5-2021

Equitable Distribution of Covid-19 Vaccines: Can Data Visualization and Optimization Help?

Zhennan Wang

Equitable Distribution of Covid-19 Vaccines: Can Data Visualization and Optimization Help?

Zhennan Wang

Business Analytics

Advisor: Dr. Pritha Dutta

Pace University

Lubin School of Business

May 2021

Abstract

Coronavirus disease (Covid-19) is known as an infectious disease that has a bigger impact on urban areas with high-density populations. In New York City, however, there are huge differences in the impacts of Covid-19 at the neighborhood level. In addition, there is a negative correlation between the Covid Case Rate and the Vaccination Rate, and areas that are affected the most have poor socio-economic conditions. To better address the equitable distribution of Covid-19 vaccines, I collected social and economic data from Community Health Profiles as well as Covid-19 data from NYC Health. Using this data I created visualizations in Tableau to identify the key factors that contribute to the disproportionate impacts of Covid-19 on different neighborhoods. From the key factors, I developed a Binary Integer Programming model in Excel to determine the optimal locations for Community-based Pop-up Vaccination Centers while minimizing costs. The results support that these vaccine sites can serve the underrepresented neighborhoods.

Table of Contents

Introduction

Business Analytics (BA) is the practice of iterative, methodical exploration of datasets with emphasis on statistical analysis (Nerker (2016)). It is the process by which one interprets data, draws conclusions, and makes decisions. The focus of analyzing data is the exploration and investigation of historical data to gain insights and make future plans and decisions. Since the selective extraction of data is the key to analytics, structured insights can be categorized as descriptive, predictive, and prescriptive (Banerjee & Banerjee, 2017). Descriptive analytics answers the “What has happened” question; it examines the data and draws conclusions. The routine of descriptive analytics is the interpretation of the numbers and the connection of the insights to the underlying problem, often it is presented through graphs and charts. It signals to us whether something is right or wrong but doesn’t explain why. It is the most simple way of analytics and is used most frequently. Predictive analytics answers the “What could happen in the future” question. Usually, one observes the trends in the data and forecasts the future outcome based on the patterns found. Prescriptive analytics answers the “What would be the best outcome based on modeling” question. It automates complex decisions and trade-offs to provide the best results based on changing events. While analytics is used extensively in businesses for decision-making, it is a powerful tool for not only profit-making businesses but can be used by government agencies to obtain valuable insights that can inform policy making.

Since the Covid-19 vaccines became available to the public in late 2020, the distribution of the vaccines has been mainly focused on the priority groups starting from elderly people age 65 and above. According to Harris, E. Jeffrey’s study (2021), there is a negative relationship between Covid Case Rate and Vaccination Rate for all the neighborhoods in New York City. Areas that have a high Covid Case Rate and a low Vaccination Rate tend to have a lower income level, a lower proportion of elderly people, and a higher population Black and Hispanic race. In this thesis, I use data visualization techniques to identify neighborhoods in New York City that are at greater risk from Covid-19 and those that have been disproportionately affected by the pandemic. Subsequently, I developed and ran an optimization model to determine the optimal location of Community-based Pop-up Vaccination Centers that would ensure equitable distribution of the vaccines while minimizing the total cost of setting up these vaccination centers.

Literature Review

In this section, I present a review of existing literature on the topics that are relevant to the work in this thesis.

Visualization

Visualization is an analytics technique that falls into the categories of descriptive and predictive analytics. By forming charts and graphs, visualization is an easy way to gain preemptive insights on datasets. By using trendlines, charts can be useful for identifying patterns. Visualization is often associated with dashboard designs for presentation; a good design can provide solutions to resolve information overload by using colors and unique shapes. A design layout is also important when clarifying the significance of the data.

According to Yap (2020), a good dashboard design should make the complex simple, reveal the critical details behind the data, and tell a story. Oftentimes, one chooses the end goal or the message that one wants to deliver with the visualization before putting in all the design elements. The focus on the different variables and tests on the relationship of the data should be prioritized. Choosing layouts, colors, and interactive elements should come in later.

Yap also dives into some techniques to improve the visualizations. First, there should be the main representation and a supporting visual. For example, a digital representation can be supported by an analog representation to enhance the visual data perceptions. Both digital and analog have their own way of delivering messages that appeals to different users. Second, different parts of inputs will lead to the various results of the outputs. Determining the most important inputs, or the data variables will help to provide better and more effective outputs. In visualization, there are many tools such as sliders, segmentation, and clustering that all improve the process of input selection. Third, the use of colors for indicators and visual synergy. Colors can attract human eyes more easily and make the data patterns stand out. The use of similar colors as background and unique colors can be a great indicator of data significance or data abnormality. The cool color of green can be related to positivity and the warm color of red will show the opposite. The synergy of using different charts, such as bar charts, pie charts, bubble charts, or even text tables, with the correct colors and indicators will make the visualization more appealing and information more apparent. Lastly, visual analytics should definitely need to fit into the context of the application. The idea of the dashboards is to keep it as simple as possible, that is, including the major variables and ignoring the less useful variables that have minimal outcomes.

Optimization

Optimization can be categorized as prescriptive analytics. Modeling is the core of optimization. By developing models using a dataset, one can extract the best solution to a specific problem. Oftentimes, modeling algorithms are time-consuming and costly, almost impossible to imitate exactly the real-life problems. In these cases, it is possible to obtain an acceptable approximate solution instead of the exact answer.

Traditional and modern optimization methods can be classified into different groups, and one of which is the Exact optimization methods, including Linear Programming (LP), Non-linear programming, Quadratic Programming, etc (Graa and Benhamida, 2020). The goal of using optimization, under the mathematics context, is to either maximize or minimize a function. If an optimization model has only the objective function, it is called the unconstrained model (Pulat, Bayyurt, & KOCAKOÇ, 2020). However, usually, an optimization model has many variables and a set of constrained equations, and this model is a constrained optimization model. The inequality constraints are typically the limitation on the types of resources in the problem, such as raw materials used in the production of finished goods, the working hours of labor, etc. The non-negativity of the intuitive constraints is usually applied to variables that need to be positive (Pulat, Bayyurt, & KOCAKOÇ, 2020). Binary Integer Programming is another useful LP model. The constraints offer to provide two choices for the decision variables. Typically, binary variables deal with the yes-or-no questions (1 or 0 in the result). A standard method of linear programming is simplex. The inequality constraints can form a polygonal region and the optimal solution typically falls onto one of the corners (Pulat, Bayyurt, & KOCAKOÇ, 2020).

Covid-19

Coronavirus disease (Covid-19) was identified in late December 2019 and the first case appears in Wuhan, China. It is an infectious disease that has gotten widespread in many countries around the world since early 2020. It is known that the Covid-19 virus spreads mainly through respiratory droplets such as breathing, coughing, or sneezing. Because of the way the disease transmits, urban areas with a high-density population will result in a bigger impact and more infections (Bewal, Minhas, Prasad, Yadav, Sreedhar, Bhasin & Kumar, 2020). In this work, I focus on New York City for the analysis.

Covid-19 affects different locations to a different extent, depending mostly on the characteristics from the number of travelers, the local policies and controls, the timing of the policies, and general public awareness and prevention of the disease (Almagro, Orange-Hutchinson, 2020). However, within New York City itself, there are large differences in Covid-19 impacts on a neighborhood level. Based on Almagro and Orange-Hutchinson's (2020) work, regions that have the highest coronavirus infection rates are found in the boroughs of Bronx, Brooklyn, and Queens. These regions include many neighborhoods that have a higher percentage of individuals who are identified as Blacks and Hispanics as well as households with a low socioeconomic status. Occupation plays a big role in these high poverty regions because they have a higher chance of getting exposure to human interactions; the crowding of shared spaces is also considered critical instead of the density of a particular location (Almagro, Orange-Hutchinson, 2020).

Policies can be issued to target the more vulnerable communities that have been impacted by Covid-19 the most. For instance, issuing more protective gear to the businesses and the workers that have more contacts with other people; setting up housing options for large households such as free hotels for quarantine and isolation to reduce large gatherings; creating more sites to get testings and vaccinations for neighborhoods that have higher infection rates.

Research Questions and Methodology

The research questions are the guides for identifying the key points that will be uncovered in this research paper. Below are the research questions studied in this thesis.

- How does Covid-19 affect NYC neighborhoods?
- How can data visualization and optimization techniques aid the selection of community-based vaccination sites for equitable distribution of Covid-19 vaccines?

For the data selection, I mainly searched for the social-economic and Covid-related information on the NYC websites. From the NYC Health website, I obtained the Community Health Profiles data from 2018 that includes all the demographics, social-economics, housing, health living, and health care datasets. The dataset is categorized based on five boroughs and fifty-nine community districts. Based on the need of this research, I extracted the demographic data, such as population, race, and age variables; social-economic data, including poverty, rent burden and unemployment, and health care data that have all the premature death rate for several diseases and major cancer rate. Table 1 illustrates all the variables that were selected from the Community Health Profiles data.

Table 1

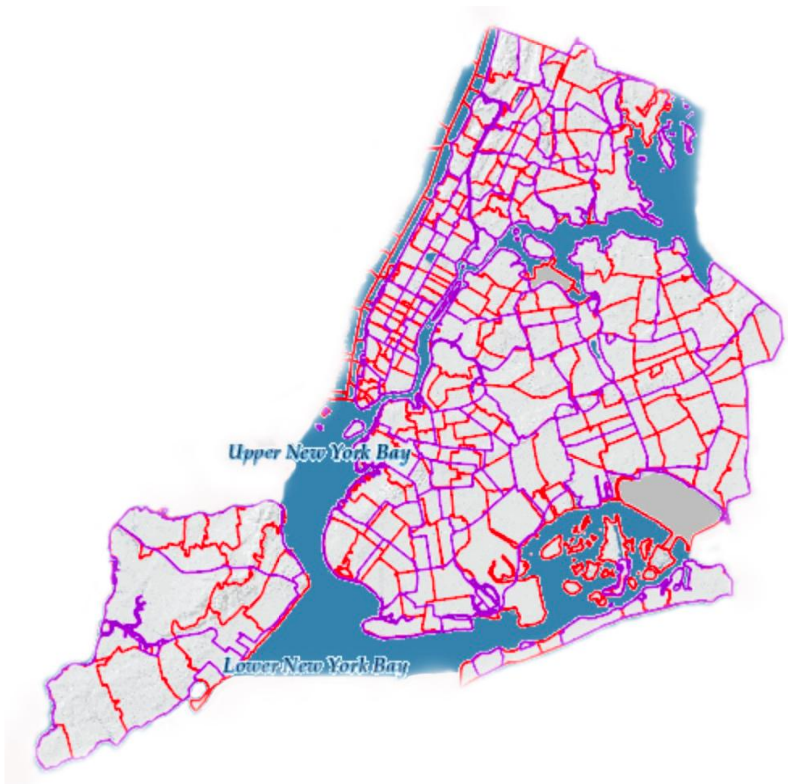
Demographics	Health Outcome
Overall_Pop	Obesity
Race_White	Diabetes
Race_Black	Hypertension
Race_Asian	HIV_Diagnoses
Race_Latino	HepC_Reports
Race_Other	Premature_Mort_Cancer_Number
Age0to17	Premature_Mort_Cancer_Rate
Age18to24	Premature_Mort_HeartDisease_Number
Age25to44	Premature_Mort_HeartDisease_Rate
Age45to64	Premature_Mort_Drug_Related_Number
Age65plus	Premature_Mort_Drug_Related_Rate
	Premature_Mort_Accidents_Number
	Premature_Mort_Accidents_Rate
Social and Economic Conditions	Premature_Mort_Diabetes_Number
Poverty	Premature_Mort_Diabetes_Rate
Unemployment	Premature_Mort_Suicide_Number
Rent_Burden	Premature_Mort_Suicide_Rate
Assault_Hosp	Premature_Mort_HIV_Number

Jail_Incarceration	Premature_Mort_HIV_Rate
Helpful_Neighbor	Premature_Mort_Stroke_Number
	Premature_Mort_Stroke_Rate
	Premature_Mort_Liver_Disease_Number
Healthy Living	Premature_Mort_Liver_Disease_Rate
Preterm_Births	Premature_Mort_Homicide_Number
Teen_Births	Premature_Mort_Homicide_Rate
Physical_Activity	Rank_1_Cancer_Type
Sugary_Drink	Rank_1_Cancer_Number
Fruit_Veg	Rank_1_Cancer_Rate
Smoking	Rank_2_Cancer_Type
Uninsured	Rank_2_Cancer_Number
Unmet_Med_Care	Rank_2_Cancer_Rate
	Rank_3_Cancer_Type
	Rank_3_Cancer_Number
	Rank_3_Cancer_Rate

The Covid related data is obtained from the NYC Health website that contains all the Covid-19 data in NYC categorized by zip code as of April 20th, 2021. The dataset contains Covid Case Count, Covid Case Rate, Covid Death Count, Covid Death Rate, Percent Positive, and Total Covid Tests. For the simplicity of this research, I decided to only use the Covid Case Rate and Covid Death Rate variables to evaluate the impact of Covid-19 Pandemic in NYC areas.

The two datasets are categorized differently, for instance, the Community Health Profiles data has geographic features of the Community District while that of the Covid data has Zip code. Below is a map that represents these two geographic features (OASIS Map).

Figure 1



From Figure 1, the purple lines represent Community District boundaries and the red lines represent Zip code boundaries. There are huge overlaps between these two geographic identifiers.

In order to join these two datasets for analysis, I decided to use the Community Districts as the main geographic feature. I set the community district areas as the baseline and take a proportion of the area on the zip codes that are included in each community district. Then I applied the proportion to the Population Denominator variable in the Covid data to convert the Covid Case Rate and Covid Death Rate categorized based on Community Districts. To clarify the process, I attach a visual representation and an image of the excel calculation shown in Figures 2 and 3.

Figure 2

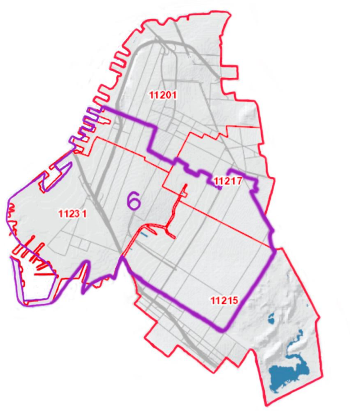


Figure 3

Community District	zipcode	percent	NEIGHBORHOOD_NAME	COVID_CASE_RATE	POP_DENOMINATOR	COVID_DEATH_RATE	New_Pop	Pop_Ratio
BK 6	11201	15	Brooklyn Heights/DUMBO/Downtown Brooklyn	2815.68	62294.06	158.92	9344.109	0.08207413
	11215	70	Gowanus/Park Slope/Windsor Terrace	2608.28	69624.37	79	48737.059	0.42808275
	11217	45	Boerum Hill/Park Slope	3403.41	40870.72	222.65	18391.824	0.16154488
	11231	100	Carroll Gardens/Cobble Hill/Red Hook	3023.28	37376.63	109.69	37376.63	0.32829824
				2889.995127		118.8407602	113849.62	

Covid Case Rate for BK 6

Covid Death Rate for BK 6

Total population for BK 6

Figures 2 and 3 show an example of the process. It is the Community District 6 in Brooklyn, and there are four zip codes that overlap in this region, which are 11201, 11215, 11217, and 11231. I estimated the percent of the area that the zip code falls in the community district and calculated the population ratio. For instance, zip code 11201 has about 15% of the area that falls in Community District 6 in Brooklyn; zip code 11215 has about 70%; zip code 11217 has about 45%; zip code 11231 completely falls inside the district or 100%. The New_Pop variable is calculated by multiplying the percent to the Population Denominator, measuring the population of each zip code that falls in the district. The population ratio can then be multiplied by its corresponding Covid Case rate and Covid Death rate and sum them all to get the final Covid Case Rate and Covid Death Rate of a particular Community District.

Using the final dataset, I explored the impacts of the Covid-19 Pandemic on different neighborhoods using visualization graphs created in Tableau. The Covid Case Rate and Covid Death Rate count in the population density in a particular area and are great identifiers on which areas were hit the hardest. The race and age variables can convey the demographic distribution in each neighborhood. The Poverty Rate and Unemployment Rate data give some clue towards the social and economic conditions in each neighborhood, but the Unemployment Rate was in 2018. Since the Unemployment Rate greatly increased during the Covid-19 pandemic, the data might not be accurate. The Disease Rate and Premature Mortality Rate are connected to the people who have potential risks of mortality once they get infected with Covid-19 disease.

From the insights obtained from these visualizations, I identified some of the major factors that contribute to the need for community-based vaccine locations and built an optimization model to decide the optimal location for Community-based Pop-up Vaccine Centers while taking into consideration the cost of establishing these vaccine sites and their daily vaccination capacities. Although I did not have the data on the cost of setting up a vaccination site and the capacity for each vaccination center, I used simulation to generate cost and capacity values for each center using uniform distribution and available cost and capacity estimates (Senese, 2021 and FEMA). Lastly, the optimal solution obtained from the optimization model is compared to the Covid-19 maps for final conclusions.

Analysis

Visualization

The goal of visualization is to explore the data preliminarily. Since Covid-19 has impacted the area with poor living conditions the most, I targeted the closely related variables and included them for visualization. The premature mortality rate caused by different diseases can convey that people in these areas usually have a high demand for hospitalization and healthcare assistance.

Figures 4 to 8 illustrate the top 10 Sub-Boroughs that have the highest Premature Mortality Rate by Cancer, Heart Disease, Diabetes, HIV, and Liver Disease from the Community Health Profiles data.

Figure 4

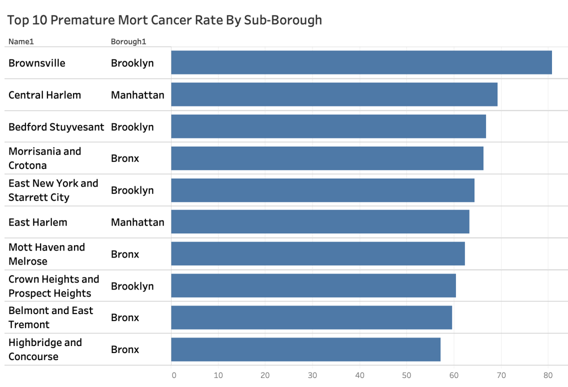


Figure 5

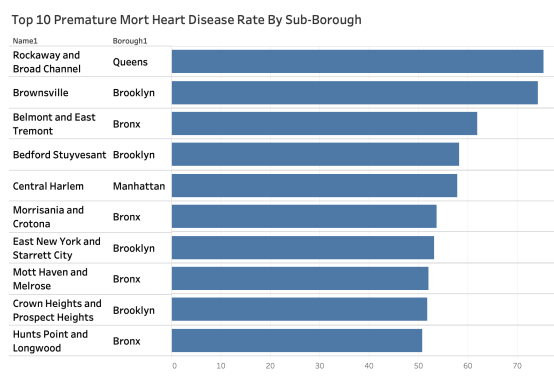


Figure 6

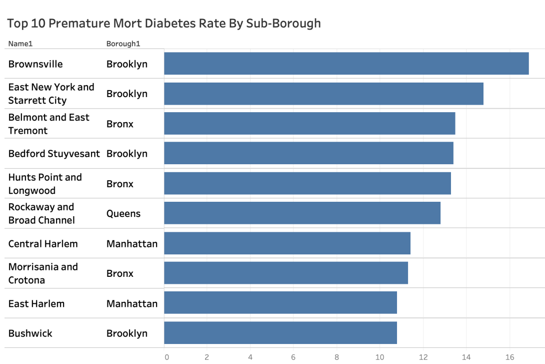


Figure 7

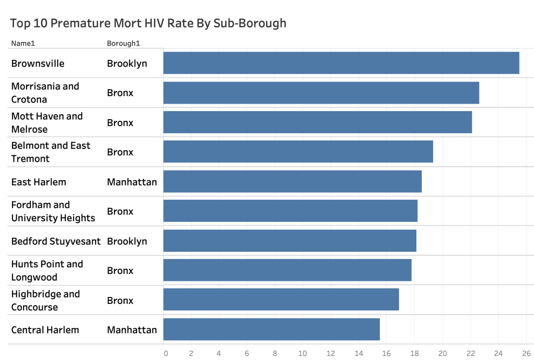


Figure 8

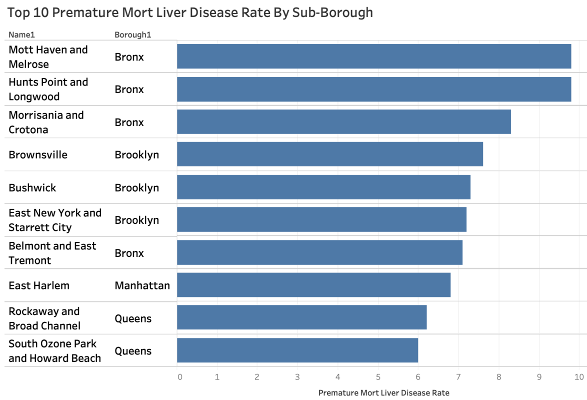


Figure 9

Average Rank of the Sub-Boroughs and the Severity of Different Diseases

Subboroughs	
Brownsville	1.80
Bedford Stuyvesant	4.50
Morrisania and Crotona	4.60
Mott Haven and Melrose	4.75
East New York and Starrett City	5.00
Belmont and East Tremont	5.20
Rockaway and Broad Channel	5.33
Central Harlem	6.00
Fordham and University Heights	6.00
Hunts Point and Longwood	6.25
East Harlem	7.00
Bushwick	7.50
Crown Heights and Prospect Heights	8.50
Highbridge and Concourse	9.50
South Ozone Park and Howard Beach	10.00

Count
1 5

Figure 9 shows summary information from Figures 4 to 8. The numeric value shows the average rank for the sub-borough regarding the top 10 list for the Premature

Death Rate of 5 diseases, and the color represents the number of appearances (the darker the color, the more frequent that sub-borough gets on the top 10 list). As the table suggests, Brownsville has the highest average rank across five Disease Death Rates (ranking 1st, 2nd, 1st, 1st, and 4th, and the average is 1.8) and appears most frequently (a total of 5 times). Sub-boroughs Morrisania and Crotona and Belmont and East Tremont have tied with the number of occurrences with Brownsville, however, their average rank is not as high. These regions have a relatively high probability of needs and visits for hospitalization, so they are the key factors when considering the vaccine locations.

Based on the highest affected sub-boroughs from Figure 9, I further break down the data by adding the age and race variables.

Figure 10

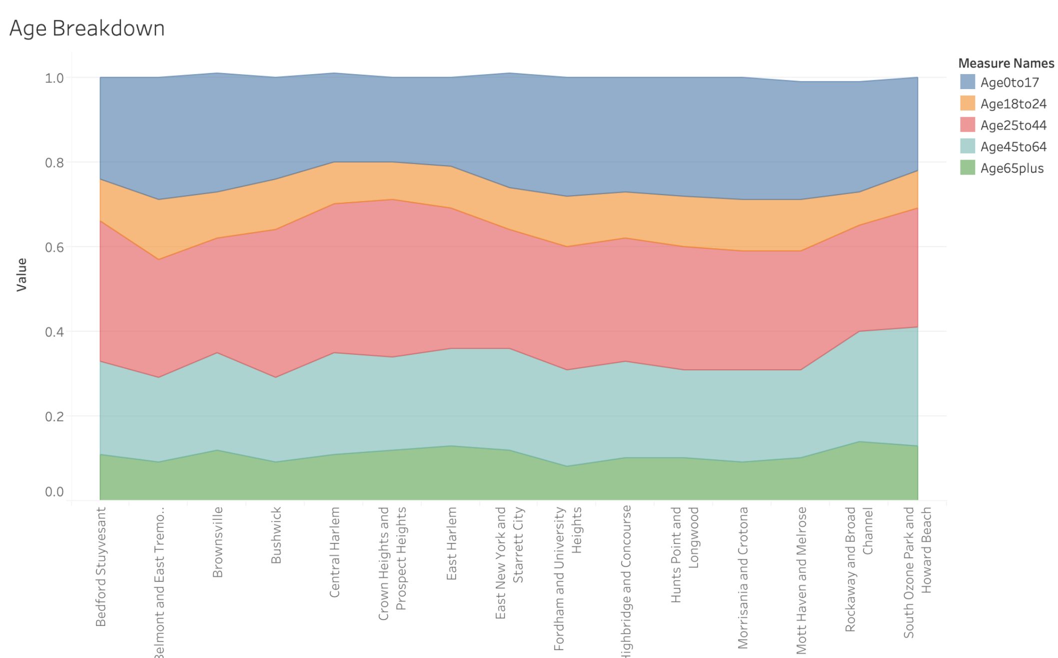
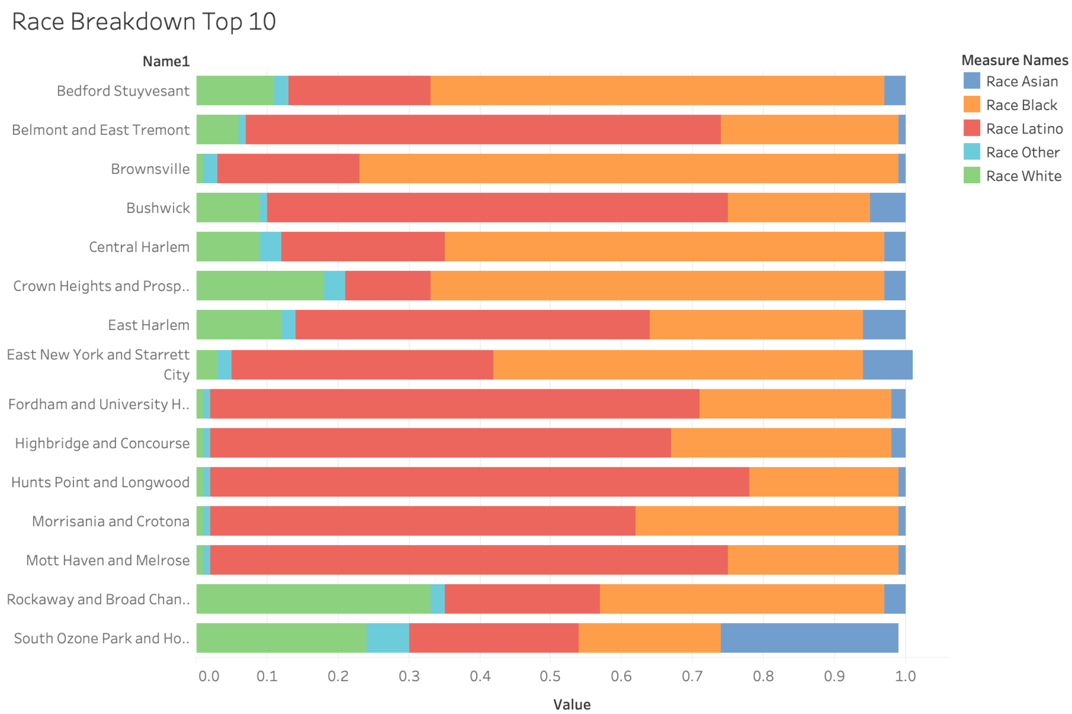


Figure 11



From Figure 10, I do not see a clear relationship between age and the severity of disease other than the population of age 0 to 17 and age 25 to 44 are slightly higher in these regions, indicating that Age is not a significant factor that contributes to the disproportionality effects of Covid-19 in New York City. In Figure 11, I can clearly see the majority of the population are either Black or Latino (Hispanic). All the regions have at least 60% of the population identified as Black or Hispanic except for South Ozone Park, which has about 40%.

I then investigated the sub-boroughs that have the least impacts by the five diseases.

Figure 12

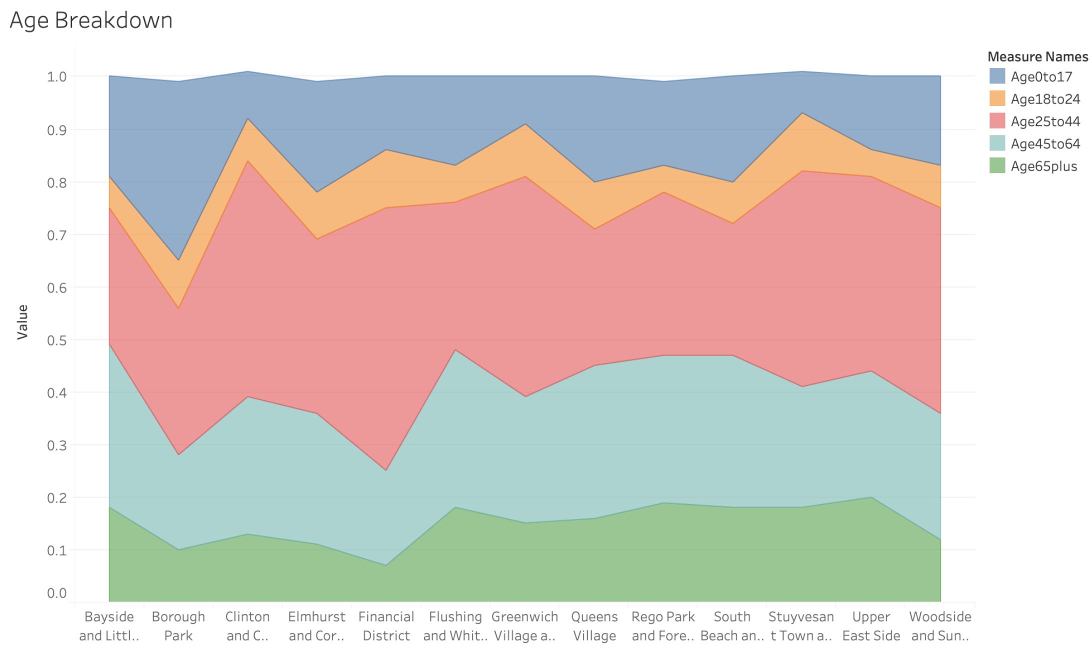
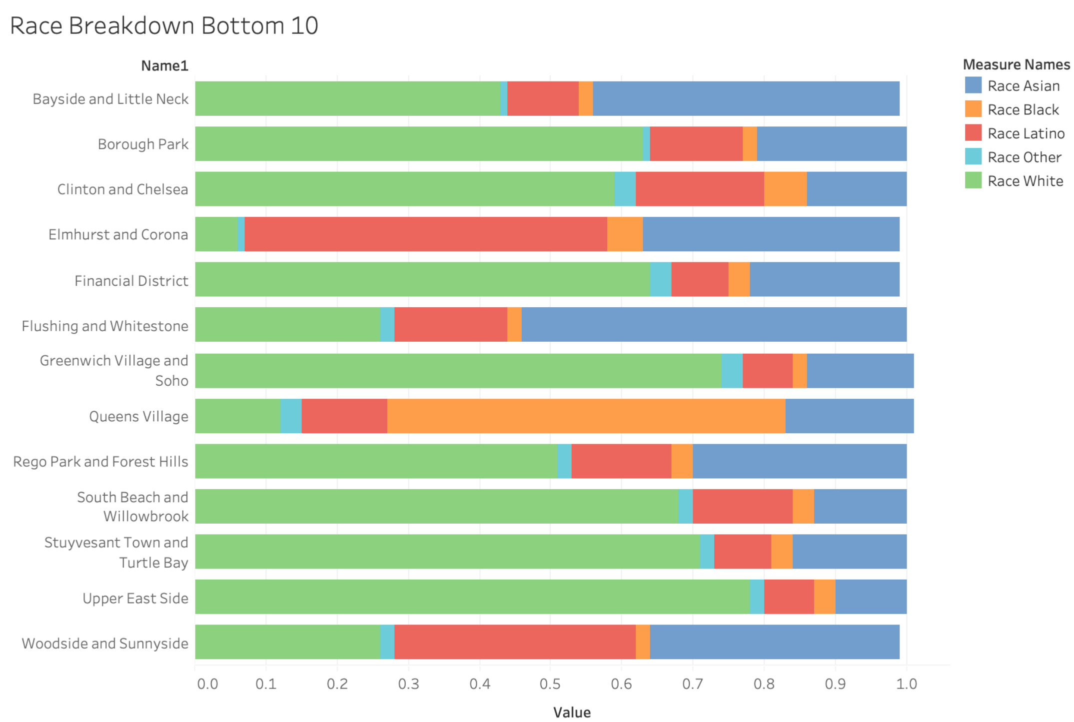


Figure 13



From Figure 12, there is a slightly higher percentage of people who are more than 45 years old, which again, shows that Age is not a significant factor. From Figure 13, these sub-boroughs have the majority of White or Asian races.

From Figures 10 to 13, I can identify the neighborhoods that have the greatest potential risks for Covid because of their high disease rate. These areas contain the majority of Hispanic and Black people, which also explains that Covid-19 hit these places the hardest. One thing to note, the age distribution factor does not depict a clear relationship with both the potential risks and the Covid infectious rate, although there is a higher Covid case rate among younger people and a higher Covid death rate among the elderly.

Next, I moved on to the Covid Case Rate and Covid Death Rate distribution using the spatial file of the Community District map in Tableau, shown in Figures 14 and 15, respectively.

Figure 14

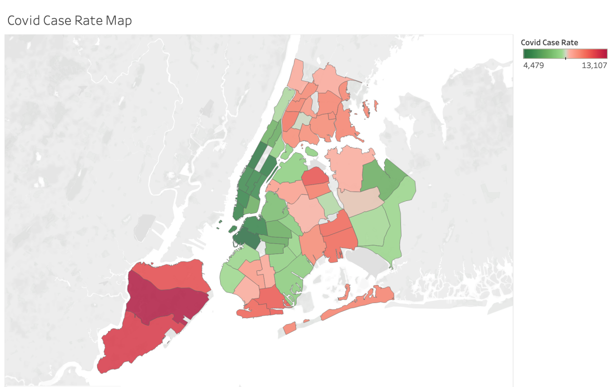
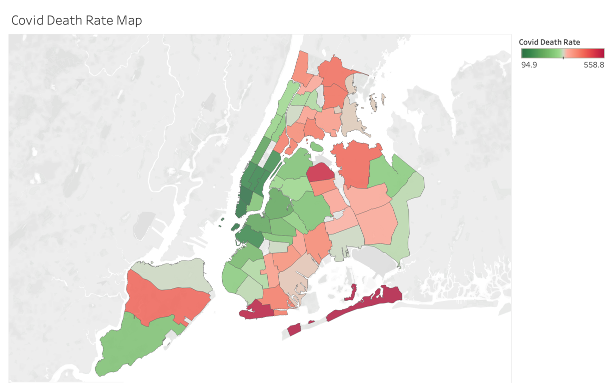


Figure 15



In Figures 14 and 15, the color green indicates a lower Covid Case Rate and Covid Death Rate while the color red indicates a higher Covid Case Rate and Covid Death Rate.

In Figure 14, neighborhoods in Staten Island have a relatively higher infectious rate in New York City while those of Manhattan have the lowest infectious rate. The majority of the Bronx and selected areas in Brooklyn and Queens have a high Covid case rate. As for Figure 15, Staten Island is on the better end, but most of the Bronx and several neighborhoods in Brooklyn and Queens are doing poorly. Specifically, South Beach and Willowbrook in Staten Island; Coney Island of Brooklyn; Rockaway, Jackson Heights, and Flushing of Queens; and East Harlem of Manhattan. Figures 16 and 17 list out the top neighborhoods sorted by highest Covid Case Rate and Covid Death Rate.

Figure 16

Top 10 Neighborhoods for Covid case rate

Neighborhoods	Borough	
South Beach and Willowbrook	Staten Island	13,107
Tottenville and Great Kills	Staten Island	12,059
St. George and Stapleton	Staten Island	11,586
Jackson Heights	Queens	11,394
Sheepshead Bay	Brooklyn	11,256
Coney Island	Brooklyn	11,041
Kew Gardens and Woodhaven	Queens	10,943
South Ozone Park and Howard Beach	Queens	10,927
Highbridge and Concourse	Bronx	10,626
Elmhurst and Corona	Queens	10,481

Figure 17

Top 10 Neighborhoods for Covid death rate

Neighborhoods	Borough	
Rockaway and Broad Channel	Queens	558.8
Coney Island	Brooklyn	547.5
Jackson Heights	Queens	524.7
South Beach and Willowbrook	Staten Island	446.4
Flushing and Whitestone	Queens	442.8
Williamsbridge and Baychester	Bronx	435.5
Morris Park and Bronxdale	Bronx	433.4
Hunts Point and Longwood	Bronx	419.4
Sheepshead Bay	Brooklyn	418.6
East Harlem	Manhattan	415.5

With the uneven distribution of the Covid case rate and Covid death rate, there are more factors that affect the severity of an individual neighborhood besides age and race. I further examine the social and economic variables, which include the Uninsured Rate, the Unemployment Rate, and the Poverty Rates, shown in Figures 18 to 20. The neighborhoods with the highest Uninsured Rate are Jackson Heights and Elmhurst/Corona by a large margin. The Unemployment Rate map and the Poverty Rate map are very similar to the severity of the Covid map: the majority of Bronx have high Unemployment and Poverty Rate as well as neighborhoods such as (Brooklyn) East New York, Brownsville, (Manhattan) East Harlem, (Queens) Jamaica, and Hillcrest. These factors all have some relation to the Covid-19 pandemic and are the features when considering the vaccine locations.

Figure 18

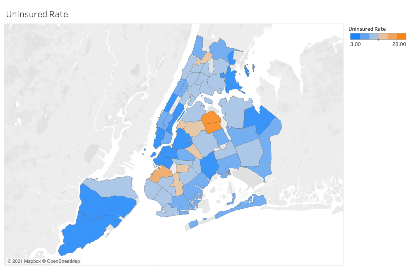


Figure 19

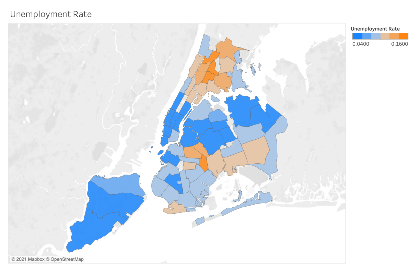
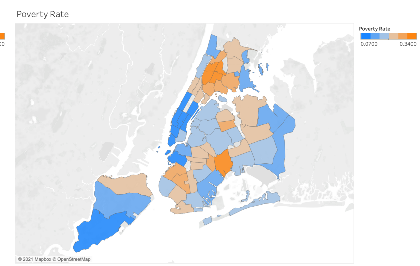


Figure 20



Optimization Model

The optimization model is developed to determine the location of Community-based Pop-up vaccine sites. Based on visualization analysis, I have determined the priorities for selecting the vaccine sites that will help neighborhoods that are affected the most by Covid-19. The factors are the Covid Case Rate, the Covid Death Rate, and the potential Disease Risks. Because there are also social-economic factors that contribute to the disproportionate impact of the pandemic, I also include the Poverty Rate in the optimization model. The neighborhoods with the highest Uninsured Rate and Unemployment Rate have already been explained by Covid Case and Death Rate, so I decided to leave these two variables out from the optimization.

I assume that based on the observations from the visualization section, the city of New York is planning to set up Community-based Pop-up Vaccination Centers to better serve the populations that are at higher risks and have been disproportionately affected by the pandemic. I assume only Community-based Pop-up Vaccine Centers because

States have been using these vaccine sites to reach out to underserved local communities (McMinn, Chatlani, Lopez, Whitehead, Talbot, & Fast, 2021).

Let there be *n* number of possible sites for vaccine centers with *j* denoting a typical site. I assume each neighborhood in New York to be a potential community-based vaccine site. I assume there is *m* number of boroughs in the city with a typical borough denoted by *i*.

Then the decision variable x_{ij} is defined as,

x_{ij} = 1, if site *j* in borough *i* is selected for vaccine center

= 0, if site *j* in borough *i* is not selected for vaccine center

for all j=1,..,n and i =1,..,m.

There is a cost associated with establishing or setting up a vaccine center in each neighborhood of each borough which is denoted by c_{ij}. Each vaccine center also has a capacity for vaccinations per day denoted by p_{ij}.

The objective here is to determine the location of vaccine sites while minimizing the total cost of setting up or establishing the sites. Vaccinating an entire population in a city or state is a cost-intensive process. The cost has been identified by several states as a major challenge in the vaccination program (New Jersey Department of Health, 2020). Hence, it is reasonable to assume that city administration would want to minimize cost while making sure that they can provide equitable distribution of the vaccine.

The optimization problem is formulated as follows:

Objective: Minimize cost = $\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$

Subject to: $\sum_{j=1}^{n} x_{ij} \geq 3,\ for\ all\ i\ =\ 1,..., m$ …………(1)

x_{ij}= 1, for those *j*s that belong to priority sets of D, C, R, and P…….(2)

$\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} x_{ij} \geq 20,000$……..(3)

Constraint (1) makes sure there are at least three vaccination sites in each borough.

Constraint (2) defines that there should be Community-based Pop-up Vaccination Centers in each neighborhood that belong to four priority sets that are defined below.

Constraint (3) defines that the community-based pop-up vaccination sites should be able to vaccinate at least 20,000 people per day. The resulting model is a Binary Integer Programming model.

Now the model is implemented in a case study based on the scenario in New York City. Here m = 5 and n = 59, indicating that there are a total of 5 boroughs and 59 neighborhoods. The priority sets as defined as follows; D is the set of 5 neighborhoods with the highest Death Rate due to Covid-19, C denotes the set of 3 neighborhoods with the highest Case Rate, R denotes the set of 5 neighborhoods with the highest Disease Rate and potential risks, and P denotes the set of 2 neighborhoods with the highest Poverty Rate.

The neighborhoods in each set are listed below:

D: East Harlem, Coney Island, Jackson Heights, Rockaway and Broad Channel, South Beach and Willowbrook

C: Highbridge and Concourse, St. George and Stapleton, Tottenville and Great Kills

R: Mott Haven and Melrose, Morrisania and Crotona, Bedford Stuyvesant, East New York and Starrett City, Brownsville

P: Highbridge and Concourse, Fordham and University Heights

When establishing or setting up a vaccine site in a neighborhood, there is a cost associated with it, including costs for storage and freezers, chairs and tables, Internet hotspots, and flyers and posters (Senese, 2021). However, due to the lack of availability of actual data on the cost of setting up a community-based pop-up vaccination center, it is assumed that cost varies by neighborhood and borough and follows a uniform distribution with an upper bound of $200,000 and a lower bound of $110,000.

In addition, to adjust the problem of capacity in each vaccination center, I based our distribution on FEMA's capacity approximation of community vaccination centers. Community-based Pop-up vaccination sites that I selected are a school, which serves approximately 1,000 vaccinations a day, and a church, which serves approximately 250 vaccinations a day (FEMA). Therefore, I assumed that the capacity of a vaccine site follows a uniform distribution with an upper bound of 1,000 vaccines and a lower bound of 250 vaccines. The capacity constraint of at least 20,000 people vaccinated a day is based on Governor Cuomo's latest announcement of 18 community-based pop-up vaccination sites able to vaccinate 8,500 people (New York State Governor, 2021).

Using the random number generator in Excel I simulated the costs and capacities for each potential site and solved the optimization problem in Excel using Solver.

Result of Optimization Model

The final optimal solution indicates a total of 26 vaccination sites should be set up. The neighborhoods that should have a Community-based Pop-up vaccine site are:

Manhattan: Lower East Side and Chinatown, Stuyvesant Town and Turtle Bay, Morningside Heights and Hamilton Heights, East Harlem.

Bronx: Mott Haven and Melrose, Morrisania and Crotona, Highbridge and Concourse, Fordham and University Heights, Belmont and East Tremont, Kingsbridge Heights and Bedford, Parkchester and Soundview, Williamsbridge and Baychester.

Brooklyn: East New York and Starrett City, Bay Ridge and Dyker Heights, Coney Island, Sheepshead Bay, Brownsville.

Queens: Jackson Heights, Hillcrest and Fresh Meadows, Bayside and Little Neck, Jamaica and Hollis, Queens Village, Rockaway and Broad Channel.

Staten Island: St. George and Stapleton, South Beach and Willowbrook, Tottenville and Great Kills.

The minimum cost incurred for setting up all these 26 vaccination sites will be $4,117,692. They can serve approximately 20,151 vaccines per day. To put all the simulated Community-based Pop-up Vaccination Centers into a clearer perspective, I mapped these locations in regards to the Covid-19 maps shown in Figures 21 and 22.

Figure 21

Covid Case Rate Map

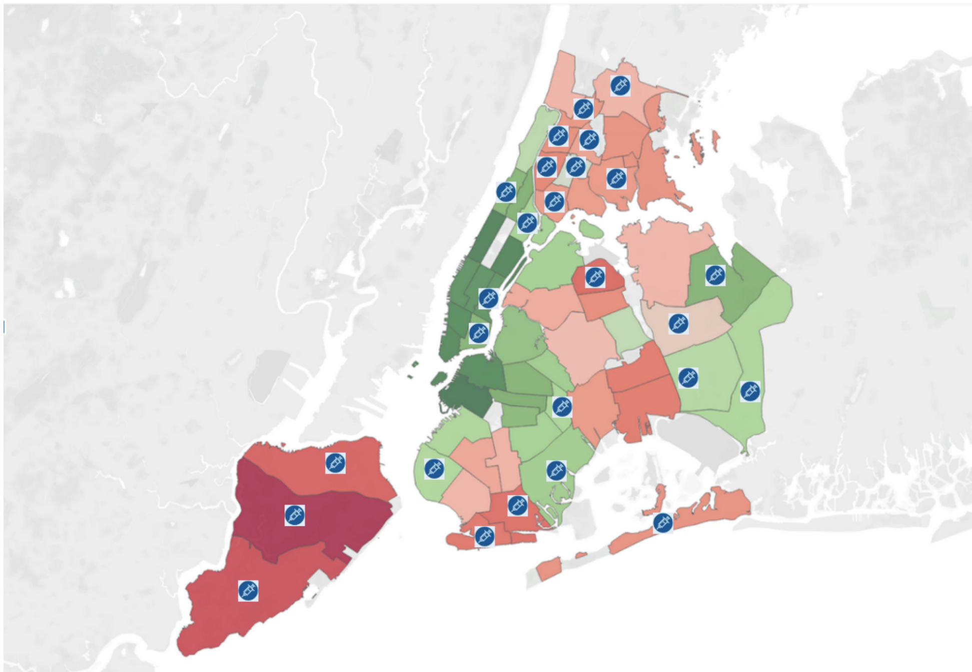
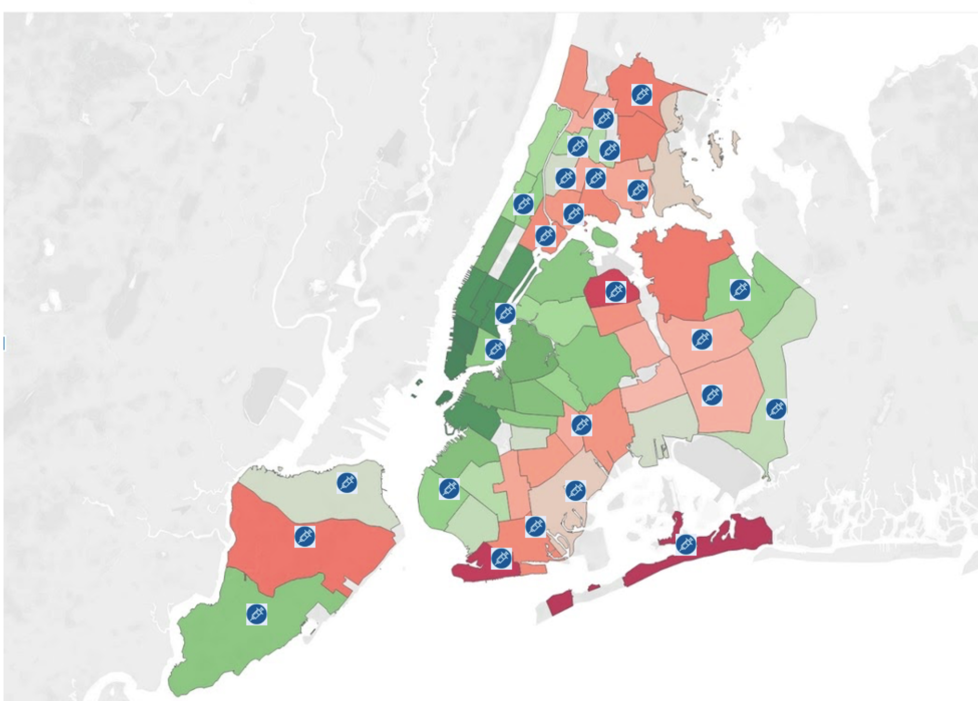


Figure 22

Covid Death Rate Map



From Figures 21 and 22, I can see that the community-based pop-up vaccination centers serve the neighborhoods that have been impacted by Covid-19 the most. They offer a more equitable distribution of the Covid-19 vaccines in New York City.

Conclusion

Covid-19 pandemic has disproportionately affected different neighborhoods within New York City. Areas that have poor socio-economic conditions have higher Covid Case Rate and Covid Death Rate. These areas also have lower Vaccination Rates. In order to provide equitable distribution of Covid-19 vaccines, I developed and solved an optimization model to determine locations for Community-based Pop-up vaccination centers to better reach neighborhoods in need.

However, since exact data on the cost of setting up a vaccination site and its respective vaccine capacity are not available, the model is based on simulation of cost and capacity using available estimates. Further research can be conducted incorporating the stochastic nature of demand for Covid-19 vaccines. I believe that by using exact data of cost, capacity, and change of demand, the obtained locations from the model will be more accurate. In addition, this research does not incorporate factors such as occupation and transportation time for people in each neighborhood as well as other types of vaccination sites that can be set up. These will be great future contributions to the optimization problem for a more equitable distribution of Covid-19 vaccines.

Reference

Almagro, Milena & Orane-Hutchinson, Angelo(2020). *The Determinants of the Differential Exposure to COVID-19 in New York City and Their Evolution Over Time*. SSRN. http://dx.doi.org/10.2139/ssrn.3573619.

Graa, Amel & Benhamida, Farid. (2020). A review on optimization methods applied to energy management system. Serbian Journal of Management, 15(2), 371–382. https://doi-org.rlib.pace.edu/10.5937/sjm15-22519.

Banerjee, A., & Banerjee, T. (2017). *Weaving Analytics for Effective Decision Making*. Sage Publications Pvt. Ltd.

Bewal, N., Minhas, S., Krishna Prasad, G., Yadav, A., Sreedhar, C., Bhasin, D., & Kumar, S. (2020). *Impact of COVID 19 lockdown on emergency admissions at a large zonal hospital.* Medical Journal of Dr. D.Y. Patil University, *13*(6), 603–607. https://doi-org.rlib.pace. edu/10.4103/mjdrdypu.mjdrdypu_382_20.

CHARVADEH, Y. K., & YI, G. Y. (2020). *Data Visualization and Descriptive Analysis for Understanding Epidemiological Characteristics of COVID-19: A Case Study of a Dataset from January 22, 2020 to March 29, 2020.* Journal of Data Science, 18(3), 526–535. https://doi-org.rlib.pace.edu/10.6339/JDS.202007_18(3).0018

Community Health Profiles. Home | Profiles. (n.d.). https://a816-health.nyc.gov/hdi/ profiles/.

Federally Supported Community Vaccination Centers. FEMA.gov. (n.d.). https://www.fema.gov/disasters/coronavirus/vaccine-support/vaccine-center.

Harris, E. Jeffrey (2021). *An Even Deeper Look at Vaccination Rates in New York City. Life Under Quarantine.* https://lifeunderquarantine.blog/2021/03/31/an-even-deeper-look-at-vaccination-rates-in-new-york-city/.

New York State Governor (2021)*. Governor Cuomo Announces 18 Community-Based Pop-Up Vaccination Sites Coming Online Over the Next Week to Vaccinate 8,500 New Yorkers*. Reimagine, Rebuild and Renew New York. (n.d.). https://www.governor.ny.gov/news/governor-cuomo-announces-18-community-based-pop-vaccination-sites-coming-online-over-next-week.

McMinn, S., Chatlani, S., Lopez, A., Whitehead, S., Talbot, R., & Fast, A. (2021). *Across The South, COVID-19 Vaccine Sites Missing From Black And Hispanic Neighborhoods*. NPR. https://www.npr.org/2021/02/05/962946721/across-the-south-covid-19-vaccine-sites-missing-from-black-and-hispanic-neighbor.

Nerkar, D. Amol (2016). *Business Analytics (BA):Core of Business Intelligence(BI).* International Journal of Advanced Engineering, Management and Science (IJAEMS). 2(12), 2176-2178.

New Jersey Department of Health (2020). *Covid-19 Vaccination Plan.* https://www.state.nj.us/health/cd/topics/New%20Jersey%20Interim%20COVID-19%20Vaccination%20Plan%20-%2010-26-20%20(1).pdf

NYChealth. (2021, May 9). nychealth/coronavirus-data. GitHub. https://github.com/nychealth/coronavirus-data/blob/master/totals/data-by-modzcta.csv.

OASIS Map. (n.d.). http://www.oasisnyc.net/map.aspx.

PULAT, Meryem, BAYYURT, Dilayla & KOCAKOÇ, DEVECİ, Ipek. (2020). *A Review on Optimization Literature Related to Operations Research Field.* Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, *22*(3), 1023–1044. https://doi-org.rlib.pace.edu/10.16953/deusosbil. 535425.

Senese, K. (2021). *How much do mass vaccine clinics cost to run? Up to $400,000 a day*. Chicago Sun Times. https://chicago.suntimes.com/coronavirus/2021/5/2/22409977/mass-vaccine-clinics-cost-chicago-illinois-springfield-quincy-columbia-data-project-coronavirus.

Yap, A. Y. (2020). Creating Business Analytics Dashboard Designs Using Visualization Methodologies: Case Methods for Innovative Analytics Pedagogy. *Information Systems Education Journal*, *18*(5), 25–33.