Executive Summary:

The purpose of this data visualization project is to gain a comprehensive understanding of the impact of the COVID-19 pandemic on public health in New York State. By analyzing three datasets from the New York State Health Department's website, this project explores the relationships between hospitalization rates, fatalities, and vaccinations in the state.

Six hypotheses were tested using different metrics to investigate the correlation between hospitalizations and fatalities, as well as the impact of vaccination rates on hospitalizations. Using appropriate graphs and charts, this project presents its findings, which offer valuable insights for policymakers and health professionals to make informed decisions and mitigate the impact of the pandemic on the state.

Additionally, this project contributes to the overall knowledge base on the COVID-19 pandemic and its impact on public health. Its results provide a comprehensive and insightful analysis of the pandemic's impact on New York State, offering valuable information to guide decision-making in managing the pandemic. Overall, this project is a vital resource for anyone interested in understanding the effects of the COVID-19 pandemic on public health in New York State.

Table of Contents

Data Description:

The project uses the following datasets obtained from the New York State Health Department to analyze the hypothesis

New York State Statewide COVID-19 Hospitalizations and Beds: The dataset contains data gathered at the reporting facility level, including information on patients who were hospitalized, admitted, discharged, and experienced fatalities, as well as information on staffed beds. The patient information was collected through the HERDS Hospital Survey and is limited to those who tested positive for lab-confirmed COVID-19. Hospitalized patients refer to those who were admitted as inpatients in either inpatient or observation beds, excluding patients who were treated and released from an Emergency Department.

Link: https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7

Data Frequency	Daily
Granularity	Facility
Coverage	State-wide
Units	Number of patients hospitalized, admitted, discharged, and fatalities, among lab-confirmed COVID-19 cases, and number of staffed beds
Rows	202K
Columns	37
Hypothesis using the data set	1, 2,4

Columns in the database

Column Name	Description	Type
As of Date	The hospital reporting date through the Health Electronic Response Data System (HERDS) survey	Date & Time
Facility PFI	Facility PFI	Plain Text
Facility Name	Hospital Name	Plain Text
DOH Region	Hospital Regional DOH Office	Plain Text

Column Name	Description	Type
Facility County	The NY county that the facility is located within	Plain Text
Facility Network	The network of the facility	Plain Text
NY Forward Region	NY Forward Region in which the facility is located	Plain Text
Patients Currently Hospitalized	How many confirmed positive COVID-19 patients does the facility have in either inpatient or observation beds at this time?	Number
Patients Admitted Due to COVID	How many patients with confirmed COVID were admitted due to COVID or complications of COVID?	Number
Patients Admitted Not Due to COVID	How many patients with confirmed COVID were admitted where COVID was not included as one of the reasons for admission?	Number
Patients Newly Admitted	How many confirmed, positive COVID-19 patients have been newly admitted since the last report?	Number
Patients Positive After Admission	How many of the positive COVID-19 patients were confirmed as positive AFTER admission AND since the last report?	Number
Patients Discharged	How many confirmed positive COVID-19 patients have been discharged from the facility since the last report?	Number
Patients Currently in ICU	How many confirmed, positive COVID-19 patients are there in the ICU at this time?	Number
Patients Currently ICU Intubated	Of the confirmed positive COVID-19 patients currently in the ICU, how many are intubated?	Number
Patients Expired	How many confirmed positive COVID-19 patients have expired in the facility since the last report? Summary level reporting by the facility.	Number

Column Name	Description	Type
Cumulative COVID-19 Discharges to Date	Cumulative Discharges	Number
Cumulative COVID-19 Fatalities to Date	The cumulative number of in-hospital fatalities to date. The reporting of cumulative in-hospital fatalities are from a patient-specific verified file reported by the hospital and may not match the summary level reporting of Patients Expired.	Number
Total Staffed Beds	Total Staffed Beds in Hospital. Data Replaced as of May 19, 2021 by Tot_Acute_Beds	Number
Total Staffed Beds Currently Available	Total Staffed Beds Currently Available in Hospital. Data Replaced as of May 19, 2021 by Tot_Acute_Occup	Number
Total Staffed ICU Beds	Total Staffed ICU Beds in Hospital. Data Replaced as of May 19, 2021 by Tot_ICU_New_Beds	Number
Total Staffed ICU Beds Currently Available	Total Staffed ICU Beds Currently Available in Hospital. Data Replaced as of May 19, 2021 by Tot_ICU_New_Occup	Number
Total Staffed Acute Care Beds	How many staffed acute care beds are currently at your hospital?	Number
Total Staffed Acute Care Beds Occupied	How many of those staffed acute care beds are currently occupied?	Number
Total Staffed ICU Beds 1	How many staffed ICU beds are currently at your hospital?	Number
Total Staffed ICU Beds Currently Occupied	How many of those staffed ICU beds are currently occupied?	Number
Total New Admissions Reported	Total New Admissions (Patients Newly Admitted + Patients Positive After Admission)	Number

Column Name	Description	Type
Patients Age Less Than 1 Year	Currently hospitalized age category less than 1 year	Number
Patients Age 1 To 4 Years	Currently hospitalized age category 1 to 4 years	Number
Patients Age 5 to 19 Years	Currently hospitalized age category 5 to 19 years	Number
Patients Age 20 to 44 Years	Currently hospitalized age category 20 to 44 years	Number
Patients Age 45 to 54 Years	Currently hospitalized age category 45 to 54 years	Number
Patients Age 55 to 64 Years	Currently hospitalized age category 55 to 64 years	Number
Patients Age 65 to 74 Years	Currently hospitalized age category 65 to 74 years	Number
Patients Age 75 to 84 Years	Currently hospitalized age category 75 to 84 years	Number
Patients Age Greater Than 85 Years	Currently hospitalized age category greater than 85 years	Number
Hospitalized Indicator	An indicator on if the sum of the age groups equals the number reported as currently hospitalized	Number

New York State Statewide COVID-19 Fatalities by Age Group

The dataset comprises information about patients with lab-confirmed COVID-19 disease, including the cumulative number and percentage of fatalities reported by healthcare facilities according to age group and reporting date. The dataset excludes fatalities related to COVID-19 that occurred outside hospitals, nursing homes, or adult care facilities. The primary objective of releasing this dataset is to provide users with healthcare facility fatality information related to lab-confirmed COVID-19 disease.

The data comes from the New York State Department of Health's (NYSDOH) Health Electronic Response Data System (HERDS), which collects daily COVID-19 surveys from hospitals, nursing homes, and adult care facilities.

The fatality figures are calculated by assigning age groups to patients and then adding up the fatalities for each group, as of each reporting date. The statewide total fatality numbers are calculated by summing up fatalities across all age groups by reporting date. The fatality percentages are calculated by dividing the number of fatalities in each age group by the statewide total number of fatalities, by reporting date. The fatality numbers represent the cumulative fatalities reported as of each reporting date.

Link: https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Fatalities-by-Ag/du97-svf7

Data Frequency	Daily
Granularity	County
Coverage	State-wide
Units	Fatalities among lab-confirmed COVID-19 cases
Rows	12.2K
Columns	4
Hypothesis using the data set	1, 5

Columns in Database

Column Name	Description	Type
Report Date	The “as of date” of the cumulative fatality number	Date & Time
Age Group	The age group of the patient fatality (includes rows for Statewide Total)	Plain Text
Fatality Count	The number of fatalities within the age group	Number
Percent	The percent of fatalities within the age group	Number

New York State Statewide COVID-19 Vaccination Data:

This dataset reports daily on the number of people vaccinated by New York providers with at least one dose and with a complete COVID-19 vaccination series overall since December 14, 2020. New York providers include hospitals, mass vaccination sites operated by the State or local governments, pharmacies, and other providers registered with the State to serve as points of distribution.

This dataset is created by the New York State Department of Health from data reported to the New York State Immunization Information System (NYSIIS) and the New York City Citywide Immunization Registry (NYC CIR). County-level vaccination data is based on data reported to NYSIIS and NYC CIR by vaccine providers. Residency is self-reported by the individual being vaccinated. This data does not include vaccine administered through Federal entities or performed outside of New York State to New York residents. NYSIIS and CIR data is used for county-level statistics. New York State Department of Health requires all New York State vaccination providers to report all COVID-19 vaccination administration data to NYSIIS and NYC CIR within 24 hours of administration.

Link: https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Vaccination-Data/duk7-xrni

Data Frequency	Weekly
Granularity	County
Coverage	Statewide
Units	COVID-19 Vaccine
Rows	53.7 k
Columns	5
Hypothesis using the data set	2,4

Column Name	Description	Type
Region	The region of the vaccinated individuals based on the Regional	Plain Text
County	The New York State County of residence of the vaccinated individuals.	Plain Text
First Dose	Represents the total number of individuals who have received at	Number

	least one dose of any COVID-19 vaccine.	
Series Complete	Represents the total number of individuals who have completed the recommend series of a given COVID-19 vaccine product (e.g., 2 doses of the 2-dose Pfizer or Moderna vaccine; 1 dose of the single dose Johnson & Johnson vaccine).	Number
Report as of	The date of the reporting.	Date & Time

Data Cleaning:

Hypothesis 1:

Data cleaning tools used for this hypothesis are Python with Pandas, Excel, and Tableau. Since the hypothesis was at a county-wise granularity, the data from multiple facilities in a county could be aggregated per day. Additionally, in order to prove the hypothesis, we only need the hospitalization per age group data. All other data could be deleted. The following steps were followed for data cleaning and data manipulation to bring it to the stage at which visualization were developed.

1. Removed all columns from dataset **New York State Statewide COVID-19 Hospitalizations and Beds** except the ones required to prove the hypothesis. The following columns were retained after this step

As of Date	The hospital reporting date through the Health Electronic Response Data System (HERDS) survey	Date & Time
Facility County	The NY county that the facility is located within	Plain Text
Patients Age 1 To 4 Years	Currently hospitalized age category 1 to 4 years	Number
Patients Age 5 to 19 Years	Currently hospitalized age category 5 to 19 years	Number
Patients Age 20 to 44 Years	Currently hospitalized age category 20 to 44 years	Number
Patients Age 45 to 54 Years	Currently hospitalized age category 45 to 54 years	Number
Patients Age 55 to 64 Years	Currently hospitalized age category 55 to 64 years	Number
Patients Age 65 to 74 Years	Currently hospitalized age category 65 to 74 years	Number
Patients Age 75 to 84 Years	Currently hospitalized age category 75 to 84 years	Number

Patients Age Greater Than 85	Currently hospitalized age category greater than 85 years	Number

2. Aggregated all patients in a county per reported date using pthon

```
import pandas as pd

df = pd.read_csv('/content/New_York_State_Statewide_COVID-19_Hospitalizations_and_Beds-2.csv')

grouped = df.groupby(['As of Date', 'Facility County']).sum().astype('int')

result = grouped.reset_index().sort_values(['As of Date', 'Facility County'], ascending=[True, True])

result.to_csv('/content/New_York_State_Statewide_COVID-19_Hospitalizations_and_Beds-2-Aggregated.csv', index=False)
```

3. Since **New York State Statewide COVID-19 Fatalities by Age Group** only identifies the fatalities on a given day, created columns to identify the cumulative death per age group on a given day using Excel functions

4. Both these datasets were imported into Tableau, and the age group data and report dates were pivoted to generate the date-wise hospitalization and fatality per age group.

Hypothesis 3:

We use Python with Pandas and Tableau as data-cleaning tools for this hypothesis.

Hypothesis 2 and 4:

Data cleaning tools used for this hypothesis are Python with Pandas, Excel, and Tableau. Since the hypothesis was at a county-wise granularity, the data from multiple facilities in a county could be aggregated per day. All other data could be deleted. The following steps were followed for data cleaning and data manipulation to bring it to the stage at which visualization was developed.

Removed all columns from the dataset New York State Statewide COVID-19 Hospitalizations and Beds except the ones required to prove the hypothesis which is:

As of Date	The hospital reporting date through the Health Electronic Response Data System (HERDS) survey	Date & Time

Patients Newly Admitted	The number of patients admitted that day all over the New York State	Number

Removed all columns from the dataset New York State Statewide Vaccination except the ones required to prove the hypothesis which is:

As of Date	The hospital reporting date through the Health Electronic Response Data System (HERDS) survey	Date & Time
next_quarter_date	The date that is exactly three months later corresponding to the date column	Date & Time
First Dose	The number of people who received their first dose of vaccination on a particular date.	Number

Data cleaning tools used for this hypothesis are Excel and Tableau. In order to verify the hypothesis, we had to modify both the Vaccination dataset & Hospitalization & Beds dataset.

We grouped the datasets by date to get the counts by quarters.

Added a new column next quarter date which is the date that is exactly three months later corresponding to the date column in the vaccination dataset.

We joined both the grouped datasets on the date column to get the dataset that we need to validate the hypothesis.

Hypothsis 5:

We use Python with Pandas and Tableau as data-cleaning tools for this hypothesis.

Hypothsis 6:

We use Python with Pandas and Tableau as data-cleaning tools for this hypothesis.

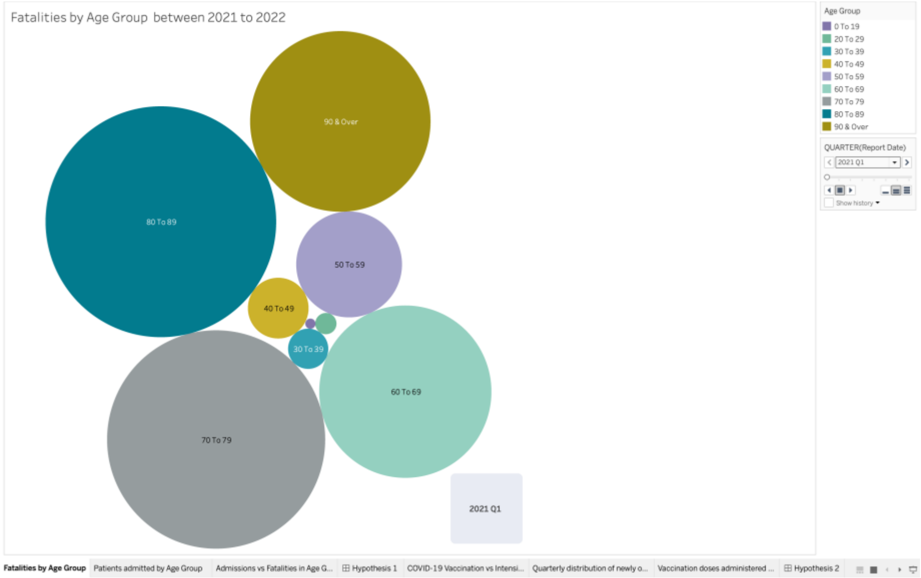
Insights and findings

We are dealing with six hypotheses where data has been analyzed and visualized leading us to either accept or reject the hypothesis.

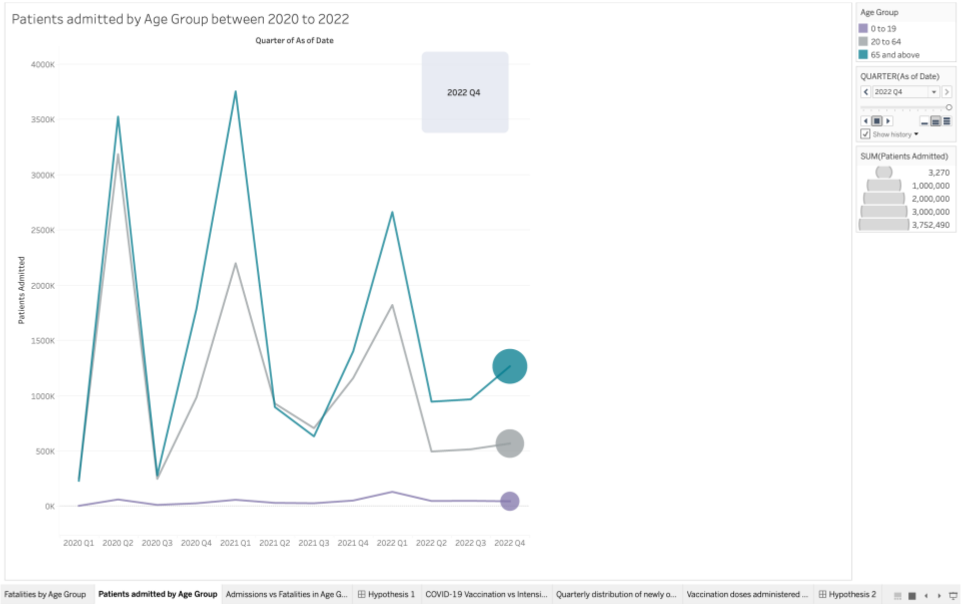
Hypothesis 1:

Covid patients aged 65+ are more likely to experience hospitalization and fatalities than patients in other age groups.

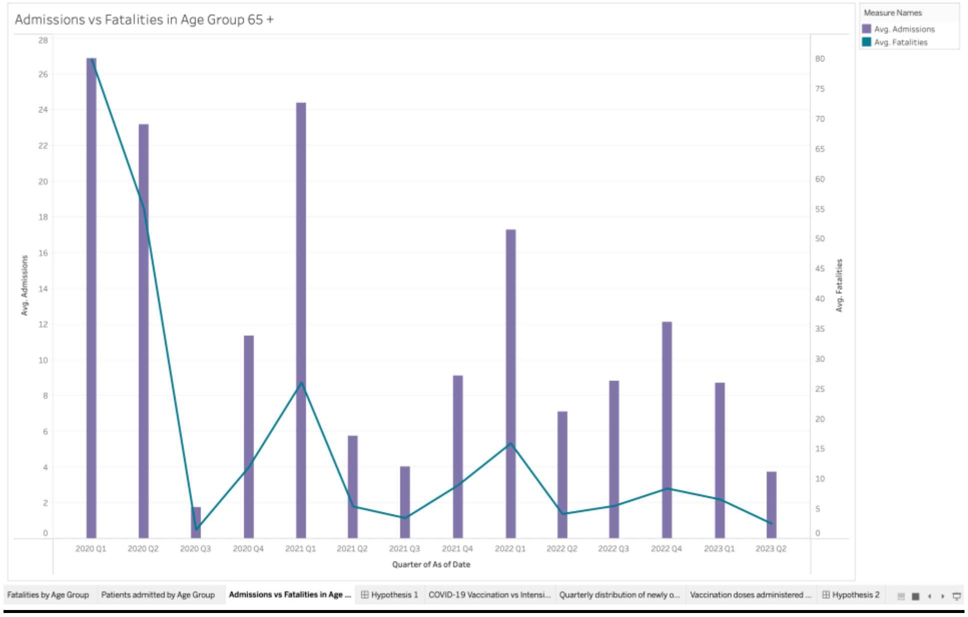
To analyze this hypothesis, we first observed the fatalities per age group over the period of Jan 2021 to Dec 2022 on a quarterly basis. Bubble charts were used to represent this data and it could be observed that most fatalities were in the age group 60 and above.



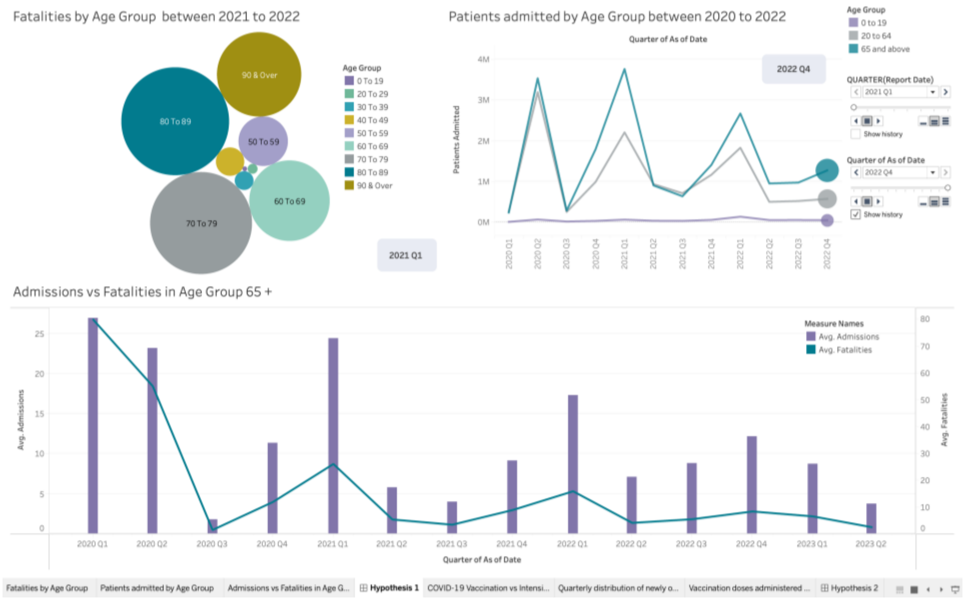
Next, we observed the hospitalization trends in the state of New York in the same duration. Line charts were used to visualize this data. Additionally, to make it more comprehensive, we grouped the age groups into three sets i.e. 0-19, 20-64 and 65 & above. This visualization further amplified the fact that hospitalization was more prominent in the age group 65 and above.



To further prove our hypothesis that fatalities in age group 65+ increased with hospitalization, we dual combination chart that plotted the hospitalization and deaths in the age group 65 and above over a period of 2 years.



The final dashboard created with the three visualizations helps us better identify the effects of COVID-19 on the age group 65 and above in the state of New York.



Fatalities by Age Group between 2021 to 2022

Patients admitted by Age Group between 2020 to 2022

Admissions vs Fatalities in Age Group 65 +

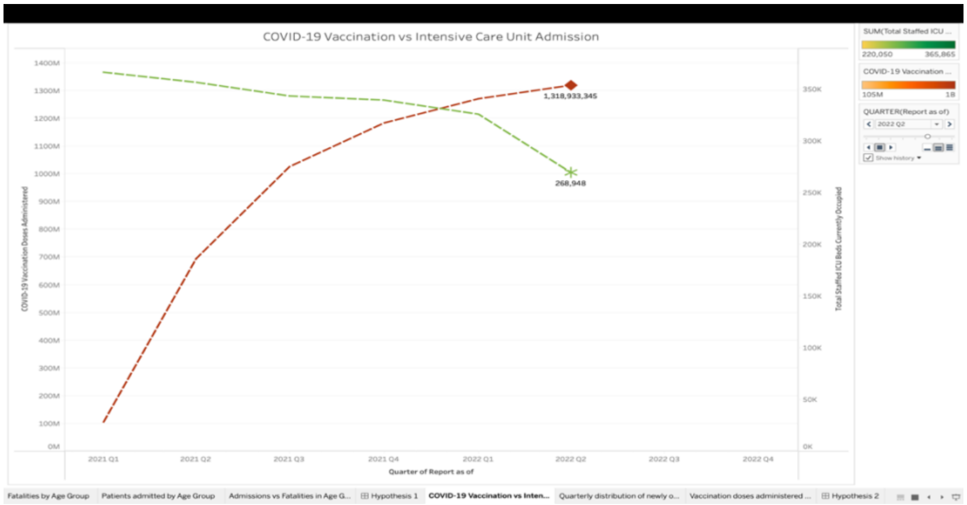
In conculsion, The visualizations indicate that COVID-19 patients aged 65 years and above experienced the highest rates of hospitalization and fatalities, providing empirical evidence to support the hypothesis.

Hypothesis 2:

Whether there is a correlation between sum of covid vaccination doses and number of icu staffed beds newly occupied in New York State.

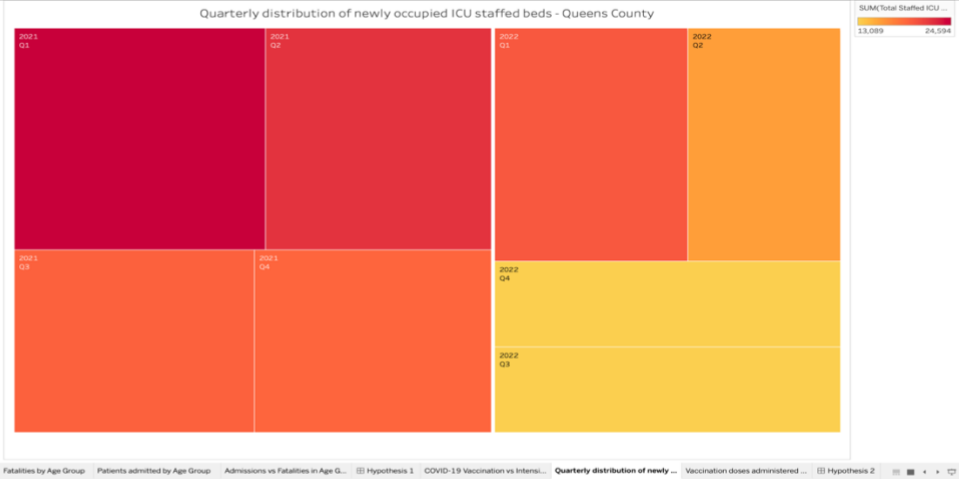
To test the hypothesis, we have used three charts which include multi-line chart, heatmap and a bubble chart.

[redacted]

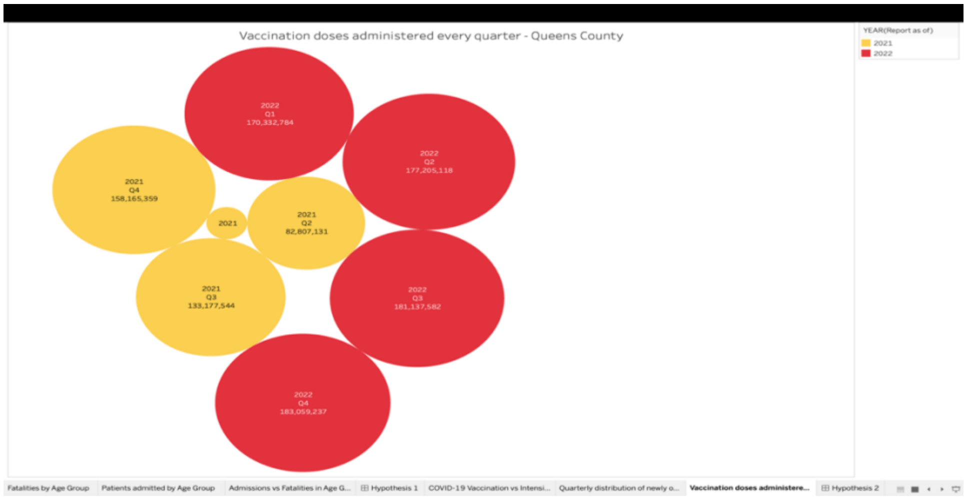


When we look at the animation, it is clear that the number of newly occupied staffed icu beds decreased gradually with the increase in total number of covid vaccination doses from first quarter of 2021 till the last quarter of 2022.

[redacted]



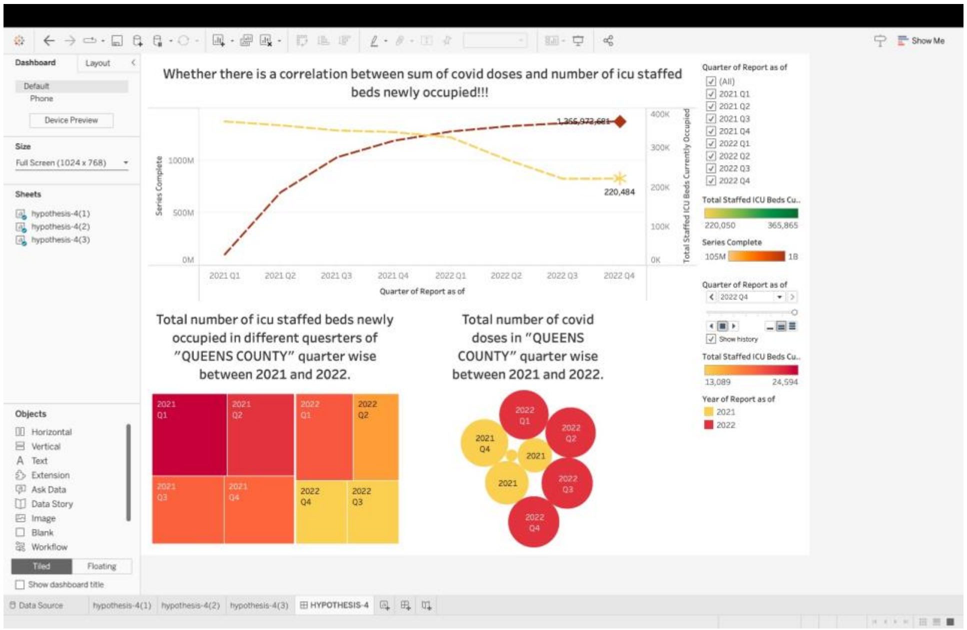
The heatmap represents the total number of newly occupied staffed icu beds in Queens County that happened between 1st quarter of 2021 till the last quarter of 2022.


[redacted]

Vaccination doses administered every quarter - Queens County

The bubble chart here shows the visualization of total number of covid vaccination doses that happened during the same time period.

It is evident that both the charts (bubble chart and heatmap) are completely contrasting with one another because very small portion of the population were fully vaccinated in quarter 1 of 2021 and its when most of the staffed icu beds were occupied whereas by the final quarter of 2022, most of the population got fully vaccinated and thus resulting in the least number of staffed icu beds being occupied.


[redacted]

Whether there is a correlation between sum of covid doses and number of icu staffed beds newly occupied!!!

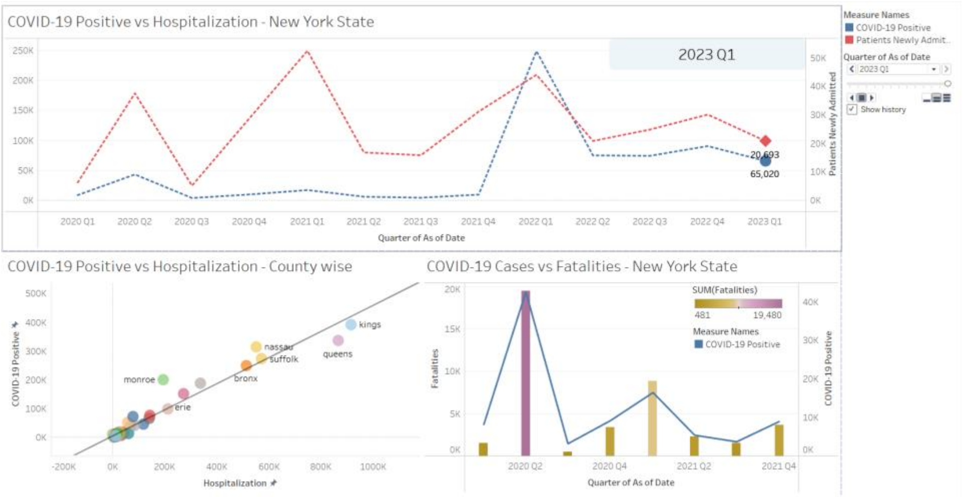
Total number of icu staffed beds newly occupied in different quesrters of "QUEENS COUNTY" quarter wise between 2021 and 2022.

Total number of covid doses in "QUEENS COUNTY" quarter wise between 2021 and 2022.

We can **conclude that, there is a strong correlation between total number of covid vaccination doses and number of newly occupied staffed icu beds.**

Hypothesis 3:

There is a positive correlation between the fatalities in a county and the population that tested COVID-19 positive in New York state

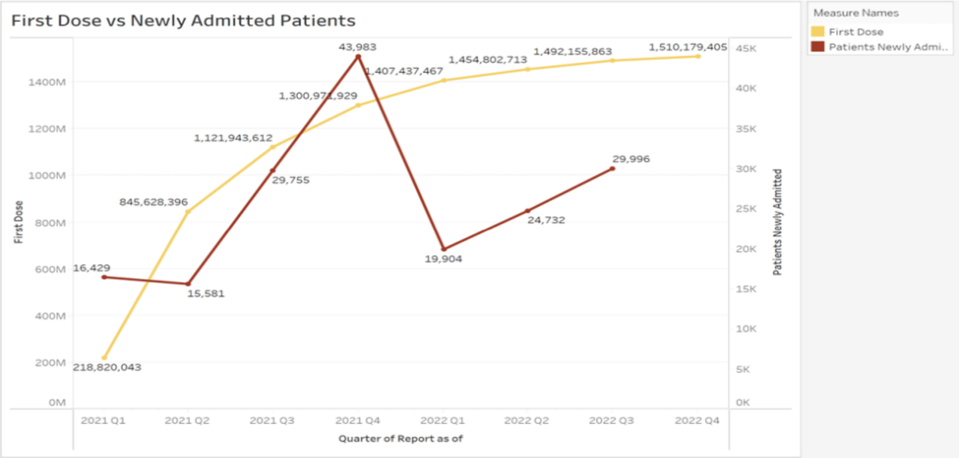


The graph at bottom left shows a scatter plot of county-wise hospitalized and COVID-19 positive counts. The plot includes a trend line, which shows a positive slope with respect to hospitalized people. This indicates that the number of people getting hospitalized is increasing with the rise in COVID-19-positive cases. The other graph in the bottom presents a dual combination graph, representing the relationship between total COVID-19 positive cases and total deaths in New York State. Each bar represents the number of patients who expired in one particular quarter, while the line represents the total positive cases. The visualization shows that the number of patients who expired in each quarter relatively increased with the increase in the number of COVID-19 positive cases recorded. This indicates that fatalities are directly proportional to the total number of positive cases. The top sheet presents an animation that shows the relationship between total COVID-19 positive cases and patients hospitalized. The animation shows that the number of patients being admitted is relatively increasing and decreasing along with the total number of positive cases recorded in a quarter. This suggests that there is a positive relationship between total COVID-19 positive cases and patients hospitalized.

The results of this analysis support the hypothesis that there is a positive correlation between the number of COVID-19 positive cases and the number of fatalities in New York State. The visualizations show that areas with a higher number of positive cases are more likely to experience a higher number of fatalities due to COVID-19. These findings are important for policymakers and healthcare professionals to better understand the impact of the pandemic and to develop effective strategies to mitigate its effects.

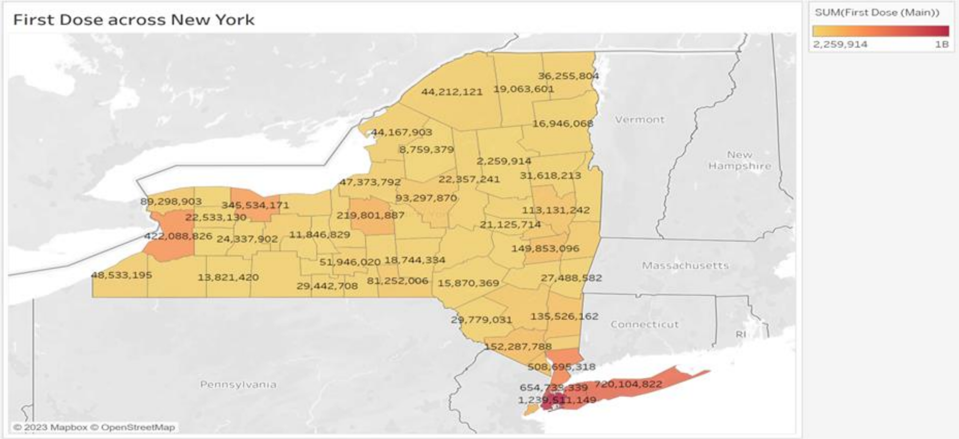
Hypothesis 4:

There exists a strong and significant correlation between the number of individuals who received their first dose of COVID-19 vaccination during a given quarter in a specific location and the number of newly admitted patients with COVID-19 in the subsequent quarter in the same location.

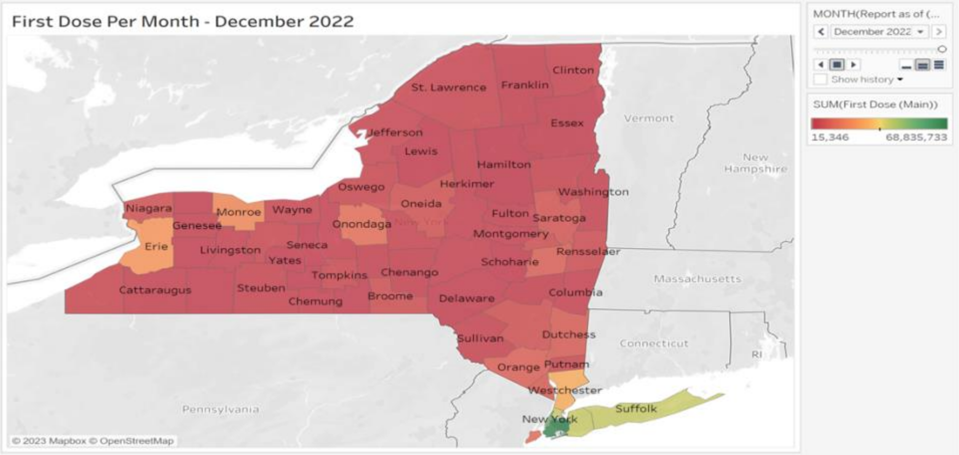


To analyze this hypothesis, we used the dual lines graph with quarters as the x-axis and the counts of both the relevant variables as the y-axis to determine correlation. From the dual lines graph, we can see that the lines for patients admitted in the successive quarter are fluctuating as opposed to the line for the first doses.

Next, we have observed the number of people that are getting the first dose in all the counties across New York State. We used symbol maps to show the number of people that are getting their first dose. Queens County and Hamilton County are respectively the counties with the highest and lowest first dose.

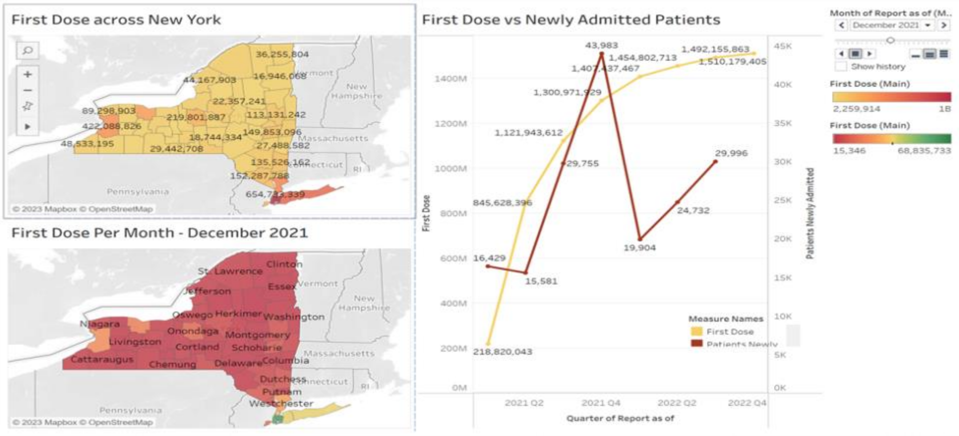


The animation shows the variation in the first dose of vaccination by each month. The map has the data color coded based on the scale shown to the right of it. The lowest values start with red which goes to orange and finally green. We can see that the majority of change is concentrated around the New York, Kings, and Queens Counties and the counties around them.



First Dose Per Month - December 2022

The final dashboard created with the three visualizations helps us better identify whether there exists a strong and significant correlation between the number of individuals who received their first dose of COVID-19 vaccination during a given quarter in a specific location and the number of newly admitted patients with COVID-19 in the subsequent quarter in the same location.



First Dose across New York

First Dose Per Month - December 2021

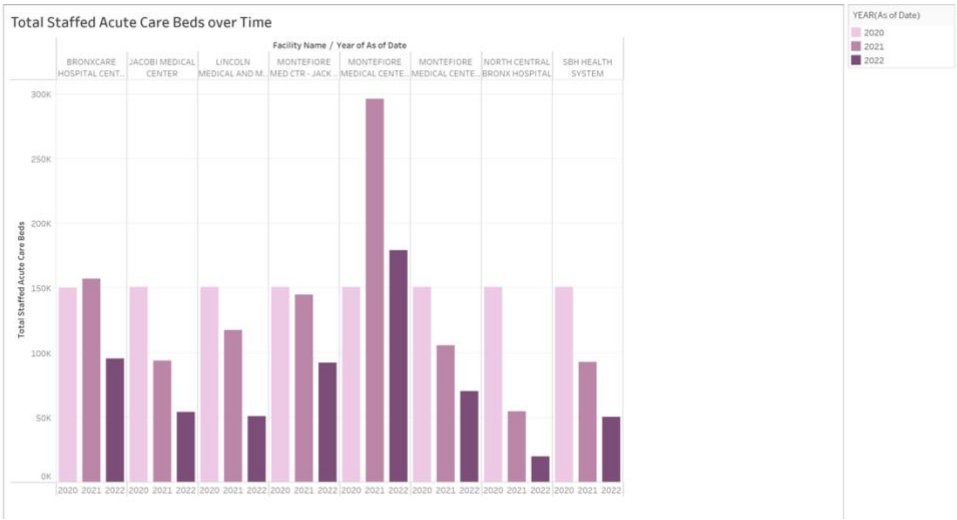
First Dose vs Newly Admitted Patients

In conclusion, there is no strong and significant correlation between the number of individuals who received their first dose of COVID-19 vaccination in a specific location during a given quarter and the number of newly admitted COVID-19 patients in the same location during the subsequent quarter.

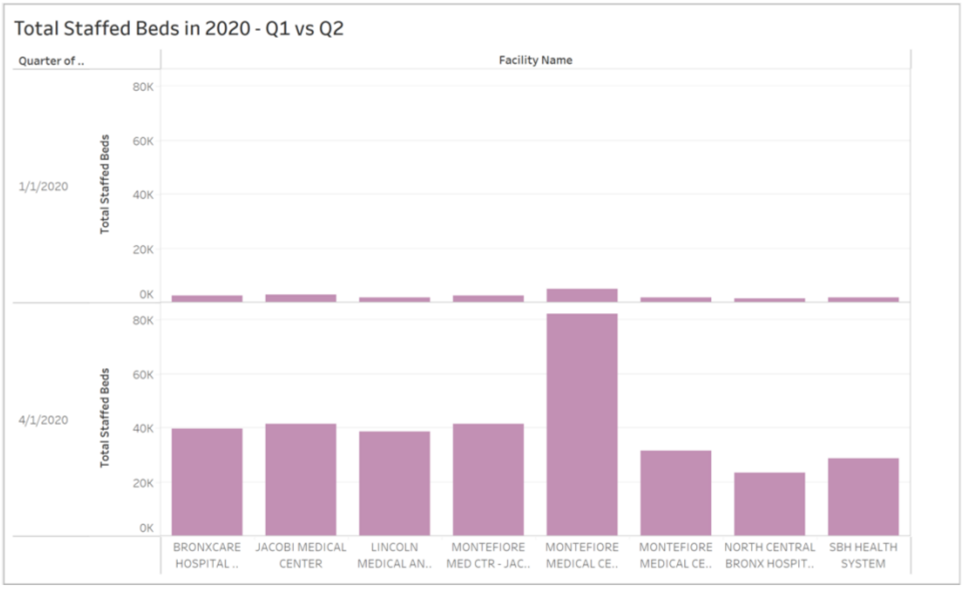
Hypothesis 5:

There is a relationship between the number of acute care beds staffed in a healthcare facility and the likelihood of patient mortality, such that facilities with more beds are less likely to have patients pass away than those with fewer staffed acute care beds?

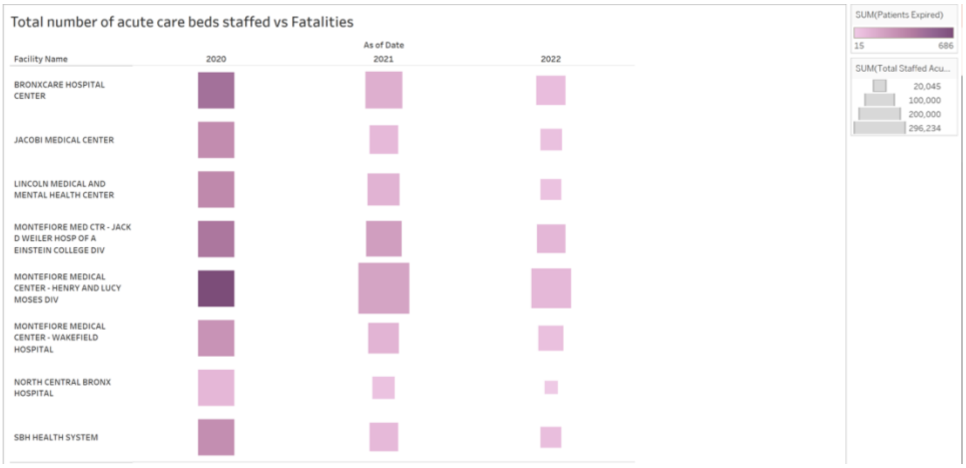
In the first Visualization we have shown the total staffed acute care beds comparing across time, here we have taken bronx county and have shown different medical care facilities in the bronx county over the years 2020, 2021 and 2022.



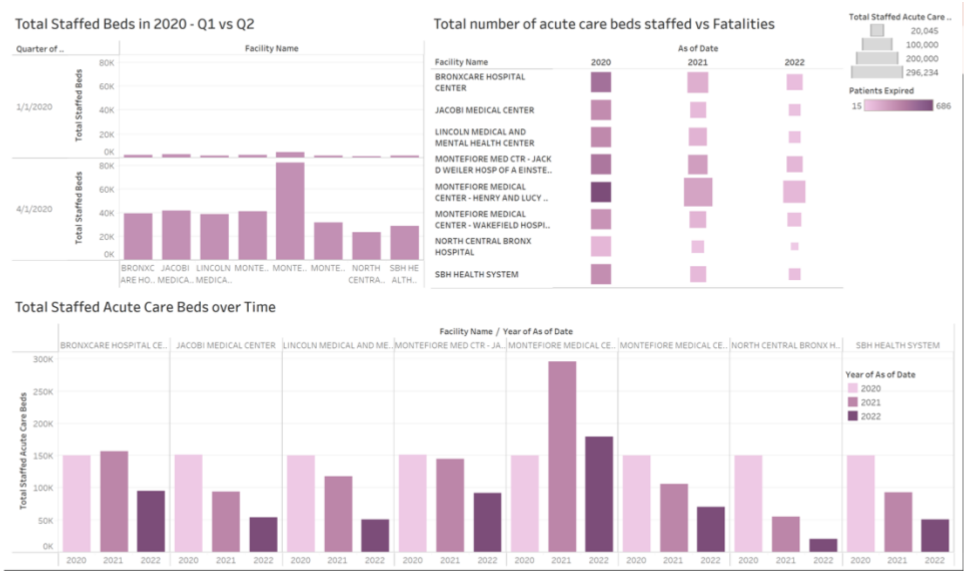
In the second visualization on the top left, we have compared it for the year 2020 for Quarter1 and Quarter 2 and we can clearly see that there is a significant increase in the staffed beds over the second quarter.



Coming to the 3rd visualization at the bottom we have shown a diagram of the total number of acute care beds staffed and the number of deaths, here when we take Montefiore medical. Even though in the year 2021 there is an increase in the beds, the deaths have decreased and in the year 2020 the number of beds is less compared to the one in 2021 but the deaths are more.



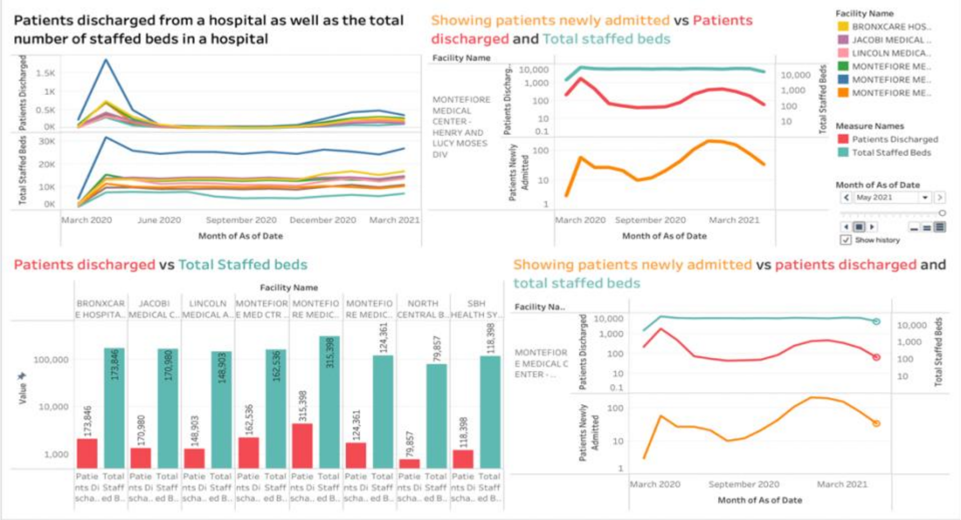
The final dashboard created with the three visualizations helps us better identify the relationship between total number of beds to deaths.



So, in conclusion we can say that the hypothesis has a contradiction and therefore has no relationship between the total number of beds to the deaths.

Hypothesis 6:

Whether the number of patients discharged in Facilities with greater number of hospital beds is more because they have a higher number of staffed beds available as compared to facilities with a smaller number of staffed beds.



The first graph, a line graph, shows the quantity of patients and the total number of staffed beds who were released from the hospital over time. It is clear that as the overall number of staffed beds increases, the number of patients released also increases until a peak is reached.

The graph definitively proves that for various healthcare facilities, if the overall number of staff beds is higher, the number of patients discharged from that facility is higher. All healthcare facilities follow this pattern.

The observations made above lead us to believe that the number of patients released in facilities with a large number of hospital beds is higher because they have a greater number of staffed beds available than facilities with a lesser number of staffed beds.

The number of discharge patients has decreased because the number of nearly admitted patients has decreased due to a decrease in the number of covid care infections over the specified time period. As a result, we may state that the Hypotheses is True,