CMP-262-26222 Final Project Summary

By Dana Tryon

The data set that I used for this project is COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths. This data set was compiled by the New York City Health Department, and I found it on data.gov website (Comma Separated Values File URL: https://data.cityofnewyork.us/api/views/rc75-m7u3/rows.csv?accessType=DOWNLOAD). The data is in CSV format and consists of 1501 rows by 55 columns. Each row is a single date starting on 2/29/20 and ending 4/8/24. Each column is a daily count of confirmed cases, probable cases, hospitalizations, and deaths for the total of New York City and then repeated for each of the 5 boroughs’ totals. Each of these columns was repeated for 7-day rolling averages. The data analysis was performed using Jupyter Notebook version 6.5.4 running Python 3.11.7 packaged by Anaconda, Inc. The data did not contain any null values or other inconsistencies and was straightforward to deal with. I only altered the original data set by removing all the 7-day rolling average total columns and I added a column called ‘month’ by assigning the corresponding row’s date value to a string of the month name. This was done for grouping purposes. I devised four questions for this data set.

When did Covid-19 peak in New York City? I was curious to see the daily ups and downs of case counts over the four-year span of the data set. I used a line graph to overlay daily case counts, daily hospitalizations, and daily deaths. The date is the x axis and daily counts is the y axis. The case counts peak for the first 3 winter seasons with an especially high peak in the 2021 to 2022 winter season probably due to the government relaxing movement restriction policies.[[1]](#footnote-1) A similar pattern is seen with hospitalizations and deaths; however, these peaks were much bigger in the first wave and lower on the second and third peaks, probably due to the role-out of vaccines.[[2]](#footnote-2)

Which boroughs were affected the most? For this question, I chose to focus on the first wave (first 100 days) of the pandemic when there was no collective (herd) immunity, and the virus was most harmful. I used line graphs to chart total daily case counts, daily hospitalization counts, and daily death counts. The total numbers were charted along with overlays of the five boroughs. Visually, the borough of Queens appeared to be the most affected and Staten Island the least affected. I calculated and displayed the sum totals of the case counts, hospitalization counts, and death counts, and this confirmed that the borough of Queens had the highest totals in each of the categories. Three ordered bar charts were created to visualize each categories’ total counts.

How severe was the Covid-19 Pandemic in New York City? This can be answered by calculating the odds of being hospitalized and dying if one contracted this highly contagious Covid-19 virus. Having to spend time for treatment in the hospital is a bad outcome and dying is the worst outcome. From the previous summed category counts of the first wave, I calculated the percentage of summed hospitalizations to summed cases and the percentage of summed fatalities to summed cases. There was a 25 percent chance of needing hospital treatment if one contracted Covid-19 and an 11 percent chance of dying if one contracted Covid-19. These are terrible odds and show how damaging this pandemic was. In hindsight, the government leaders made the correct decision to shut down society to the bare functioning minimum to minimize the amount of harm caused.

Is there a particular time of the year with the most impact due to Covid-19? This question would be most useful to hospital administrators for planning staffing requirements. According to the graphs used in the first question, there seemed to be a pattern of the winter months having peak cases and hospital visits. This is typical of respiratory illnesses that spread from exhaled viral substances. The total summed hospital visits for the entirety of the data set will be grouped by the month that they occurred and displayed on a bar chart. The results show a maximum of hospital visits caused by the Covid-19 virus occurred in the month of January with the months of December through April being the busiest time. The months spanning May through November were a reprieve. I created a second graph to demonstrate some of the worst days of the pandemic and what months these worst of days occurred in. A scatterplot graph of daily hospitalization counts versus daily case counts grouped by coloring each point by the month of its occurrence show a pattern. Each point is a day describing the number of hospital visits and confirmed cases that occurred on that day. Most of the points are clustered in the lower left quadrant and it is not easy to tell much from this. There is a grouping of points of high hospitalization counts around the daily case count line of about 5000. This pattern of points has the coloring of the months of March and April which corresponds to the bad days of the first wave of the pandemic when hospitalization counts were about 25 percent of the case counts. The second pattern of interest in this graph is the December and January colored dots that move up into the upper right quadrant. This corresponds to the third wave when case counts soared after the restrictions were eased in the later part of 2021. Just due to the high volume of case counts, this probably caused hospitalizations to go up, but at a lower overall percentage. So, in conclusion to this question, hospital administrators probably will want all hands-on deck for the months of December and January and again in March and April during a pandemic of a respiratory nature.

One thing I learned from this data set is that public policy should be data driven. What really stood out to me from the first graph is the number of daily case counts that occurred during the third wave in the winter of 2021-2022 after the restrictions were relaxed. This is after two years of the pandemic and vaccinations were available to everyone which resulted in much lower percentages of hospitalizations and deaths. What would happen if government officials didn’t have the courage to shut down the economy in the early days of the pandemic for the sake of public health? One could extrapolate that if case counts soared in the first wave like they did in the third wave with the same odds for survival faced in the beginning before herd immunity was in place, fatalities could have been ten times higher, and the health care system would have collapsed.

1. [Which COVID-19 Rules Still Apply In New York City? (ny1.com)](https://ny1.com/nyc/all-boroughs/news/2021/04/01/which-covid-19-rules-still-apply-in-new-york-city-) [↑](#footnote-ref-1)
2. [COVID-19 Vaccines | HHS.gov](https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html) [↑](#footnote-ref-2)