RECOMMENDER SYSTEM FOR HEALTH CARE

1. Introduction

1.1 Background

New York (NY or N.Y.) is a state located in the Northeastern United States. New York was one of the original Thirteen Colonies that formed the United States. With an estimated 20 million residents in 2019, it is the fourth most populous state.

The world now is in a war with the invisible enemy (COVID-19). During the 2019-20 pandemic, the first case of COVID-19 in the U.S. state of New York was confirmed on March 1, 2020. As of April, 2020, there are more than 240,000 confirmed cases (an increase of 6,000 from the day before) in the state, and of those 17,000 people have died (up 540 overnight). New York has the highest number of confirmed cases of any state in the United States, with three times as many cases as neighboring New Jersey (the state with the second-most confirmed cases) and eight times that of neighboring Pennsylvania (which has the fifth-most).

1.2 Problem

The medical centers in every part of NY is overcrowded because of this pandemic. In such situation, a regular medical checkup for pregnant women or elderly people would be a problem. Finding the right place, to get treated in the best possible way with less risk is what every person expects. So, this can be possible through finding the nearest hospitals or clinics. So there are few questions that must be addressed, such as:

1. What kind of medical centers are available (such as Hospital, Maternity Clinic, Eye Doctor, Dentist etc.) in a particular neighborhood?
2. What are the most frequently visited/popular medical centers in the neighborhood?
3. Where are these medical centers located?

To answer the above questions, a map must be displayed based on a person's location and find the available medical centers in their neighborhood. Expectations from this system is to get answer for the above questions.

Another aspect of this project is for a health care professional who is looking to open his/her own practice center. Our goal is to find a location that is having fewer health care centers that are similar to what the professional is about to open. This project will help him/her as well as the people in the neighborhood. For this, we consider the following:

1. Density of population in the neighborhoods of interest
2. Number of specific kind of health care centers (example- Hospital) in those neighborhoods

1.3 Interest

The two groups of target audience who will be interested in our project are:

- People who are trying to find a medical center for a general or critical examination/treatment.

- Health care professionals who are looking for an area to open a new medical center

2. Data acquisition and cleaning

2.1 Data sources

We will use the neighborhood information of New York City that is available from the server. From this we get a json file consisting of Boroughs and Neighborhoods of New York city. It has a total of 5 boroughs and 306 neighborhoods as well as latitude and longitude information for each neighborhood. We use 'wget' command to access newyork_data.json from https://cocl.us/new_york_dataset .

We will get Manhattan population data from https://data.cityofnewyork.us/City-Government/Manhattan-populations-by-neighborhood/8m6s-esnp and scrape this website to get population density of each neighborhood in Manhattan. More population density with few hospitals might help a health care professional to open a new medical center.

2.2 Data cleaning and feature selection

After getting the neighborhood data, we will use Foursquare API to explore neighborhoods in New York City. We now search of all the nearby venues for each neighborhood and filter the data for different medical centers such as Hospital, Emergency Room, Maternity Clinic, Urgent Care Center and Dentist's Office.

Here is the example data containing information about some medical centers (such as Dentist's office, Hospital, Physical Therapist) in some neighborhoods of Manhattan:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Address	Venue Category
3	Marble Hill	40.876551	-73.910660	Marble Hill Dental Office	40.875783	-73.908366	112 W 228th St	Dentist's Office
41	Chinatown	40.715618	-73.994279	Chinatown Medical & Dental Group	40.716229	-73.992755	58 Eldridge St	Dentist's Office
77	Washington Heights	40.851903	-73.936900	Washington Heights Physical Therapy	40.850043	-73.938506	452 Fort Washington Ave Apt 7	Physical Therapist
79	Washington Heights	40.851903	-73.936900	Washington Heights Dental Clinic	40.849570	-73.939161	427 Fort Washington Ave	Dentist's Office
105	Washington Heights	40.851903	-73.936900	New Heights Dental Office, P.C	40.849755	-73.938761	436 Fort Washington Ave Apt 1B	Dentist's Office

3. Methodology and exploratory data analysis

In this project, we will direct our efforts on detecting areas of Manhattan that have low hospital density.

In first step we have to collect the required **data: location and type (category) of every hospital in Manhattan** and cleanse the dataframe to make it more clear to understand and visualize the data.

Second step in our analysis will be calculating and exploring **'Hospitals'** across different areas of Manhattan - we will use **geopy** to identify nearest hospital to a neighborhood.

In third step, we will visualize the nearest hospitals to a neighborhood using **folium maps**. We will also focus on the areas where there are very fewer or no hospitals so that a health care professional can open his/his own practice center. We will present map of all such locations.

3.1 Calculating distance from neighborhood to the nearest hospital

Distance between neighborhood and hospital is not a feature in the dataset, and has to be calculated. Let us use Python geopy library to calculate the distance.

Firstly, let us add a feature distance from neighborhood to nearest hospital column in the dataframe. This column is completely empty as shown in the below figure.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Address	Venue Category	Distance from Neig to Venue
3	Marble Hill	40.876551	-73.910660	Marble Hill Dental Office	40.875783	-73.908366	112 W 228th St	Dentist's Office	
41	Chinatown	40.715618	-73.994279	Chinatown Medical & Dental Group	40.716229	-73.992755	58 Eldridge St	Dentist's Office	
78	Washington Heights	40.851903	-73.936900	Washington Heights Physical Therapy	40.850043	-73.938506	452 Fort Washington Ave Apt 7	Physical Therapist	
80	Washington Heights	40.851903	-73.936900	Washington Heights Dental Clinic	40.849570	-73.939161	427 Fort Washington Ave	Dentist's Office	
102	Washington Heights	40.851903	-73.936900	New Heights Dental Office, P.C	40.849755	-73.938761	436 Fort Washington Ave Apt 1B	Dentist's Office	

Later we calculate the distance using geopy library. Then we find the least distance. And this least distance i.e. the nearest hospital from that neighborhood is stored in the dataframe. Then we also update the dataframe by taking the name of the hospital, venue address, latitude and longitude of only nearest hospital.

	Dist to nearest Hospital	Name of hospital	Address	Lat_Long	Latitude	Longitude
Marble Hill	2.219362	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	(40.849922, -73.93441860000001)	40.8499	-73.9344
Chinatown	0.430749	CityMD Lower East Side	138 Delancey St.	(40.7184065, -73.9869387)	40.7184	-73.9869
Washington Heights	0.188641	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	(40.849922, -73.93441860000001)	40.8499	-73.9344
Inwood	1.407498	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	(40.849922, -73.93441860000001)	40.8499	-73.9344
Hamilton Heights	0.217358	CityMD W 146th St.	3556 Broadway	(40.826753000000004, -73.949572)	40.8268	-73.9496

Here Lat_Long column has been dropped as we got Latitude and Longitude in separate columns. We also need to reset the index.

	Neighborhood	Dist to nearest Hospital	Name of hospital	Address	Latitude	Longitude
0	Marble Hill	2.219362	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	40.8499	-73.9344
1	Chinatown	0.430749	CityMD Lower East Side	138 Delancey St.	40.7184	-73.9869
2	Washington Heights	0.188641	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	40.8499	-73.9344
3	Inwood	1.407498	CityMD West 181st Urgent Care - NYC	617-625 W 181st St	40.8499	-73.9344
4	Hamilton Heights	0.217358	CityMD W 146th St.	3556 Broadway	40.8268	-73.9496
5	Manhattanville	0.791695	CityMD W 146th St.	3556 Broadway	40.8268	-73.9496
6	Central Harlem	0.177776	Harlem Hospital Center	506 Malcolm X Blvd	40.8144	-73.9405
7	East Harlem	1.092570	CityMD E 86th St	336 East 86th Street	40.7771	-73.9504

As seen above, some of the hospitals are repeated as they must be the nearest hospitals to those neighborhoods. To represent this data in a more meaningful way and for our convenience, we shall group the data.

Name of hospital	Address	Latitude	Longitude	Neighborhood	Dist to nearest Hospital
CityMD E 86th St	336 East 86th Street	40.777145	-73.950431	Carnegie Hill	0.409843
				East Harlem	1.092570
				Upper East Side	0.538698
				Yorkville	0.192963
CityMD Financial District Urgent Care - NYC	24 Broad Street	40.706494	-74.011266	Financial District	0.052776
CityMD Lenox Hill	1150 3rd Avenue	40.766831	-73.963067	Lenox Hill	0.237784
				Roosevelt Island	0.797290

3.2 Showing the health care professionals where a medical center can be established

We have to find the neighborhoods which have fewer or no hospitals. This can help the health care professionals to setup their centers in those locations / neighborhoods. Get the population data of each neighborhood in Manhattan, convert it to a dataframe.

	Borough	Year	FIPS County Code	NTA Code	NTA Name	Population
0	Manhattan	2010	61	MN01	Marble Hill-Inwood	46746
1	Manhattan	2010	61	MN03	Central Harlem North-Polo Grounds	75282
2	Manhattan	2010	61	MN04	Hamilton Heights	48520
3	Manhattan	2010	61	MN06	Manhattanville	22950
4	Manhattan	2010	61	MN09	Morningside Heights	55929

The above data is useful to plot a map to visualize the population density of each neighborhood in Manhattan.

Manhattan Population
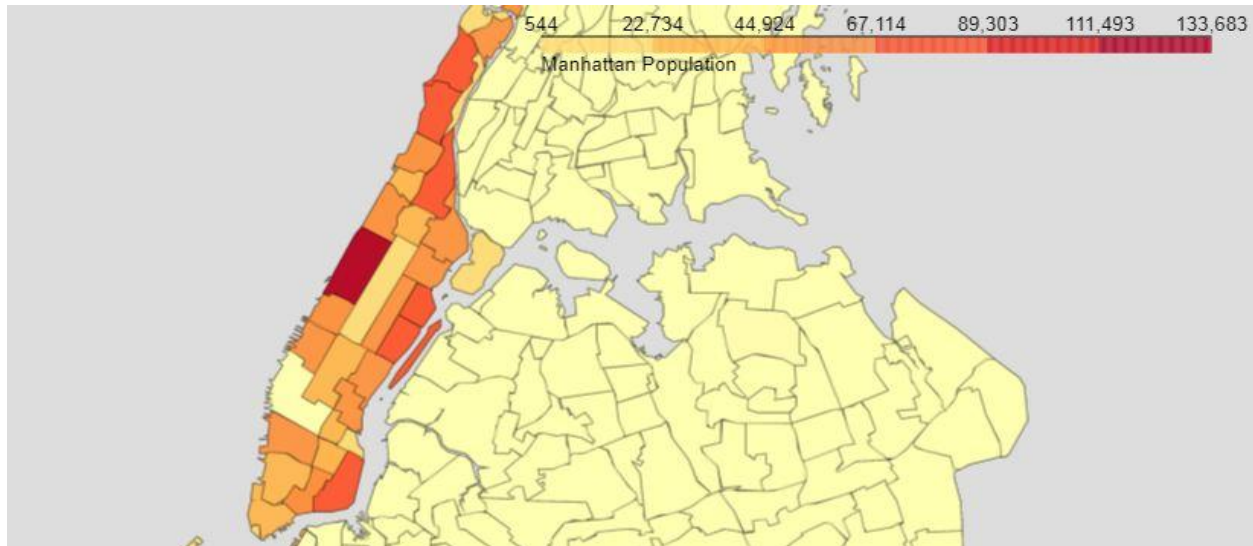
Now let's get the count of number of hospitals in every neighborhood in Manhattan to understand clearly and plot a map of number of hospitals present in a neighborhood, help health care professionals to establish a new medical center.

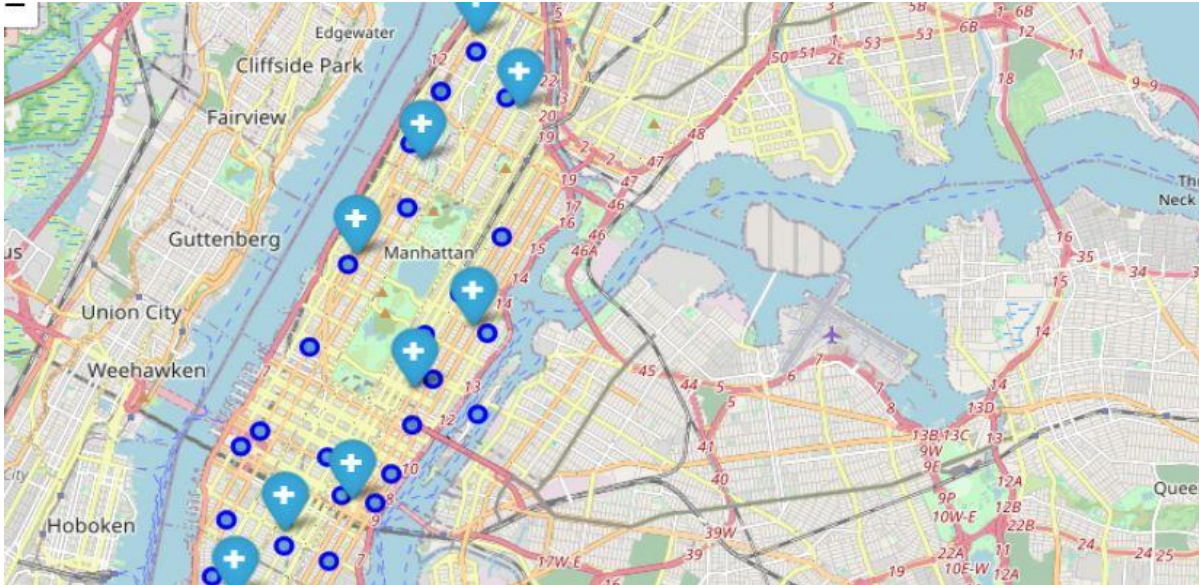| Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Address	Venue Category	Distance from Neig to Venue
Central Harlem	40.815976	-73.943211	4	4	4	4	4	4
Financial District	40.707107	-74.010665	2	2	2	2	2	2
Flatiron	40.739673	-73.990947	1	1	1	1	1	1
Greenwich Village	40.726933	-73.999914	1	1	1	1	1	1
Hamilton Heights	40.823604	-73.949688	1	1	1	1	1	1
Lenox Hill	40.768113	-73.958860	1	1	1	1	1	1
Lower East Side	40.717807	-73.980890	1	1	1	1	1	1
Morningside Heights	40.808000	-73.963896	1	1	1	1	1	1
Murray Hill	40.748303	-73.978332	1	1	1	1	1	1
Tudor City	40.746917	-73.971219	1	1	1	1	1	1

4. Results

We have two types of Target audience:

- ➢ Health care professionals who are about to start new medical centers.
- ➢ People who are in need of critical health care as well as regular health examination.

Let's visualize the data to find whether the above criteria are satisfied or not.

People who are in need of critical health care as well as regular health examination at this time go to the safest and nearest hospital possible. So to visualize the data with nearest hospitals from the neighborhoods a folium map is plotted as shown.

The above map is useful for first set of target audience
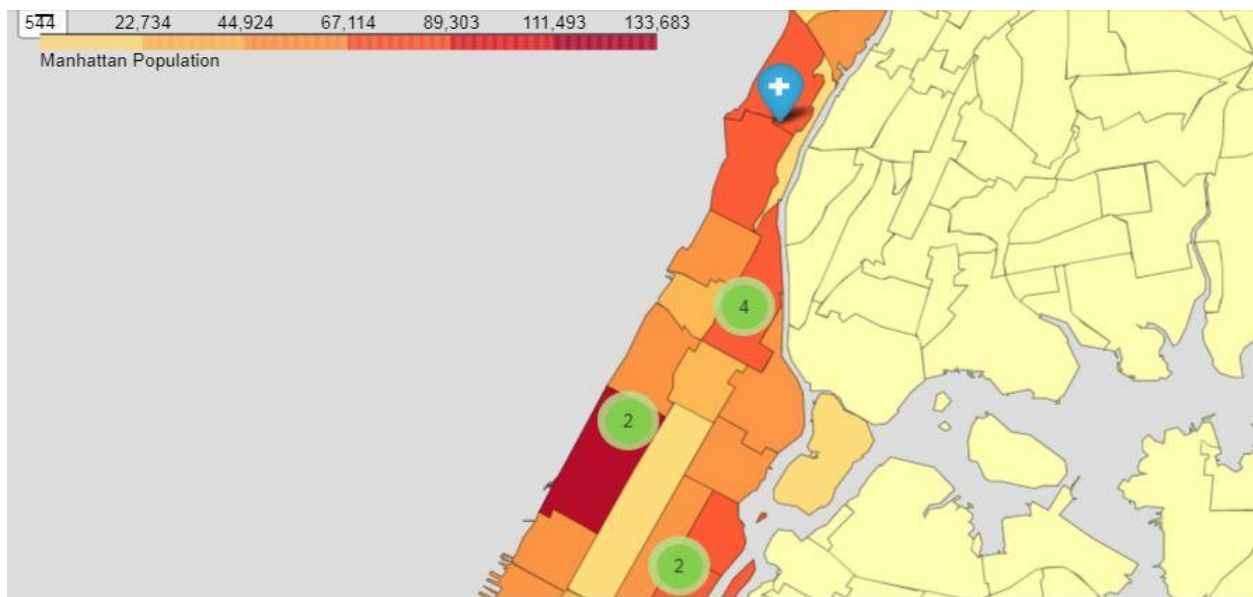
In the above figure, it is clearly seen that neighborhoods which are overcrowded have less number of hospitals. This might help medical professionals to establish a medical center.
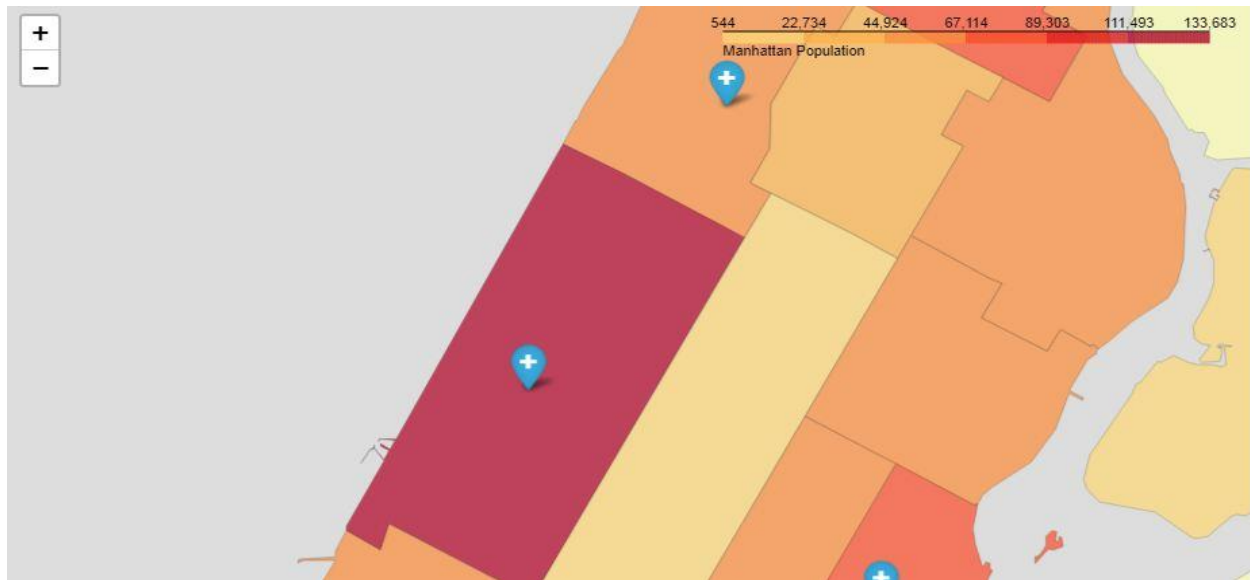
On zooming the above map, it is clear that crowded neighborhood has only one hospital, as shown above. **This might help our second set of target audience.**

5. Conclusions

In this study, we identified medical centers in Manhattan using Foursquare and plotted the nearest hospitals to a neighborhood using folium map to help a person to find the nearest hospital around him, along with the hospital address. We also found the population density of each neighborhood and plotted number of hospitals along with population density on a map, to establish a new hospital where there are few, by a health care professional. For example - Upper West Side has highest population density, but has only one hospital. A person may not be able to get treated if the hospital is crowded and they may be in search of another nearest hospital. So, a health care professional may want to establish a medical center in such densely populated areas.

6. Future directions

This project is mainly focused on nearest hospitals for a neighborhood and on places where new hospital can be opened. The study can be expanded to other types of medical centers such as Emergency Rooms, Urgent Care Centers, etc. Also, the facilities in a hospital might also contribute to a patient's visit to hospital. For example - if a hospital has less medical equipment, this may affect the quality of the hospital and person might not want to visit that hospital. This kind of data is obviously more difficult to extract and quantify. More data, especially data related to the number of medical staff, quality and quantity of medical equipment, types of available treatments/tests, would help to improve this project significantly.