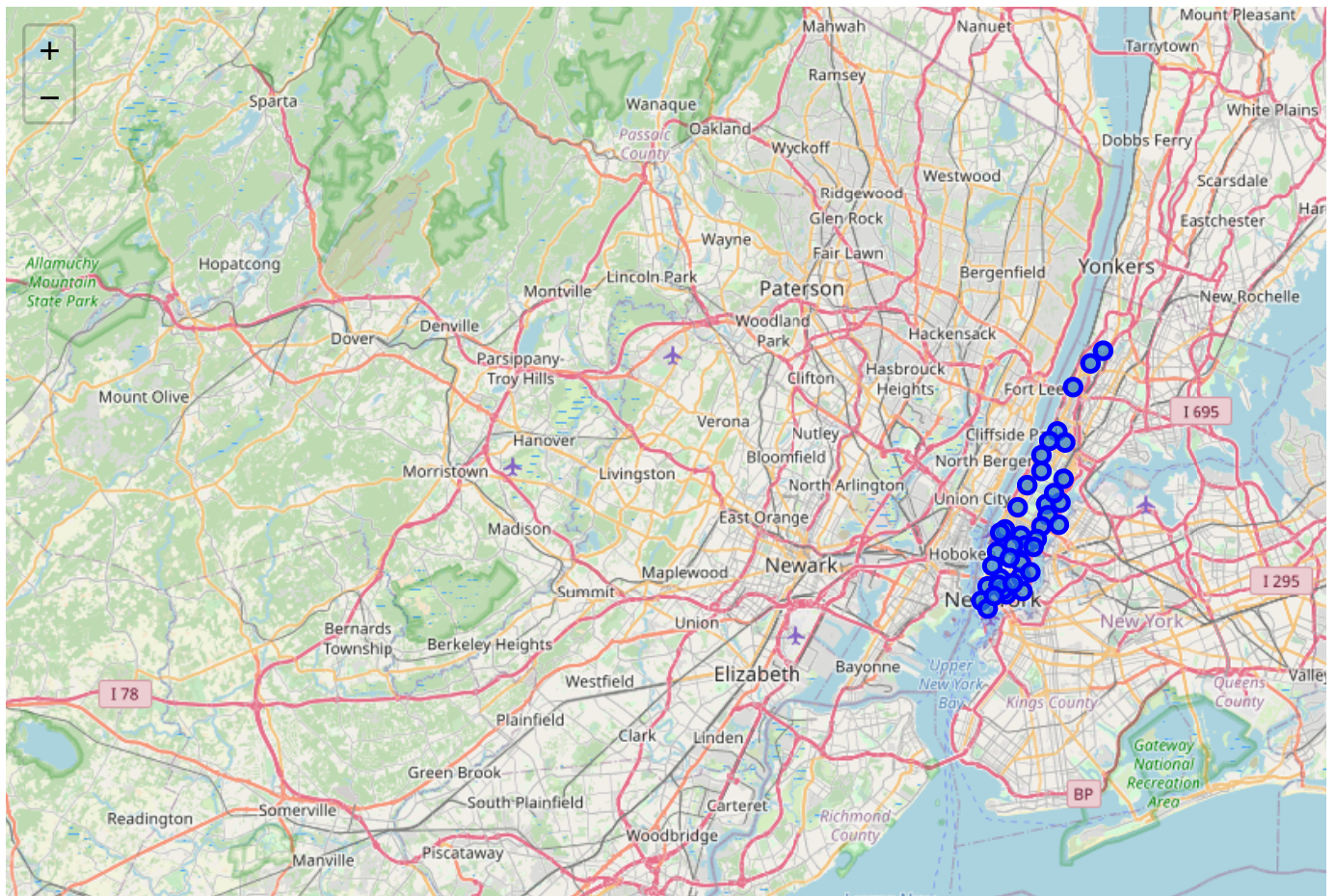
geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}'.format(latitude, longitude))
```

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

```
import folium
# 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)
    popup = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=popup,
        color='blue',
        fill=True,
        fill_color='blue',
        fill_opacity=0.7,
        parse_html=False).add_to(map_newyork)

map_newyork
```



▼ Step 5: Find Manhattan Daycares using Foursquare

You'll need to set up a developer account on Foursquare to access the API and venue data. Getting your CLIENT_ID AND CLIENT_SECRET are straightforward. Follow the directions closely to get the ACCESS_TOKEN...

```
CLIENT_ID = '4I1D0A5U1RA0JTYKCWWJMTAVCB0JK0PWFULYSI2XYN1INUGN' # your Foursquare ID
CLIENT_SECRET = 'B0XNB JV00UJ20BHUUW402LMR52KXQ4QULXERMVKA30N5WK' # your Foursquare Secret
ACCESS_TOKEN = 'GK4JPTTRTIYQ2DIV0BMCK1XG1LQ0I2AVQBPYUTNEXFD3BG5U' # your FourSquare Access Token
VERSION = '20180604'
LIMIT = 30
print('Your credentials:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

```
Your credentials:
CLIENT_ID: 4I1D0A5U1RA0JTYKCWWJMTAVCB0JK0PWFULYSI2XYN1INUGN
CLIENT_SECRET: B0XNB JV00UJ20BHUUW402LMR52KXQ4QULXERMVKA30N5WK
```

Once I had credentials, you can begin querying the Foursquare database. You can query using GET requests through the URL.

I defined a function getNearbyVenues to build the query string from my credentials and my query parameters (in this case, categoryId). You can find the IDs of a huge list of Foursquare venue types here:

<https://developer.foursquare.com/docs/resources/categories>

```

import urllib
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

            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)

```

Let's execute our query building function getNearbyVenues with 'Daycare' as our categoryID parameter.

```

#https://developer.foursquare.com/docs/resources/categories
#Daycare = 4f4532974b9074f6e4fb0104
neighborhoods = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)

```

```
newyork_venues_daycare = getNearbyVenues(names=neighborhoods['Neighborhood'], latitudes=neighborhoods['Latitude'],
newyork_venues_daycare.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude
0	Marble Hill	40.876551	-73.910660	The Learning Experience Riverdale	40.882727
1	Marble Hill	40.876551	-73.910660	Bright Horizons at Riverdale	40.885396
2	Chinatown	40.715618	-73.994279	MetroKids - Southend	40.712081
3	Chinatown	40.715618	-73.994279	First Steps Academy	40.720756
4	Chinatown	40.715618	-73.994279	Chung Pak Day Care Center	40.717325

▼ Step 6: Plot the Daycare Data

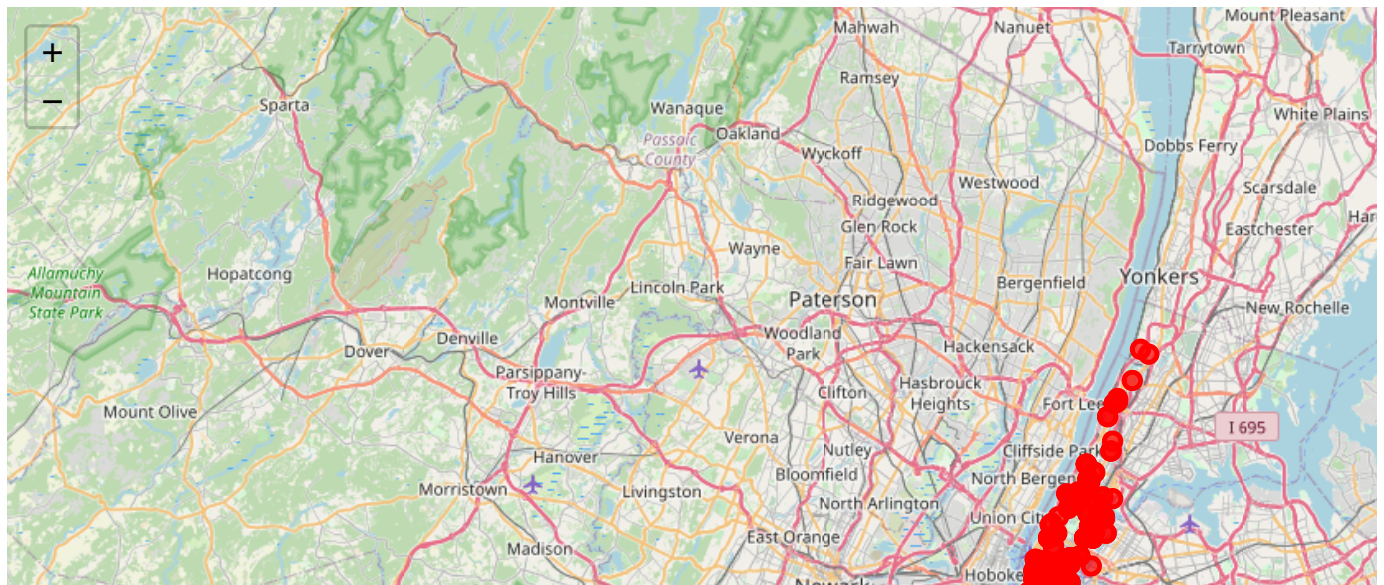
Now let's define a function, `addToMap`, to plot the daycares on a map of Manhattan.

```
def addToMap(df, color, existingMap):
    for lat, lng, local, venue, venueCat in zip(df['Venue Latitude'], df['Venue Longitude'], df['Neighborhood'], df['Venue Category'], df['Venue Name']):
        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)
```

Plotting the daycare locations yields this map:

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

map_newyork_daycare
```

We can see that far more daycare options exist in the southern and middle parts of Manhattan, with far fewer daycare options north of Central Park and into the Bronx. We can do a count of the daycares by neighborhood as follows:

```
manhattan_grouped = newyork_venues_daycare.groupby('Neighborhood').size().nlargest(50)
manhattan_grouped
```

Neighborhood	
Civic Center	19
Little Italy	15
Flatiron	15
Chelsea	13
Midtown South	13
Gramercy	13
Soho	12
Financial District	12
Battery Park City	10
Upper West Side	10
Tribeca	10
Murray Hill	10
Manhattan Valley	10
Yorkville	10
Chinatown	9
Greenwich Village	9
West Village	8
Tudor City	8
Upper East Side	7
Carnegie Hill	7
Clinton	6
Noho	6
Morningside Heights	6
East Village	6
Midtown	6
Washington Heights	6
Lincoln Square	6
Lenox Hill	6
Turtle Bay	6
Roosevelt Island	5
Lower East Side	5
Stuyvesant Town	5
Hudson Yards	4
Sutton Place	4
Central Harlem	3
East Harlem	3
Marble Hill	2
Hamilton Heights	2

```
Manhattanville    1
Inwood            1
dtype: int64
```

Results

There are a wealth of daycare options in southern Manhattan, which peter out in the north (Harlem and the Bronx). The best-served neighborhoods are

Neighborhood # of Daycares Civic Center 19 Little Italy 15 Flatiron 15 Chelsea 13 Midtown South 13 Gramercy 13 Soho 12 Financial District 12 Battery Park City 10 Upper West Side 10 Tribeca 10 Murray Hill 10 Manhattan Valley 10 Yorkville 10

And the underserved neighborhoods are

Neighborhood # of Daycares Roosevelt Island 5 Lower East Side 5 Stuyvesant Town 5 Hudson Yards 4 Sutton Place 4 Central Harlem 3 East Harlem 3 Marble Hill 2 Hamilton Heights 2 Manhattanville 1 Inwood 1

Discussion

There is a big disparity in daycare availability across the island of Manhattan, with many more daycares in the southern part of the island than in the north. In particular, the Bronx and Harlem are underserved, and the Upper West Side and Midtown are well-served.

Conclusion

I feel comfortable making this basic observation about daycare inequity because Manhattan is very densely populated across its entire area; even if fewer people live and work in Inwood than in the Financial district, there is no way that there exists 19 times more need for daycare service in the Financial District.

Generally speaking, daycares are more available in the neighborhoods where wealthier people live, and more white-collar work are happening: midtown, downtown, and the Upper West Side. Further study could reveal how strongly daycare availability correlates with an area's income.