

EXECUTIVE SUMMARY:

To uncover hospitalization-related insights on COVID-19 in New York in 2021, we gathered two data sets, one dataset is about hospitalization in New York counties, and another is about COVID testing taken in 2021. To generate specific insights to support the overall hospitalization analysis, two counties (Erie and Bronx) were used for most of the hypotheses. We have taken the admission of current patients in terms of quarters, seasons, age-wise Covid-19 impact, discharge rate, staffed beds, and ICU admission to dive further into the criteria. Further to strengthen According to the data, counties with a higher population suffered more than counties with a lower population since the hospitalization facilities could not accommodate the demand. When we looked more closely at the data, we saw that many hospitals upgraded their facilities moving ahead and gave patients the right care.

Seasons may have played a role in the spread of COVID-19, according to our analysis of many dimensions. In the meantime, we attempted to analyze the severity of the patients admitted in various age groups, which was found to be closely linked to the actions made by the government to restrict the environment by enforcing lockdowns and curfews.

Overall, a range of methods was used to test our assumptions to gather information about hospitalization across two sizable counties as a result of Covid-19.

INDEX

Data Description.....	Page 4
Data Cleaning.....	Page 4
Data Cleaning.....	Page 4
General Introduction.....	Page 7
Insights and Findings.....	Page 8
Conclusion.....	Page 14

DATA DESCRIPTION

This dataset contains statistics on patients who were hospitalized, admitted, released, and fatalities at the reporting facility level. Additionally, it contains details on staffed beds. Lab confirmation of the COVID-19 positivity in the patient data obtained as part of the HERDS Hospital Survey. Hospitalized refers to patients who have been treated and released from an emergency department, not to patients who have been admitted as inpatients in either inpatient or observation beds. The initial name of this dataset was COVID-19 Hospitalizations and Beds from the Hospital Electronic Response Data System (HERDS) Hospital Survey.

One drawback of this data set is that no features were provided to support the governmental actions impacting the admission of covid-19 patients, which would have provided strong evidence of an increase or decrease in the Covid-19 effect.

All things considered, we had adequate data to compare and display the variables influencing hospitalization in the counties of Bronx and Erie.

DATA CLEANING

The dataset underwent a manual first round of cleaning in Excel because it contained few parameters and no data that contributed to the hypotheses, which was eliminated.

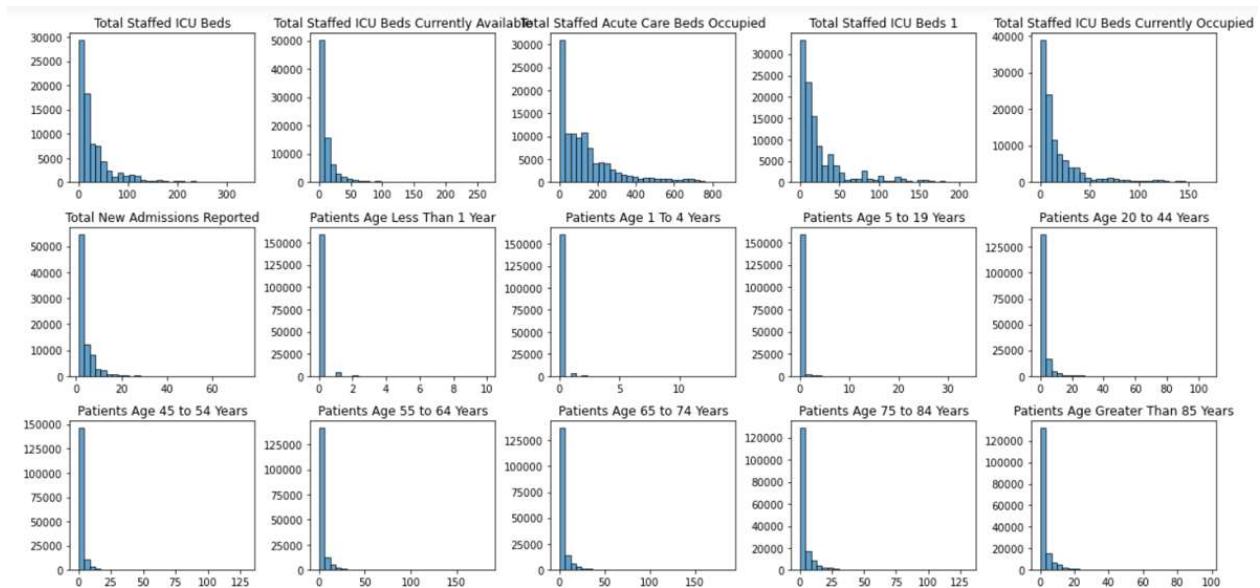
The dataset was then further cleaned in the Python Language using Jupyter notebook. First, the data type of the variables was discovered to determine the percentage of missing values in the data set. Values of the category and numeric data types were absent. Using the mode of the category value and the median of the numerical value (Since the dataset was skewed to the right), a statistical technique was developed to fill in the missing value. With the aid of looping statements, the columns were filled with the

found

statistical

value.

- The skewness of the dataset was found to be right skewed:



- Cleansing code:

```
In [11]: #For filling numerical data
num = ['Patients Currently Hospitalized', 'Total Staffed ICU Beds Currently Occupied', 'Total New Admissions Reported',
       'Total Staffed ICU Beds 1', 'Total Staffed Acute Care Beds Occupied', 'Patients Admitted Due to COVID',
       'Patients Admitted Not Due to COVID', 'Patients Newly Admitted', 'Patients Positive After Admission',
       'Patients Discharged', 'Patients Currently in ICU', 'Patients Currently ICU Intubated', 'Patients Expired',
       'Total Staffed ICU Beds', 'Total Staffed ICU Beds Currently Available', 'Patients Age Less Than 1 Year',
       'Patients Age 1 To 4 Years', 'Patients Age 5 to 19 Years', 'Patients Age 20 to 44 Years', 'Patients Age 45 to 54 Years',
       'Patients Age 55 to 64 Years', 'Patients Age 65 to 74 Years', 'Patients Age 75 to 84 Years',
       'Patients Age Greater Than 85 Years']

for column in num:
    Hospital_beds_df[column].fillna(round(Hospital_beds_df[column].median(),0), inplace=True)

In [12]: #For filling categorical data
cat = ['Total Staffed Beds', 'Total Staffed Acute Care Beds', 'Total Staffed Beds Currently Available']
for column in cat:
    Hospital_beds_df[column].fillna(Hospital_beds_df[column].mode()[0], inplace=True)
```

- After Cleansing the Primary dataset:

```
In [13]: #Check the cleansed data
Hospital_beds_df.isna().sum()
```

```
Out[13]: As of Date 0
Facility PFI 0
Facility Name 0
DOH Region 0
Facility County 0
Facility Network 0
NY Forward Region 0
Patients Currently Hospitalized 0
Patients Admitted Due to COVID 0
Patients Admitted Not Due to COVID 0
Patients Newly Admitted 0
Patients Positive After Admission 0
Patients Discharged 0
Patients Currently in ICU 0
Patients Currently ICU Intubated 0
Patients Expired 0
Cumulative COVID-19 Discharges to Date 0
Cumulative COVID-19 Fatalities to Date 0
Total Staffed Beds 0
Total Staffed Beds Currently Available 0
Total Staffed ICU Beds 0
Total Staffed ICU Beds Currently Available 0
Total Staffed Acute Care Beds 0
Total Staffed Acute Care Beds Occupied 0
Total Staffed ICU Beds 1 0
Total Staffed ICU Beds Currently Occupied 0
```

- After Cleansing the secondary dataset:

```
In [39]: COVID_Testing_df.isna().sum()
```

```
Out[39]: Test Date      0
          County        0
          New Positives  0
          Cumulative Number of Positives  0
          Total Number of Tests Performed  0
          Cumulative Number of Tests Performed  0
          Test % Positive  0
          Geography      0
          dtype: int64
```

GENERAL INTRODUCTION

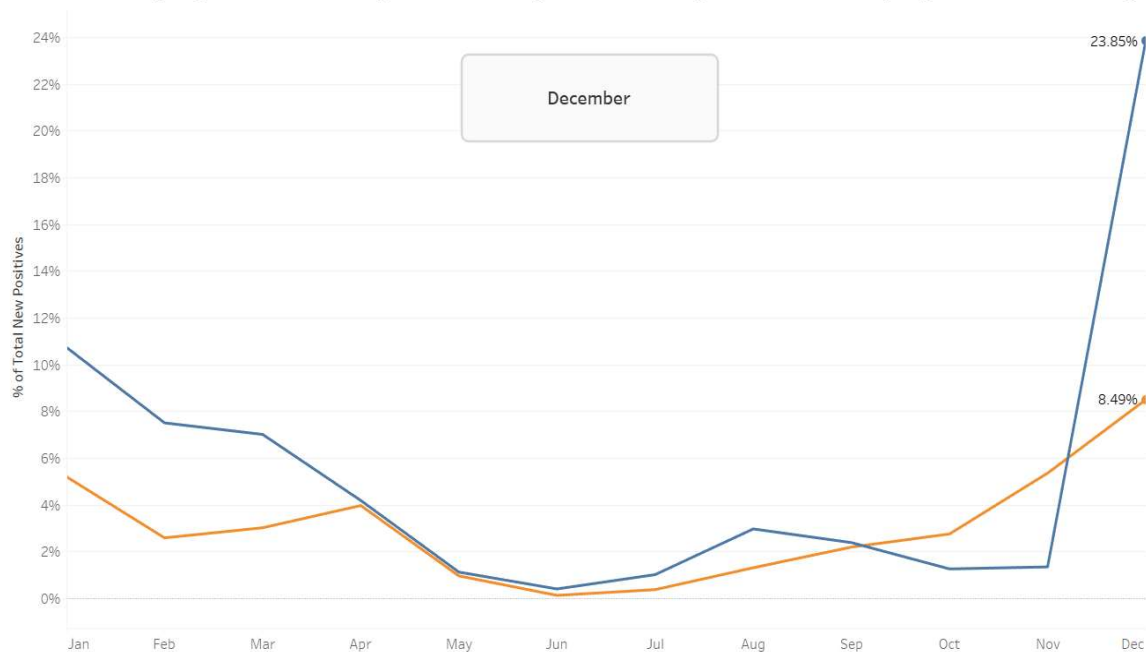
In the first month of 2021, there was a huge, unexpected raise in terms of patient admission across both counties which shocked the hospital industry in providing proper treatments to the patients as there were not many facilities to cure. People were forced into lockdowns, social distancing, mandatory masks in public, and a variety of other changes because of covid-19 spread over the two counties.

We decided to find insights into New York during the peak pandemic and how it gradually decreased, also found the trends of other factors influencing the patient's admission. Due to the initial spread of the disease, we have considered discharge rates and the patient's expired and staffed beds and we have justified different hypotheses pertaining to the factors considered. Overall, all the visualization would provide a detailed analysis/study of the hospitalization.

INSIGHTS AND FINDINGS

Hypothesis 1:

Whether the people in Erie County are less likely to affected by COVID-19 than people in Bronx County



The first hypothesis is that whether the people in Erie County are less likely to affected by COVID-19 than people in Bronx County. The X-axis represents the month, which is from January to December in 2021 and the Y-axis stands for the total percentage of the people which are tested positive. The blue line is for Bronx County and the Orange line is for Erie County.

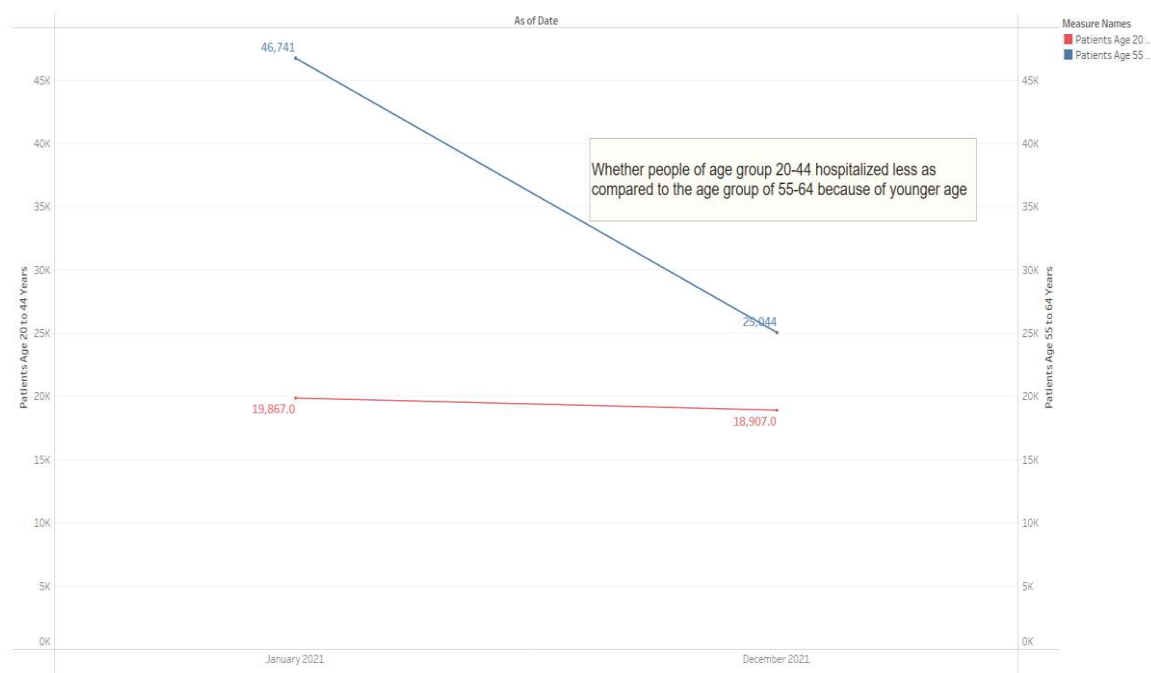
As you can see from the line chat, Bronx County initially had 11% people affected and Erie County had 5% people affected. As the time goes by, both counties decreased the percentage of affected people and stayed below 1% at June 2021. This might because of mandatory mask policy created for both counties and people were following the rules and took prevention. While they both increased slowly during the month of July to October.

As you can see here, in November and December, they both increased dramatically, especially for Bronx County, which increased to 23.85%, compared to 8.49% for Erie County.

As a result, we cannot reject the hypothesis 1, and conclude that Erie County are less likely to affected by COVID-19 than people in Bronx County.

Hypothesis 2:

Whether people of age group 20-44 hospitalized less as compared to the age group of 55-64 because of younger age



The 2nd hypothesis is stated as:

Whether people of age group 20-44 are hospitalized less as compared to the age group of 55-64 because of younger age

To prove this hypothesis, we have used a **slope chart**. Slope charts are simple graphs that quickly and directly show **transitions, changes over time, absolute values, and**

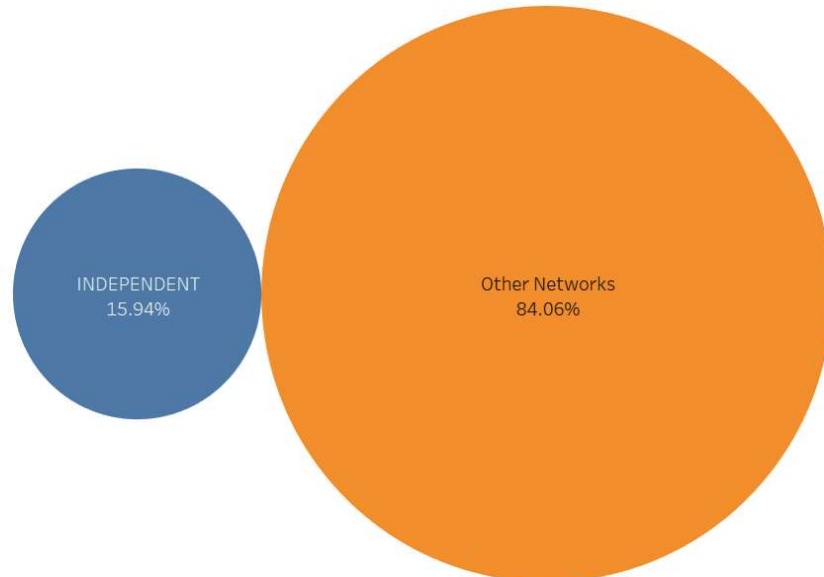
even rankings. Their main use is in **before and after storytelling** since they show very easily what changes have occurred in the analyzed time interval.

People of age group 20-44 are represented by red line slope and people of the age group 55-64 are represented by blue line slope. We have used a dual axes graph to show the trends between both the groups for the year 2021.

As we can see from this visualization, more people from the age group of 55-64 were hospitalized since the beginning of the year. Gradually, the number decreases, but it is still greater than the number of patients in 20-44 age group. Hence, we accept this hypothesis.

Hypothesis 3:

Whether INDEPENDENT hospital network in Erie County have less treatment scale than CATHOLIC HEALTH network and KALEIDA HEALTH network



The Hypothesis 3 is "Whether the INDEPENDENT hospital network in Erie County has less treatment scale than the CATHOLIC HEALTH network and KALEIDA HEALTH network"

We used packed bubble chart for this hypothesis.

To categorize the hospital networks, we used grouping in Tableau to group CATHOLIC HEALTH NETWORK and KALEIDA HEALTH network named as 'Other networks'.

After grouping, we create a calculated field to identify TREATMENT SCALE:

Treatment scale = ([Discharges to Date]/([Discharges to Date]+[Fatalities to Date]))/([Currently Hospitalized]+[Newly Admitted]+1)

Using the formulae, we were able to get the treatment scale metric.

Using the sum of the Treatment scale in size and label in marks, we get the visualization.

From the visualization, we can determine that the patients getting treatment in the independent hospital network is significantly fewer than the getting treatment

The other networks have 84.06% Treatment scale and INDEPENDENT networks have 15.94%.

From this visualization, we can conclude that people in ERIE county trust the CATHOLIC HEALTH NETWORK and KALEIDA HEALTH network than INDEPENDENT

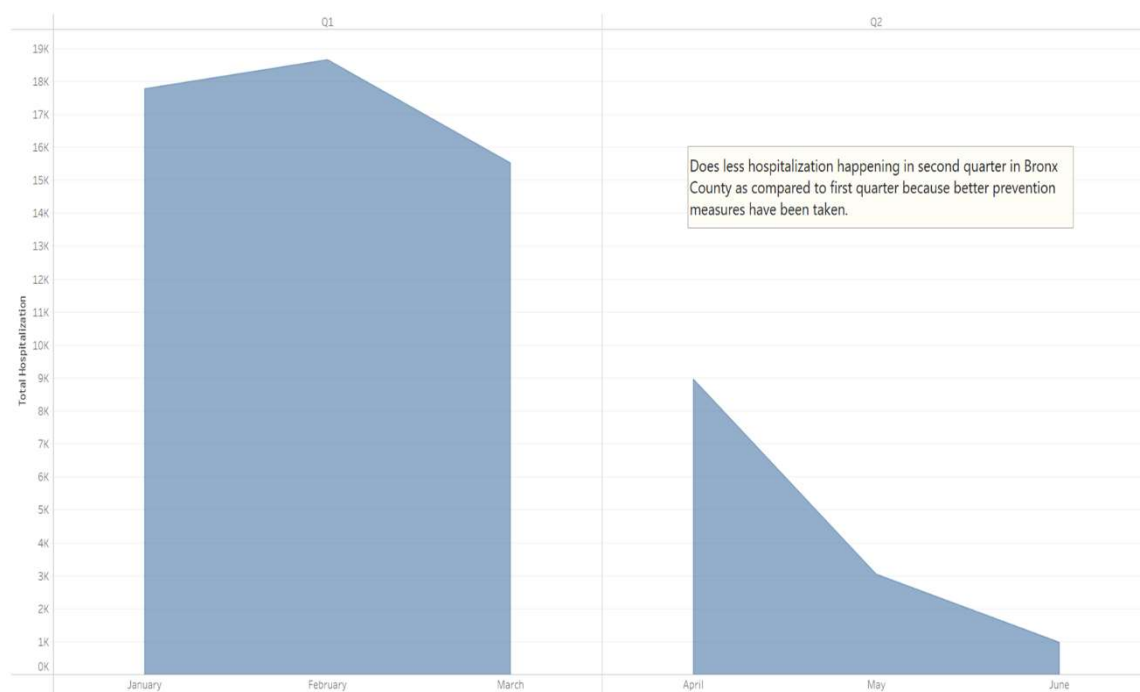
This also shows that the hospitals in that network have good standing with the people of ERIE county and the beds are getting filled in the Other first than the INDEPENDENT network hospitals.

Hence, we do not reject this hypothesis.

Hypothesis 4:

Hypothesis 4 states that does less hospitalization happening in second quarter in Bronx County as compared to first quarter because better prevention measures have been taken. First quarter comes under the months of January, February and march and second quarter comes under the April, May and June. The area charts are used to visualize the

hypothesis. Y- axis shows the people who were admitted to the hospital. Getting in detail February has the peak point where the maximum number of people admitted to the hospital than January. In month of April number of people number of people admitted to the hospital Were reduced where we can consider it as a mid-point and the June month shows the least point where 974 people admitted to the hospital. We can clearly observe that a downward trend in the second quarter compared to the first quarter. To conclude, because of taking better prevention measures in Bronx county during covid less hospitable happened in the second quarter when compared to the first quarter.

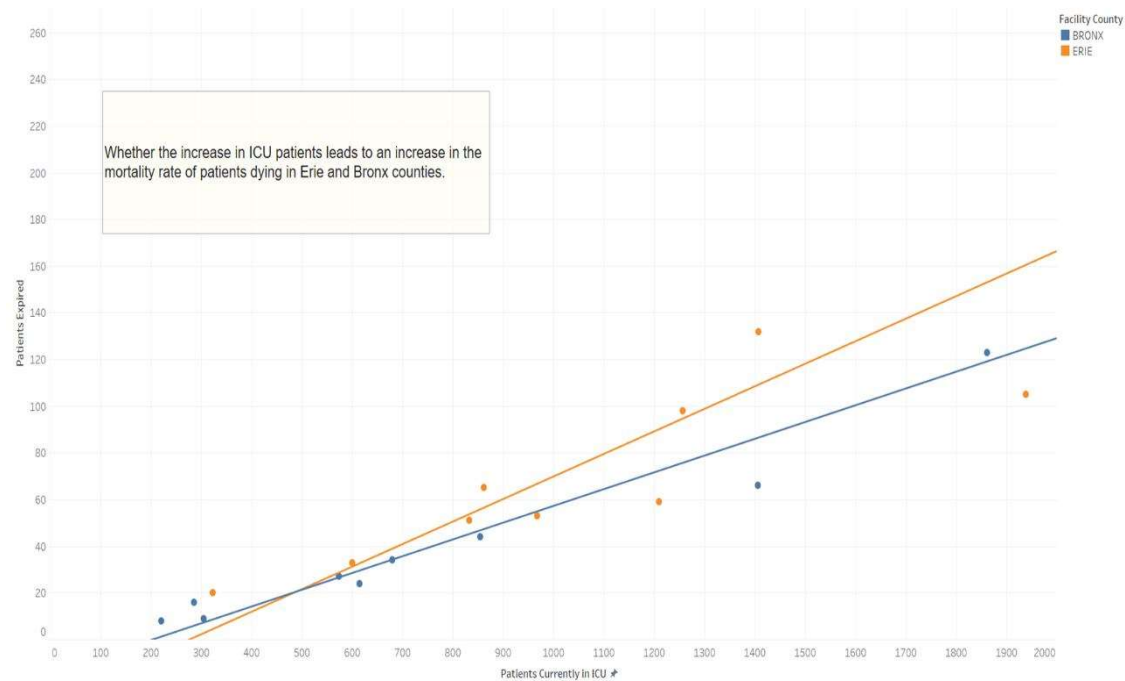


Hypothesis 5:

The hypothesis 5 is “Whether the increase in ICU patients leads to an increase in the mortality rate of patients dying in Erie and Bronx counties”.

The “Scatterplot” and “Trend Line Chart” visualizations are used to represent the data for year 2021 with “Patients currently in ICU” on x-axis and “Patients expired” on y-axis. The blue color indicates “Bronx” county and orange color indicates “Erie” county. The straight trend line derived on scatterplot for both counties indicates that as the number of patients in ICU increased, the expiry rate of the patients also went up. Since the similar trend is

observed in both counties, it can be deduced that there is a strong correlation between patients admitted to ICU and patients who died. The reasons for this trend can be there might be the same kind COVID-19 variant is prevalent in both counties and weather conditions might be same in both the areas. Thus, the hypothesis is true for both Erie and Bronx County.



CONCLUSION

From our findings we can conclude the following observations:

1. Unpredictable future
2. Preventive measures
3. With the number of patients in hospitals, we can observe that older people are more prone to Covid 19 when compared to younger people.
4. People in ERIE county trust CATHOLIC HEALTH NETWORK and KALEIDA HEALTH network than INDEPENDENT NETWORKS.
5. Services provided by hospitals in Bronx is better than that provided in Erie County, as proved by hypothesis 1 and 5.

REFERNCES

1. Primary Dataset: <https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Hospitalizations/jw46-jpb7>
2. Secondary Dataset: <https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing/xdss-u53e/data>